Inaugural dissertation
for
obtaining the doctoral degree
of the
combined Faculty of Mathematics, Engineering and Natural Sciences of the
Ruprecht-Karls-University
Heidelberg

Presented by
M.Res. Rebecca Emma Wagner
born in: Cambridge, UK
Oral examination: 22.02.2024

Unraveling Transcriptional Vulnerabilities in Cell Cycle and DNA Repair Genes

Referees:
Professor Michaela Frye
Dr. Karin Müller-Decker

Abstract

All cells in a multicellular organism contain the same genetic information, yet individual cell types express different genes in order to fulfil their specific functions. Expression patterns are regulated dynamically throughout the cell cycle and in response to internal and external stimuli. Improper regulation of gene expression patterns is associated with diseases such as cancer, where developmental gene expression profiles are often hijacked.

The serine - arginine rich splicing factor (SR) proteins are a conserved family, consisting of 12 members in humans. They were first described as splicing factors, promoting the inclusion of weak splice sites and shown to work as a network, regulating their own and each others expression in a cell type dependent manner. Various SR proteins have also been shown to have additional RNA processing functions, including regulation of transcription, translation initiation and non-sense mediated decay.

Other than splicing the precise mechanisms by which SR proteins modulate gene expression were unknown. My study focused on family member SRSF2 and a recurrent mutant form of SRSF2 (P95H) commonly found in cancer.

I directly compared the effects of conditional deletion or mutation of SRSF2 in mouse skin. Both perturbations caused a loss of cellularity and cell cycle arrest accompanied by impaired differentiation and increased DNA damage. I showed that SRSF2 acts as an important regulator of transcription and that the P95H mutation results in a loss of protein function.

To decipher the molecular mechanisms of SRSF2 action it was knocked down in primary human keratinocytes, an established *in vitro* model for normal skin, and two squamous cell carcinoma cell lines. In all cell lines analysed SRSF2 depletion caused cell cycle arrest and increased DNA damage. In an attempt to look at the relative contributions of splicing and transcription both processes were analysed. No conserved changes to splicing were observed across cell lines, while analysis of transcription rates via SLAM-sequencing and RNA polmerase II (Pol II) occupancy via Cut&Run experiments revealed that depletion of SRSF2 was associated with reduced global nascent transcription due to enhanced stalling of Pol II at transcription start sites, suggesting a direct effect on transcription is the main cause of the effects of SRSF2. DNA replication and repair associated genes were most affected by loss of SRSF2 due to their specific genomic organisation as bi-directional gene pairs.

These results provide insights into why SRSF2 is an essential gene, but how mutant forms of SRSF2 contribute to carcinogenesis remained unexplained. Therefore I generated cancer cell clones expressing reduced levels of SRSF2 using CRISPR-cas9

mediated gene editing approaches. Cancer cells with reduced SRSF2 function can cycle but accumulate mutations faster and thereby enable faster tumour evolution. Introducing the *Srsf2* P95H mutation into one allele of the mouse skin inhibited tumour formation following exposure to carcinogens. However, within SRSF2 P95H established clones cell growth was not affected and clone size tended to be larger than that of control cells.

In summary, my studies showed for the first time that the P95H mutation causes a loss of function and that SRSF2's role in transcription, and not splicing, is required for proper expression of DNA replication and repair genes due to their organisation as bi-directional gene pairs.

Zusammenfassung

Alle Zellen in einem mehrzelligen Organismus enthalten dieselben genetischen Informationen, aber einzelne Zelltypen exprimieren unterschiedliche Gene, um ihre spezifischen Funktionen auszuführen. Diese Expressionsmuster werden dynamisch im Zellzyklus und als Reaktion auf interne und externe Reize reguliert. Eine fehlerhafte Regulation von Genexpressionsmustern ist mit Krankheiten wie Krebs verbunden, bei denen häufig Expressionsprofile aus der Embryonalentwicklung wiederverwendet werden.

Die serin-arginin-reichen Spleißfaktor (SR)-Proteine bilden eine konservierte Familie, die in Menschen aus 12 Proteinen besteht. Sie wurden zunächst als Spleißfaktoren beschrieben, die die Inklusion schwacher Spleißstellen fördern, und es wurde gezeigt, dass sie als Netzwerk funktionieren, das ihre eigene und die Expression untereinander in zelltypabhängiger Weise reguliert. Verschiedene SR-Proteine haben zusätzliche Funktionen bei der RNA-Verarbeitung, einschließlich der Transkriptionsregulation, Translationseinleitung und der Nonsense-vermittelten RNA-Degradation.

Abgesehen vom Spleißen waren die genauen Mechanismen, durch die SR-Proteine Genexpression modulieren, unbekannt. Meine Studie konzentrierte sich auf das Familienmitglied SRSF2 und eine rezessiv mutierte Form von SRSF2 (P95H), die häufig in Krebszellen vorkommt.

Unter Verwendung der Haut als Modellsystem habe ich direkt verglichen, wie die gezielte Deletion oder Mutation von SRSF2 in epidermalen Stamm- und Vorläuferkompartimenten zu einem Verlust der Zellzahl und Zellzyklusarrest führte, begleitet von beeinträchtigter Differenzierung und vermehrten DNA-Schädigungen. Ich konnte zeigen, dass SRSF2 ein wichtiger, genomweit agierender Transkriptionsregulator ist und dass die P95H-Mutation in SRSF2 zu einem Funktionsverlust des Proteins führt.

Um die molekularen Mechanismen durch die SRSF2 wirkt zu entschlüsseln, habe ich es in primären menschlichen Keratinozyten, einem etablierten *in vitro*-Modell der gesunden Haut, und zwei Zelllinien von Plattenepithelkarzinomen herunterreguliert. In allen analysierten Zelllinien führte die Depletion von SRSF2 zu einem Zellzyklusarrest und vermehrten DNA-Schädigungen. Ich habe weiter die relativen Beiträge des Spleißens und der Transkriptionsregulation zu dem SRSF2-Phänotyp analysiert. Ich konnte keine konsistenten Änderungen beim Spleißen in allen Zelllinien beobachten, während die Analyse der Transkriptionsgeschwindigkeit durch SLAM-Sequenzierung und RNA-Polymerase-II (Pol II)-Besetzung über Cut&Run-Experimente ergab, dass die Depletion von SRSF2 mit einer verringerten globalen naszierenden Transkription aufgrund eines verstärkten Stockens von Pol II an Transkriptionsstartstellen verbunden war, was auf einen direkten Einfluss auf die Tran-

skription als Haupfunktion von SRSF2 hindeutet. Interessanterweise waren Gene, die mit DNA-Replikation und -reparatur assoziiert sind, am stärksten von einem Verlust von SRSF2 betroffen, aufgrund ihrer spezifischen genomischen Organisation als Genpaare an bidirektionalen Promotoren.

Diese Ergebnisse geben Einblicke, warum SRSF2 ein essenzielles Gen ist, aber wie mutierte Formen von SRSF2 zur Karzinogenese beitragen, blieb unerklärt. Daher habe ich durch CRISPR-Cas9-vermittelte Knockouts klonale Krebszellinien generiert, die reduzierte SRSF2-Levels exprimieren. Diese Krebszellen mit eingeschränkter SRSF2-Funktion können den Zellzyklus normal durchlaufen, aber Mutationen akkumulieren schneller und ermöglichen so eine beschleunigte Tumorevolution. Die Einführung der *Srsf2* P95H-Mutation in ein Allel inhibierte die Tumorformation in der Haut nach Exposition gegenüber Karzinogenen im Mausmodell. Innerhalb der etablierten SRSF2-P95H Zelllinien wurde das Zellwachstum jedoch nicht beeinträchtigt, und die Klone neigten dazu, größer zu sein als die Kontrollzellen.

Zusammenfassend zeigten meine Studien zum ersten Mal, dass die P95H-Mutation einen Funktionsverlust verursacht und dass SRSF2 nicht primär im Spleißen erforderlich ist, sondern direkt in der Transkription und für die ordnungsgemäßen Genexpression von DNA-Replikations- und reparaturgenen, durch ihre Organisation als Genpaare mit bidirektionel Promotoren.

Dedication

This thesis is dedicated to all those who have guided me along this path

"...a journey of a thousand miles begins with a single step..."

Loazi

Acknowledgements

The first thank you has to go to Ela, my supervisor. Your support started even before the PhD did and I am so grateful for the opportunities you have helped me take advantage of! Our weekly meetings, sometimes short and sometimes long were always a joy for me and I will miss discussing results with you. Thank you for taking a chance on me!

Next I would like to thank Martyna, you welcomed me into the lab and showed me what it's like to do Science everyday! Thank you for always taking time to answer my questions and teach me so many techniques! Thank you also for supporting me in finding my PhD position and for talking to me about life while doing experiments in the middle of the night! Being the midnight fairies cleaning the department has a very special place in my heart.

Roberto, Bobby! You were re-Becky's first supervisor and I am so grateful for everything you taught me, although I still scare easily. Thank you for the endless discussions and pretending to know the songs on the radio.

Nicole thank you for being my lab big sister and all the fun Fridays! I wish we had had more time together, although I think much less work would have been done so maybe it all worked out for the best.

To all other members of the Cambridge addition of the Frye lab, Tomasso, Susanne, Aga, Sandra thank you it was a pleasure to work along side you all, even if it was just very briefly!

Next I have to thank Sylvain and Mika, you are colleagues, friends, an adopted family! Thank you for the endless support and encouragment and a lot of fun!! I cannot believe that we wont be moving together again, I dont know how to do that without you both, but i'm sure we will all see each other again, maybe at Mika's Grandmother's place?

To Frye members 2.0; Sophia, Dany, Lilla, Fu, Tim, Shafagh, Cornelius, Veronica, Sabrina, Liana, Milica, Leona, Beatrice and honorary member Alex along with all those who just passed through, thank you for making work fun, I know you are all going to do amazing things! A special thank you to Leonie, Anke and Faidra who had to put up with working with me at even closer proximity. Thank you for always being up for a challenge and riding the SRSF2 wave with me!

None of this work would have been possible without the help of the animal house technicians, members of the sequencing open lab and flow cytometry core facility. I would especially like to thank Gaby for cutting every block in record time and Damir for teaching me so much about microscopy!

A big thank you also to Thiago, Ali and Etienne for all the time and energy you invested in the project and always being up for discussing ideas!

Next up I want to thank Anastasia, Ricca, Xico, Rafi, Monica and all the extended Papavasiliou lab for welcoming me to Heidelberg and showing me all the sites, like Untere Strasse... Also to Yogi Evi for brightening my Tuesday evenings and teaching me about eating the tail of the Donkey.

I also want to thank the TGIF girls Larissa Eirini, Bea, Kati and Sofia we organised amazing happy hours, I mean Santa Claus doesnt turn up to just any happy hour! I hope we will stay in touch. An extra thank you to Eirini, I will really miss the walks, talks, coffees and ice-creams, but hope there will be many many more, maybe in different parts of the world!

I am also very grateful to the summer schools that let me travel to Greece for several consecutive years and the things I learnt but also the amazing people that I met there! Special shout out to the party crew Margit, Kadi, Elsa, Ben, Alessia, Kim, Simon and Pablo, the week we took over Hydra is burnt in my memory, mostly, and I have loved randomly catching up on zoom and hearing about how you're all doing! A very special thank you to Pablo and Simon for continuing the party in Heidelberg, oh and for introducing me to your colleague...

To all the friends in Heidelberg and around the world, thank you for making me laugh, for supporting me and exploring the world with me.

A special thank you for J, you're support and love means the world. Thanks for telling me when I'm a princess and helping me to always see the positive side!

Last, but definitely not least to my family I could not have done any of this, or even anything at all without you, thank you for your unwavering support, love and guidance, I would be lost without you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Skin and its appendages

The skin is the body's largest organ and protects from external stresses such as UV radiation, microorganisms and physical insult. It also prevents water loss and allows for adaptation to changes in temperature through its appendages; hair follicles and sweat glands (Figure 1.1 A). Skin can broadly be divided into three layers: the epidermis, the dermis and the hypodermis (Janes and Watt, 2006). The epidermis is the outer-most layer and consists of the interfollicular epidermis (IFE) and appendages including the hair follicle, sweat glands and sebaceous glands (Cédric Blanpain and Fuchs, 2009, Janes and Watt, 2006). The epidermis is connected to the dermis via the basement membrane which is a layer of specialised extracellular matrix (ECM) secreted by the epidermis, which plays an important role in maintaining a niche for the stem cells of the IFE (Cedric Blanpain et al., 2004, U. B. Jensen, Lowell, and Watt, 1999, J. Li et al., 2022). The dermis consists of various fibroblast populations, connective tissue, vessels and nerves. Below the dermis is the hypodermis, which largely consists of adipose tissue (Janes and Watt, 2006).

1.1.1 The interfollicular epidermis

The IFE is a stratified squamous epithelium made up of undifferentiated basal and differentiated suprabasal keratinocytes (Figure 1.1 B) (Cédric Blanpain and Fuchs, 2009). The basal layer encompasses the mitotically active compartment of the IFE, made up of stem and progenitors cells that divide symmetrically to self-renew and asymmetrically to produce progeny that leave the basal layer and undergo differentiation (Cockburn et al., 2022, Ya and Fuchs, 2022). The precise mechanisms by which self-renewal and differentiation are regulated have not been fully elucidated, however, it is clear that the ability to divide is coupled to the location of cells within the basal layer and that crowding may force cells out of this layer, which triggers differentiation. The cells of the basal layer secrete the basement membrane and maintain contact with it through adhesion molecules such as integrins (U. B. Jensen, Lowell, and Watt, 1999). When cells lose contact with the basement membrane they leave the basal layer and are destined to undergo terminal differentiation (Tumbar et al., 2004). The basement memebrane along with lower calcium concentrations

provide a niche which supports proliferation and maintains an undifferentiated state, essential for maintainance the IFE. Differentiation is supported by increasing calcium concentrations and involves extensive changes to gene expression, which impacts cell morphology and function. Cells become larger and flatter and eventually loose their nucleus, forming a cornified envelope (Cédric Blanpain and Fuchs, 2009, Ya and Fuchs, 2022).

Keratins are the major structural proteins of cells in the IFE, they form heterodimers and their expression patterns can be used to define the various cell types, which are also spatially restricted, present in the IFE (Cohen et al., 2022, Moll, Divo, and Langbein, 2008). Basal cells express keratin 14 (KRT14) and Keratin 5 (KRT5), while cells in the intermediate suprabasal (spinous) layer express keratin 10 (KRT10) and keratin 1 (KRT1). Cells in the outer suprabasal (granular) layer express other proteins, such as loricrin (LOR) and transglutaminase 1 (TGM1) (Figure 1.1 B).

Under homeostatic conditions the IFE is maintained by stem and progenitor cells in the basal layer, however, upon injury stem cell populations from the hair follicle are activated, proliferate and undergo trans-differentiation in order to contribute to the IFE, ensuring that the barrier is maintained (Cedric Blanpain et al., 2004, Tumbar et al., 2004, Y. Ge et al., 2017).

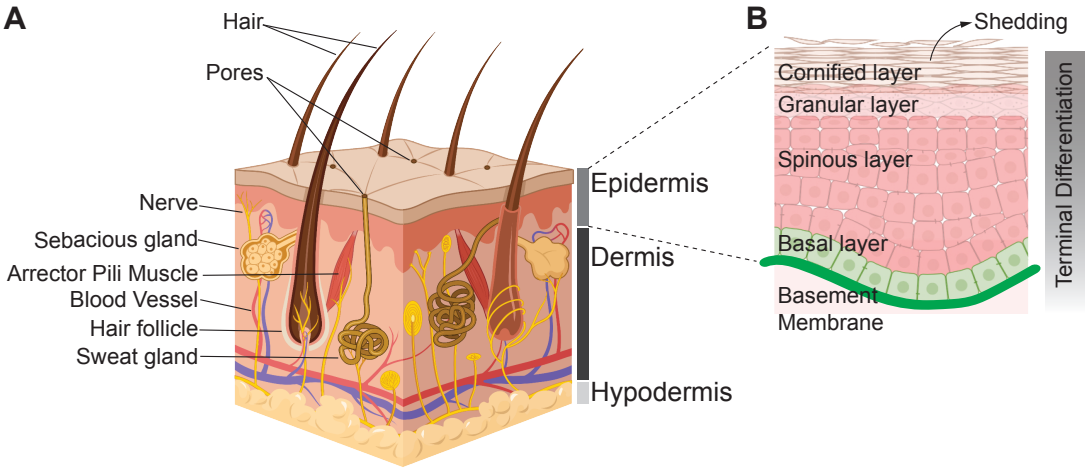

Figure 1.1: **Skin Architecture**
A Skin cross section showing the three major layers: Epidermis, dermis and hypodermis and the appendages that it contains; hair follicles, sweat glands and sebacious glands. Nerves and bloods vessels are located under the skin. Figure adapted from Encyclopedia Britannica 2013. **B** Representation of the epidermis composed of undifferentiated keratinocytes in the basal layer (green), which differentiate and move out of this layer. The epidermis is seperated from the dermis via the basement membrane. Figure adapted from Janes and Watt, 2006 and made using Biorender.

1.1.2 The hair follicle and hair cycle

The hair follicle is another self-renewing compartment of the skin, which produces hair in cyclic bouts of growth (anagen), regression (catagen) and rest (telogen).

In mouse hair follicle (HF) formation begins during embryogenesis when mesenchymal cells colonise the skin to form the underlying dermis (Schmidt-Ullrich and

Paus, 2005, Nowak et al., 2008, Cédric Blanpain and Fuchs, 2009). Mesenchymal cells in the dermis interact with epidermal cells triggering them to begin forming hair follicles. This begins at embryonic day 14 with placode formation and involves proliferation of cells in the epidermal sheet and their migration into the dermis (Figure 1.2 A). Following placode formation specialised fibroblasts cluster at the base of the placode forming the hair germ (Schmidt-Ullrich and Paus, 2005). These fibroblasts will become the dermal papilla (DP), which has a blood and nerve supply, thus nourishing the hair follicle (Nowak et al., 2008, Schmidt-Ullrich and Paus, 2005). Between embryonic day 16.5 and 18.5 epidermal cells in the placode continue to proliferate and migrate into the dermis. Cells closest to the DP grow around it and begin to differentiate into cell populations found in the mature hair follicle, such as the inner root sheath (IRS) and outer root sheath (ORS), eventually forming a cylindrical structure (Schmidt-Ullrich and Paus, 2005).

Hair follicle development continues postnatally; at postnatal day 6 the hair follicle nears the bottom of the dermis. Proliferative cells, termed matrix cells, continue to divide producing progeny that terminally differentiate forming hair (Alonso and Fuchs, 2006, Schmidt-Ullrich and Paus, 2005). Hair follicle morphogenesis continues until postnatal day 19 and the end of development is marked by the first catagen, followed by a short telogen phase. Hair growth is synchronised for the first 6 weeks of mouse life, occurring in waves that move in a head to tail direction, after which hair cycles asynchronously (Mueller-Roever et al., 2001).

Skin homeostasis and response to stress has been well studied meaning the molecular mechanisms and how the tissues function are well understood, this makes it an excellent model system for studying proteins with unknown functions.

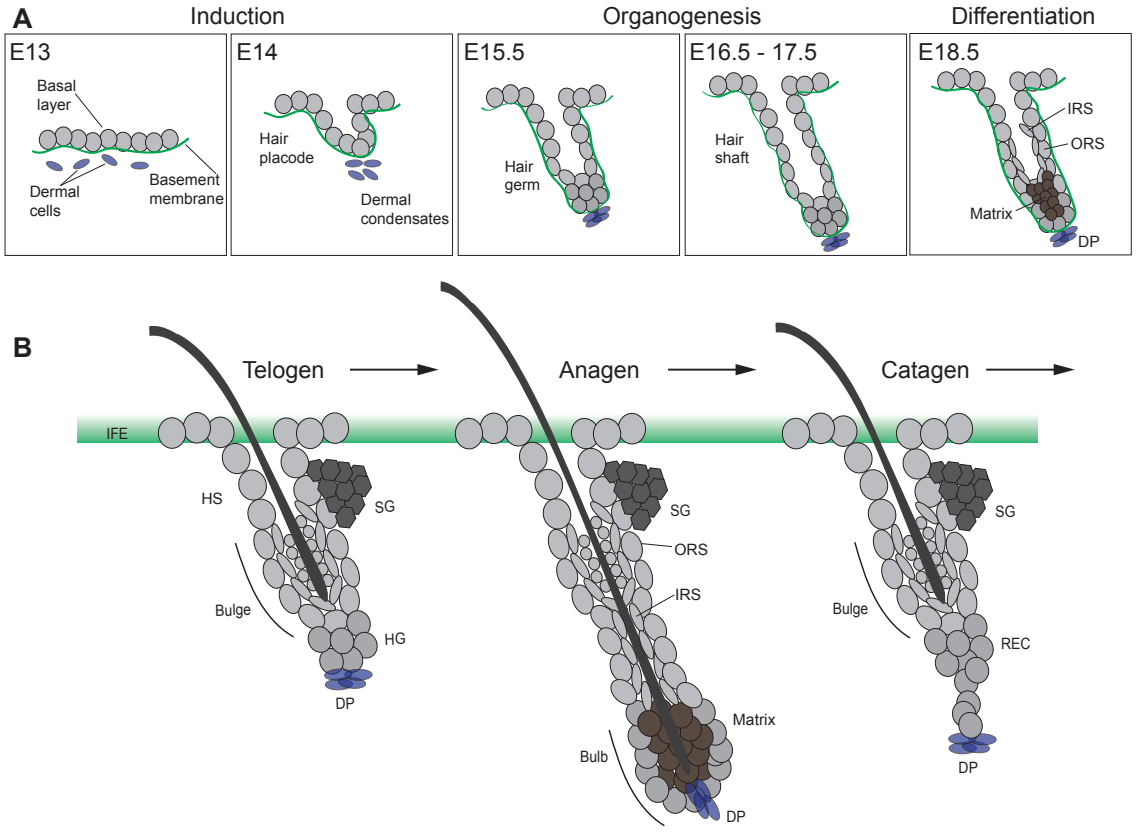

Figure 1.2: **Hair morphogenesis and the hair cycle**
A Illustration of hair follicle morphogenesis. Embryonic day 13 (E13) and 14 (E14) encompass the induction phase. Embryonic day 15.5 (E15.5) to embryonic day 17.5 (E17.5) consistute the organogenesis phase and embryonic day 18.5 is when cells of the hair shaft differentiate into various cell types present in the adult hair follice. Figure adapted from Cédric Blanpain and Fuchs, 2006 and Schmidt-Ullrich and Paus, 2005. **B** Representation of changes to the hair follicle that occur during the hair cycle. Telogen is the hair follicle at rest, anagen is the hair follicle during growth and catagen is the hair follicle regressing, via apoptosis. Figure adapted from Jaks et al., 2008.
IFE-interfollicular epidermis, HS-hair shaft, SG-sebaceous gland, HG-hair germ, DP-dermal papilla, ORS-outer root sheath, IRS-inner root sheath, REC-regressing epithelial column

1.1.3 Skin Cancer

Tumours with an epithelial origin account for between 80 and 90% of cancer cases (Siegel et al., 2023). These cancers arise from tissues including the intestine, lung, prostate, breast and skin. Skin cancer can originate from melanocytes which are neural crest derived pigment producing cells, or epithelial cells (Centeno, Pavet, and Marais, 2023). Cancers originating from melaonocytes are called melanomas and although these only account for approximately 5% of skin cancer cases they are responsible for 75% of deaths due to their much higher propencity to metastasise (Centeno, Pavet, and Marais, 2023). Epithelial cell derived skin cancers have been further categorised into basel cell carcinoma (BCC), squamous cell carcinoma (SCC) and Merkel cell carcinoma (MCC) (Siegel et al., 2023).

BCC is the most frequently occuring type of skin cancer in humans and has been shown to arise from several cell types in the skin, including the basal cells of the IFE

and the hair follicle (Atwood, Whitson, and Oro, 2014). A key feature is constitutive activation of hedghog signalling (Atwood, Whitson, and Oro, 2014, Atwood, Sarin, et al., 2015). In most cases surgery is sufficient to treat BCC. However, occasionally BCC can metastasise, or surgery may not been an option due to the location of the tumour for these cases Hedgehog antagonists have been developed. However, drug resistance has limited their usage (Atwood, Whitson, and Oro, 2014, Whitson et al., 2018).

SCCs can occur in various locations in the body, but often occur in sun exposed areas. The tumours have an epithelial origin and share common mutations and express markers associated with squamous epithelia. They have been extensively studied via chemical carcinogenesis protocols (E. L. Abel et al., 2009, Nassar et al., 2015).

Following the aquisition of mutations, most frequently in Ras and p53 (Nassar et al., 2015), cells begin to proliferate in an uncontrolled manner giving rise to benign papillomas, which are not invasive (E. L. Abel et al., 2009, Nassar et al., 2015). However, if they are allowed to persist they gain additional mutations and can transform into SCCs. These have the capcity to invade neighbouring tissues and metastasise to distal sights.

The existence of cancer stem cells and whether the cell of origin impacts the features of a tumour has been hotly debated. Experiments where p53 was deleted and oncogenic Ras was expressed in various cell types in the skin showed that the cell of origin does influence features of the resulting tumour (Sanchez-Danes and Cedric Blanpain, 2018). If mutations were induced in the progenitor populations of the IFE and hair follicle no tumour would occur, conversly when these mutations were induced in stem cell populations, of either the IFE or the hair follicle, tumours formed. Interestingly, the tumours which formed from the bulge stem cells of the hair follicle were more heterogenious and aggresive, containing cells which appeared to be undergoing epithelial to mesenchymal transition (EMT), which is associated with the ability to metastasise (Sanchez-Danes and Cedric Blanpain, 2018).

1.2 Cell cycle regulation and dysregulation in cancer

Cell division is a fundamental process which encompasses the steps required for a cell to copy its genetic material and divide into two cells. Normal cell division is particularly important in tissues with high turnover rates, such as the skin and intestine to ensure they continue to function fully throughout the lifetime of an organism.

The cell cycle has been extensively studied and can broadly be divided into two main phases; synthesis (S phase), when DNA is replicated in a semi-conservative manner (Meselson and Stahl, 1958), and mitosis (M phase), when DNA is segregated (Figure 1.3) (Morgan, 2007, Basu et al., 2022). These two events are bridged by gap phases, G1 and G2, which encompass crucial regulatory steps (Matthews, Bertoli, and R. A. d. Bruin, 2022). During G1 cells decide whether they are fit to begin replication or whether to re-enter G0, while in G2 cells check whether replication has occured faithfully in order for duplicated cell contents to be divided between two daughter cells (Matthews, Bertoli, and R. A. d. Bruin, 2022).

The eukaryotic cell cycle is orchestrated by cyclin dependent kinase (CDK) activity. These are heterodimeric enyzmes with a protein kinase subunit and a cyclin subunit (Hochegger, Takeda, and Hunt, 2008, Basu et al., 2022). The levels of CDKs remain relatively constant throughout the cell cycle while the levels of different cyclins vary, depending on the cyclin-CDK complex that forms different substrates are phosphorylated and this allows progression through the cell cycle (Hochegger, Takeda, and Hunt, 2008).

1.2.1 Cell cycle progression

The majority of cells in a tissue are in G0, meaning they are not dividing. In order to commit to replication cells enter G1, which is associated with D-type cyclin-CDK activity. If this activity is lost cells re-enter G0. Along with D-type cyclin-CDK activity the E2F-dependent transcriptional network must be active for cells to commit to replication. The E2F family consists of 8 members and they function with the pocket proteins; retinoblastoma (RB), p130 and p107 (Matthews, Bertoli, and R. A. d. Bruin, 2022).

RB inhibits E2F-dependent transcription during G1 but is inactivated by CDK mediated phosphorylation. E2F-dependent transcription is activated by mitogen associated protein kinase (MAPK) signalling and Myc acitivity, which leads to expression of crucial regulators of cell cycle progression, such as E-type cyclins, which corroborate CDK activity required for replication. CDK inhibitors are also inactivated, which ensures CDK activity (Matthews, Bertoli, and R. A. d. Bruin, 2022).

Replication initiation then additionally requires A-type cyclin activity. These positive feedback loops ensure that DNA replication is only triggered under the correct circumastances and that it only happens once per cycle (Matthews, Bertoli, and R. A. d. Bruin, 2022). DNA is replicated by the replication machinery which consists of DNA polymerase and other factors, including helicases, which unwind double stranded DNA, allowing DNA polymerase to function (Masai et al., 2010)

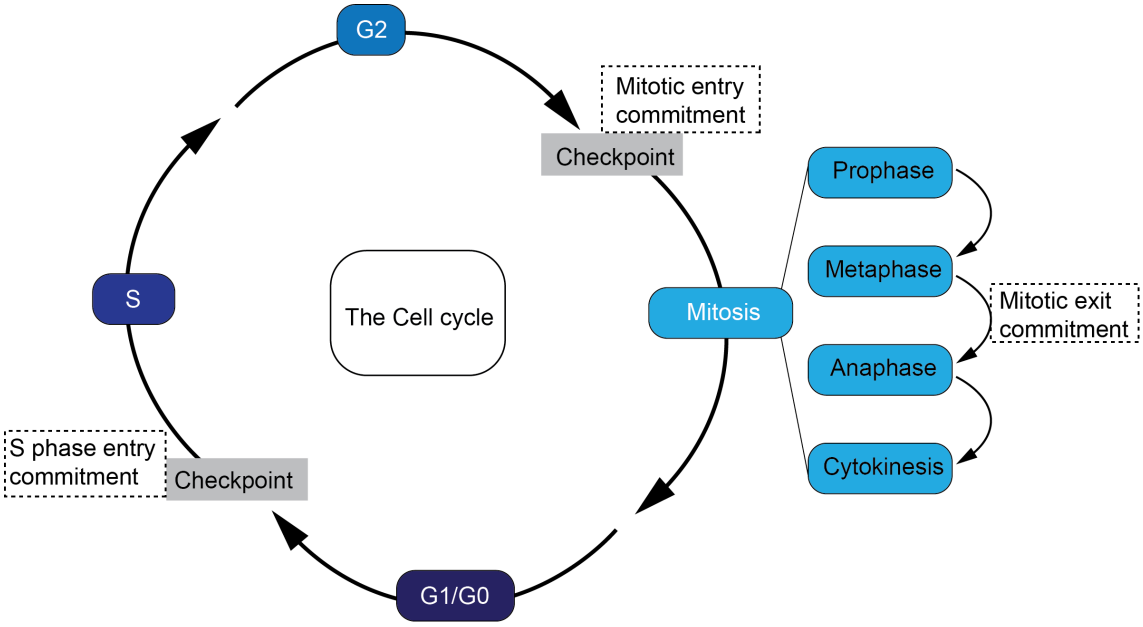

Figure 1.3: **The eukaryotic cell cycle**
Cell contents is duplicated during interphase, which encompasses G1, S, when DNA is replicated, and G2. Duplicated DNA is divided between two daughter cells during mitosis. There are regulatory steps before entry into S phase and before entry into mitosis. There is also a regulatory step during mitosis to ensure that chromosomal alignment has taken place properly before mitosis is completed. After mitosis cells can either re-enter the cell cycle or enter G0. Figure adapted from Matthews, Bertoli, and R. A. d. Bruin, 2022.

During S phase A- and B-type cyclins accumulate, and these interact with cyclin dependent kinase 1 (CDK1) (Matthews, Bertoli, and R. A. d. Bruin, 2022). CDK1 is inactive when phosphorylated, but phosphotase activity of cell division control 25 (CDC25) leads to active CDK1, which is required for mitosis (Matthews, Bertoli, and R. A. d. Bruin, 2022). CDK1 activity promotes a wave of kinase activity as it activates other kinases, such as (serine/threonine-protein kinase) PLK1, Aurora A and Aurora B (Matthews, Bertoli, and R. A. d. Bruin, 2022). This increased level of phosphorylation alters cell morphology priming the cells for division (Matthews, Bertoli, and R. A. d. Bruin, 2022). During mitosis phosphatase activity increases, which eventually inhibits CDK1 activity and leads to exit from mitosis. The daughter cells are then ready to re-enter G1 or remain in G0 (Morgan, 2007).

1.2.2 Cell cycle check points

The cell cycle is tightly regulated and dependent on feedback loops to ensure progression through the cell cycle happens in a robust and timely manner. Transition into and out of each phase of the cell cycle is regulated by numerous check points (Figure 1.3) (Murray, 1992). These ensure that each phase is completed properly and prevents cells harbouring errors persisting and giving rise to prodgeny.

During interphase DNA damage is sensed by the MRN complex, which consists of Mre11, Rad50 and Nbs1 (Matthews, Bertoli, and R. A. d. Bruin, 2022). It is a DNA damage sensor complex, which plays an important role in processing double-strand

breaks to enable repair. It also activates Ataxia-telangiectasia mutated (ATM), which then phosphorylates p53 and other proteins, including H2A histone family member X (H2AX) and breast cancer type 1 susceptibility protein (BRCA1), which are essential for coordinating the DNA damage response (Matthews, Bertoli, and R. A. d. Bruin, 2022). It also induces DNA repair proteins, both transcriptionally and post-transcriptionally, thus priming cells for repair. Active p53 recruits p21, a CDK inhibitor which prevents entry into S phase, allowing DNA repair to take place before the DNA is replicated (Jackson and Bartek, 2009).

S phase specific checkpoints ensure that DNA is replicated faithfully and regulates three main aspects of replication (Bertoli, Skotheim, and R. A. D. Bruin, 2013, Matthews, Bertoli, and R. A. d. Bruin, 2022, Murray, 1992). Firstly it prevents late origin firing, which ensures that DNA replication only occurs once per cell cycle. If DNA damage is present the replication fork is slowed or stalled, in order to allow DNA repair before replication. Therefore, stalled replication forks are also stabalised to allow repair to take place. Replication fork stalling leaves single stranded DNA exposed, which leads to ataxia telangiectasia and Rad3-realted protein (ATR) and checkpoint kinase 1 (CHK1) activity, ensuring cell cycle arrest, DNA repair or cell death (Matthews, Bertoli, and R. A. d. Bruin, 2022).

Mitosis can be divided into four steps (Figure 1.3) (Matthews, Bertoli, and R. A. d. Bruin, 2022, Earnshaw and Pluta, 1994, Crncec and Hochegger, 2019). In animal cells it begins with Prophase, here chromosome condensation occurs and mitotic spindle formation begins, while transcription ceases and does not begin again until G1/G0 (Earnshaw and Pluta, 1994, Murray, 1992). Prometaphase follows prophase, here the nuclear envelope breaks down and kinetochore microtubules connect with chromosome kinetochores. Once this has occured the two centrosomes, located at opposite poles of the cell pull the chromosomes apart, this stage is called metaphase. The chromosomes then align along the metaphase plate and the metaphase checkpoint ensures that kinetochores are properly attached to the mitotic spinlde and that chromosomes are properly aligned (Matthews, Bertoli, and R. A. d. Bruin, 2022, Earnshaw and Pluta, 1994, Murray, 1992). Mitosis then progresses to anaphase, here the cohesins that bind sister chromatids are cleaved, leading to two identical daughter chromosomes, which are then pulled apart. Finally cytokinesis occurs which completes division into two daughter cells (Earnshaw and Pluta, 1994).

During mitosis the spindle assembly checkpoint (SAC) ensures chromosomes segregate properly between two daughter cells. The SAC machinery is a multi-protein complex which is recruited to kinetochores not bound to microtubules, which delays mitosis until segregation is complete (Murray, 1992, Matthews, Bertoli, and R. A. d. Bruin, 2022).

Activation of any cell cycle checkpoint is sufficient to prevent further cell cycle progression (Murray, 1992). Cell fate after damage is variable, depending on the cell type and the extent of the damage, but there are three main outcomes. Firstly, cells may become senescent, in this case they permanently exit the cell cycle. Secondly, cells may become quiescent, this means they will not divide but could, if they recieve the appropriate signalls. Finally, if the damage cannot be repaired cells undergo programmed cell death (Matthews, Bertoli, and R. A. d. Bruin, 2022). These events are all triggered by p53 activity, which makes it unsurprising that this is the most

commonly mutated gene in cancer (Matthews, Bertoli, and R. A. d. Bruin, 2022, Lane, 1992).

1.2.3 Dysregulation of the cell cycle in cancer

The number of times a cell can divide depends on the cell type, for example, stem cells have the largest capacity to divide, while progenitors are, in part, defined by having a more limited capacity to divide and differentiated cells are unable to undergo cell division (Liu et al., 2019). In line with this DNA repair capacity also varies depending on the cell type, with stem cells showing enhanced repair, compaired to more differentiated cells (Liu et al., 2019).

Cancer is associated with uncontrolled proliferation and the ability to divide indefinitely (Hanahan and Weinberg, 2000). Cells acquire features that prevent them responding to anti-growth signals and produce their own growth promoting signals (Hanahan and Weinberg, 2000). Furthermore, cancer cells often acquire telomerase activity which increases the number of times a cell can divide as telomeres typically shorten with every cell division as a mechanism to prevent unlimited division. When telomeres reach a critical length a catastrophic level of DNA damage occurs, which triggers apoptosis (Jackson and Bartek, 2009). Telomerase activity removes this limit by continuously lengthening telomores (Hanahan and Weinberg, 2000).

Interestingly, components of the cell cycle involved in preventing cell cycle exit are mutated more frequently than components that drive progression through the cell cycle in cancer (Matthews, Bertoli, and R. A. d. Bruin, 2022). This suggests that the highly regulated nature of the cell cycle does not allow uncontrolled cell division and confines proliferation, even in cancer cells.

E2F-dependent transcription is frequently dysregulated in cancer, this enables cells to increase the window of time in which they can enter S phase (Matthews, Bertoli, and R. A. d. Bruin, 2022, Vermeulen, Bockstaele, and Berneman, 2003). Interestingly, ATR and CHK1, key players of the replication stress checkpoint are not mutated in cancer, highglighting that a functional replication stress response is always essential (Matthews, Bertoli, and R. A. d. Bruin, 2022). SAC mutations have also been observed but are very rare and there is some conflicting evidence, as it has also been reported that members of the SAC are over-expressed in cancer (Cahill et al., 1998, Castro, Cárcer, and Malumbres, 2007, Sarkar et al., 2021). This also suggests that a functional SAC is essential. Interestingly, normal cells appear to be less sensitive to SAC inhibition than cancer cells, which may provide a possible therapeutic window (Matthews, Bertoli, and R. A. d. Bruin, 2022, Vermeulen, Bockstaele, and Berneman, 2003).

1.3 DNA damage and repair

DNA damage can occur from both extrinsic and intrinsic sources. Extrinsic sources include mutagens such as ultraviolet (UV) and ionising irradiation and compounds that intercalate into DNA such as 5-fluorouracil (5FU) and 1,3-Dimethylbutylamin (DMBA) (Luch, 2005, Jackson and Bartek, 2009). Intrinsic sources of DNA damage include reactive oxygen species (ROS), produced during oxidative phosphorylation,

and damage that occurs during replication and transcription as a result of collisions and formation of features such as R-loops (Saxena and Zou, 2022, Tubbs and Nussenzweig, 2017, Lans et al., 2019, Petermann, Lan, and Zou, 2022).

Estimates suggest that approximately 70,000 lesions occur in a cell every day, therefore repair mechanisms are essential to ensure that excessive numbers of cells are not lost and that cells carrying deleterious mutations do not persist (Tubbs and Nussenzweig, 2017).

Different DNA repair pathways exist to respond to various types of DNA damage and are also differentially active depending on the cell type and cell cycle stage (Jackson and Bartek, 2009). For example, long-lived stem cells have been shown to use more accurate mechanisms of DNA repair than other somatic cells (Vitale et al., 2017, Jackson and Bartek, 2009).

There are five main pathways that can be activated after DNA damage occurs. Although they repair different types of damage or are active at different times during the cell cycle the basic steps are the same. First a lesion is detected, then signals are established to show where the damage has occured and this leads to recruitment of the repair machinery (Jackson and Bartek, 2009).

1.3.1 Base excision repair

Damage that does not distort the DNA double helix is repaired via base excision repair (BER). A damaged base is recognised by a DNA glycosylase enzyme, which then mediates excision, creating an abasic site that is recognised by apurinic/apyrimidinic (AP) endonuclease 1 (APE1), which then recruits DNA polymerase beta and other proteins required for repair, finally a ligase ensures that double stranded DNA is re-formed (Vitale et al., 2017, Jackson and Bartek, 2009).

1.3.2 Nucleotide excision repair

Damage that causes adducts and structures that distort the double helix is repaired by nucleotide excision repair (Vitale et al., 2017, Jackson and Bartek, 2009). Damage is excised as a 22-30 base oligonucleotide, which produces single stranded DNA. DNA polymerases along with other factors replace the DNA in a stepwise manner. Finally ligase activity ensures double stranded DNA is restored (Vitale et al., 2017). Thera are two types of NER; transcription coupled NER acts specifically on lesions blocking transcription, while global NER acts on lesions beyond those impairing transcription (Jackson and Bartek, 2009). Interestingly, NER has been shown to only occur in transcribed regions in differentiated cells, in un-transcribed regions DNA damage is not repaired (Nouspikel, Hyka-Nouspikel, and Hanawalt, 2006, Jackson and Bartek, 2009).

1.3.3 Mismatch repair

Mismatches and small insertions or deletions are repaired via DNA mismatch repair (MMR) (Jackson and Bartek, 2009). The incorrect base is excised, leading to a single stranded DNA, this is then repaired through the nuclease, polymerase and ligase activity (Vitale et al., 2017, Jackson and Bartek, 2009).

1.3.4 Double strand break repair

DNA double-strand breaks (DSBs) can be repaired by non-homologous end joining (NHEJ) or homologous recombination (HR). NHEJ operates throughout the cell cycle but often results in errors (Jackson and Bartek, 2009). HR can only occur when a template strand of DNA is present, as is the case during S and G2 phases of the cell cycle (Jackson and Bartek, 2009). Due to the use of a template it is the most accurate mechanism of repair (Jackson and Bartek, 2009). HR is triggered by ssDNA which are produced by the MRN complex (Matthews, Bertoli, and R. A. d. Bruin, 2022, Jackson and Bartek, 2009). HR is also active when replication forks resume activity after stalling and is important for resolving DNA cross links, where it acts with members of the fanconi's anaemia pathway (Jackson and Bartek, 2009).

1.4 Mechanisms regulating gene expression

All cells of a multicellular organism contain the same DNA, yet the genes expressed vary depending on the cell type and this largely dictates cellular function and morphology (Rué and Arias, 2015). Therefore, gene expression must be tightly regulated to ensure that cell identity is maintained and that cells are able to respond to both internal and external stimuli.

Gene expression can be regulated at several levels, firstly chromatin architecture and epigenetic features such as histone and DNA modifications influence which genes are accessible to RNA polymerase II (Pol II) and modulate it's activity (Tessarz and Kouzarides, 2014, Allshire and Madhani, 2018, Schneider and Grosschedl, 2007). Pol II activity is also modulated by interactions with transcription factors (TFs), which bind promoter or enhancer sequences in order to promote or repress Pol II activity. They can stabilise or prevent Pol II binding directly and they can also alter histone modifications to sign post gene activation or repression either through independent catalytic activitiy or by recruiting proteins with this capacity (F. X. Chen, Smith, and Shilatifard, 2018). Pol II activity also depends on the modifications present on its C-terminal domain (CTD), phosphorylation, acetylation and ubiquitiniation are all important for regulating Pol II levels and activity (L. Chen et al., 2018, Gilchrist et al., 2012, Peterlin and Price, 2006, Phatnani and Greenleaf, 2006). Furthermore, transcription rates impact co-transcriptional processes such as alternative splicing, which determines the protein isoform that will be produced from the transcript or whether the transcript will be degraded due to the inclusion of premature termination codons (PTC), or other features that target a transcript for degradation (Bentley, Guthrie, and J. Steitz, 2005).

1.4.1 Gene organisation

The organisation of genes in the genome is a highly conserved mechanism that allows both uni-cellular and multi-cellular organisms to regulate gene expression in response to stimuli in a coordinated manner (Adachi and Lieber, 2002, Zinani, Keseroğlu, and Özbudak, 2022).

Genes can be organised in operons, clusters or pairings (Zinani, Keseroğlu, and

Özbudak, 2022). Operons allow multiple genes to be transcribed from one promoter, which results in a single mRNA. They are very common in prokaryotes, but are also found in lower order eukaryotes including, C. elegans, trypanosomes and flatworms (Zinani, Keseroğlu, and Özbudak, 2022). Clusters and pairings also allow genes to be transcribed in a coordinated manner, but may also rely on other mechanisms, such as chromatin architecture to ensure fine tuning of expression (Zinani, Keseroğlu, and Özbudak, 2022).

Adjacent genes are co-expressed regardless of their orientation in the gnome (Trinklein et al., 2004, Zinani, Keseroğlu, and Özbudak, 2022). Divergently paired genes (DPGs), or bi-directional gene pairs, which are genes oriented in a head-to-head direction in close proximity, less than 1kb, apart, make up 10% of the human genome (Zinani, Keseroğlu, and Özbudak, 2022, Wakano et al., 2012). Interestingly, bi-directional genes are functionally related and genes important for meriad functions have been shown to be organised in this way in various species, including those involved in DNA repair, ribosome biogenesis and metabolism (Zinani, Keseroğlu, and Özbudak, 2022, Wakano et al., 2012). Organisation in close proximity extends further than bi-directional genes and in humans 16% of genes have been shown to be organised into functional clusters (Zinani, Keseroğlu, and Özbudak, 2022).

Beyond direct close proximity in the genome topologically associated domains (TADs) and phase seperation highlight that bringing specific regions of the genome into close proximity is essential for proper regulation of gene expression (Zinani, Keseroğlu, and Özbudak, 2022, Y. E. Guo et al., 2019, Boija et al., 2018, Hnisz et al., 2017). Placing functionally related genes in close proximity has clear advantages, such as balancing the dosage of the proteins in a complex and minimising the variability of expression of genes needed for a specific function. However, it has been shown that the mutational burden in bi-directional genes is higher and that they are found mutated in cancer more frequently (Wakano et al., 2012, Zinani, Keseroğlu, and Özbudak, 2022).

1.4.2 The transcription cycle

Pol II mediated transcription produces messenger RNA (mRNA), which can then be translated into proteins, as well as some long and small non-coding RNAs (Vervoort et al., 2022). The CTD of Pol II is made up of heptad repeats of Tyr1-Ser2-Thr4-Ser5-Ser7, the number of repeats varies depending on the class of organism, in mammals there are a total of 52 (Figure 1.4 A). Transcription occurs through a series of steps, which can be distinguished by the modifications present on the CTD of Pol II (Figure 1.4 B, C) (Williams et al., 2015).

The phosphorylation status of Serine 5 (Ser5) and Serine 2 (Ser2) of Pol II's CTD have been extensively studied (Nojima et al., 2015, Vervoort et al., 2022, Phatnani and Greenleaf, 2006, Peterlin and Price, 2006). Initially hypo-phosphorylated Pol II binds a promoter, along with general transcription factors (GTFs) leading to the formation of the pre-initiation complex (PIC) (Heidemann et al., 2013). Transcription initiation then requires Ser5 of the CTD to be phosphorylated and this is carried out by cyclin dependent kinase 7 (CDK7), the kinase domain of tran-

scription factor IIH (TFIIH) (F. X. Chen, Smith, and Shilatifard, 2018, Vervoort et al., 2022). This prompts promoter clearance, but Pol II halts within 200 bases of the transcription start site (TSS). This is a regulatory step known as "promoter proximal pausing" (Yamaguchi, Shibata, and Handa, 2013). Pol II is maintained in this paused state by two key factors; negative elongation factor (NELF) and DRB-sensitivity inducing factor (DSIF). Elongation can only resume after Ser2 of the CTD, NELF and DSIF are phosphorylated, which is carried out by cyclin dependent kinase 9 (CDK9), the kinase domain of the positive transcription elongation factor b (PTEFb). Following phosphorylation NELF detaches and DSIF becomes a positive elongation factor, promoting elongation (Yamaguchi, Shibata, and Handa, 2013. The hyper-phosphorylated form of the CTD is also broader and more rigid enabling RNA processing factors to bind (Vervoort et al., 2022). Phosphorylation of Tryrosine 1 (Tyr1) of the CTD prevents termination factors binding prematurely, phosphatase activity then removes this modification allowing termination factors to bind and promote transcription termination (Heidemann et al., 2013). Further phosphatase activity restores Pol II to a hypophosphorylated state garnering it ready to re-engage in transcription initiation.

Promoter proximal pausing is the final regulatory step before elongation and the importance of this step is highlighted by NELF knock out, which leads to malformation of the inner cell mass and pre-implantation lethality during embryonic development (Williams et al., 2015). PTEFb is essential for regulating pause release, however, exactly how PTEFb activity is regulated remains largely unknown. PTEFb can be sequestered in a complex with the non-coding RNA *RN7SK* and hexamethylene bisacetamide-inducible (HEXIM) protein, which inhibit PTEFb activity (F. X. Chen, Smith, and Shilatifard, 2018). How and why PTEFb dissociates from this complex is not fully understood. Two members of the Serine arginine rich splicing factor (SR) protein family, SRSF1 and SRSF2, also associate with the *RN7SK* complex (Ji et al., 2013). SRSF2 specifically has been shown to promote the dissociation of PTEFb from the complex through affinity association assays (Ji et al., 2013, Lin, Coutinho-Mansfield, et al., 2008). PTEFb is then active and can phosphorylate the CTD of Pol II and other pausing associated factors (Ji et al., 2013, Zaborowska, Egloff, and Murphy, 2016, Lin, Coutinho-Mansfield, et al., 2008).

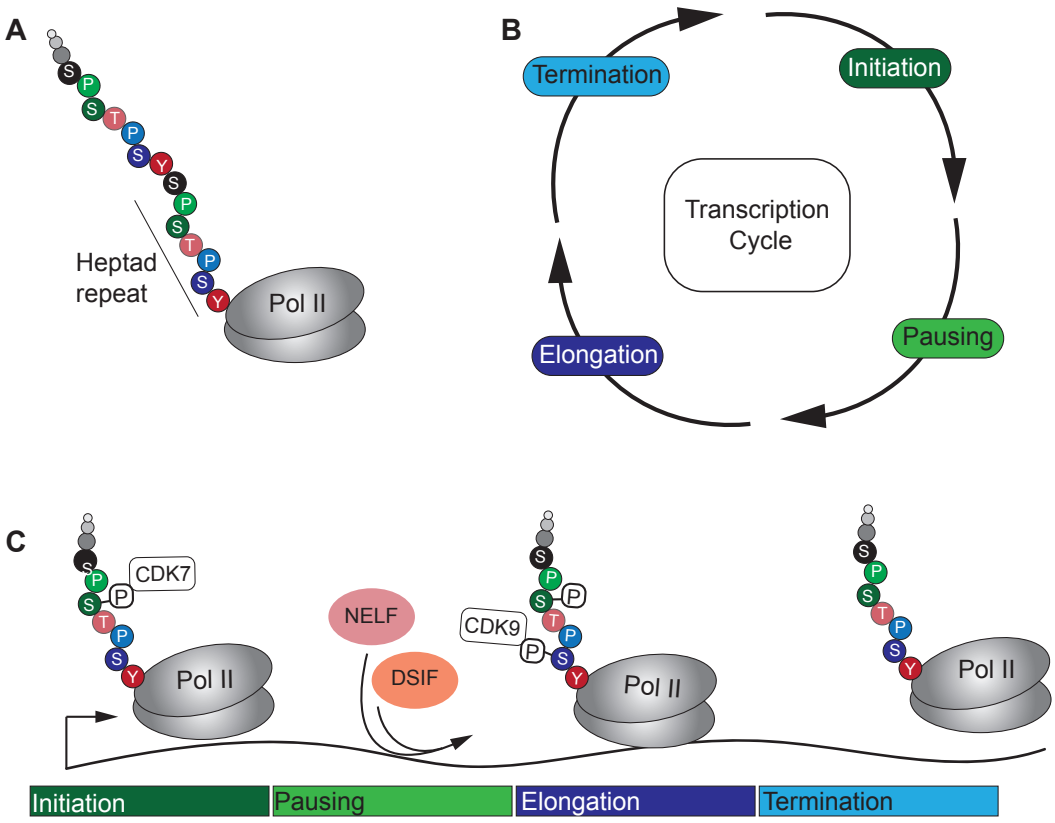

Figure 1.4: **The Transcription Cycle**
A Illustration of RNA polymerase II (Pol II) with the C-terminal domain heptad repeats; tyrosine (Y), Serine (S), Threonine (T), Serine (S), Serine (S). **B-C** Production of mRNA by Pol II happens in a cyclical manner. Transcription inition occurs when Pol II binds a promoter, Ser5 is phosphorylated. However, binding of NELF and DSIF causes pausing which is resolved by phosphorylation and leads to elongation. Termination the requires dephosphorylation. Figure adapted from Vervoort et al., 2022. Pol II-RNA polymerase II, CDK-cyclin dependent kinase, NELF-negative elongation factor, DSIF-DRB sensitivity inducing factor, P-phospho group.

1.4.3 Non-sense mediated decay (NMD)

Gene expression can also be regulated post-transcriptionally through mechanisms that regulate transcript stability, such as RNA modifications and non-sense mediated decay (NMD), or via translation (Huang et al., 2011, Hentze and Izaurralde, 2013, Blanco et al., 2016).

NMD is a regulatory mechanism that acts on gene expression in two ways (Rahman et al., 2020, Hentze and Izaurralde, 2013, Wagner and Frye, 2021). Firstly, it is a surveillance mechanism ensuring that transcripts with errors, such as those containing premature termination codons (PTCs), are not translated by orchestrating transcript degradation. PTCs can arise as a result of splicing, improper RNA editing or DNA mutations, if transcripts with PTCs persist they may have deleterious effects (Huang et al., 2011). Secondly, NMD can also act as a broader mechanism for regulating gene expression post-transcriptionally and was shown to regulate the abundance of various transcripts (Lykke-Andersen, Shu, and J. A. Steitz, 2000).

NMD pathway function depends on the exon junction complex (EJC), which plays a key role in determining the fate of processed mRNAs, not only through

regulating NMD but also in determining mRNA structure, composition, export and initial round of translation (Wagner and Frye, 2021, Maquat, Tarn, and Isken, 2010, V. N. Kim, Kataoka, and Dreyfuss, 2001, Rahman et al., 2020). It is a large complex consisting of four proteins: eIF4A3 the anchor, Y14/MAGOH or Y14/MAGOH B heterodimer and BARENTSZ. The EJC is deposited on 80% of exon-exon junctions as a consequence of splicing (Rahman et al., 2020). It acts as an anchor for the NMD machinery, UPF3B initially binds spliced mRNA by interacting with Y14 and escorts the mRNA-protein complex to the cytoplasm. UPF2 interacts with UPF3B and binds the complex, which is then either bound by the cap-binding complex CBP80-CBP20 or eIF4E undergoes translation which is a pre-requisite for NMD, (Gudikote et al., 2005, Maquat, Tarn, and Isken, 2010). When translating ribosomes encounter stop codons the transient SURF complex, consisting of SMG1, UPF1 and the release factos eRF1 and eRF3 are recruited to the mRNA-protein complex. A PTC is identified by the presence of one or more EJCs downstream and SMG1 then phosphorylates UPF1, which promotes degradation of the mRNA through endo- and exo-nucleases (Gudikote et al., 2005, Maquat, Tarn, and Isken, 2010, Rahman et al., 2020.

1.4.4 Alternative Splicing

Alternative splicing occurs on approximately 95% of human multi-exon genes (M. Chen and Manley, 2009), the extent of alternative splicing highlights that it is a fundamental component of gene expression in higher organisms. Transcriptomic and proteomic diversity contribute to cells ability to carry out specialised functions.

Splicing is largely a co-transcriptional process which can occur at several stages of transcription. Two models have been proposed for how Pol II may regulate splicing. The recruitment model suggests that Pol II's interactions with transcription factors and splicing factors alters their activity, while the kinetic model suggests that the rate of transcription elongation determines the recruitment of the splicing machinery and therefore splicing (Zaborowska, Egloff, and Murphy, 2016, F. X. Chen, Smith, and Shilatifard, 2018).

Splicing is executed by the spliceosome, which is made up of 5 small nuclear ribonuclear protein particles (snRNPs) and auxiliary factors that assemble in a step-wise manner (Figure 1.5) (M. Chen and Manley, 2009). Initially the 5' splice site of an exon is recognised by the snRNP U1. Splicing factor 1 (SF1) binds the branch point of an exon while the U2 auxiliary factor (U2AF) binds the polypyrimidine tract and 3' terminal AG of the splice site in an ATP-independent step to form the E comlex. SF1 is then replaced by the snRNP U2 in an ATP-dependent step forming the A complex. Finally, the tri-snRNP U4/U5-U6 complex leads to formation of the B complex, which is converted to the catalytically active C complex through conformational changes which ultimately leads to splicing (M. Chen and Manley, 2009).

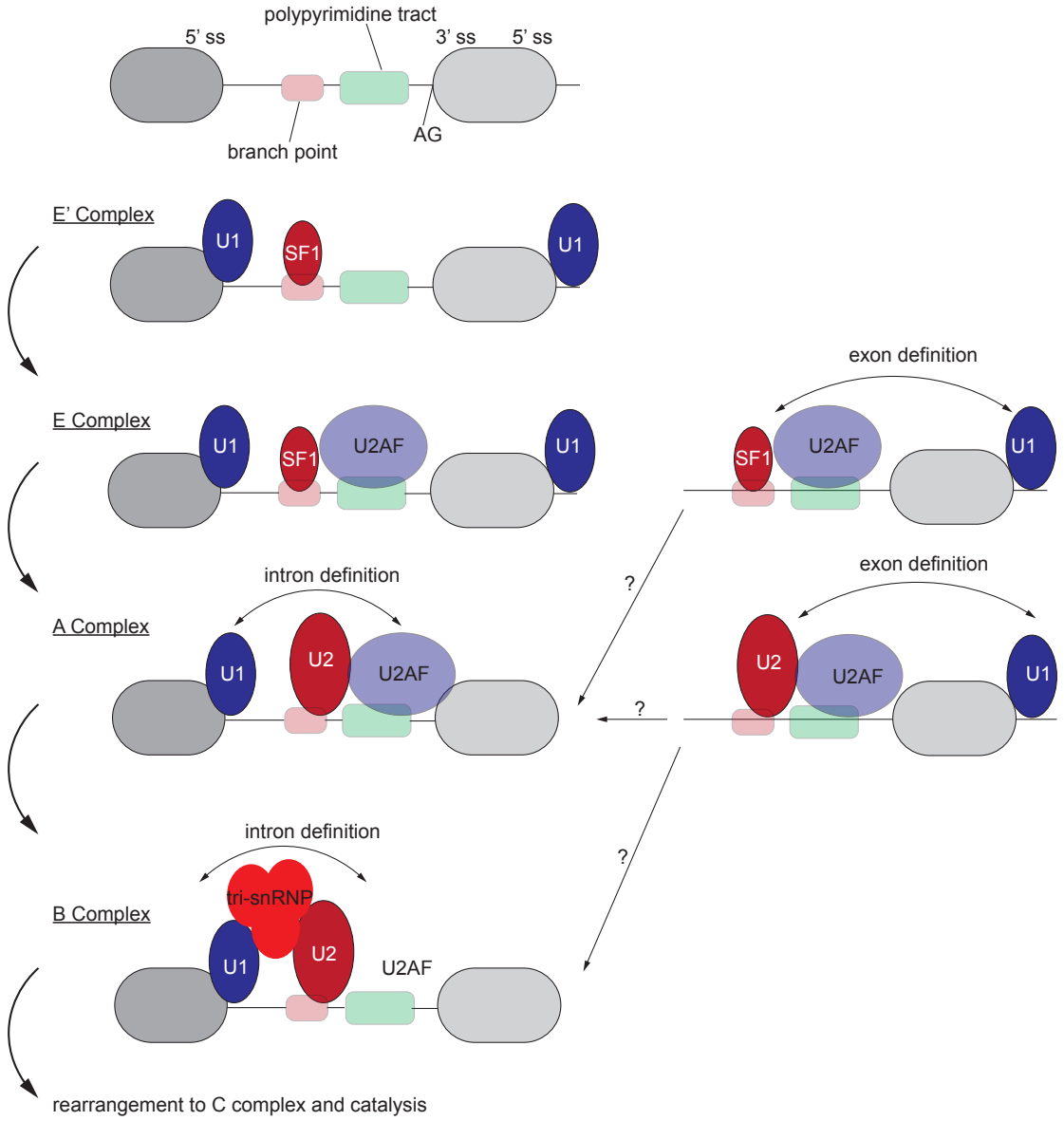

Figure 1.5: **Mechanism of spliceosome assembly**
Spliceosome assembly and splicing occur in a series of steps. snRNP U1 binds to the 5' splice site (5'ss) and splicing factor 1 (SF1) binds the branch point forming the E complex. U2 auxiliary factor (U2AF) binds the polypyrimidine tract and the 3' terminal AG, a dinucleotide required for splice site recognition, SF1 is then replaced by snRNP U2 forming the A complex. The U4/U6-U5 tri-snRNP is then recruited to the complex leading to the formation of the B complex. The catalytically active C complex then forms through extensive remodelling, which enables splicing. Figure adapted from M. Chen and Manley, 2009.

1.5 The Serine-Arginine rich splicing factor family

The SR protein family are a highly conserved family of RNA processing factors (Wagner and Frye, 2021, Longman, Johnstone, and Cáceres, 2000, Rué and Arias, 2015, Califice et al., 2012, Richardson et al., 2011). In humans there are 12 family members which all have a common structure with an RNA recognition motif (RRM)

domain at the N-terminal, required for RNA binding and an RS domain, which is made up of Arginine (R) and Serine (S) dipeptide repeats, at the C-terminal, which is important for protein-protein interactions, these domains are linked by a glycine-arginine rich spacer (Figure 1.6 A). Variable RS domain lengths and additional domains lead to 12 unique family members (Manley and Krainer, 2010, Wagner and Frye, 2021).

The first family member was discovered by two groups simultaneously and eleven further family members were identified over the next decade and named in order of their discovery (Zahler et al., 1992, Krainer, Conway, and Kozak, 1990, H. Ge and Manley, 1990, X. D. Fu and T. Maniatis, 1992, Cowper et al., 2001, J. Wang, Takagaki, and Manley, 1996).

Interestingly, the SR proteins have a single ancient origin and more distantly related proteins cooperate more, while more closely related family members compete more (Califice et al., 2012, Leclair et al., 2020, Wagner and Frye, 2021). However, how this is carried out by different SR proteins in different cell types remains largely unknown.

1.5.1 Canonical functions of SR proteins in splicing

They were first identified as splicing factors and shown to promote the inclusion of weak splice sites by binding exonic splicing enhancer (ESE) sequences and enhancing spliceosome assembly (X. D. Fu and T. Maniatis, 1992, H. Ge and Manley, 1990, Krainer, Conway, and Kozak, 1990). Interestingly, their role in splicing is linked to where they bind a transcript (Howard and Jeremy R. Sanford, 2015, Björk et al., 2009). For example, SR proteins bound to exons often recruit U1 and U2 snRNPs which then leads to splicing, while SR proteins bound to intronic sequences inhibit recruitment of U1 and U2 snRNPs.

SR proteins act as a network regulating their own and each others expression (Leclair et al., 2020, Lareau et al., 2007, Sureau et al., 2001). If two SR proteins bind adjacent exons they compete to recruit snRNPs (J. Han et al., 2011, Pandit et al., 2013). The splicing activity of SR proteins is antagonised by heterogenous nuclear ribonuclear proteins (hnRNPs), these bind intronic sequence silencers (ISSs) and exonic sequence silencers (ESIs) competing with SR proteins for spliceosome activity (M. Chen and Manley, 2009).

There appears to be redundancy in the splicing function of the SR protein family (Long and Caceres, 2009, Howard and Jeremy R. Sanford, 2015, Pandit et al., 2013), however, they are highly conserved and many family members are essential proteins (Zahler et al., 1992, J. Wang, Takagaki, and Manley, 1996, Salamonsen et al., 2001, R. Jia et al., 2010, Jumaa, Wei, and Nielsen, 1999, Karni et al., 2007). This suggests that the precise roles which garner them essential have not been fully elucidated.

1.5.2 Noncanonical functions of SR proteins

Many SR proteins are also important for other cellular processes including transcription, genome integrity, NMD, mRNA export and translation (Figure 1.6 B) (J. R. Sanford, Ellis, and Càceres, 2005, Zhong et al., 2009, Ji et al., 2013, L. Chen

et al., 2018, Lin, Coutinho-Mansfield, et al., 2008, Xiao et al., 2007, Bapat et al., 2018). For example, SRSF3 and SRSF4 have been implicated in regulating transcription termination and SRSF3 has additionally been shown to interact with the exosome and may therefore have a role in mRNA quality control (Wagner and Frye, 2021, Cui et al., 2008).

All SR proteins continuously shuttle between the nucleus and the cytoplasm, with the exception of SRSF2 and SRSF5, which have nuclear retention signals, preventing them from shuttling (Botti et al., 2017, Twyffels, Gueydan, and Sr, 2011). The majority of the SR protein pool has been shown to be nuclear, however, altering transcription rates has been shown to impact the extent of shuttling for some family members (Cáceres, Screaton, and Krainer, 1998).

The SR protein family is also a highly regulated group of proteins and their activity is modulated intricately through several mechanism. Firstly, splicing factor kinases and phosphotases act on SR proteins and phosphorylation status has been linked to function and localisation (Edmond et al., 2013, Aubol et al., 2003, Gout et al., 2012). SR proteins have also been shown to be sequestered by non-coding RNAs, such as *MALAT1*, *VTRNA1.1* and *RN7SK* (Tripathi et al., 2012, Sajini et al., 2019, Ji et al., 2013). *MALAT1* holds SR proteins in nuclear speckles, which are hubs of processes such as transcription and RNA processing. *VTRNA1.1* has been shown to interact specifically with SRSF2. When SRSF2 is bound *VTRNA1.1* cannot be processed into vault-derived regulatory RNAs, which are required for differentiation of epidermal cells (Sajini et al., 2019). Interestingly, the methylation status of VTRNA1.1 influences SRSF2 binding affinity and thus its processing, which suggests that a complex regulatory mechanism is at play. SRSF1 and SRSF2 have been shown to bind *RN7SK* and may be important in regulating transcription (Ji et al., 2013). SR proteins also regulate their own and each others expression levels dynamically by coordinating the inclusion or exclusion of a PTC, when included the resulting transcript will be targetted for NMD (Lareau et al., 2007). All SR protein family members contain a PTC and these are highly conserved suggesting that this is an important means of regulating their own and each others expression (Leclair et al., 2020, Lareau et al., 2007). SRSF3 in particular has been shown to regulate the expression of SRSF2, SRSF5 and SRSF7 (Wagner and Frye, 2021, Anko et al., 2012).

Some SR proteins have been shown to interact with the NMD machinery directly. SRSF1 interacts with UPF1, stabilising its binding to mRNA in an RNA-dependent manner (Aznarez et al., 2018). SRSF2 has also been shown to be important in coordinating NMD, however, unlike SRSF1 it acts in an EJC dependent manner (Rahman et al., 2020).

SRSF1, and SRSF3 have been shown to have roles in translation (Koenigs et al., 2020, Jeremy R Sanford et al., 2004). SRSF1 has been shown to enhance translation of reporter genes and is recruited to stress granules, which house stalled translation pre-initiation complexes (Zappa et al., 2019). By contrast SRSF3 has been shown to repress translation by recruiting mRNA's to P-bodies, organelles where silenced mRNAs are often deposited (Wagner and Frye, 2021, J. Kim et al., 2014). SRSF5, SRSF6 and SRSF7 have been shown to autoregulate their expression levels on a translational level (Koenigs et al., 2020). Overexpression of SRSF7 led to the in-

clusion of a PTC that splits the full transcript into two, containing either the RRM domain or the RS domain. These smaller transcripts are then translated, producing proteins that are not fully functional but can compete with the fully functional SR proteins by binding either RNA or proteins. However, as they are not fully functional they are unable to mediate splicing which means the PTC is no longer included and normal protein levels are restored (Wagner and Frye, 2021, Koenigs et al., 2020).

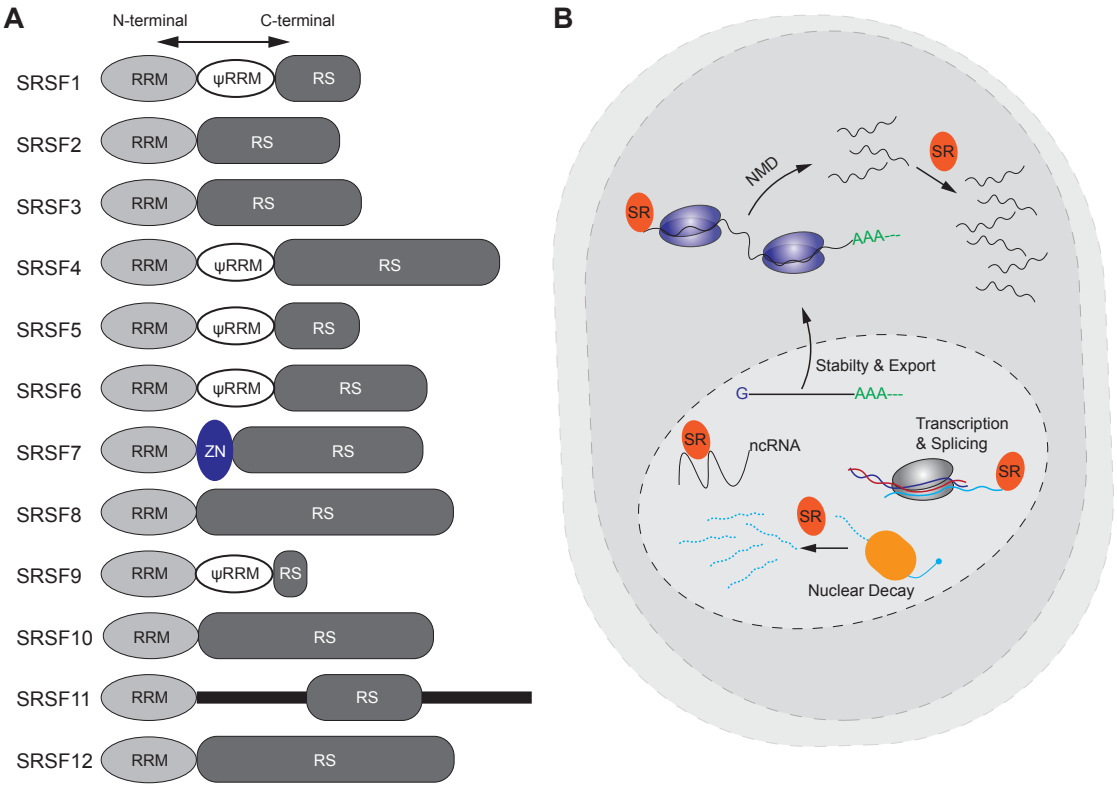

Figure 1.6: **SR protein structure and functions**
A SR proteins have a common structure, with an RRM domain at the N-terminal and a variable RS domain at the C-terminal (black bars are undefined regions). **B** SR proteins function in several steps of RNA processing. All SR proteins act as splicing factors, specific family members have additional roles in transcription, translation, mRNA export and non-sense mediated decay (NMD). Figure adapted from Wagner and Frye, 2021.

1.5.3 SR proteins in development and disease

Several SR proteins have been shown to be essential for normal development in various organisms (J. Wang, Takagaki, and Manley, 1996, Xiao et al., 2007, Longman, Johnstone, and Cáceres, 2000, Jumaa, Wei, and Nielsen, 1999, Salamonsen et al., 2001, H.-Y. Wang et al., 2001). All knock out approaches tested were embyonic lethal in Drosophila and mice, while only SRSF1 deletion was lethal in caenorhabditis elegans. (C. elegans) (Longman, Johnstone, and Cáceres, 2000). However, depletion of the C. elegans SR specific kinase (ceSRPK) was lethal, therefore, proper regulation of SR protein function is also essential for development in C. elegans.

Conditional deletion of SRSF1, SRSF2 and SRSF3 also led to developmental deficits in a variety of tissues (Wagner and Frye, 2021, J. Wang, Takagaki, and Manley, 1996, Cheng et al., 2016, Ding et al., 2004, Xiao et al., 2007). The precise mechanisms through which SR proteins confer their essential functions remains largely unknown. Processes such as differentiation, proliferation and genome stability were altered by SR protein loss in a variety of organisms and cell types (Wagner and Frye, 2021). These are all essential processess, which require tight regulation to maintain homeostasis or enable proper development.

Due to SR proteins first being described as splicing factors studies have largely focused on the alterations to splicing that occur after SR protein loss, however, it is concievable that their additional functions also contribute to their essential nature. SRSF2 and SRSF3 have been linked to maintaining pluripotency in embyronic stem cells (Jumaa, Wei, and Nielsen, 1999, Lu et al., 2014) although SRSF2 was shown to contribute via alternative splicing, SRSF3 was shown to contribute by regulating mRNA export of key pluripotency regulators such as Nanog (Jumaa, Wei, and Nielsen, 1999).

In the brain SRSF3, SRSF6 and SRSF7 have been shown too be required for normal development and health as they are involved in splicing of tau protein (Y. Wang et al., 2005, D. v. Abel et al., 2011), which is important for maintaining microtubule stability in neurons, and has been implicated in Alzheimer's disease (Gao et al., 2007, Y. Wang et al., 2005, Yu, J. Guo, and Zhou, 2004, Fernández-Nogales et al., 2014).

As SR proteins are important for development it is unsurprising that they have also been shown to contribute to disease progression. Most studies have implicated SR proteins in various types of cancer (Wagner and Frye, 2021). SRSF1 and SRSF3 overexpression was sufficient to immortalise fibroblasts (Karni et al., 2007, R. Jia et al., 2010, L. Han et al., 1995, Wagner and Frye, 2021). However, the mechanisms through which they enable this remain unknown.

SRSF1, SRSF2, SRSF3, SRSF4, SRSF5, SRSF6, SRSF7 and SRSF9 have been shown to contribute to cancer progression through both their splicing functions and noncanonical functions (Gout et al., 2012, Y. Fu et al., 2013, Edmond et al., 2013, J. Kim et al., 2014, E. Kim et al., 2015, Wagner and Frye, 2021).

SRSF2 and SRSF3 are overexpressed in head and neck squamous cell carcinoma cell lines (Sharma et al., 2011, André et al., 2017) and SRSF2 is mutated in approximately 30% of myelodysplastic syndromes (MDS), making it one of the most commonly mutated genes (Papaemmanuil et al., 2013). The most frequently observed mutation is a mis-sense mutation that leads to an amino acid substitution at position 95, which is located in the linker region of the protein. As the mutation does not occur in a domain with known function the precise effect of the mutation on protein function remains unknown. However, the mutation has been shown to alter SRSF2's RNA motif recognition, leading to mis-splicing (E. Kim et al., 2015). Specifically mis-splicing of enhancer of zest homolog 2 (EZH2), which is required for differentiation, therefore cells with mutant SRSF2 maintain a more undifferentiated state, which provides cancer cells with an advantage (E. Kim et al., 2015). Interestingly, splicing factors are often mutated in MDS and Chen et al proposed a unifying mechanism by which splicing factor mutants contribute to disease progres-

sion. They showed that R-loop formation was increased when splicing factors were mutated (L. Chen et al., 2018) and this was advantageous for cancer progression as increased DNA damage allows for faster clonal evolution.

SRSF1 is also highly expressed in acute lymphoblastic leukemia (ALL) and enhances proliferation (L. Xu et al., 2018). However, SRSF1 depletion has been shown to cause DNA damage via increased R-loop formation (X. Li and Manley, 2005). This highlights the context dependent and multifaceted nature of SR protein function. The precise mechanisms through which SR proteins contribute to a variety of diseases remains largely unknown. Increasing our understanding of their functions will improve our understanding of a variety of fundamental cellular processes such as proliferation and differentiation and may provide targets for drug development in the future.

1.6 Aims

Skin protects an organism through its barrier function and allows temperature regulation and prevents dehydration, which are all essential functions for life. Maintenance of the skin requires coordinating proliferation and differentiation. If the balance between these is lost it can lead to loss of tissue architecture and loss of the barrier function, both of which may be deleterious for the organism. Furthermore, in case of injury a tissue needs to be able to orchestrate a wound healing response to regenerate the tissue that was lost and ensure the barrier is repaired as quickly as possible and can function fully.

Normal skin turnover has been well studied, however, the underlying molecular mechanisms have not been fully elucidated, here I focused on studying the role of the serine arginine rich splicing factor 2 (SRSF2) in skin homeostasis using mouse back skin and human cell lines as models. I aimed to address the following:

1. Elucidate the role of SRSF2 in skin homeostasis.
2. Further understand the mechanism by which SRSF2 controls transcription in a cell type independent manner.
3. Decipher whether reduced SRSF2 levels can be advantageous during various stages of carcinogenesis.

These questions have not been previously addressed and the answers will provide a better understanding of SRSF2 function.

Chapter 2

Materials and Methods

2.1 *Mouse Procedures*

2.1.1 *Ethical Statement*

All mouse experiments were carried out according to protocols that were approved by the internal welfare committee of the DKFZ (Heidelberg, Germany) and the governmental authority (Regerungspresädium Karlsruhe, G-215/19, G-351/19, G-96/21).

2.1.2 *Mouse Lines*

Mice were housed in pathogen free conditions in individually ventilated cages in the central animal Laboratory at the DKFZ. Mice were fed *ad-libitum* and kept under a 12 hour light/dark cycle. Mice were kept on a mixed background and both genders were used for experiments, unless otherwise specified.

Krt14CreER$^{\mathbf{tam}}$: these mice allow for Cre mediated recombination via tamoxifen under the human keratin14 promoter. Crossing with mice containing loxp sites enables keratinocyte specific deletion. (Jackson Laboratory stock:005107)

Lgr5-eGFP-IRIS-CreER$^{\mathbf{T2}}$: these mice contain a GFP-IRIS-CreER^{T2} cassette in the endogenous Lgr5 translational start site. When crossed with a strain containing loxP sites it generates offspring with tamoxifen inducible deletions in Lgr5 expressing cells. (Jackson Laboratory stock: 008875)

Rosa26-LSL-tdTomato: these mice contain a Cre reporter allele that has a loxP flanked stop cassette which prevents transcription of a CAG promoter driven fluorescent protein expression. This construct has been inserted into the Rosa26 locus. Cre activation leads to expression of the fluorescent protein tdTomato. (Jackson Laboratory stock:007905)

Srsf2$^{\mathbf{fl/fl}}$: These mice have loxp sites flanking exon 1 and 2 of the SRSF2 gene. Cre activation leads to excision of the gene. Jackson Laboratory stock:018019)

Srsf2$^{\mathbf{P95Hfloxed}}$: these mice have loxp sites flanking exon 1 and 2 of the SRSF2 gene, additionally they have a downstream coding region which contains the point mutation. Upon cre activation the endogenous gene is excised and this leads to expression of the mutant. (Jackson Laboratory stock:028376)

NOD.Cg-Prkdc^{scid}Il2rg$_{tm1Wjl}$ISzJ: These mice are immunodeficient and carry mutations on the NOD/ShiLtj; a null allele of the IL2 recepter common gamma chain and the scid mutation in the DNA repair complex protein Prkdc, which renders the mice B and T cell deficient. (jackson Laboratory stock:001976)

2.1.3 *Genotyping*

Genomic DNA Extraction

Mouse ear biopsies were used as a source of genomic DNA. Biopsies were incubated in 25µl 25mM NaOH and 0.2mM EDTA for 15 minutes at 95°C. To neutralise the solution 25µl 40mM Tris-HCl pH 5 was added.

Polymerase Chain Reaction (PCR)

In order to amplify the genomic loci of interest genomic DNA was assayed via polymerase chain reaction (PCR). Primers provided by the Jackson Laboratoy were used and ordered through Sigma Aldrich and are listed in table 2.1. Each reaction consisted of 1µl of extracted genomic DNA, 0.1µM of each primer, 12.5µl GoTaq Green master mix (M7823, Promega) and nuclease free water for a final reaction volume of 25µl. Amplification was performed on a standard PCR cycler (C1000, Touch Thermal Cycler, Biorad).

- The following PCR program was used for Lgr5 and SRSF2$^{\prime\prime P95H}$ and Krt14CreER:
  94°C for 2 minutes, 94°C for 20 seconds, -0.5°C per cycle until 68°C is reached, 68°C for 10 seconds for 9 additional cycles, 94°C for 15 seconds, 60°C for 15°C, 72°C for 10 seconds, 27 cycles from step 6, 72°C 2 minutes.

- PCR program for detecting Rosa26-LSL-tdTomato:
  94°C for 3 minutes, 94°C for 20 seconds, 61°C for 30 seconds, 72°C for 30 seconds, repeat for 34 cycles and final extension at 72°C for 2 minutes.

- PCR program for Srsf2$^{fl/fl}$:
  95°C for 5 minutes, 95°C for 30 seconds, 58°C for 30 seconds, 72°C for 1 minute repeat for 39 cycles, final extension at 72°C for 10 minutes.

Gel Electrophoresis

All PCR products were resolved on a 1.5% (w/v) agarose (3810.3, Roth) in Tris, Acetate, EDTA (TAE) buffer (24710030, Thermo Fisher) gel containing 7.5µl/100 ml gel mix SafeView Nucleic acid stain (NBS-SV5, NBS Biologicals). Gels were run in TAE buffer at 120 volts for 40 minutes using a standard elecrophoresis system (Biorad). PCR products were visualised using the Chemidoc system (Biorad).

2.1.4 *Administration of Tamoxifen*

4-hydroxy tamoxifen (4-OHT) (H6278, Sigma Aldrich) was dissolved in Acetone (11348415, Fisher Scientific) at room temperature to a final concentration of 7.14

Primer Name	**Sequence (5' to 3')**	**Product size (bp)**
Krt14CreER		
Ctr_Fwd	CTAGGCCACAGAATTGAAAGATCT	wt = 324
Ctr_Rvs	GTAGGTGGAAATTCTAGCATCATCC	
Cre_Fwd	CGCATCCCTTTCCAATTTAC	cre = 169
Cre_Rvs	GGGTCCATGGTGATACAAGG	
Lgr5-eGFP-IRIS-CreER^{T2}		
common	CTGCTCTCTGCTCCCAGTCT	
Wt_Rvs	GTCTGGTCAGAATGCCCTTG	wt = 386
Mut_Rvs	CTGAACTTGTGGCCGTTTAC	mut = 119
Rosa26-LSL-tdTomato		
Rosa26_Fwd	AAGGGAGCTGCAGTGGAGTA	rosa26 = 297
Rosa26_Rvs	CCGAAAATCTGTGGGAAGTC	
tom_Fwd	CTGTTCCTGTACGGCATGG	tomato = 196
tom_Rvs	GGCATTAAAGCAGCGTATCC	
Srsf2^{P95H}		
common	TCCAGAAGAAGAGGGAGCAG	
Wt_Rvs	CAGACGTTGGATAAAAACAGCA	wt = 289
Mut_Rvs	ACAGCAACAAAATGGCACA	mut = 200
Srsf2$^{fl/fl}$		
Srsf2_Fwd	GTTATTTGGCCAAGAATCACA	wt = 323
Srsf2_Rvs	TAGCCAGTTGCTTGTTCCAA	ko = 363

Table 2.1: Primers used for genotyping; their sequence and product size

mg/ml and stored at 20°C, in the dark. Prior to use the solution was warmed at 37°C to aid solubility. 100µl of the solution was applied topically to mouse back skin, which had been shaved, and tail skin every other day for up to 28 days.

The dose of the drug had been optimised by previous Frye group members to give sufficient cre-recombinase activation (Gkatza, 2017). All experimental mice were treated.

Tamoxifen (85256, Sigma Aldrich) was dissolved in corn oil (C8267, Sigma Aldrich) and 100% ethanol, vortexed and incubated at 55°C until dissolved. The solution was then aliquoted and stored at -20°C. Mice were given up to three Intraperitanial (IP) injections in five days two weeks prior to further procedures.

2.1.5 *Orthotopic Transplantation*

70,000 cells were injected into the tongue of female NSG mice. Cancer development was measured for 28 days and luciferase signal was measured to monitor tumour growth. Luciferase bioluminescent signal was measured immediately after injection (T0) and at least once a week for the duration of the experiment using the Xenogen IVIS imaging system-100 (Caliper Life Sciences). In order to visualise the

tumour cells mice recieved IP injections of 50µl luciferin (Promega; 5mg/ml in PBS). Isoflurane gas was used to anaesthetise mice during imaing. Data were quantified with Live Image Software v4.4 (Caliper Life Sciences).

2.1.6 *Chemical Carcingogenesis*

Female mice between five and seven weeks old were given three IP injections of 100µl tamoxifen (85256, Sigma Aldrich) in corn oil (C8267, Sigma Aldrich), each of which contained 500µg. A week later mice were shaved and treated with DMBA(D3254, Sigma Aldrich) (400nmol in 200µl), the following week promotion with TPA (P8139, Sigma Aldrich) (10nmol in 100 µl) began three times a week for up to 20 weeks.

2.2 *Histology*

2.2.1 *Tissue Collection*

Mouse back skin was collected after the stated time of topical treatment with 4-OHT. Skin was excised and stretched on blotting paper. Pieces were then cut for either embedding in paraffin or OCT. This allows sagital sections to be cut and stained to analyse skin architecture.

2.2.2 *Embedding*

Paraffin Embedding

Skin pieces were incubated in paraformaldehye (sc-281692, Santa Cruz) (PFA 4% (w/v) in PBS) overnight at room temperature. Pieces were then transferred to 70% (v/v) ethanol and stored at 4°C until embedding. After embedding in paraffin skin was section at 5µm thickness.

OCT Embedding

Skin pieces were frozen in OCT compound (14291, VWR Chemicals) on dry ice and stored at -80°C. Sections were cut between 10 and 100 µm thick and also stored at -80°C.

2.2.3 *Haematoxylin and Eosin staining (H&E)*

Paraffin embedded sections were de-waxed by incubating in Xylene for 2 times 10 minutes. Sections were then rehydrated by incubating them in decreasing concentration of ethanol in water (100% 4 mintutes, 96% 4 minutes, 70% 4 minutes and water 2 minutes. Slides were then placed under running tap water for 2 minutes. Sections were then incubated in haematoxylin solution for 3 minutes, followed by 4 minutes under running tap water. Sections were then dipped in HCl-EtOH solution for 30 seconds. Sections were then placed under running tap water for 5minutes. Sections were then incubated in Eoisin solution for 1 minute, followed by running

tap water for 2.5 minutes. Sections were then dehydrated by incubating them in increasing concentrations of ethanol in water (70% 4 minutes, 96% 4 minutes, 100% 4 mintutes) and finally Xylene for 10 minutes. Slides were then mounted.

2.2.4 *Immunofluorescence (IF)*

Paraffin embedded sections were de-waxed by incubating in Xylene (158692, MP Biomedicals) for 10 minutes. Sections were then rehydrated by incubating them in decreasing concentration of ethanol in water (100% 3 mintutes, 70% 3 minutes, 40% 3 minutes and water 2 minutes. Antigen retrieval was performed by boiling samples for 20 minutes in 10mM sodium citrate buffer pH 6. Slides were then dried and sections circled with a PAP pen (R3777, Thermo Fisher Scientific). Blocking and permiabilisation was then done by incubating sections for 1 hour with blocking solution; 10% FBS and 0.1% tween20 (P1379, Sigma Aldrich). Primary antibodies were diluted in blocking solution and sections were incubated overnight at 4°C. Primary antibodies are listed in table 2.2. The following day slides were washed three times with PBS and incubated with secondary antibody and DAPI for 1 hour at room temperature. There after slides were washed three times with PBS and mounted using fluorescence mounting medium (S302380-2, Agielnt). Frozen sections were briefly thawed and fixed with 4% PFA (sc-281692, Santa Cruz) for 10 minutes. Sections were then permeabilised for 5 minutes in 0.5% triton (T8787, Sigma Aldrich), if necessary. Slides were then washed with PBS three times. Blocking was performed using 10% FBS for one hour. Primary antibodies were then incubated in 1% FBS overnight at 4°C. Sections were then washed three times with PBS and incubated with secondary antibodies, diluted 1:1000 and DAPI, 1:3000, for one hour at room temperature. Slides were then mounted using fluorescence mounting medium (S302380-2, Agielnt).

Antibody	**Dilution**	**sample type**	**Species**	**Provider**	**Cat. Nr.**
Keratin 14	1:1000	paraffin	mouse	Abcam	AB7800
Keratin 10	1:100	paraffin	rabbit	Abcam	AB76318
Loricrin	1:700	paraffin	rabbit	Biolegend	905103
Keratin 6	1:200	paraffin	mouse	Abcam	AB18586
Ki67	1:400	paraffin	rabbit	CST	9129
PCNA	1:500	paraffin	mouse	Abcam	AB29
RFP	1:1000	paraffin	rabbit	Rockland	600-401-379
γH2AX	1:100	paraffin/cells	rabbit	CST	9718
E-cadherin	1:200	paraffin/cells	rabbit	CST	3195
N-cadherin	1:200	cells	mouse	CST	14215S
p63	1:900	paraffin	rabbit	CST	39692S
S9.6	1:200	cells	mouse	Kerafast	ENH001

Table 2.2: Antibodies used for immunofluorescence

2.2.5 *Image Acquisition and Analysis*

Images were acquired using a widefield cell observer (Zeiss). Images were processed using Adobe photoshop and arranged using Adobe illustrator. Image quantifications were done using ImageJ.

2.3 *Cell Culture*

Culturing of mammalian cells was performed using aseptic technique under a laminar flow hood. Cells were maintained at 37°C in humidified incubators with a 5% CO_2 atmosphere. Cells were grown on various tissue culture grade plastic flasks or dishes, depending on the experiment (Fisher Scientific, Nunc and Corning). Medium was changed every two to three days and cells were passaged according to the requirements of each experiment. In order to maintain stocks of the various cell lines used, cells were collected in 1ml aliquots containing cell culture medium supplemented with 10% tissue culture grade dimethyl-sulfoxide (DMSO, D5879; VWR) and frozen slowly using a CoolCell freezing container (Fisher Scientific) at -80°C. Cells were then transferred to liquid nitrogen for long term storage. To thaw cells aliquots were placed in a water bath set to 37°C. Once partly thawed cells were transferred to pre-warmed culture medium in an appropriate dish or flask. The following day the medium was changed.

2.3.1 *Cell Lines*

Primary human keratinocytes were obtained from Promocell after being isolated from neonatal foreskin. Keratinocytes were maintained in Basal 3 medium (C-20021, Promocell) supplemented with 1% penicillin-streptomycin (15140122; Life Technologies). SCC25 cells were obtained from the ATCC and maintained in DMEM/F12 (31330095, Life Technologies) supplemented with 10% FBS (16140071, Life Technologies) and 1% penicillin-streptomycin (15140122; Life Technologies). FaDu cells were obtained from ATCC and maintained in Eagle's minimum essential medium (EMEM, 30-2003; ATCC) supplemented with 10% FBS (16140071, Life Technologies) and 1% penicillin-streptomycin (15140122; Life Technologies).

In order to passage SCC25 and FaDu cells were washed with PBS and incubated with trypsin-EDTA (0.25%) (25200056, Life Technologies) for up to six minutes at 37°C. Trypsin was then quenched using a defined trypsin inhibitor (R007100, Life Technologies) and cells were centrifuged for three minutes at 300g, the supernatant was then removed and cell pelets were diluted to the desired concentration using the appropriate medium. Keratinocytes were passaged using the detachkit (C-41210, promocell). Briefly cells were washed with HEPES BSS and incubated with Trypsin/EDTA for up to six minutes at room temperature. Trypsin was then quenched with trypsin neutralising solution (TNS), cells were centrifuged at 250g for three minutes. The supernatant was then removed and cells were resuspended in medium.

2.4 *Transfection and Transduction*

2.4.1 siRNA transfection

All cell lines were cultured as described previously and treated with either a negative siRNA control (1027281, Qiagen) or siRNA pools specific to SRSF1, SRSF2, SRSF5 or INTS3 (siTools). siRNA was delivered via Lipofectamine RNAiMAX transfection reagent (13778150, Life Technologies), following manufacturer's instructions. Cells were transfected when they were between 30% and 40% confluent. siRNA and lipofectamine were prepared in seperate tubes and diluted in either the appropriate medium or Opti-MEM (31985062, Life Technologies). 0.6µl siRNA (10µM) for 1µl RNAiMAX. The lipofectamine was then mixed with the siRNA and the mixture was incubated at room temperature for 10-15 minutes. The mixture was then added to cells in a dropwise fashion. Cells were incubated with siRNA for 48 to 72 hours before various assays were carried out.

In order to achieve efficient knock down in FaDu cells the concentrations of siRNA and lipofectamine were doubled.

2.4.2 CRISPR Cas9 genome editing

FaDu cells were cultured as previously described. Guide RNAs were designed using Benchling (https://www.benchling.com) and ordered as DNA oligos from Sigma Aldrich. They were then cloned into the pSpCas9(BB)-2A-GFP plasmid (48138; Addgene). Plasmids either containing one of the three guide RNAs or nothing were transfected using Lipofectamine2000 (11668019, Life Technologies) following manufacturer's instructions. 48 hours post transfection cells were collected and single cell sorted. Clones were then grown out and genotyped to select those where targetted genome editing was successful.

2.4.3 Transduction

Clones were transduced with a retroviral vector expressing luciferase-mCherry (pCDH-EF1-Luc2-P2A-tdTomato; 72486; Addgene). Briefly, Pheonix A cells were cultured in DMEM (D6429, Sigma Aldrich) with 10% FBS, 1% non-essential amino acids (11140050, Life Technologies), penicillin-streptomycin (15140122; Life Technologies) and 1% glutamate (35050038, Life Technologies). Pheonix A cells were then transfected with 20µg of plasmid using calcium and HBS when they were approximately 50% confluent. Medium was changed 6 hours after transfection and 48 hours later virus was harvested and clones were transduced using polybrene (TR-1003-G, Sigma Aldrich) at a final concentration of 4µg/ml. This process was repeated the following day. Puromycin selection was started 24 hours later and carried out for three days.

2.5 *Flow Ctyometry*

2.5.1 *Cell Cycle Analysis*

Cell cycle was assayed in all cell lines. Supernatant was collected. Cells were then rinsed with PBS (14190250, Life Technologies) and trypsinised as described above. Cells were then pelleted and resuspended in ice-cold 70% ethanol. Cells were stored at 4°C at least over night. Cells were then spun down at 1300g for three minutes and the supernatant was aspirated. Cells were then resuspended in 500µl of PBS (14190250, Life Technologies) with DAPI (10236276001, Sigma Aldrich) (1:3000). Fluorescence was measured on the flow cytometer (BD FACS Canto, BD Biosceinces) at 450/50 405nm. DAPI fluorescence was visualised as a histogram and cell cycle phases were determined by the peaks.

2.5.2 *Cell Viability Assay*

Cell viability was assayed using the AnnexinV/PI kit (556419, BD Biosceinces), following manufacturer's instructions. Supernatant was collected and cells were rinsed with PBS (14190250, Life Technologies) and trypsinised as described above. After pelleting cells were resuspended in 1x binding buffer and stained with 5µl FITC AnnexinV and 10µl propidium Iodide (PI). Cells were then incubated at room temperature in the dark for 15 minutes before being analysed within an hour. Fluorescence was measured at 530/30 488nm and 610/20 561nm. To analyse the data measurements were visualised as scatter plots and compensation and quadrants were set up based on control samples of unstained cells and cells stained for only FITC AnnexinV or PI.

2.5.3 *Click-it Chemistry*

One hour prior to collection EdU (10mM stock) was added to each sample to a final concentration of 10µM. Supernatant was collected and cells were then rinsed with PBS (14190250, Life Technologies) and harvested as described previously. Cell pellets were resuspended in 1% PFA (sc-281692, Santa Cruz) diluted with PBS (14190250, Life Technologies) and kept on ice in the dark for 15 minutes. Samples were then washed with PBS (14190250, Life Technologies). The Click-it reaction was then carried out according to manufacturer's instructions. Briefly, cells were permeabilised by being incubated in permeabilisation buffer, which consisted of PBS (14190250, Life Technologies) supplemented with 3% FBS (10500064, Life Technologies) and 0.1% Saponin (S7900, Sigma Aldrich) at room temperature for 5 minutes in the dark. The cells were then incubated with a mixture from the Click-iT cell reaction buffer kit (C10269, Life Technologies) and 5µM Alexafluor 555-Azide (A20012, Life Technologies). The cells were then incubated at room temperature in the dark for 30 minutes. After this cells were washed twice in permeabilisation buffer before being resuspended in PBS (14190250, Life Technologies). Fluorescence was then measured at 585/15 561nm and raw fluorescence values extracted for downstream analysis.

2.5.4 Isolation of Keratinocytes from Mouse Back Skin

Keratinocytes were isolated as described in K. B. Jensen, Driskell, and Watt, 2010. Briefly, if necessary back skin was shaved and removed. Subcutaneous fat was scraped off using a scalpel. Back skin was then floated in trypsin without EDTA (25200056, Life Technologies) epidermal side up for two hours at 37^{O}C with gentle shaking. The epidermis was then seperated from the dermis using a scalpel and chopped into small pieces. The cells were collected in Calcium free DMEM (21068028, Life Technologies) and filtered through a 70µm strainer and spun at 800g for five minutes. Cells were then resuspended in PBS (14190250, Life Technologies) supplemented with BSA (1mg/ml) (A7030, Sigma Aldrich).

2.5.5 Data Acquisition and Analysis

Flow cytometry analysis was performed on a FACSCanto analyser (BD Biosceinces). Data was analysed using Flowjo software. All samples were gated using forward versus side scatter to eliminate debris and ensure only single cells were analysed.

2.5.6 Alkaline Comet Assay

Comet assays were performed in collaboration with Dr. Ali Bakr.

250,000 cells were collected in 250 µl of the appropriate medium. 50 µl of cell suspension was mixed with 350 µl of 0.7% low melting agarose (850111, Biozym) and the mixture was spread on comet slides (4250-200-03, R&D systems). Slides were then left on ice until the agarose solidified. Cells were then lysed by placing slides in lysis buffer (2.5M NaCl, 10mM Trizma Base, 100mM EDTA disodium, 1% sodium N-Laurylsarcosin, pH 10) 1% triton and 10% DMSO were added fresh,overnight at 4^{O}C. The following day slides were placed in an electrophoresis chamber and submerged in electrophoresis buffer (24g NaOH, 0.744g disodium-EDTA in 2L distilled H_2O)for 20 minutes before electrophoresis was performed for 20 minutes at 25V and 200mA. Slides were then placed in 100% ethanol for fixation before being air dryed. DNA was stained using SYBR-Green (S7563, Life Technologies) dilutred 1:10000 in TE buffer. Olive tail moment (OTM) was evaluated using fluorescence microscopy at 400x magnification, approximately 100 cells were counted per condition.

2.6 *mRNA Expression Analysis*

2.6.1 *RNA Extraction*

Medium was removed from cells and they were washed with PBS (14190250, Life Technologies) before being lysed in Trizol reagent (15596018, Life Technologies) and incubated at room temperature for 5 minutes. In order to extract RNA 200µl of chloroform was added per ml of Trizol and samples were vortexed for 10 seconds. Samples were then centrifuged at 4°C at 1200g for 15 minutes. The aqueous phase was then transferred to a new tube and placed on ice. 500µl of isopropanol per ml of Trizol was then added. Tubes were inverted several times and then incubated at room temperature for 10 minutes. Samples were then centrifuged at 4°C at 1200g

for 10 minutes. The majority of the supernatant was then removed and RNA pellets were washed with 75% ethanol. Samples were then centrifuged at 4°C at 9000g for 5 minutes. All liquid was then removed and pellets were air dried for up to 15 minutes at room temperature before being resuspended in an appropriate amount of nuclease free water.

2.6.2 *Reverse Transcription*

cDNA synthesis was performed using up to 1µg of RNA. The superscript III reverse transcriptase (18080085, Invitrogen) kit was used according to manufacturer's instructions. RNA was mixed with 0.5µl of random primers (10mM stock) (C1181, Promega) and 1µl of dNTP solution (10mM stock) (N0447L, NEB) to a final volume of 11µl. Samples were then incubated at 65°C for 5 minutes in a standard thermal cycler. 4µl of first strand synthesis buffer, 1µl 0.1M DTT, 1µl RNAse-OUT inhibitor (10777019, Life Technologies) and 1 µl superscript III enzyme were then added and the reverse transcription PCR was run: 25°C for 5 minutes, 50°C for 60 minutes , 70°C for 15 minutes. A further 20µl of nuclease-free water was then added.

2.6.3 *Quantitative PCR (qPCR)*

1µl of cDNA was used for qPCR using either sybr green or taqman reagents. For sybr green reactions 5µl of sybr green master mix was mix with 3µl of nuclease free water and 0.5µl of each primer. For taqman reactions 5 µl of Taqman fast universal PCR master mix (4444557, Life Technologies) was mixed with 3.5µl of nuclease free water and 0.5µl of pre-designed probe sets. For sybrgreen reactions 5µl of Sybrgreen master mix was mixed with 3µl of nuclease free water and 0.5µl of each primer. Probes or primers against 18S were used to normalise samples using the δCT method. Probes and primers used can be found in table 2.3 and 2.4. Data acquisition was carried out using a Quantstudio 3 applied biosystems real-time PCR system (Thermo Fisher).

2.7 *Western Blotting*

2.7.1 *Protein Extraction*

Total protein was isolated from cells using RIPA buffer (50 mM Tris-HCl pH 7.4, 1% NP-40, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate) supplemented with cOmplete™ mini EDTA-free protein inhibitor cocktail tablets (11873580001, Sigma Aldrich) and PhosphoSTOP phosphatase inhibitor cocktail tablets (4906837001, Sigma Aldrich), which were added fresh. Cell scrapers were used to collect the lysates, which were then placed on ice for 5 minutes before being centrifuged at maximum speed at 4°C for 10 minutes. Supernatants were then transferred to a new tube and stored at -80°C. The concentration of lysates was measured via Pierce BCA protein assay kit (23225, Thermo Fisher) according to manufacturer's instructions and measured using the GloMax spectrophotometer.

Target gene	Probe
CDK1	Hs00938777_m1
CDC25	Hs00156411_m1
CDC45	Hs00907337_m1
MCM10	Hs00218560_m1
ITGα6	Hs01041013_m1
KRT10	Hs00166289_m1
TGM1	Hs00165929_m1
TGM2	Hs01096681_m1
INTS3	Hs00224892_m1
SRSF1	HS00199471_m1
SRSF5	Hs00951036_g1
18S	Hs99999901_s1
SRSF2	Mm00448705_m1
18S	Mm03928990_g1

Table 2.3: Taqman probes used to assay gene expression

Target gene	Primer sequence
18S_Fwd	GTAACCCGTTGAACCCCATT
18S_Rvs	CCATCCAATCGGTAGTAGCG
Srsf2_Fwd	CCTAATTTGTGGCCTCCTGA
Srsf2_Rvs	TCAATCTCTTGACAGCTTTAGGC

Table 2.4: Expression primers

2.7.2 *Western Blot*

Protein lysates were mixed with NuPAGE LDS sample buffer (4x) (NP0007, Invitrogen) and NuPage reducing agent (10x) (NP0004, Life Technologies) and denatured for 10 minutes at 70°C. Appropriate pre-cast gels (Biorad) were used and gels were run in 1x SDS running buffer (13.35g Tris-HCl, 18.5g Tris-base, 144.13g Glycine, 100 ml 10% SDS and water to 1L, for 10x buffer) at 120 volts for 80 minutes. After this time proteins were transferred to a nitrocellulose membrane (10600002, GE Healthcare) in 1x transfer buffer (13.35g Tris-HCl,18.5g Tris-base, 144.13g Glycine, and water to 1L, for 10x buffer) supplemented with 20% methanol at 0.2 amper for 2 hours. Membranes were then blocked in 10% (w/v) skimmed milk in TBS-T (24g Tris-HCl, 5.6g Tris-base, 88g NaCl and water to 1L for 10x (1x TBS and 0.1% Tween-20). Membranes were then incubated with primary antibody, according to manufacturer's suggestion overnight at 4°C. Primary antibodies used are listed in table 2.5 The following day membranes were washed three times in TBS-T and then incubated for 1 hour at room temperature with the appropriate horseradish peroxidate (HRP)-conjugated secondary antibody. Membranes were then washed a further three times in TBS-T. Detection was done using using ECL

(1705062, Biorad) and membranes were visualised using a ChemiDoc (Biorad).

Antibody	Dilution	Species	Supplier	Cat. Nr.
SRSF2	1:1000	rabbit	Abcam	AB229473
Pol II total	1:1000	rabbit	CST	14958S
Pol II Ser5	1:1000	rabbit	Abcam	AB5131
Pol II Ser2	1:1000	rabbit	Abcam	AB5095
γH2AX	1:1000	rabbit	CST	9718S
SRSF1	1:1000	rabbit	CST	14908S
SRSF5	1:1000	mouse	abcam	ab67175
INTS3	1:1000	rabbit	Bethyl-Laboratories	A302-050A
HSP90	1:1000	mouse	Santa Cruz	sc-13119
GAPDH	1:1000	mouse	Santa Cruz	sc-47724

Table 2.5: Antibodies used for western blotting

2.8 *Sequencing*

2.8.1 *RNA sequencing*

RNA was extracted using Trizol reagent (15596018, Life Technologies) following manufacturer's instructions. DNase treatment was carried out for 37 minutes at 37°C using TURBO DNAse (AM2239, Life Technologies) and clean up was done via phenol-chloroform extraction. Libraries for total RNA sequencing were prepared using kits for Illumina sequencing. Three to five replicates were sequenced for all samples.

2.8.2 Thiol(SH)-linked alkylation (SLAM) sequencing

4SU labeling and RNA isolation were performed according to manufacturer's instructions (061.24, Lexogen). In brief, FaDu cells were cultured and transfected as previously described. 4SU was added to the culture medium at a concentration of 500 µM for 15, 30 or 60 minutes, after which time cells were harvested with Trizol reagent (15596018, Life Technologies). RNA was extracted under reducing conditions in the dark and quantified via nanodrop. 5 µg of RNA was then treated with iodoacetamide and incubated at 50°C for 15 minutes. RNA was then precipitated and libraries were prepared using the QuantSeq 3' mRNA sequencing library prep kit (015.96, Lexogen).

2.8.3 Cut&Run sequencing

Antibody bound DNA was isolated according to manufacturer's instructions (86652S, Cell Signalling Technologies). Briefly, cells were trypsinised and 100,000 cells were used per replicate. Cells were then mixed with activated concanavalin

A beads. Cell-bead suspension was then mixed with antibody against total RNA polymerase II (2629, Cell Signalling Technologies) and samples were rotated at 4°C overnight. The following day cells were washed and mixed with pAG-MNase for 1 hour at 4°C. Calcium chloride was then added, the mix was then incubated at 4°C for 30 minutes. Following incubation stop buffer was added and samples were incubated at 37°C for 30 minutes. Samples were then centrifuged at 4°C for 2 minutes at 16,000g. The supernatant was kept and DNA was extracted using the DNA purifcation spin columns (14209S, Cell Signalling Technollogies). Libraries were then prepapred using the SimpleChIP ChIP-seq DNA Library Prep Kit for Illumina (56795S, Cell Signalling Technologies).

2.9 *Bioinformatic Analysis*

Bioinformatic analysis were performed by Dr. Anke Heit-Mondryzk, Dr. Thiago Britto-Borges and Etienne Sollier.

2.9.1 RNA sequencing analysis

Data were processed by the DKFZ One Touch Pipeline (OTP), using the RNA-seq workflow version 1.3.0, in combination with the workflow management system Roddy version 3.5.8 or 3.5.9. Briefly, data was aligned to the appropriate reference genome; 1KGRef_PhiX (generated from the 1000 Genomes assembly, based on hs37d5 and including decoy sequences merged with PhiX contigs to be able to align spike in reads) or GRCm38mm10_PhiX (based on GRCm38 merged with PhiX contigs to be able to align spike in reads) using the STAR aligner version 2.5.3a. Duplicate marking was performed using Sambamba version 0.6.5 and quality control was performded using Samtools version 1.6 flagstat as well as rnaseqc version 1.1.8. FeatureCounts of the subread package version 1.5.1 was used for gene specific quantification of reads on the appropriate gene annotation; GENCODE version 19 or GENCODE version M12 in the strand-specific counting mode.

Raw gene count values were then used as input for differential gene expression analysis using DESeq2 version 1.28.1 with R version 4.0.0, which was done using default conditions besides fitType = "local" for each experiment seperately. Results tables were generated for each treatment compared to an appropriate control.

Gene fusions were detected using Arriba version 0.8. The fusions that pass all of Arriba's filters were further filtered to only contain high confidence fusions and from those unique events were counted.

2.9.2 Splicing Analysis

StringTie was run to build de novo transcriptomes for control and SRSF2 depleted conditions in HK, SCC25 and FaDu cells (M. Pertea, G. M. Pertea, et al., 2015, M. Pertea, D. Kim, et al., 2016, Kovaka et al., 2019).

Differential splicing analysis with LeafCutter (v0.2.7) and Majiq (v2.2) was carried out using the Baltica Framework using default parameters. In brief, LeafCutter

differential splicing used a minimum samples per group of two and a minimum samples per intron of two. For Majiq voila tsv the threshold was set to 0.1. Individual introns were examined and significant events were defined as those with an adjusted p-value of 0.05 (Leafcutter) and a probability of non-changing greater than 0.95 (Majiq). Gene ontology for genes with differential exon inclusion was conducted with GOrilla.

2.9.3 SLAM sequncing analysis

Adapter sequences were removed and reads filtered for a minimum length of 21 using cutadapt version 3.4, with the following parameters: -a "AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC" and -m 21.

The "all" option of the SlamDunk package version 0.4.3 was used to align sequencing reads and perform read counting of total transcripts as well as newly transcribed RNAs, using the following parameters: -5 12, -n 100, -c 1, -m, -rl 101, –skip-sam, as reference GRCh38.dna.primary_assembly was used, in combination with a bed file based on Ensembl Genes version 104 that was modified by in-house scripts in conjunction with bedtools merge version 2.29.2 to only include unique, non-overlapping 3'UTRs per annotated gene.

The count values of ReadCounts (total mRNA) and TcReadCounts (newly transcribed mRNA) were then summed up, per gene, for each sample. These were used as input for differential expression analysis using DESeq2 version 1.28.1, within R version 4.0.0. Total mRNA and newly transcribed mRNA were considered seperately and differential expression analysis was then done using default conditions besides fitType = "local". Results were generated for each time point seperately and used to assess gene expression in a quantitative manner.

2.9.4 Bi-directional transcription analysis

The distance to the closest neighbour of protein coding genes in any orientation was calculated as previoulsy described in Bandiera et al. In short protein coding genes were extracted from the GENCODE version 38 annotation file and then used as input for bedtools version 2.24.0 closest command with varying parameters depending on direction.

2.9.5 Cut&Run analysis

FastQ files were processed using the nf-core CUT&RUN pipeline (Philip et al., 2020). Briefly, adaptors were trimmed using TrimGalore and reads were aligned to the human reference genome Hg19 (target) or the yeast reference genome R64-1-1 (spike-in) using Bowtie2. Bedtools were used to generate bedGraph files and bedGraphToBigWig was used to produce BigWig files.

Differentially expressed genes were identified from SLAM sequencing data using DESeq2, genes were filtered to have a mean FPKM$>$ 5 and an adjusted pvalue of 0.05. For each gene its Ensembl canonical transcript was selected. The number of reads aligned a 500 bp window centered on the transcription start site (TSS) and

the number of reads aligned within the gene body (from transcript start -250bp to transcript end +250bp) were counted. A pausing index was then computed as the ratio of reads at the TSS divided by the number of reads in the gene body.

Two methods were used to compute significance of the pausing index between knock down and control. First a t-test was performed per gene for knock down replicates and control replicates. These pvalues were then combined using Fisher's method. Second the log ratio of the mean TSS enrichment for knock down replicates was divided by the log ratio of the mean TSS enrichment for the controls an a t-test was then performed on these log ratios.

2.9.6 SRSF2 eCLIP data analysis

Data for wild type SRSF2 and P95L mutant SRSF2 were obtained from Wheeler et al., 2022. Bed files were downloaded and analyzed with ChIPseeker (v1.37.0). Peak profiling was executed with the plotPeakProf function to transcript bodies, using a relative distance from -20% TSS to 20% the TES. 95% confidence interval was obtained via bootstrapping and the number of bins set to 800. The peak score was taken in account.

2.10 Data Analysis and Statistical tests

Data analysis and plotting were performed with RStudio packages, Python and GraphPad Prism Software. Gene ontology enrichment analysis was performed using GOrilla online tool (Eden, Navon, et al., 2009, Eden, Lipson, et al., 2007) or Enrichr (Xie et al., 2021, Kuleshov et al., 2016, E. Y. Chen et al., 2013). Statistical tests were performed with GraphPad prism and RStudio.

Chapter 3

The role of SRSF2 in skin homeostasis

Analysis of RNA sequencing data in this chapter was assisted by Dr. Anke Heit-Mondrzyk.

3.1 *Aims of this chapter*

SRSF2 is essential for embryonic development (H.-Y. Wang et al., 2001, Xiao et al., 2007). Inducible knock out models have shown that it regulates key cellular processes, such as proliferation, differentiation and genome stability (Bapat et al., 2018, Xiao et al., 2007). These functions have been described in various tissue types, including liver, heart, kidney and blood (Cheng et al., 2016, E. Kim et al., 2015, Ding et al., 2004, Kon et al., 2018, Bapat et al., 2018, H.-Y. Wang et al., 2001, Xiao et al., 2007). However, elucidating the molecular mechanisms that SRSF2 acts through has been hampered by its multifaceted nature as an RNA processing factor, whereby it has roles in splicing, activation of transcription and non-sense mediated decay (X. D. Fu and T. Maniatis, 1992, L. Han et al., 1995, Lin, Coutinho-Mansfield, et al., 2008, Ji et al., 2013, Rahman et al., 2020). Many of the phenotypes that occur as a result of SRSF2 loss have been linked to mis-splicing of specific transcripts (Cheng et al., 2016, Ding et al., 2004, E. Kim et al., 2015, Kon et al., 2018, J. Zhang et al., 2015, H.-Y. Wang et al., 2001, Yoshimi et al., 2019), how SRSF2's additional roles contribute to its function remains largely unknown.

SRSF2 is frequently mutated in various types of cancer, but has been best described in blood cancers (Papaemmanuil et al., 2013, Yoshida et al., 2011, Yoshida et al., 2011, André et al., 2017). The most common mutation is a mis-sense mutation that occurs at position 95 causing a Proline to Histidine (P95H) substitution (Haferlach et al., 2014, Thol et al., 2012, Yoshida et al., 2011 Wu et al., 2013). Despite the frequency with which the mutation occurs the specific consequence of the mutation on protein function remains unclear.

Previous work from the Frye group demonstrated that SRSF2 is required for differentiation in epithelial cells *in vitro* (Sajini et al., 2019) and that conditional SRSF2 deletion has a macroscopic effect on skin (Gkatza, 2017). However, the

molecular mechanisms through which SRSF2 acts in skin have not been investigated. Here, the effects of SRSF2 knockout or knock in of the P95H mutation on skin homeostasis were explored. The specific aims of this chapter were as follows:

1. Determine whether expression of mutant SRSF2 is important for skin homeostasis and determine whether the effects are different to those seen when SRSF2 is deleted. In order to address this a conditional mouse model was established to induce knock in of the P95H mutant in the basal layer of the interfollicular epidermis and conditional deleteion of SRSF2 was also used over a time course of up to 28 days.

2. Characterise the effect of loss of wild type SRSF2 or induction of the P95H mutation on skin architecture. To address this skin architecture and marker expression was examined by immunohistochemistry and immunofluorescence of various markers over the time course.

3. Understand the molecular mechanisms by which SRSF2 acts in the epidermis. This was done by assaying the transcriptomes of skin where SRSF2 was deleted or the P95H mutant was expressed.

3.2 *Results*

3.2.1 *Establishing mouse models to understand the role of SRSF2 in skin*

In order to elucidate the role of SRSF2 in skin I first generated conditional mouse models in which SRSF2 could be deleted or mutated in the basal layer of the IFE (Figure 3.1 A). This was achieved by crossing mice bearing a Cre-recombinase with the tamoxifen responsive hormone-binding domain of the estrogen receptor (Cre-ERtam) under the control of the human keratin 14 promoter, a tdTomato construct under the control of the Rosa26 locus and either the SRSF2 domain containing a pair of loxP sites flanking exon 1 and 2, or an SRSF2 domain consisting of loxP sites flanking exon 1 and 3, which allows expression of the P95H mutant version of SRSF2 (Figure 3.1 B).

Previous work from the Frye group has shown that conditional SRSF2 deletion has a macroscopic effect on the IFE after 28 days whereby skin becomes thickened and flaky (Gkatza, 2017). In order to understand the underlying molecular mechanisms, and avoid secondary effects, I examined the consequences of SRSF2 knockout on skin for up to 12 days (Figure 3.1 C). To elucidate the role of the P95H mutation I used a time course of up to 28 days because the effects on protein function were unknown (Figure 3.1 D).

Firstly, I confirmed successful activation of the Cre-recombinase by staining for the reporter tdTomato (RFP) expression (Figure 3.2 A). Secondly, I assayed SRSF2 mRNA levels in back skin where SRSF2 deletion should have occurred. SRSF2 expression was reduced after 4OHT treatment (Figure 3.2 B). SRSF2 expression was not entirely lost, however, RNA was isolated from total skin, which includes cells where recombination may not have been successful and other cell types found in skin which also express SRSF2. Thirdly, I sorted reporter positve cells from mouse back skin after 10 days of 4OHT treatment and performed RNA sequencing, SRSF2 was the most significantly down regulated gene (Figure 3.2 C). Transcriptomic data generated from back skin of mutant mice confirmed that the P95H mutation was expressed (Figure 3.2 C). RNA was isolated from total skin which explains why the mutant form of SRSF2 is not the only transcript detected. Therefore, I successfully established mouse models to conditionally delete or mutate SRSF2 in the basal layer of the IFE. I next wanted to characterise the effects of SRSF2 loss and the presence of the P95H mutant on the IFE.

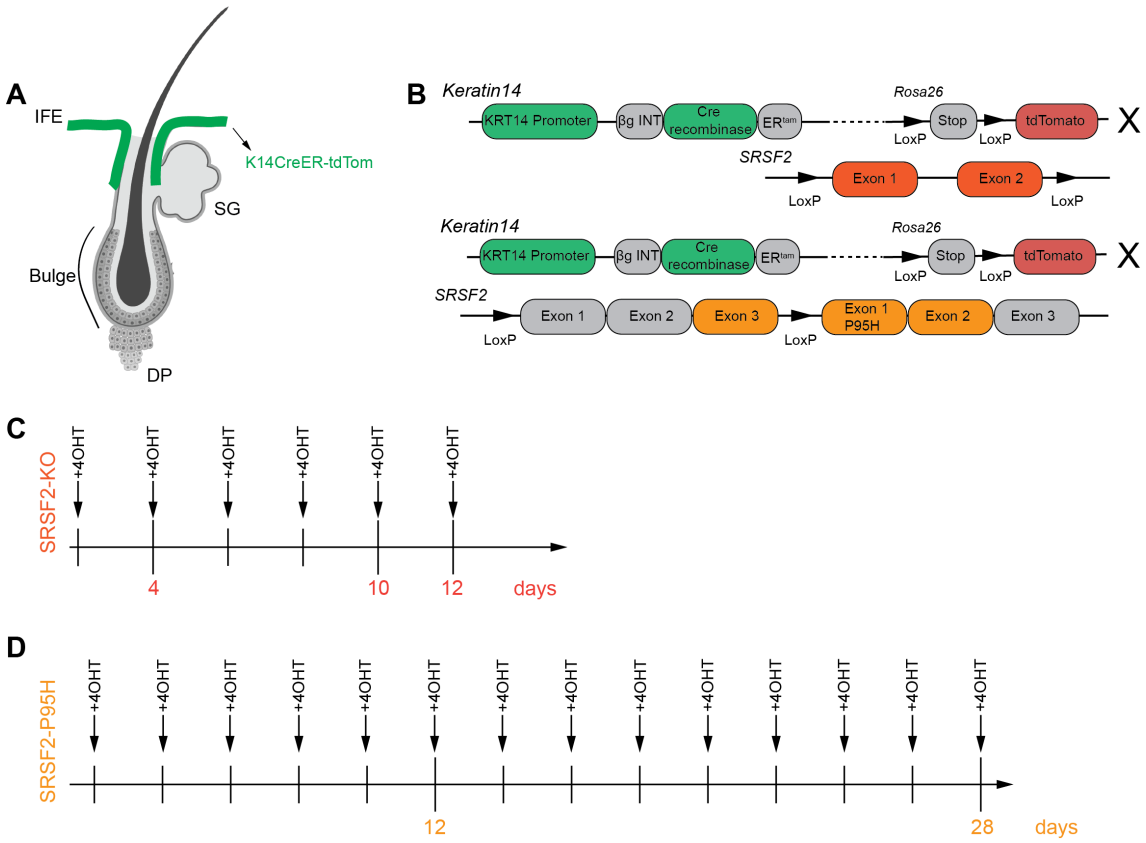

Figure 3.1: **Experimental outline of SRSF2 knockout and knock in of the P95H mutation, in mouse back skin**
A Schematic representation of skin: Green area marks Keratin 14 (KRT14) expressing area, where SRSF2 was deleted or mutated. **B** Schematic representation of transgenes to induce SRSF2 knockout (SRSF2-KO) and knock-in of mutant SRSF2 (SRSF2-P95H). **C** Outline of time course with 4-hydroxy tamoxifen (4OHT) treatment used to induce SRSF2 knockout for up to 12 days. **D** Outline of time course with 4-hydroxy tamoxifen (4OHT) treatment used to induce mutant SRSF2 for up to 28 days. IFE-interfollicular epidermis, SG-sebaceous gland, DP-dermal papilla, 4OHT-4-hydroxy tamoxifen.

3.2.2 *Functional SRSF2 is required for skin homeostasis*

No macroscopic alterations to the skin were observed, therefore, immunohistochemisty was performed on skin sections to decipher the effects of SRSF2 loss and the P95H mutant on the IFE. Haematoxylin and Eosin (H&E) staining revealed that after 12 days of SRSF2 knockout or expression of the P95H mutant the thickness of the epidermis was significantly increased (Figure 3.3 A, B). Surprisingly when the cellularity of the basal layer was probed a decrease of approximately 50% was observed in both knockout and mutant back skin (Figure 3.3 C). An increase in the number of cells in the basal layer may lead to increased thickness as these cells are mitotically active which could lead to an imbalance in proliferation and differentiation thus causing thickening (Cédric Blanpain and Fuchs, 2009). As this was not the case I wondered whether individual cell size was increased. Quantification of the nuclear size of cells in the basal layer revealed that they were larger in back skin lacking SRSF2 or expressing the P95H mutation (Figure 3.3 D). Cells double their contents before dividing into two daughter cells, the increased nuclear size suggests

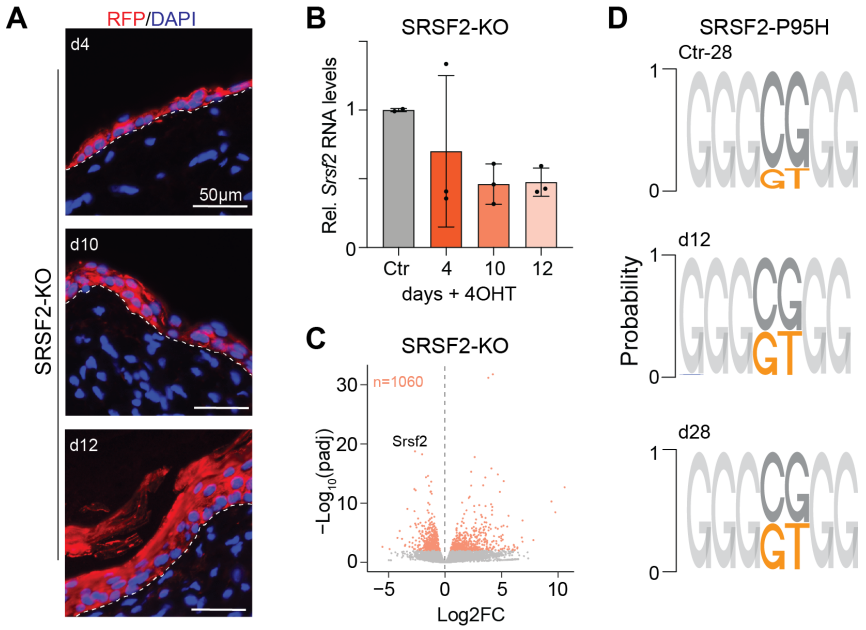

Figure 3.2: **Validation of in vivo Cre-activation in the interfollicular epidermis**
A Immunofluorescence staining for tdTomato (RFP) and counter stained with DAPI in mouse back skin, after 4, 10 and 12 days of 4-hyrdoxy tamoxifen (4OHT) treatment. Dotted line represents the basement membrane. **B** Relative (Rel.) SRSF2 expression in mouse back skin, expression levels normalised to 18S relative to control, each data point represents a biological replicate and the average of three technical replicates. **C** Volcano plot, genes with a padj<0.01 are coloured red, n= number of genes. **D** Frequency of reads that contain the nucleotide substitutions responsible for the P95H point mutation (orange) in total RNA collected from mouse back skin of control (Ctr) and after 12 and 28 days of 4OHT treatment. d-day, 4OHT-4-hydroxy tamoxifen

that cells may be unable to complete division when SRSF2 has been deleted or mutated.

Cell division and differentiation are tightly linked in the IFE (Ya and Fuchs, 2022, Cédric Blanpain and Fuchs, 2006, Cédric Blanpain and Fuchs, 2009). In order to undergo differentiation cells must leave the basal layer, once they have exited they have a limited capacity to divide and undergo terminal differentiation (Cockburn et al., 2022). Therefore, I wondered whether differentiation was affected. I co-stained for the markers; keratin (KRT) 14 (KRT14), which marks the undifferentiated cells of the basal layer, and KRT10 a marker of early differentiation, or Loricrin (LOR), a terminal differentiation marker. The skin of control mice showed normal architecture, with KRT14 expression restricted to the basal layer and KRT10 and LOR marking suprabasal cells. These markers were mutually exclusive. However, when SRSF2 was knocked out or the P95H mutant was expressed a population of cells expressing both undifferentiated and differentiated markers emerged (Figure 3.4 A, B).

The cytoplasmic area of basal cells was larger when SRSF2 was knocked out or mutated, quantification was done by measuring KRT14 as this is an intermediate filament so can be used to cell size. In SRSF2 knock out or mutant cells cell size was approximately double that of control cells (Figure 3.4 C). This further suggests that cells are unable to complete cell division and suggests that they arrest during

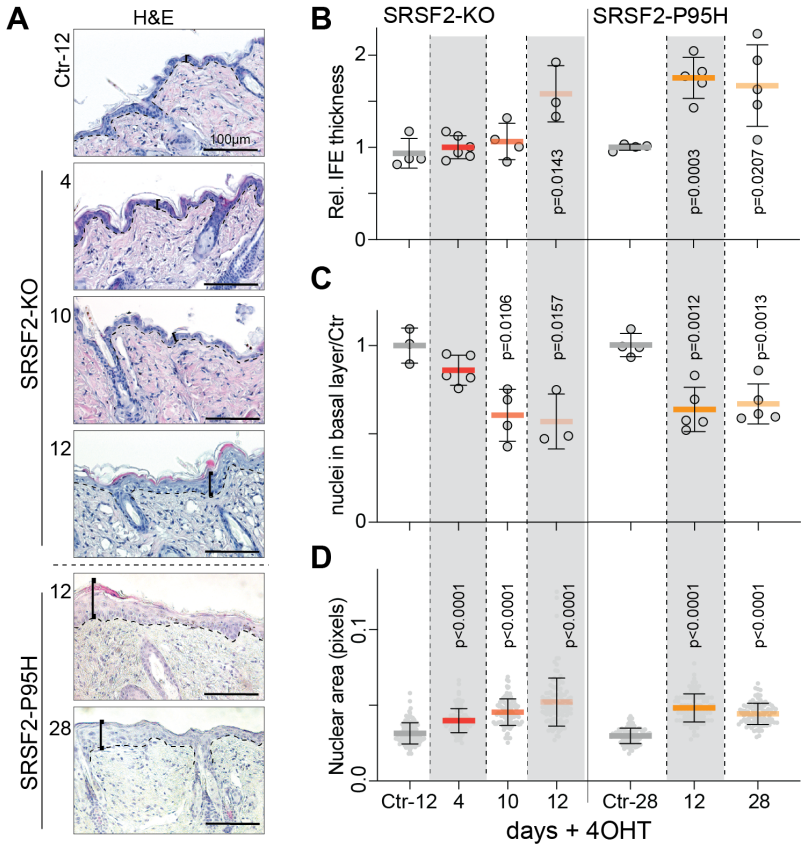

Figure 3.3: **SRSF2 is required for maintenance of the interfollcular epidermis**
A Haematoxylin and Eosin (H&E) staining of mouse back skin in control (Ctr) and after SRSF2 deletion (SRSF2-KO) for 4, 10 and 12 days, or induction of the point mutation (SRSF2-P95H) for 12 and 28 days. Bracket indicates interfollicular (IFE) thickness. **B** Quantification of IFE thickness, normalised to control, each point represents one biological replicate and the average of three measurements from each of five seperate images. **C** Quantification of basal layer cellularity, normalised to control, each point represents one biological replicate and the average of three areas from each of five seperate images. **D** Quantification of nuclear area in the basal layer of the IFE, each point is one cell from three to five biological replicates, a minimum of 100 cells were measured per condition. (**B-D** Shown is mean± SD, p= unpaired t-test comparing to the appropriate control. IFE-interfollicular epidermis, 4OHT-4-hydroxy tamoxifen

G2 or M phase, when both the DNA and cellular contents for two daughter cells is present.

In order to rule out that proliferation was driving the increase in IFE thickness I stained for Ki67, which marks cells in S phase (Figure 3.5 A). This confirmed that loss of functional SRSF2 does not trigger proliferation as thickened areas were negative for Ki67. Interestingly, cells of the hair follicle were sometimes positive for Ki67, which was reminiscent of a wound healing response, when the HF is triggered to contribute to maintenance of the IFE (Y. Ge et al., 2017). To test this I stained for Keratin 6 (KRT6) (Figure 3.5 A). This is a marker of hyperproliferation and expressed upon wound healing (Wojcik, Bundman, and Roop, 2000). Fascinatingly, KRT6 expression could be obsereved four days after the first treatment in knockout mice and was also expressed in back skin of mice expressing the P95H mutation (Figure 3.5 A. The loss of functional SRSF2 leads to a loss of cellularity and this

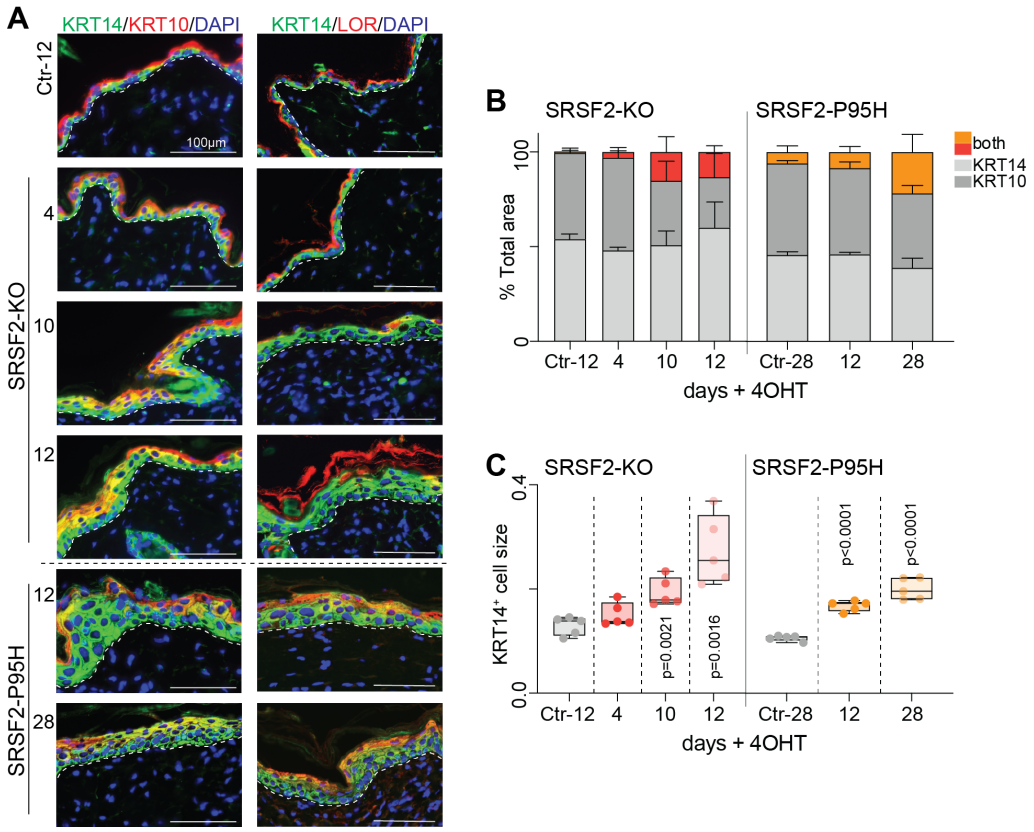

Figure 3.4: **Functional SRSF2 is required for differentiation**
A Immunofluorescence staining of keratin 14 (KRT14) and keratin 10 (KRT10) (left panel) or KRT14 and Loricrin (LOR) (right panel) in control (Ctr) and after SRSF2 deletion (SRSF2-KO) for 4, 10 and 12 days, or induction of the point mutation (SRSF2-P95H) for 12 and 28 days. **B** Quantification of KRT14 and KRT10 positive cells, measurements are the average from five images and a minimum of three biological replicates. **C** Quantification of KRT14 positive cell size, each point represents one biological replicate and the average of minimum 150 cells. (**B, D**) Box plot shows minimum, first quartile, median, third quartile, and maximum (**C**) p= unpaired t-test. KRT-Keratin, LOR-Loricrin, 4OHT-4-hydroxy tamoxifen

may trigger a wound healing response.

The keratin family has 30 members, of which 18 are expressed in skin (Moll, Divo, and Langbein, 2008). They form specific heterodimers and are important for cell adhesion. KRT14 forms a heterodimer with KRT5, while KRT10 dimerises with KRT1. In order to explore the changes to expression patterns of the various keratins known to be important for IFE differentiation or response to stress I analyzed transcriptomics data from skin where SRSF2 was deleted or mutated. This revealed that after 10 days of treatment in the knockout expression of KRT10 and KRT1 are down regulated, and are replaced by KRT6a and KRT6b, which is known to occur during wound healing (Figure 3.5 B) (Navarro, Casatorres, and Jorcano, 1995, Wojcik, Bundman, and Roop, 2000). KRT14 expression was also increased, which reflects the differentiation block. In the P95H mutant KRT6a and KRT6b were also upregulated. However, after 28 days of treatment the increase was weaker, suggesting that a new equilibrium may be established.

My findings suggest that the P95H mutant is a loss of function mutation and that

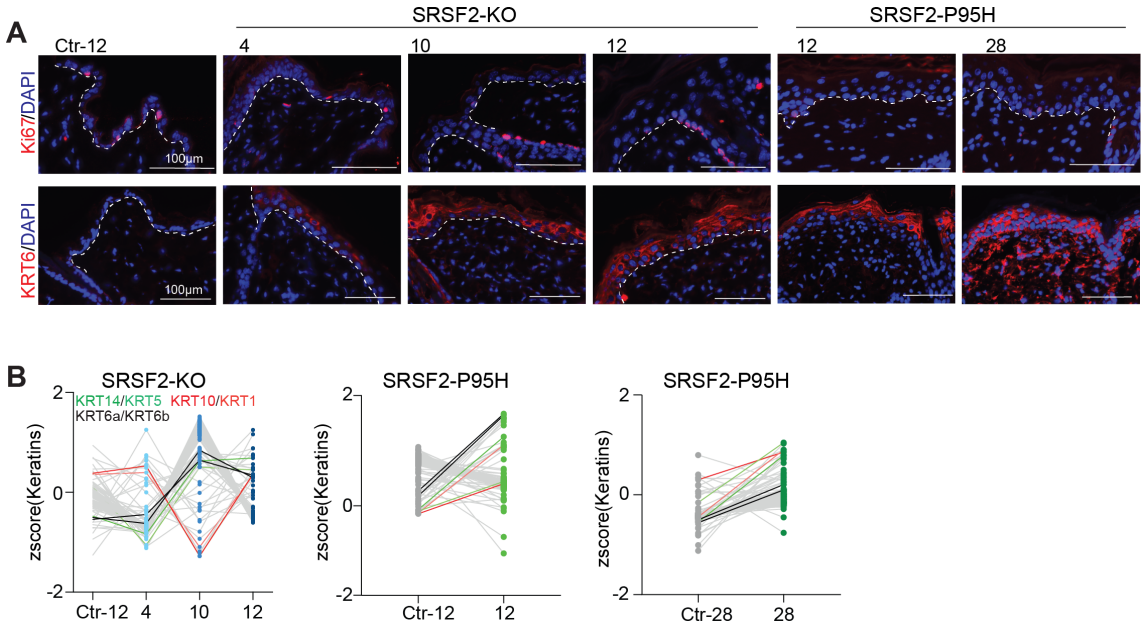

Figure 3.5: **Loss of functional SRSF2 invokes a wound healing response**
A Immunofluorecence of Ki67 (top) and KRT6 (bottom) in mouse epidermis; control (Ctr) and after SRSF2 deletion (SRSF2-KO) for 4, 10 and twelve 12 days, or induction of the point mutation (SRSF2-P95H) for 12 and 28 days, dotted line represents the basement membrane. **B** zscores of keratin expression in mouse back skin, control (Ctr) and with either SRSF2 knockout (SRSF2-KO; left) or mutation (SRSF2-P95H; right). KRT14 and KRT5 are highlighted in green, KRT10 and KRT1 are highlighted in red and KRT6a and b are highlighted in black.

functional SRSF2 is essential for maintaining homeostasis of the IFE. However, the underlying mechanisms through which SRSF2 acts remain unknown. To explore this I began by analzying SRSF2 expression in a published sinlge cell atlas of mouse skin (Joost et al., 2016). This revealed that SRSF2 expression was significantly changed across differentaiton and its expression increases as cells begin to differentiate (Figure 3.6 A). As SRSF2 has known roles in regulating splicing and transcription I next explored whether these processes were also changed across differentiation (Figure 3.6 B). This revealed that both transcription and splicing change significantly across differentiation, cornification was included as a positive control (Figure 3.6 B).

I next tested whether the transcriptome was affected by loss of SRSF2 using RNA sequencing data I generated from reporter positive cells sorted from mouse back skin (Figure 3.6 C). Gene ontology analysis revealed that genes involved in regulation of transcription were amongst those most affected by loss of functional SRSF2 (Figure 3.6 D). The appearance of terms associated with histone modifications was also interesting and suggests that SRSF2 deletion may lead to alterations to the entire gene regulatory landscape. Finally, terms related to cell death and differentiation were also down regulated, which is in line with work in other tissues and reflects the loss in cellularity and differentiation block I observed (E. Kim et al., 2015, Kon et al., 2018, Cheng et al., 2016, Ding et al., 2004).

These findings suggest that SRSF2's role in transcription may be important in regulating homeostasis of the IFE. As SRSF2 has been very well characterised as a splicing factor it was surprising that splicing was not amongst the top ontological

processes affected by loss of SRSF2. However, SRSF2's splicing activity is promiscuous and it is known that there is redundancy in splicing function between SR protein family members (Pandit et al., 2013, Leclair et al., 2020, Lareau et al., 2007), therefore, SRSF2 may not be essential for splicing in the IFE. Furthermore, if alterations to splicing were driving the phenotype loss of SRSF2 or induction of the P95H mutation should have had different consequences on the skin as the point mutation has been shown to alter SRSF2's RNA motif recognition (E. Kim et al., 2015, Kon et al., 2018). This was not the case and the P95H mutant phenocopied SRSF2 loss, interestingly it has also been shown to alter transcription in a manner comparable to when SRSF2 is absent (Yoshimi et al., 2019). A common function which is lost in both cases must be driving the alterations to the IFE, suggesting that SRSF2's role in transcription is its essential function.

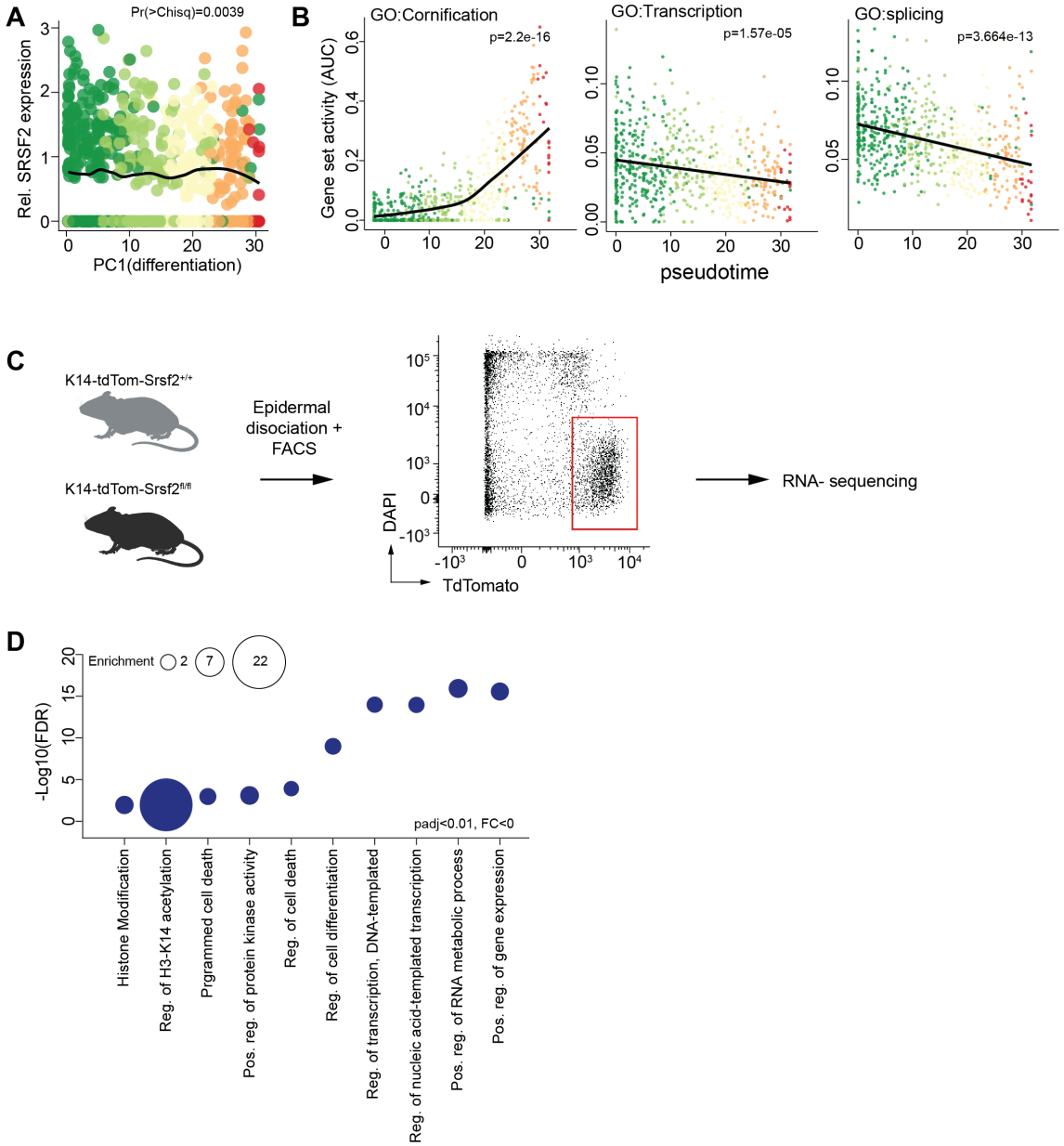

Figure 3.6: **SRSF2 expression changes across differentiation**
A Relative SRSF2 expression across differentiation, each point represents one cell. Colours indicate differentiation status, green is undifferentiated and red is differentiated, pr=Chi squared (Chisq). **B** Gene set activity calculated by AUCell (Aibar et al., 2017) across psuedotime, focusing on Cornification (left), transcription (middle) and splicing (right), p= linear model **C** Scheme of work flow and gating stratergy for sorting tdTomato positive cells from mouse back skin. **D** Gene ontology analysis of significantly down regulated genes when SRSF2 was deleted compared to reporter positive control cells, using GOrilla GO process Eden, Lipson, et al., 2007, Eden, Navon, et al., 2009.
Reg.-regulation, Pos.-positive, GO-gene ontology.

3.2.3 *Summary, discussion and future directions*

The results presented in this chapter show that the P95H mutation leads to a loss of function and that functional SRSF2 is required for maintaining homeostasis in the IFE. In the absence of functional SRSF2 there was a loss of cellularity accom-

panied by an increase in the cytoplasmic and nuclear size of the remaining cells in the basal layer, which caused an increase in the overall thickness of the IFE. Differentiation was also affected; when functional SRSF2 was absent a population of cells expressing both undifferentiated and differentiated markers emerged. These markers are usually mutually exclusive, however, when skin is forced to divide rapidly this double positive population has been found and shown to represent a population of cells with limited proliferative capacity (Aragona et al., 2020, (Cohen et al., 2022, S. Wang et al., 2020). This population may emerge when functional SRSF2 is lost in order to maintain the skin barrier, which is in line with the observation that loss of functional SRSF2 triggered a wound healing response in the skin.

Using a single cell atlas from mouse skin (Joost et al., 2016) revealed that SRSF2 is upregulated during differentiation in the IFE, which suggests that SRSF2 may be required for execution of the differentiation program and hence the lack of functional SRSF2 leads to a block in differentiation. This is in line with a previous study which showed that SRSF2 is required for proper differentiation of epidermal cells *in vitro* (Sajini et al., 2019). Further work is required to fully elucidate the molecular mechanisms SRSF2 acts through to coordinate differentiation.

Total RNA sequencing of cells sorted from mouse back skin revealed that terms related to regulation of Pol II, differentiation and cell death were down regulated when SRSF2 was absent. This suggests that SRSF2 is required for proper regulation of Pol II in the IFE and that this mis-regulation impacts differentiation and viability. Differentiation status and proliferative capacity are tightly linked in the IFE, when a cell in the basal layer divides crowding forces cells out of the basal layer. Increasing calcium concentrations and loss of contact with the basement membrane pushes cells to stop dividing and differentiate. Several mechanisms drive cells to differentiate when they leave the basal layer of the IFE, which makes the differentiation block that occurs when functional SRSF2 is lost surprising. Interestingly, in myelodysplastic syndromes (MDS) the reason for the high frequency of SRSF2 mutations has also been linked to it contributing to a differentiation block via missplicng of enhancer of zeste homolog 2 (EZH2), which causes it to be degraded. EZH2 is part of the polycomb repressive 2 (PRC2) complex and important for coordinating differenetiation (Kon et al., 2018, Bapat et al., 2018). This is likely to be an over simplification, however, it suggests that SRSF2 plays a role in differentiation in several cell types, although the underlying molecular mechanisms remain a pressing question for future work.

In order to address the molecular mechanisms through which SRSF2 coordinates differentiation single cell transcriptomics could be performed. This will allow the position along the differentiation trajectory at which the block occurs to be fully elucidated. This would also enable the question of whether cells arrest in a specific stage of the cell cycle to be addressed.

SRSF2 has also been implicated in regulating transcription (Lin, Coutinho-Mansfield, et al., 2008, Ji et al., 2013) and how changes to transcription rates directly contribute to determining cell states, as has been shown with translation (Blanco et al., 2016), remains unknown. SRSF2 loss or mutation has been shown to cause proliferation and differentiation defects in other tissues (Bapat et al., 2018, Cheng et al., 2016, Kon et al., 2018, H.-Y. Wang et al., 2001, Xiao et al., 2007). However,

these studies focused on SRSF2's role in splicing, it would be interesting to remove functional SRSF2 in other stem cell compartments in the skin, or other epithelial tissues, such as intestine, and see whether similar phenotypes are observed with a focus on identifying cell type independent mechanisms that SRSF2 acts through. This would help to elucidate specific functions of SRSF2 and also to characterise a broader functional role for SRSF2.

Splicing is also impacted by transcription rates and is important in regulating gene expression (Bentley, Guthrie, and J. Steitz, 2005, Moore and Proudfoot, 2009, E. T. Wang et al., 2008) and the role splicing plays in differentiation is also largely unknown. It is possible that there are various isoforms of keratin proteins which have different RNA stabilities and perhaps loss of functional SRSF2 contributes to isoforms that are more stable, thus giving rise to cells that express both KRT14 and KRT10 protein. For this question having both a knockout and the P95H mutant will be useful as it allows changes to splicing that occur as a result of having no SRSF2 and those that occur due to the mutation to be disentangled. This could be done using long read sequencing which is enabling alterations to splicing to be addressed in a more quantitative manner as whole transcripts can now be sequenced.

Chapter 4

SRSF2 is essential for cell division and maintenance of the SR protein network

Parts of the work presented in this chapter have been published in a similar form in Nature Communications (Sajini et al., 2019). Some of the experiments done with SCC25 cells were done in collaboration with Leonie Arnetzl a Master's student whom I supervised. Analysis of RNA-sequencing data in this chapter was assisted by Dr. Anke Heit-Mondrzyk.

4.1 *Aims of this chapter*

In the previous chapter I showed that functional SRSF2 is required for proliferation and differentiation in skin. SRSF2 has also been shown to be required for proliferation and differentiation in a range of other organs, including blood, liver and heart (Kon et al., 2018, Cheng et al., 2016, Bapat et al., 2018, Ding et al., 2004). However, the underlying molecular mechanisms remain largely unknown.

SR proteins were first identified as splicing factors (Krainer, Conway, and Kozak, 1990, H. Ge and Manley, 1990, Long and Caceres, 2009, Howard and Jeremy R. Sanford, 2015, Björk et al., 2009, J. Han et al., 2011, Pandit et al., 2013, Zahler et al., 1992, Richardson et al., 2011, Graveley and Tom Maniatis, 1998). They have been shown to function as a network, in which they regulate each others expression and also compete and collaborate to carryout splicing (Lareau et al., 2007, Leclair et al., 2020). Interestingly, the proportions of different family members vary depending on the cell type and the cell state (Leclair et al., 2020). However, how this contributes to cell identity is not well understood.

The SR proteins have very well described functions in splicing, but are RNA processing factors and contribute to all aspects of RNA metabolism, including mRNA export, NMD and transcription (Ji et al., 2013, Lin, Coutinho-Mansfield, et al., 2008, Rahman et al., 2020, Bapat et al., 2018, Xiao et al., 2007, Tripathi et al., 2012, Ratnadiwakara et al., 2018, Reed and Hurt, 2002, Cáceres, Screaton, and Krainer, 1998, Cazalla et al., 2002, Botti et al., 2017, Aznarez et al., 2018, Singh

et al., 2012, Hir et al., 2001, Sureau et al., 2001, J. R. Sanford, Ellis, and Càceres, 2005, Zhong et al., 2009).

Previous work from the Frye lab has shown that SRSF2 mediates differentiation of epidermal cells *in vitro* by altering the cleavage of the vault RNA *VTRNA1.1* (Sajini et al., 2019). However, how SRSF2's role in transcription and splicing influences epithelial cell identity and function remains unknown. The aims of this chapter were as follows:

1. Establish SRSF2 knock down in various epithelial cell lines. To achieve this three cell lines will be transfected with an siRNA pool containing 20 different siRNAs targetting SRSF2. Transfection efficiency will then be assayed by measuring mRNA and protein levels, via qPCR and western blot.

2. Characterise the cell type independent effects of SRSF2 knock down. Differentiation, proliferation and viability will be assayed through a combination of qPCR and flow cytometry based assays over a time course following SRSF2 knock down in three epithelial cell lines

3. Explore whether the SR protein network is altered by depletion of one family member. This will be achieved by generating whole transcriptome data and analysing how SR protein family members expression changes after SRSF2 depletion.

4.2 *Results*

4.2.1 *SRSF2 expression is highest in the most proliferative cell line*

In order to further understand how SRSF2 regulates differentiation and cell cycle progression I made use of three epithelial cell lines; a primary human keratinocyte cell line (HK), a well established *in vitro* model for the IFE (Rheinwald and Green, 1975), and two squamous cell carcinoma (SCC) cell lines; SCC25 and FaDu cells (Figure 4.1 A).

SRSF2 knock out is lethal (H.-Y. Wang et al., 2001), therefore I used an siRNA stratergy to deplete SRSF2 for up to 72 hours (Figure 4.1 B). To further characterise the chosen cell lines I assayed SRSF2 mRNA levels (Figure 4.1 C). HK and SCC25 cells express similar levels of SRSF2, while FaDu cells, which are more proliferative, express higher levels of SRSF2.

Knock down was efficient in all cell lines, leading to a reduction in both mRNA and protein levels (Figure 4.1 D-G).

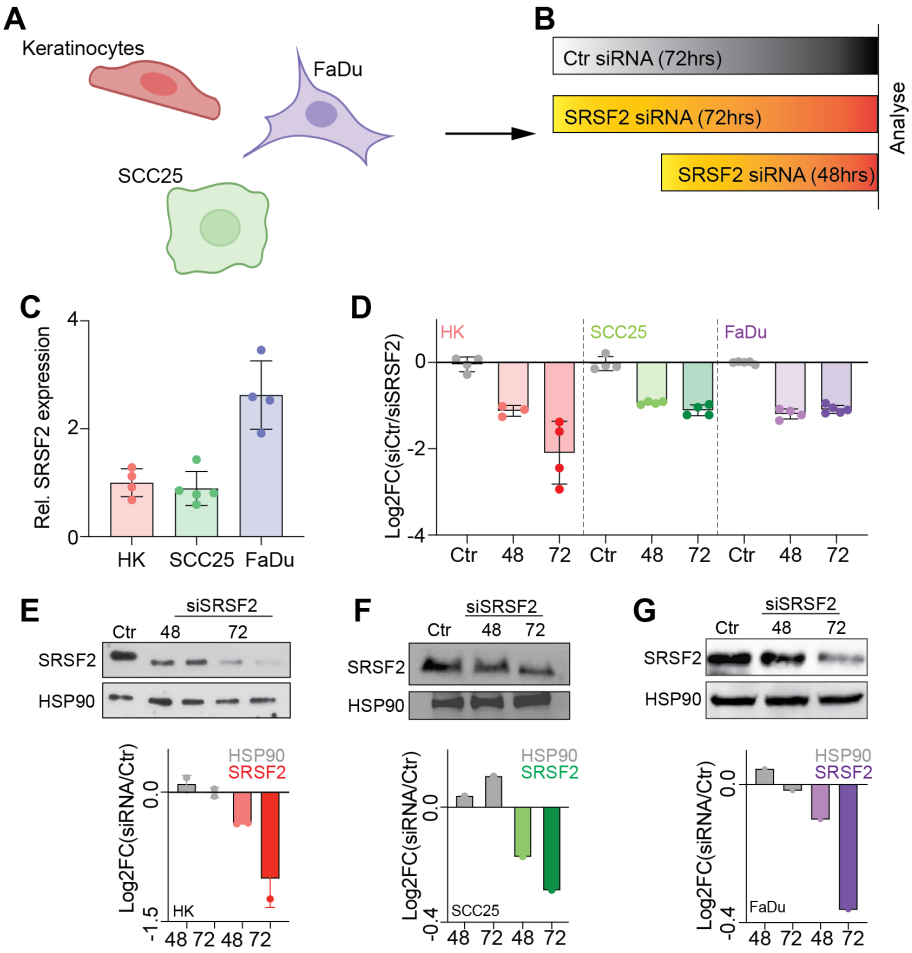

Figure 4.1: **Establishing SRSF2 knock down in epithelial cell lines**
A Schematic representation of the cell lines used for *in vitro* experiments.**B** Illustration of the experimental design used to deplete SRSF2 for 48 or 72 hours.**C** Log2 fold change (Log2FC) of SRSF2 expression in keratinocytes (HK), SCC25 and FaDu cells, each point represents one replicate and the average of three technical replicates.**D** SRSF2 knock down for 48 and 72 hours, each point represents one transfection and the sum of three technical replicates.(**E**-**G**) SRSF2 protein levels (top) and quantification of band intensity, relative to control (bottom) in **E** keratinocytes (HK), **F** SCC25 and **G** FaDu cells, loading control=HSP90.

4.2.2 *SRSF2 knock down alters epithelial cell adherence*

After 72 hours of SRSF2 depletion HK and FaDu cells were no longer adherant, this was not the case in SCC25 cells (Figure 4.2 A). A loss of adherence was unexpected as epidermal cells usually differentiate in response to stress, which is associated with cells becoming rounder and flatter and adhering strongly to the surface they are on.

That cells do not differentiate upon SRSF2 depletion is in line with the *in vivo* findings outlined in the previous chapter, along with studies in other tissues (Kon et al., 2018, H.-Y. Wang et al., 2001, Cheng et al., 2016, Xiao et al., 2007). In the mouse epidermis I also observed that cells increased in size, which I thought may be indicative of cell cycle arrest.

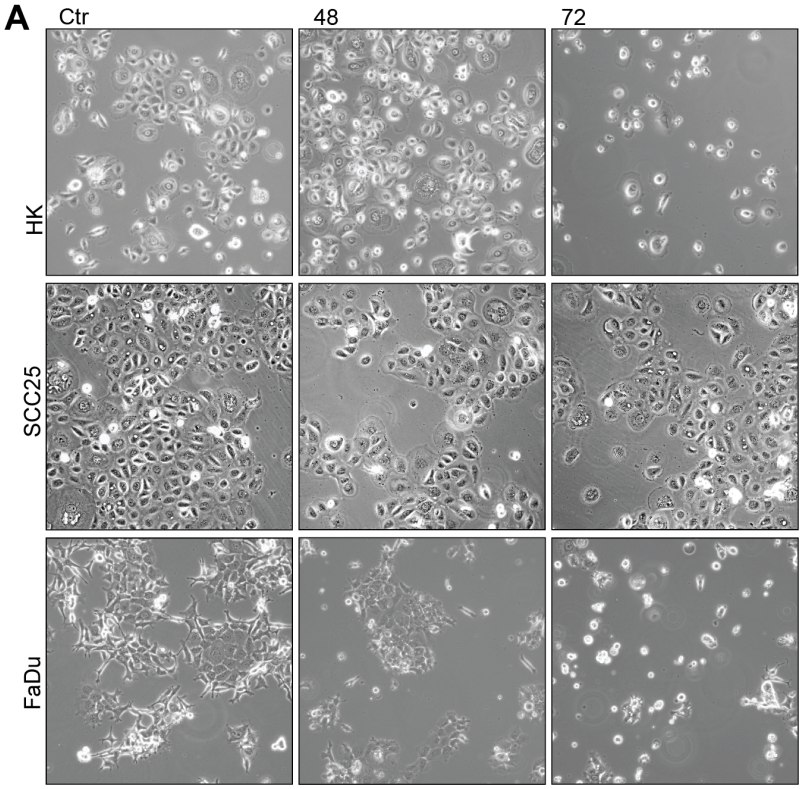

Figure 4.2: **SRSF2 alters the adherence capacity of epithelial cells**
A Bright field images of control (Ctr; left), 48 (middle) and 72 (right) hours after knock down in keratinocytes (HK) (upper), SCC25 (middle) and FaDu (lower) cells. Representative image of four images.

4.2.3 *SRSF2 knock down leads to cell cycle arrest*

In order to ascertatin whether SRSF2 regulates cell cycle progression in epidermal cells I quantified cell cycle distribution in control and SRSF2 depleted cells 48 and 72 hours after transfection (Figure 4.3). A reduction in SRSF2 levels led to a decrease in the proportion of cells in S phase and an increase in the proportion of cells in SubG0 (Figure 4.3 A-C). This was the case for all cell lines and suggests that SRSF2 is required for replication. In order to confirm this the proportion of cells replicating was assayed directly via EdU incorporation. There was a 20% reduction in proliferating cells after SRSF2 depletion in all cell lines (Figure 4.3 D-F).

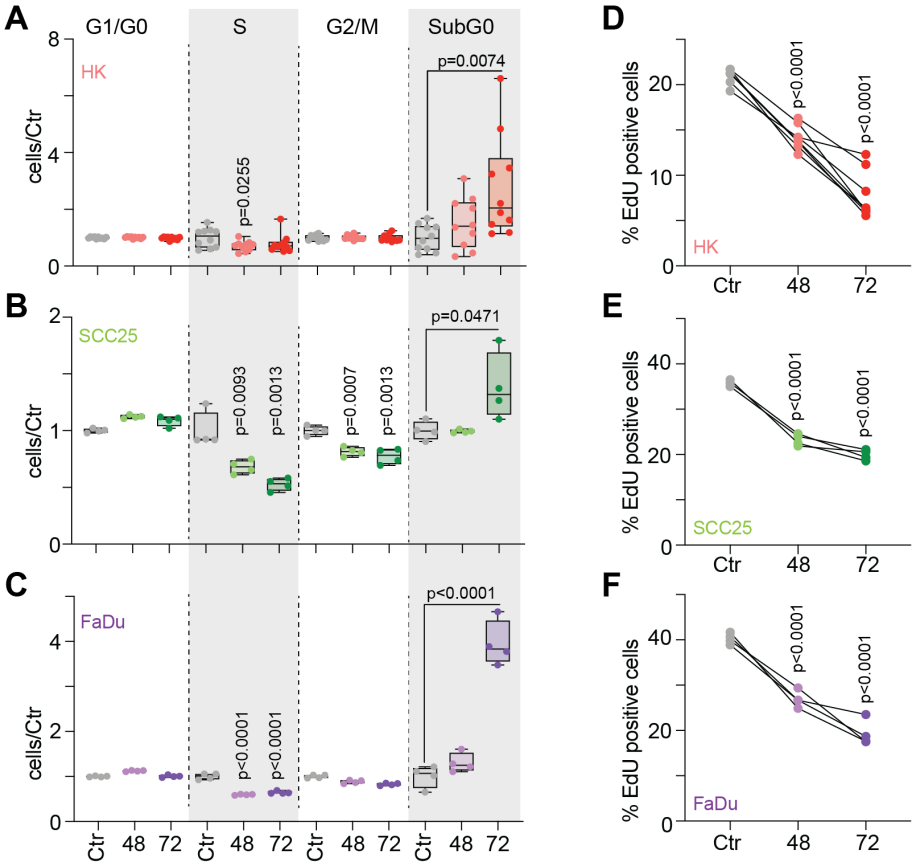

Figure 4.3: **SRSF2 is required for replication**
(**A-C**) Proportion of cells in the different phases of the cell cycle relative to the average in control cells in control (Ctr), and 48 or 72 hours after SRSF2 knock down, in HK (**A**), SCC25 (**B**) and FaDu cells (**C**). (**D-E**) Percentange (%) of cells proliferating in, control (Ctr), 48 and 72 hours after SRSF2 knock down in (**D**) HK, (**E**) SCC25 and (**F**) FaDu cells. (**A-C**) Each point represents one transfection, Box plots show minimum, first quartile, median, third quartile, and maximum. p= unpaired t-test.

SRSF2 has been linked to cell division in various tissues (Bapat et al., 2018, Xiao et al., 2007, Kon et al., 2018), however, the precise molecular mechanisms that depend on SRSF2 remain unknown. In order to address this I began by assaying expression of selected cell cycle genes (Figure 4.4 A-C). In HK and SCC25 cells all assayed cell cycle genes decreased after 48 hours of SRSF2 depletion. In FaDu cells only cell division cycle 45 (CDC45) showed a significant reduction in expression. CDC45 is important for regulating DNA replication because it is a target for the

checkpoint kinase-1 (CHK1) dependent DNA damage checkpoint signal transduction pathway. It's consistent down regulation makes it a potential candidate for a mechanism through which SRSF2 regulates cell division.

To confirm that SRSF2 depletion was not triggering differentiation I assayed gene expression of selected differentiation markers (Figure 4.4 D-F). In HK differentiation marker expression was mildly increased after 72 hours of SRSF2 depletion. This was not the case in either SCC cell line.

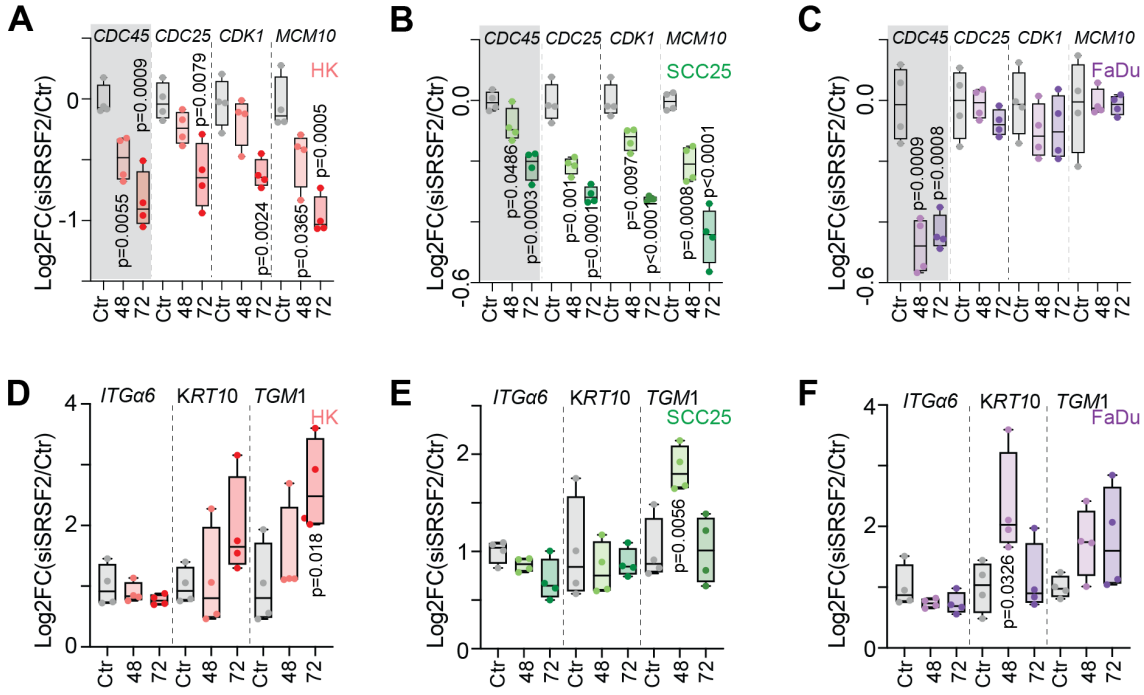

Figure 4.4: **SRSF2 depletion alters expression of cell cycle genes**
(**A-C**) Log2 fold change (Log2FC) of cell division cycle (CDC45), cell division cycle 25 (CDC25), cyclin dependent kinase 1 (CDK1) and minichromosome maintenance 10 (MCM10) RNA levels in control (Ctr), and 48 or 72 hours after SRSF2 knock down in **A** Keratinocytes (HK), **B** SCC25 and **C** FaDu cells. (**D-F**) Log2 fold change (Log2FC) of Integrin alpha 6 (Itgα6), Keratin 10 (KRT10) and Transglutaminase 1 (TGM1) RNA levels in control (Ctr) and 48 or 72 hours after SRSF2 depletion in **D** Keratinocytes (HK), **E** SCC25 and **F** FaDu cells. Each point represents one transfection and the average of three technical replicates. Box plots show minimum, first quartile, median, third quartile, and maximum. p= unpaired t-tests.

As SRSF2 depletion led to a loss of adherence I wondered whether they were undergoing apoptosis (Figure 4.5 A,B). SRSF2 depletion reduced viability of both HK and FaDu cells (Figure 4.5 C, E). Interestinlgy SCC25 cells were the least affected (Figure 4.5 E).Therefore, although SRSF2 depleition does affect cell viability it does not trigger classical apoptosis. In order to minimise measuring secondary effects due to reduced viability all subsequent experiments were performed after 48 hours of SRSF2 depletion.

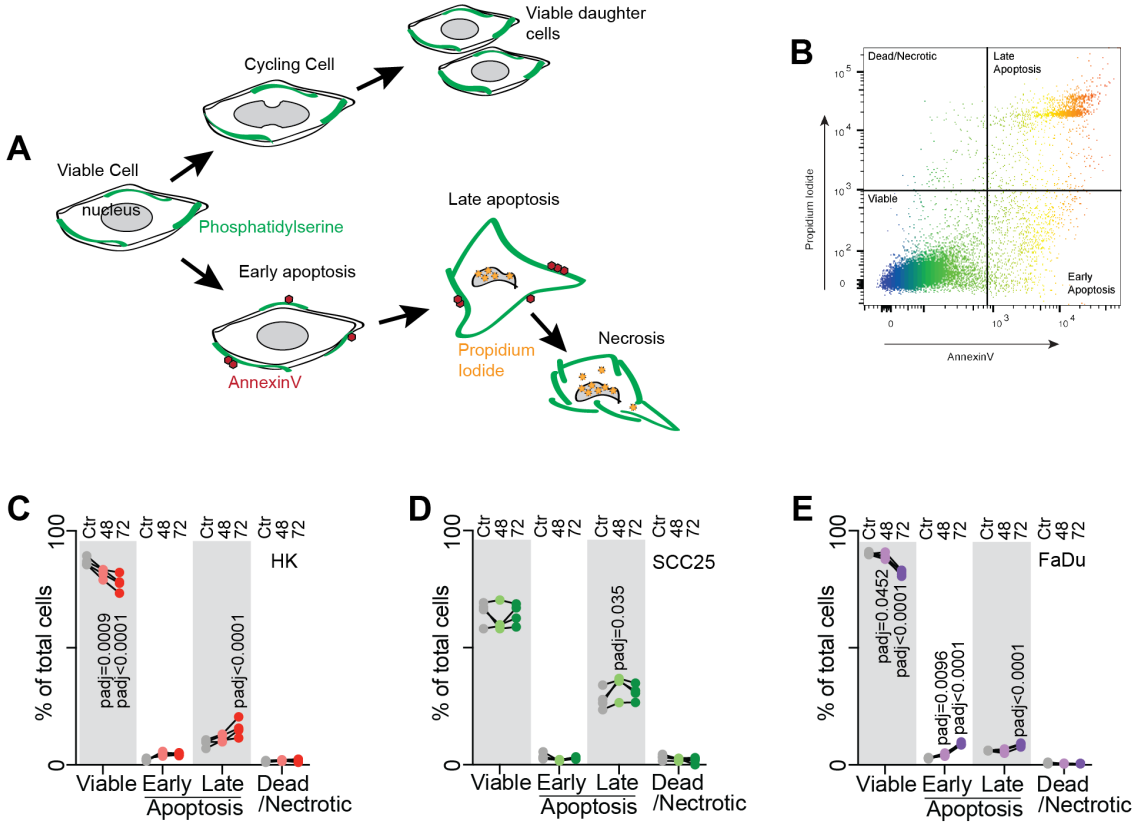

Figure 4.5: **SRSF2 depletion does not trigger apoptosis**
A Apoptosis triggers changes to cell membrane composition, which can be measured by flow cytometry. **B** Gating strategy for Annexin/PI viability assay. (**C**-**E**) Percentage (%) of viable cells, cells undergoing early and late apoptosis and dead or nectrotic cells in control (Ctr) or cells where SRSF2 was depleted for 48 or 72 hours in **C** keratinocytes (HK), **D** SCC25 and **E** FaDu cells. Each point represents one transfection. Each point represents one transfection. padj= Dunnett's multiple comparisons test.

4.2.4 *SRSF2 depletion changes mRNA expression levels of other SR proteins*

It is known that SR proteins act in a network, but this network is dynamic functioning in a cell type and cell state dependent manner (Leclair et al., 2020, Lareau et al., 2007). How changing the expression of one family member alters the expression of other family members is unknown. However, cell line specific alterations to expression of all SR proteins may contribute to cell line specific changes seen after SRSF2 depletion. In order to determine whether SR protein gene expression was altered and whether specific family members were consistanly affected by SRSF2 depletion I generated total RNA sequencing data from all three cell lines.

The expression of other SR proteins was signficantly altered as a result of SRSF2 depletion in all cell lines (Figure 4.6). HK cells showed the most dysregulation, seven of the twelve other family members were significantly upregulated as a result of SRSF2 depletion (Figure 4.6 A). One family member was signifcantly upregulated after SRSF2 depletion in SCC25 (Figure 4.6 B), while five were significantly upregulated in FaDu cells (Figure 4.6 C).

SRSF5 was significantly upregulated in all three cell lines as a result of SRSF2

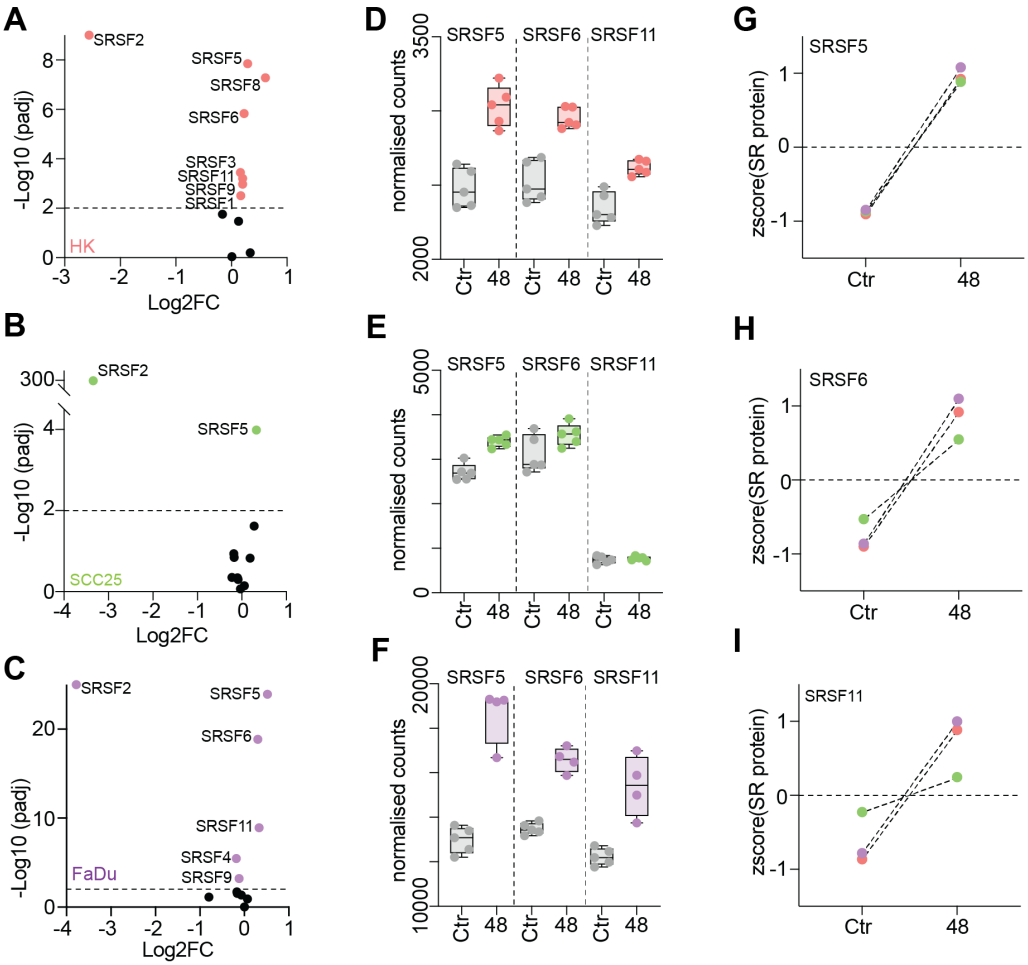

Figure 4.6: **SRSF2 depletion alters expression of other SR proteins in a cell line dependent manner**
(**A-C**) SR protein expression in **A** keratinocytes (HK), **B** SCC25 and **D** FaDu cells. (**D-F**) Normalised counts in control (Ctr) and SRSF2 knock down (48) for SRSF5, SRSF6 and SRSF11 in **D** HK, **E** SCC25 and **F** FaDu cells. **E** Normalised counts in control (Ctr) and SRSF2 knock down (48) in SCC25 cells of SRSF5, SRSF6 and SRSF11. (**G-I**) zscores showing **G** SRSF5, **H** SRSF6 and **I** SRSF11 expression changes in HK (pink), SCC25 (green) and FaDu (purple).

depletion. This is the only other SR protein that does not shuttle between the nucleus and the cytoplasm, due to a nuclear retention signal (Cáceres, Screaton, and Krainer, 1998, Cazalla et al., 2002, Lin, Xiao, et al., 2005, Sapra et al., 2009), and so perhaps competes more with SRSF2 relative to other family members. Additionally, SRSF6, SRSF9 and SRSF11 were all upregulated in HK and FaDu cells. Interestingly, expression levels of these three SR proteins were five-fold higher in FaDu cells than HK, which is in line with studies showing that SR protein expression is dysregulated in cancer (Figure 4.6 D-I) (Wagner and Frye, 2021).

4.2.5 *Summary, discussion and future directions*

I successfully used an siRNA pool to deplete SRSF2 in a primary keratinocyte cell line and two cancer cell lines, which both have p53 mutations. This revealed that SRSF2 is required for replication and functions in a p53 independent manner. In order to further understand why cells undergo cell cycle arrest when SRSF2 was depleted expression of selected cell cycle genes was assayed. CDC45 expression was consistently down regulated when SRSF2 was depleted. CDC45 is part of the helicase complex, CDC45-MCM-GINS (CMG), which is important for DNA unwinding and replication initiation (Rzechorzek et al., 2020, Baris et al., 2022, Guillou, Coloma, and Montoya, 2009). This makes it a promising candidate for a specific mechanism through which SRSF2 plays an essential role in coordinating cell division.

Cells were unable to replicate when SRSF2 was depleted but this was not associated with differentiation. Furthermore, cell viability was also only mildy reduced after SRSF2 depletion and cells were not undergoing apoptosis. However, 72 hours after SRSF2 depletion both HK and FaDu cells were no longer adherent. The reason SRSF2 is required for cells to maintaine adherence remains unclear.

I also showed that SRSF2 depletion alters the expression of other SR proteins and that the extent varied depending on the cell line. This is in line with previous work showing that SR protein expression varies depending on the cell type and cell state (Lareau et al., 2007, Leclair et al., 2020).

SRSF5 was consistently upregulated in all cell lines after SRSF2 depletion. There is little known about SRSF5's function, but it has been shown to be associated with increased prostate cancer risk (Narla et al., 2005). Furthermore, SRSF5 and SRSF2 are the only SR proteins that do not shuttle between the nucleus and the cytoplasm (Botti et al., 2017). More closesly related family members have been shown to compete more (Lareau et al., 2007), which could be the reason that SRSF5 expression is consistently increased after SRSF2 depletion.

n HK and FaDu cells SRSF6 and SRSF11 were also upregulated. SRSF6 has previously been shown to be important for IFE homeostasis (M. A. Jensen, Wilkinson, and Krainer, 2014). When overexpressed in skin the phenotype described was remarkbly similar to what I observed when functional SRSF2 was lost in the epidermis. The authors find that fibronectin is mis-spliced leading to expression of isoforms seen in psoriasis (M. A. Jensen, Wilkinson, and Krainer, 2014). The consequences of SRSF6 over expression or SRSF2 deletion on the expression of other SR protein family members in skin remain unknwon, this would be an interesting question for future work. Comparing the SR protein pool in different contexts and different tissues may elucidate more about the functions of individual SR proteins and how the network acts as a whole.

In the future it would be interesting to over-express CDC45 and assess whether this rescues the cell cycle arrest that occurs after SRSF2 depletion. CDC45 would also be an interesting candidate to examine whether SRSF2 functions via splicing or transcription as changes to splicing and transcription could be assessed directly via minigene (Poonian, Mccomas, and Nussbaum, 1977). SRSF2 depletion largely resulted in modest upregulation of SR protein gene expression. However, the precise role SRSF2 has in regulating the expression of other SR proteins is unknown. It

would be interesting to deplete and overexpress SRSF5 or SRSF6 and look at how this changes SRSF2 expression and what phenotypic affect this has on epidermal cells.

Fully elucidating the roles of SR proteins in epidermal differentiation remains a pressing question, using calcium switch assays, where cells are cultured with high concentrations of calcium would be an excellent model for exploring how SR proteins contribute to differentiation in the IFE.

Chapter 5

SRSF2 is required for expression of DNA repair and replication genes

Analysis of RNA and SLAM sequencing data was assisted by Dr. Anke Heit-Mondrzyk, splicing data and eCLIP data from Wheeler et al., 2022 was analysed by Dr. Thiago Britto-Borges and analysis of Cut&Run data was assisted by Etienne Sollier.

5.1 *Aims of this chapter*

Altering transcription is an important aspect of a cellular responses to both extrinsic and intrinsic stress (Mahat et al., 2016). It has several regulatory steps, which allow changes to gene expression to occur rapidly and to ensure that the correct gene expression program is executed (Gilchrist et al., 2012). The final regulatory step before productive elongation is called "promoter proximal pausing". This occurs between 50 and 200 base pairs down stream of the transcription start site (TSS). Pol II is held in a paused state by the negative elongation factor (NELF) and DRB-sensititvy inducing factor (DSIF) and phosphorylation of these proteins and Ser2 of the CTD of Pol II by PTEFb is required for productive elongation (Gilchrist et al., 2012, Yamaguchi, Shibata, and Handa, 2013, Adelman and Lis, 2012).

PTEFb is held in an inhibitory complex by the proteins HEXIM and LARP7, along with the non-coding RNA, *RN7SK* (Yamaguchi, Shibata, and Handa, 2013). SRSF1 and SRSF2 both associate with the *7SK* complex, but only SRSF2 has been shown to influence transcription (Ji et al., 2013, Lin, Coutinho-Mansfield, et al., 2008, Tripathi et al., 2012). It has been hypothesised that SRSF2 has a higher affinity for the nascent RNA that emerges from Pol II than it does for the *RN7SK* complex, it therefore dissociates, which leads to PTEFb dissociation (Lin, Coutinho-Mansfield, et al., 2008, Ji et al., 2013). Whether SRSF2 regulates PTEFb at a specific group of genes or genome wide remains unknown.

SRSF2 has been shown to have roles in both splicing and transcription, two processes which are intricately linked (Graveley and Tom Maniatis, 1998, X. D. Fu

and T. Maniatis, 1992, L. Han et al., 1995, Ji et al., 2013, Lin, Coutinho-Mansfield, et al., 2008, Aslanzadeh et al., 2018, Bentley, Guthrie, and J. Steitz, 2005). However, how SRSF2's roles in both these processes contributes to the phenotypes that have been associated with loss or mutation of SRSF2 remains largely unknown. The dual nature of SRSF2's function has not been explored fully, whether its roles in transcription and splicing feed back on each other leading to a more severe phenotype is yet to be fully elucidated. The aims of this chapter are as follows:

1. Understand the effect of SRSF2 loss on the transcriptome. This was done by performing RNA sequencing in three cell lines and looking at genes that are differentially expressed in all cell lines.

2. Look at the specificty of SRSF2 function in transcription. This was done by assaying nascent RNA and Pol II occupancy genome wide using SLAM sequencing and Cut&Run respectively.

3. Examine how splicing is altered by the absence of SRSF2. This was done using bioinformatics tools, that quantify changes to splicing in total RNA sequencing.

5.2 *Results*

5.2.1 *SRSF2 depletion leads to down regulation of gene expression*

To elucidate the direct effects of SRSF2 depletion on gene expression I knocked down SRSF2 in three cell lines; HK, SCC25 and FaDu cells (Figure 5.1 A, B). Bulk RNA sequencing was then performed to assess the affects of SRSF2 depletion on the transcriptome. When comparing the three cell lines 21% of differentially expressed genes were mis-expressed (Figure 5.1 C). I noted that there were consistently more down regulated genes when SRSF2 was depleted, which may be due to SRSF2's role as a transcriptional activator (Figure 5.1 D) (Lin, Coutinho-Mansfield, et al., 2008, Ji et al., 2013).

It was previously shown that when SRSF2 was mutated, in acute myeloid leukemia (AML), integrator subunit 3 (INTS3) was mis-spliced, leading to a reduction in its expression (Yoshimi et al., 2019). INTS3 expression was also down regulated in the three epithelial cell lines after SRSF2 depletion (Figure 5.1 E). A closer look at the INTS3 transcript revealed that intron retention occurs after SRSF2 depletion, as described in AML (Figure 5.1 F) (Yoshimi et al., 2019).

To ascertain which cellular processes require SRSF2 gene ontology analysis was performed on genes that were down regulated in all three cell lines. This revealed that genes involved in DNA repair and replication are consistently down regulated when SRSF2 was depleted in epithelial cell lines (Figure 5.1 G).

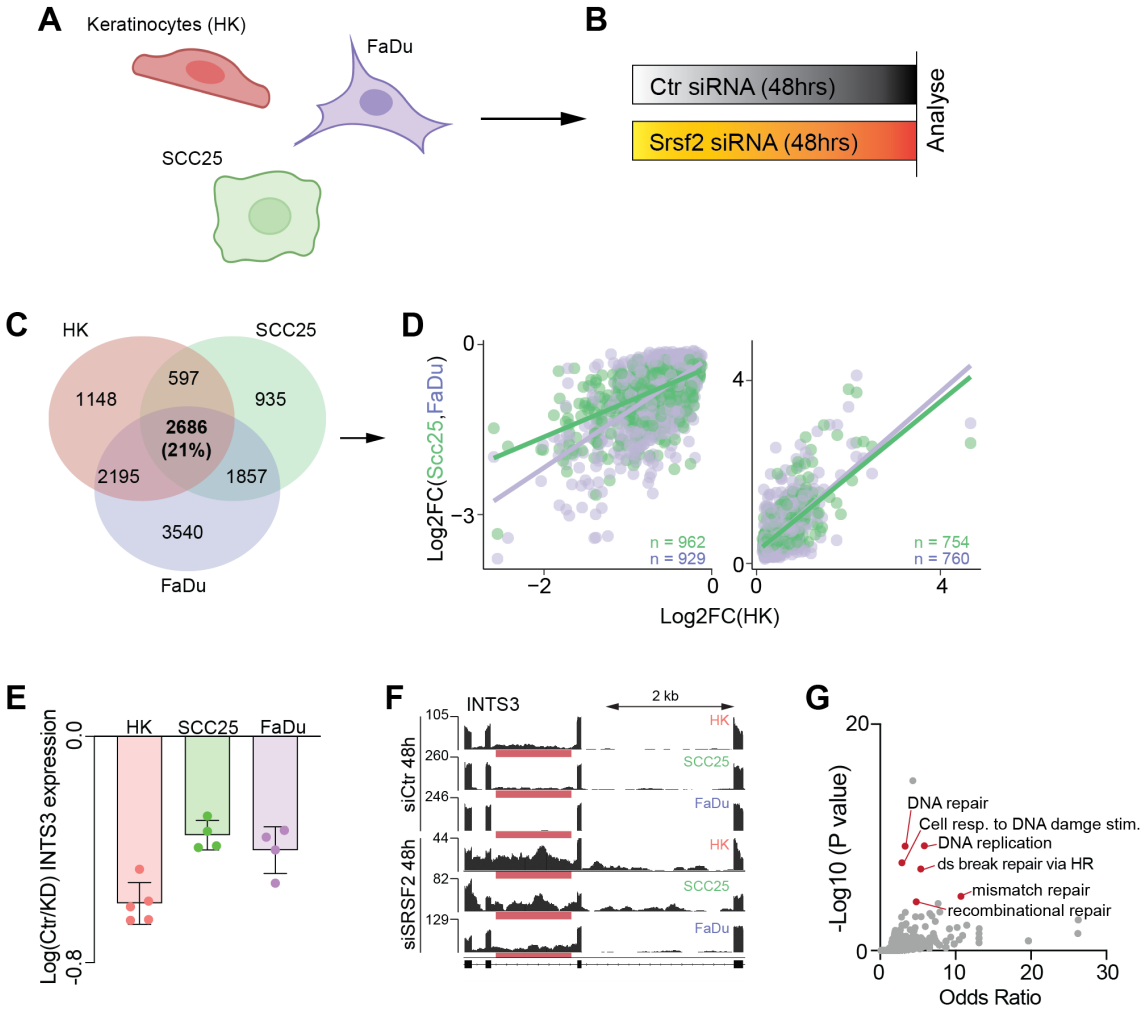

Figure 5.1: **SRSF2 depletion leads to gene repression**
A Schematic representation of the cell lines used for *in vitro* experiments. **B** Illustration of the time course used to assay transcriptomic changes due to SRSF2 depletion. **C** Venn diagram of the transcriptomic changes observed after SRSF2 knock down in HK, SCC25 and FaDu cells. **D** Commonly down regulated (left) and up regulated (right) genes after SRSF2 depletion. **E** Relative INTS3 expression in HK, SCC25 and FaDu cells, each point represents one RNA isolation and the average of three technical replicates. **F** Genome browser view of INTS3, intron retention is highlighted via red bar. **G** Gene ontology analysis of commonly down regulated genes done with EnrichR (Xie et al., 2021. Kuleshov et al., 2016, E. Y. Chen et al., 2013. (**C, D, G**) padj< 0.01, n= gene number

5.2.2 *SRSF2 is required for active transcription*

Whether DNA repair and replication genes require SRSF2 for proper expression due to its function in splicing or transcription was unclear. To address this I performed thiol(SH)-linked alkylation for the metabolic sequencing of RNA (SLAM-seq), which unlike bulk RNA sequencing, allows direct transcriptional affects to be analysed (Figure 5.2) (Herzog et al., 2017, Neumann et al., 2019).

Cells were incubated with 4-thiouridine (4sU), a uridine analog which is easily taken up by cells and incorporated into RNA (Herzog et al., 2017). It has specific properties which make it reactive to thiol modification, which leads to mis-incorporation during reverse transcription. Therefore when total RNA was se-

quenced nascent RNA could be identified as reads that have thymine to cytosine (T to C) mutations (Figure 5.2 A). Transcription rates were then assessed by quantifying the number of transcripts with mutations per gene, as previously described by Herzog et al., 2017.

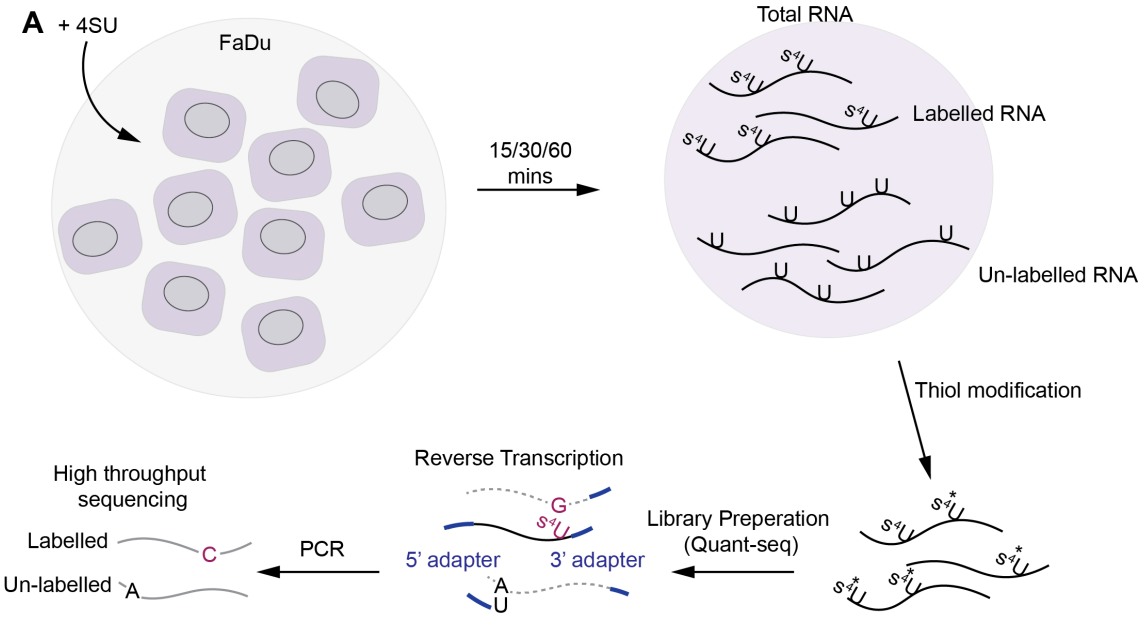

Figure 5.2: **Experimental outline for SLAM-sequencing**
A. Cells were incubated with 4SU for 15, 30 or 60 minutes, which is incorporated into nascent RNA. Total RNA is then extracted and thiol modification is performed. Libraries are prepared and sequenced, the nascent transcriptome can be identified as reads containing T-C mutations. Figure adapted from Herzog et al., 2017.

In order to establish an appropriate incubation time with 4sU, I carried out a time course where cells, transfected with a scrambled control or the siRNA pool targetting SRSF2, were incubated with 4sU for 15, 30 or 60 minutes (Figure 5.2 A). As expected the majority of mutations were T to C mutations, caused by 4sU incorporation (Figure 5.3 A-C). After 15 minutes of 4sU incubation the mutation rate was less than 1%, which was used as background in downstream analysis (Figure 5.3 D). After 30 minutes of 4sU incubation the mutation rate was 1% and after 60 minutes the mutation rate was 3% (Figure 5.3 D). The levels of incorporation did not vary between control and SRSF2 depleted cells (Figure 5.3 D). Therefore the number of differentially transcribed genes was investigated. After 15 minutes of incubation the most genes showed differential incorporation, however, this is likely to be due to a high signal to noise ratio (Figure 5.3 E). After 30 minutes more than twice as many genes showed less incorporation when SRSF2 was depleted (Figure 5.3 F) and after 60 minutes there were five times as many genes with less incorporation after SRSF2 knock down (Figure 5.3 G).

Differential expression analysis of newly transcribed RNA revealed that less RNA was synthesised when SRSF2 was depleted (Figure 5.4 A). When this analysis was done with the pre-existing RNA the global effect was masked (Figure 5.4 B). A reduction in incorporation was further confirmed by looking at cummulative frequency, which showed there was a significant reduction of 4sU incorporation after 30 and

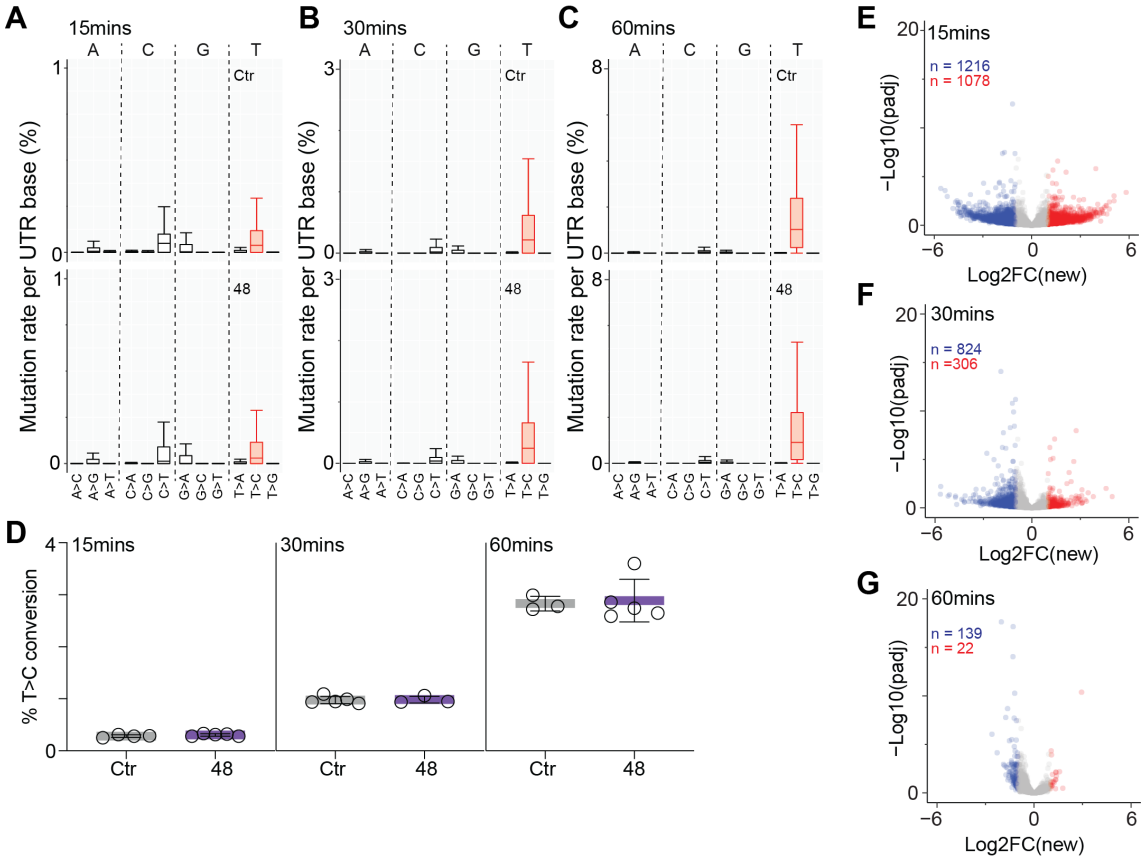

Figure 5.3: **Establishing SLAM-seq in FaDu cells**
(**A-C**) Quantification of mutation rates per transcript in control (Ctr, upper panel) and after 48 hours of SRSF2 depletion (48, lower panel) after **A** 15, **B** 30 and **C** 60 minutes of 4sU incorporation . **D** Percentage (%) T to C conversion in control (Ctr) and SRSF2 depleted (48) conditions after 15 (left), 30 (middle) and 60 (right) minutes each point represents one transfection, shown is mean± SD. (**E-G**) Volcano plots showing the number of genes that show differential incorporation after **E** 15, **F** 30 and **G** 60 minutes of 4sU incorporation, n= genes where the log2 fold change (Log2FC)< -1 (blue) or > 1 (red).

60 minutes (Figure 5.4 C, D). Gene ontology analysis of the most affected genes again revealed that the majority of differentially transcribed genes were involved in DNA replication and repair, along with transcription and histone acetylation (Figure 5.4 E). These results suggest that gene down regulation was due to alterations to transcription.

To ascertain whether the changes to gene expression that occur after SRSF2 depletion are due to Pol II activity and not mRNA degradation, Pol II phosphorylation status was investigated (Figure 5.5 A-C). This revealed that upon SRSF2 depletion there is a reduction in the Ser2 phosphorylated form of Pol II, which is associated with productive elongation (Figure 5.5 B, C), which was in line with previously published data (Lin, Coutinho-Mansfield, et al., 2008, Ji et al., 2013, Yoshimi et al., 2019). I next asked how Pol II occupany was affected by SRSF2 depletion genome wide via Cut&Run (Meers et al., 2019, Skene, J. G. Henikoff, and S. Henikoff, 2018).

This provides measurements of Pol II binding across the length of a gene, promoter proximal pausing can be estimated by computing the ratio of reads in the

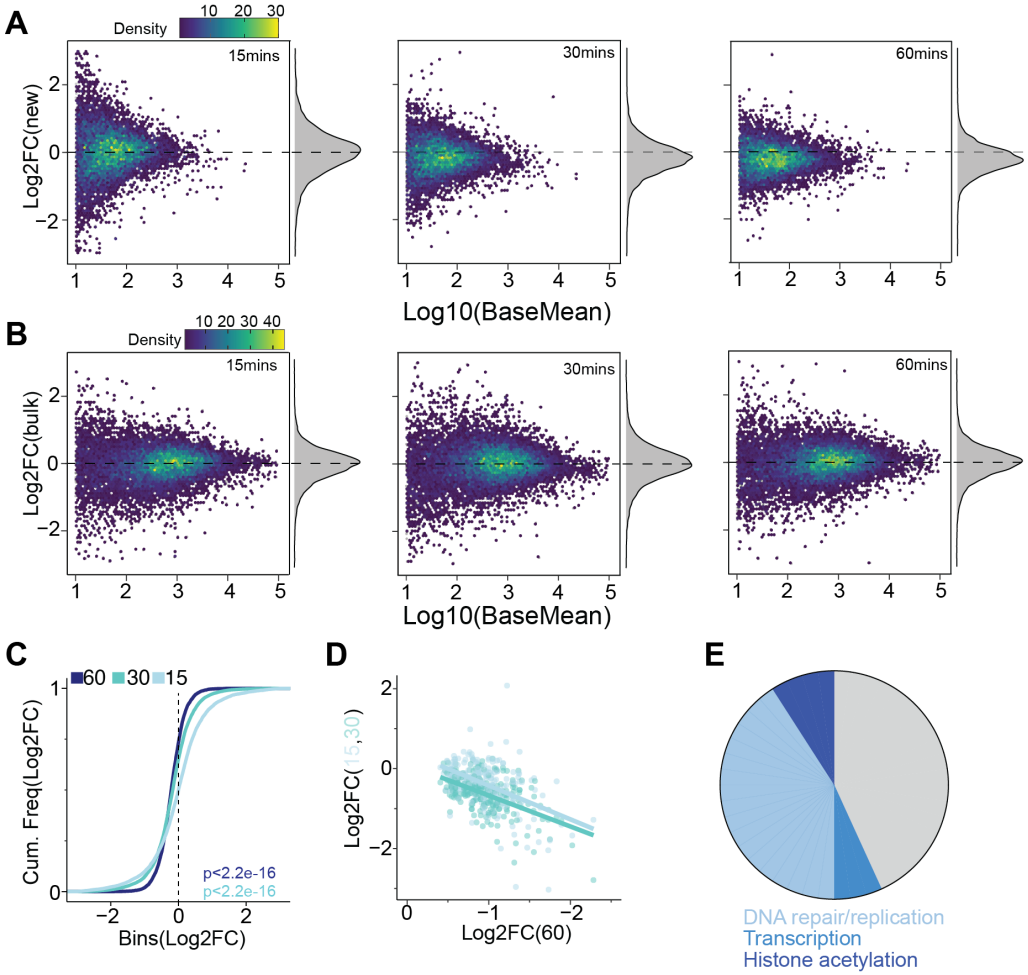

Figure 5.4: **SRSF2 depletion leads to a decrease in transcription**
A Density plots and histograms showing $\log_2$ fold change (Log2FC) of newly transcribed RNA after 15 (left) 30 (middle) and 60 (right) minutes of 4sU incubation. **B** Density plots and histograms showing $\log_2$ fold change (Log2FC) of pre-existing RNA after 15 (left) 30 (middle) and 60 (right) minutes of 4sU incubation. **C** Cumulative frequency (Cum. Freq) plot showing changes to incorporation observed after 15, 30 and 60 minutes (mins) of 4SU incorporation, p= Wilcoxon rank sum test with continuity correction. **D** Scatter plot showing genes with less incorporation at all three time points, genes were filtered padj$<$ 0.05 and $\log_2$ fold change$<$ 0. **E** Pie chart showing gene ontology of genes transcribed less after SRSF2 depletion, genes were filtered padj$<$ 0.05 and log2FC$<$ 0, analysis was done with GOrilla (Eden, Lipson, et al., 2007, Eden, Navon, et al., 2009).

promoter, defined as the first 250 base pairs of a gene, compared to the gene body, which has been termed the "pausing index" (Bandiera et al., 2021, Adelman and Lis, 2012). SRSF2 depletion led to an increase in the pausing index, therefore, more Pol II was bound at the promoter relative to the gene body (Figure 5.5 D). Focusing on Pol II occupancy in the transcription start site (TSS), confirmed that significantly more Pol II was bound to the TSS when SRSF2 was depleted (Figure 5.5 E, F).

Both SLAM sequencing and Cut&Run experiments suggested that SRSF2 is playing an essential role in mediating Pol II pause release genome wide as Pol II was stalled and less nascent RNA was produced.

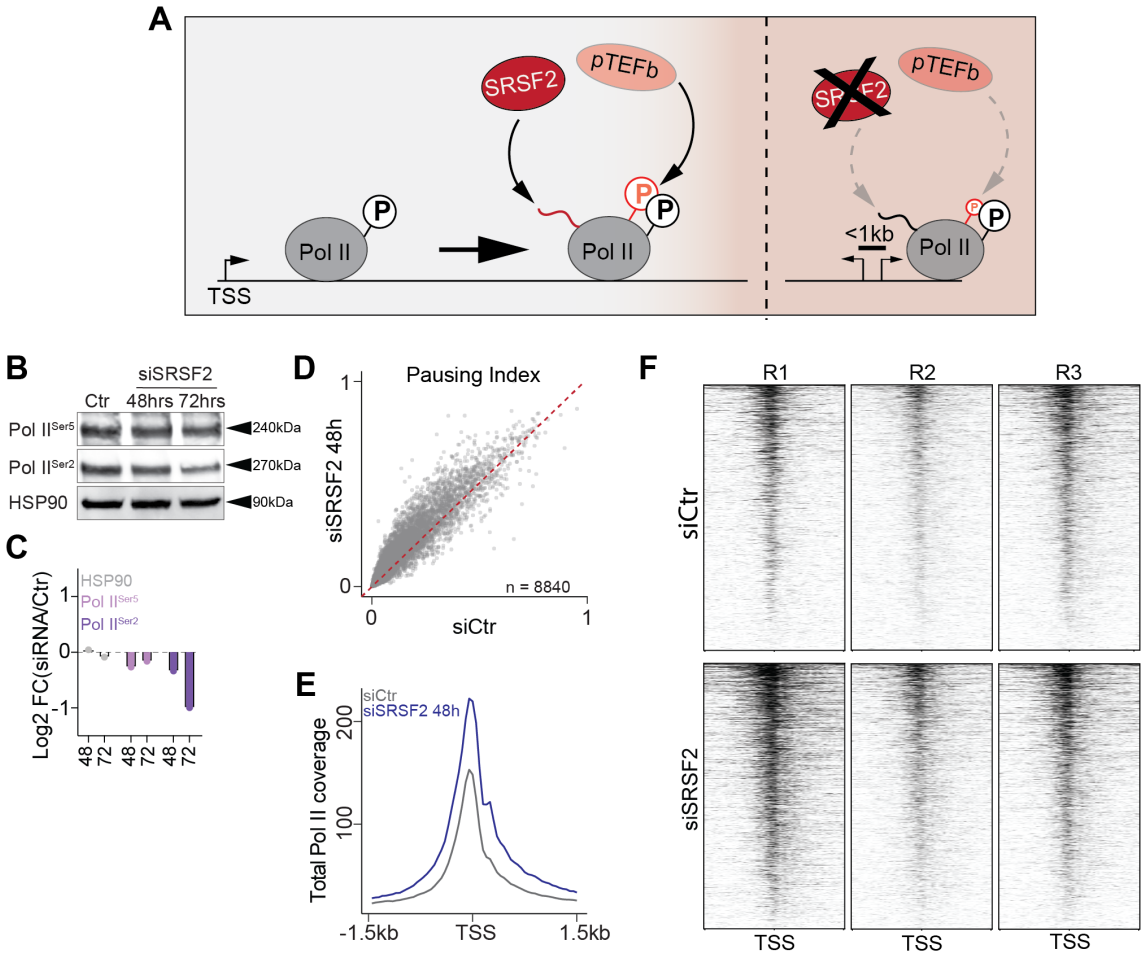

Figure 5.5: **Genome wide Pol II stalling occurs after SRSF2 depletion**
A SRSF2 depletion alters the release of PTEFb from an inhibitory complex, Pol Ii is not phosphorylated and stalls close to promoters. **B** Western blot of Pol II serine 5 (Ser5) and Serine 2 (Ser2) in control (Ctr) and SRSF2 knock down (48hrs) and (72hrs) (upper panel) and quantification (lower panel). **D** Pol II pausing index, n= number of genes. **E** Mean Pol II coverage at the transcription start site (TSS) in control (siCtr) and SRSF2 depleted (siSRSF2 48h) conditions. **F** Heatmaps of the TSS from all control (siCtr; upper panel) and SRSF2 depleted (siSRSF2, lower panel) replicates (R1-3).

5.2.3 *Splicing is altered by SRSF2 depletion*

SRSF2 has been well described as a splicing factor, however, in the previous chapter splicing was not amongst the most dysregulated processes when SRSF2 was deleted. Nevertheless, whether SRSF2 regulates the expression of a specific group of genes by splicing remains unknown. Addressing this question is complex as splicing has been shown to be impacted by transcription rates. This is likely to be a further reason why identifying specific splicing events that are regulated by SRSF2 across cell types has been challenging (Moore and Proudfoot, 2009, Pandit et al., 2013).

The transcriptome as a whole was assessed by building de novo transcriptomes for control and SRSF2 depleted conditions. These were then compared to assess whether SRSF2 depletion leads to a global change in the strucutre of the transcriptome. This was not the case, which indicates that SRSF2 is likely to be important for splicing of a specific genes (Figure 5.6 A, B). In order to assess the extent to which

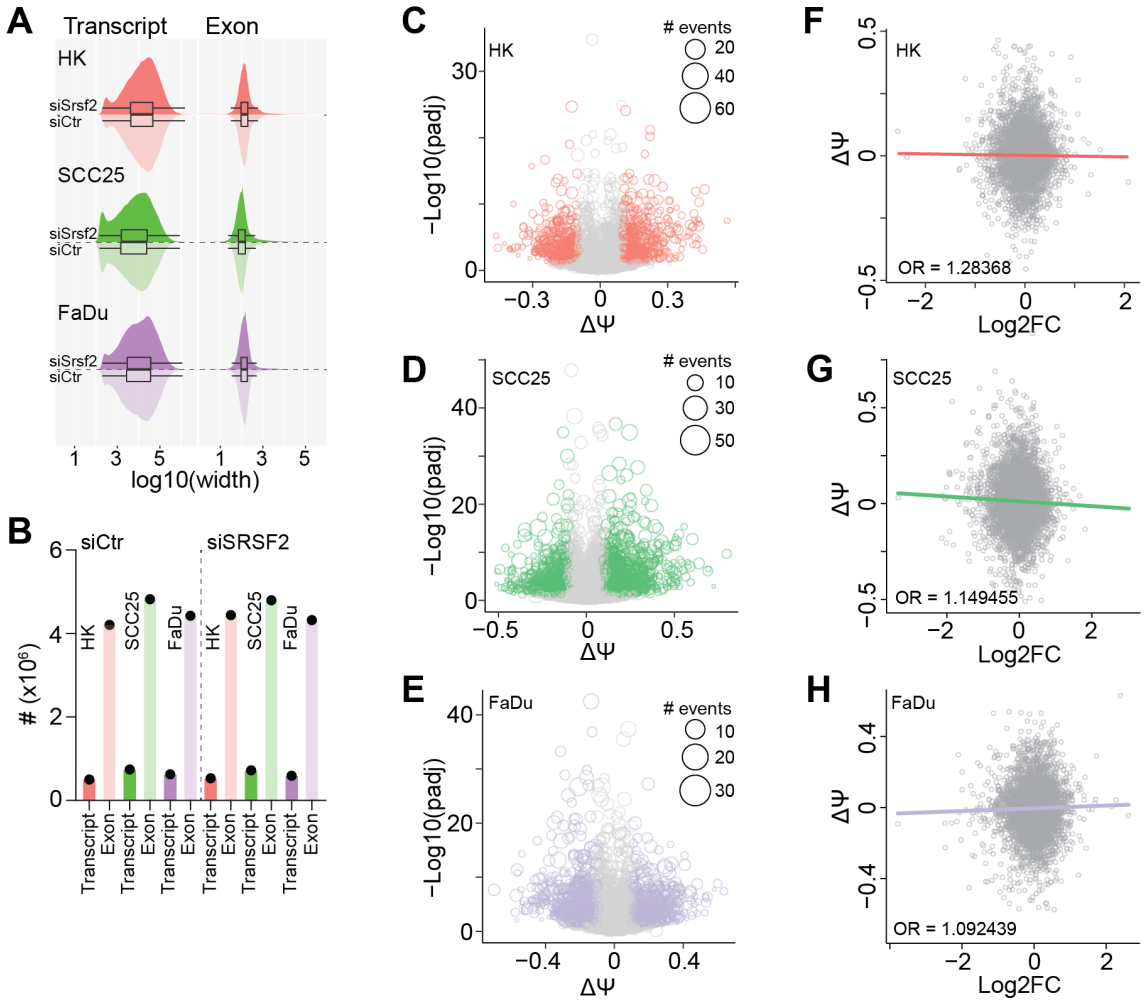

Figure 5.6: **SRSF2 depletion affects splicing in epidermal cell lines**
A The transcriptome and exons in HK (pink), SCC25 (green) and FaDu (purple) cells. **B** The number of transcripts and exons in control (siCtr) and after SRSF2 depletion (siSRSF2) in HK, SCC25 and FaDu cells. (**C-E**) Deltapsi calculated with leafcutter comparing control and SRSF2 depleted splice junction usage in **C** HK, **D** SCC25 and **E** FaDu cells. Events padj$<$ 0.01 and delatapsi $<$ -0.1 or $>$ 0.1 are coloured. (**F-G**) Correlation of deltapsi with Log_2 fold change (Log2FC) in **F** HK, **G** SCC25 and **H** FaDu cells, OR = odds ratio.

splicing alterations occur the $\delta\psi$, a measure of differences in splice junctions present between two conditions was calculated using leafcutter (Y. I. Li et al., 2018). The deltapsi was calculated from total RNA-sequencing data from three cell lines HK, SCC25 and FaDu. This revealed that SRSF2 depletion consistently led to changes in splice junction usage (Figure 5.6 C-E). To address whether these changes altered gene expression I looked at the correlation between differential expression and the deltapsi (Figure 5.6 F-H). Surprisingly, there was no correlation between changes to gene expression and splicing.

SRSF2's splicing function mainly involves promoting exon inclusion, therefore in the absence of SRSF2 exon skipping should be the main form of mis-splicing that occurs. The bioinformatic tool Majiq was used to quantify specific mis-splicing events (Vaquero-Garcia et al., 2016). Independent of cell type exon skipping (ES)

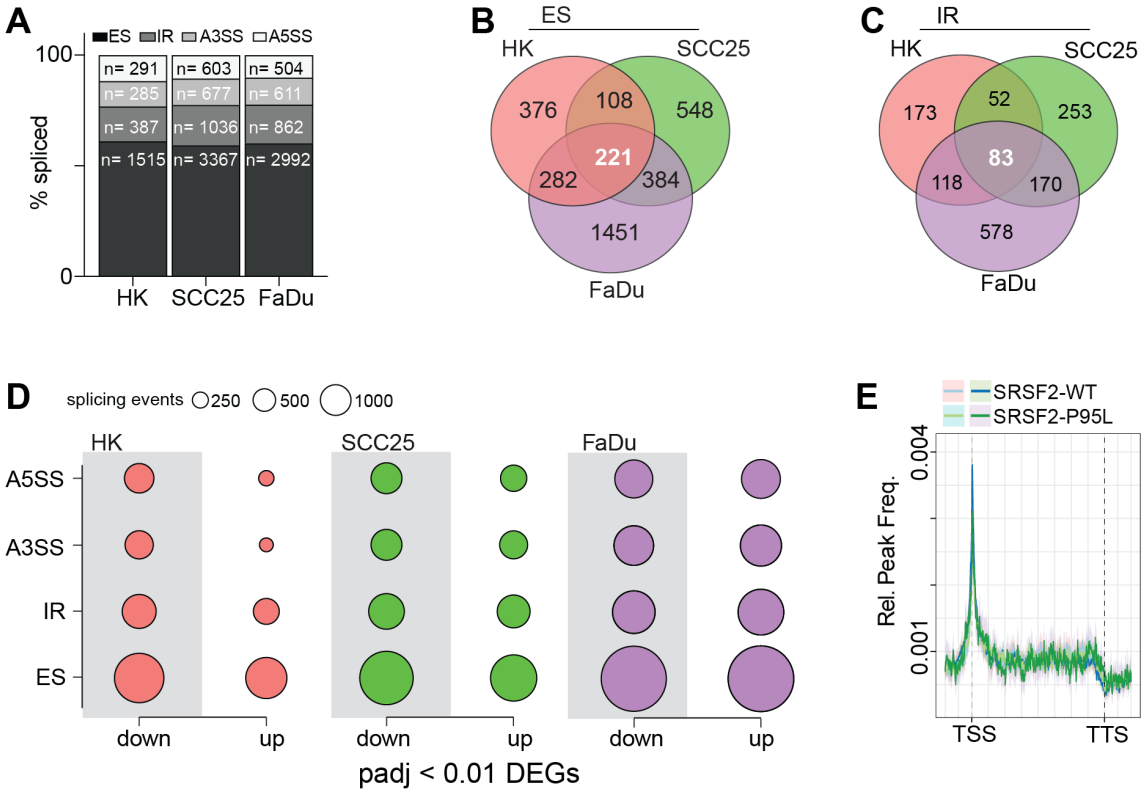

Figure 5.7: **Changes to gene expression are not driven by mis-splicing after SRSF2 depletion** **A** Percentage (%) of splicing events; exon skipping (ES), intron retention (IR), alternative 3' splice site (A3SS) and alternative 5' splice site (A5SS) in keratinocytes (HK), SCC25 and FaDu cells after SRSF2 depletion, n = number of events. (**B, C**) Venn Diagram showing common **B** exon skipping (ES) and **C** intron retention (IR) events after SRSF2 knock down. **D** Correlation between splicing events and alterations to gene expression in HK (left), SCC25 (middle) and FaDu (right). **E** Metagene plots showing wild type (SRSF2-WT) and mutant (SRSF2-P95L) SRSF2 binding from the transcription start site (TSS) to the transcription termination site (TTS).

accounted for approximately 60% of mis-splicing events (Figure 5.7 A). However intron retention (IR), alternative 3' splice site (A3SS) usage and alternative 5' splice site (A5SS) usage were also observed (Figure 5.7 A). Interestingly, exon skipping and intron retention that occured in all cell lines was limited and gene ontology analysis did not reveal any significant enrichment (Figure 5.7 B, C).

I then asked whether any of these specific splicing events correlated with changes to gene expression (Figure 5.7 D). This was largely not the case, however, in HK A3SS and A5SS usage correlated with gene down regulation (Figure 5.7 B, C). Finally, published eCLIP data looking at wildtype and mutant SRSF2 binding was analysed (Wheeler et al., 2022). This revealed that SRSF2 was most enriched at transcription start sites, where splicing does not take place (Figure 5.7 E). This highlights that the transcriptional changes that occur as a result of SRSF2 depletion are likely to drive the associated phenotypes and changes to gene expression.

5.2.4 *Summary, discussion and future directions*

Using two sequencing methods; SLAM-seq, which allow the nascent transcriptome to be assayed (Herzog et al., 2017 and Cut&Run which allows Pol II occupancy to be assessed Meers et al., 2019, Skene, J. G. Henikoff, and S. Henikoff, 2018), revealed that transcription is attenuated globally after SRSF2 depletion. The genes most affected were those involved in DNA replication and repair, which correlates with the cell cycle arrest that occurs after SRSF2 depletion. Furthermore, SRSF2 depletion led to extensive alterations to splicing, however, changes to splicing did not correlate with changes to gene expression and were mostly cell line specific, this is in line with previous work which found no correlation between SRSF2 binding and the changes to splicing that occured when SRSF2 was depleted (Pandit et al., 2013). Finally, I confirmed that there is strong enrichment of SRSF2 at exon 1, where splicing is not possible, which indicates that SRSF2's main function may be the regulation of active transcription.

In the future, it would be interesting to look at transcription in a broader context, including chromatin architecture and histone modifications in order to ascertain whether cells try to compensate for SRSF2 loss using other regulatory mechanisms. This would provide a better understanding of the consequences of SRSF2 depletion on transcription and on how far reaching the effects are.

SRSF2 appears to regulate Pol II activity genome wide. However, changes to splicing have been much harder to decipher, this may be in part due to the fact that transcription rates also impact spliceosome function and so splicing (Moore and Proudfoot, 2009). SRSF2 may be important for key splicing events but these are masked by the abberant transcription that results when SRSF2 is depleted. In order to better understand SRSF2's role in splicing several parameters need to be measured. For example, it would be important to account for which other SR proteins expression are affected in the specific cell type or cell line being investigated. Furthermore, it would be interesting to alter transcription rates to different extents, via treatment with 5,6-dichlorobenzimidazole 1- β-D-Ribofuranoside (DRB), and then measure splicing. This would provide information on how Pol II dynamics alter splicing in a systematic way. Changes to splicing that occur due to different levels of Pol II stalling could then be overlayed with changes to splicing that occur due to SRSF2 depletion and this may enable splicing events that occur due to SRSF2 alone to be deciphered. Finally, it would be interesting to use long read sequencing to probe the changes to splicing that occur as a result of SRSF2 depletion. This may help to better understand changes to splicing as they can be assessed in an annotation free manner.

Chapter 6

SRSF2 depletion increases DNA damage

Comet assays were performed in collaboration with Dr. Ali Bakr. Analysis of bi-directional gene pairs was assisted by Dr. Anke Heit-Mondrzyk using scripts published in Nature Communications Bandiera et al., 2021.

6.1 *Aims of this chapter*

Mutant SRSF2 has been shown to cause Pol II stalling (L. Chen et al., 2018, Yoshimi et al., 2019). I have shown that Pol II stalling also occurs when SRSF2 is depleted and expression of genes involved in DNA repair and replication are most affected by SRSF2 loss. However, what makes these genes particularly susceptible to SRSF2 loss remains unclear. Gene organisation in the genome offers a possible explanation. DNA repair and replication genes are often found as bi-directional pairs (Trinklein et al., 2004, Braastad, Leguia, and Hendrickson, 2002, Connelly et al., 1998, Galgóczy, Rosenthal, and Platzer, 2001, Platzer et al., 1997, C. F. Xu, Chambers, and Solomon, 1997, Ronen and Glickman, 2001, Wood et al., 2001). Bi-directional genes can be transcribed together or reciprocally but how this is coordinated remains largely unknown (Trinklein et al., 2004). The presence of elements such as TATA boxes and non-coding RNAs, such as the noncoding RNA, *RN7SK*, have been implicated in contributing to regulation of gene expression from bi-directional gene pairs (Bandiera et al., 2021, Trinklein et al., 2004, Flynn et al., 2011, Ji et al., 2013). SRSF1 and SRSF2 bind *RN7SK* (Ji et al., 2013, L. Chen et al., 2018, Ji et al., 2013) and may be further factors that contribute to coordinating proper transcription at bi-directional gene pairs.

SRSF2 depletion led to cell cycle arrest and DNA replication genes were down regulated. DNA repair genes were also down regulated and DNA damage can also lead to cell cycle arrest. DNA damage occurs continuously as a result of transcription and replication collisions, along with other DNA damaging agents such as reactive oxygen species (ROS). The presence of mutant SRSF2 has been shown to increase R-loop formation (L. Chen et al., 2018). Whether SRSF2 depletion has the same affect and why DNA damage accumulates remains unknown. The aims of this chapter are as follows:

1. Explore how proximity to another gene affects changes to gene expression after SRSF2 depletion. This was done by categorising genes based on distance to the closest neighbour in all possible orientations and correlating this with expression.

2. Quantify whether damage accumulates after SRSF2 depletion. This was done by determining the amount of double strand breaks via γH2AX and quantifying DNA damage via comet assay.

3. Assay the repair capacity of cell lacking SRSF2. This was done by irradiating human keratinocytes and tracking repair, using olive tail moment as measured by comet assay as a readout.

6.2 *Results*

6.2.1 *Gene organisation is conserved*

In order to assess whether organisation in the genome was a factor which made expression of genes more likely to be dysregulated after SRSF2 depletion the distance for each gene to its closest neighbour was calculated for all protein-coding genes in all possible orientations; upstream sense, upstream anti-sense, downstream sense and downstream anti-sense, as previously described by Bandiera et al., 2021 (Figure 6.1 A). When looking at gene pairs in an upstream anti-sense orientation the majority of genes appear to had a neighbour within 10 and 100kb, which suggests that there is an optimal distance for genes to be spaced (Figure 6.1 B). However, there was a group of approximately 1500 genes with a neighbour within 1kb, these genes have previously been referred to as bi-directionally transcribed gene pairs.

Gene ontology analysis of bi-directionally transcribed genes revealed that they are enriched for genes associated with DNA repair (Figure 6.1 B, C), which is in-line with previous findings (Braastad, Leguia, and Hendrickson, 2002, Connelly et al., 1998, Galgóczy, Rosenthal, and Platzer, 2001, Platzer et al., 1997, C. F. Xu, Chambers, and Solomon, 1997, Ronen and Glickman, 2001, Wood et al., 2001, Adachi and Lieber, 2002). Genes where the nearest neighbour was more than a 100kb away tended to be involved in processes that are cell type specific, such as cell adhesion and differentiation (Figure 6.1 D).

In order to confirm that gene organisation was conserved I also looked at the mouse genome; a similar proportion of genes was also bi-directionally transcribed (Figure 6.1 E). Gene ontology analysis also showed that these genes were involved in DNA repair (Figure 6.1). Genes where the closest gene was more than 100kb away were again involved in differentiation (Figure 6.1).

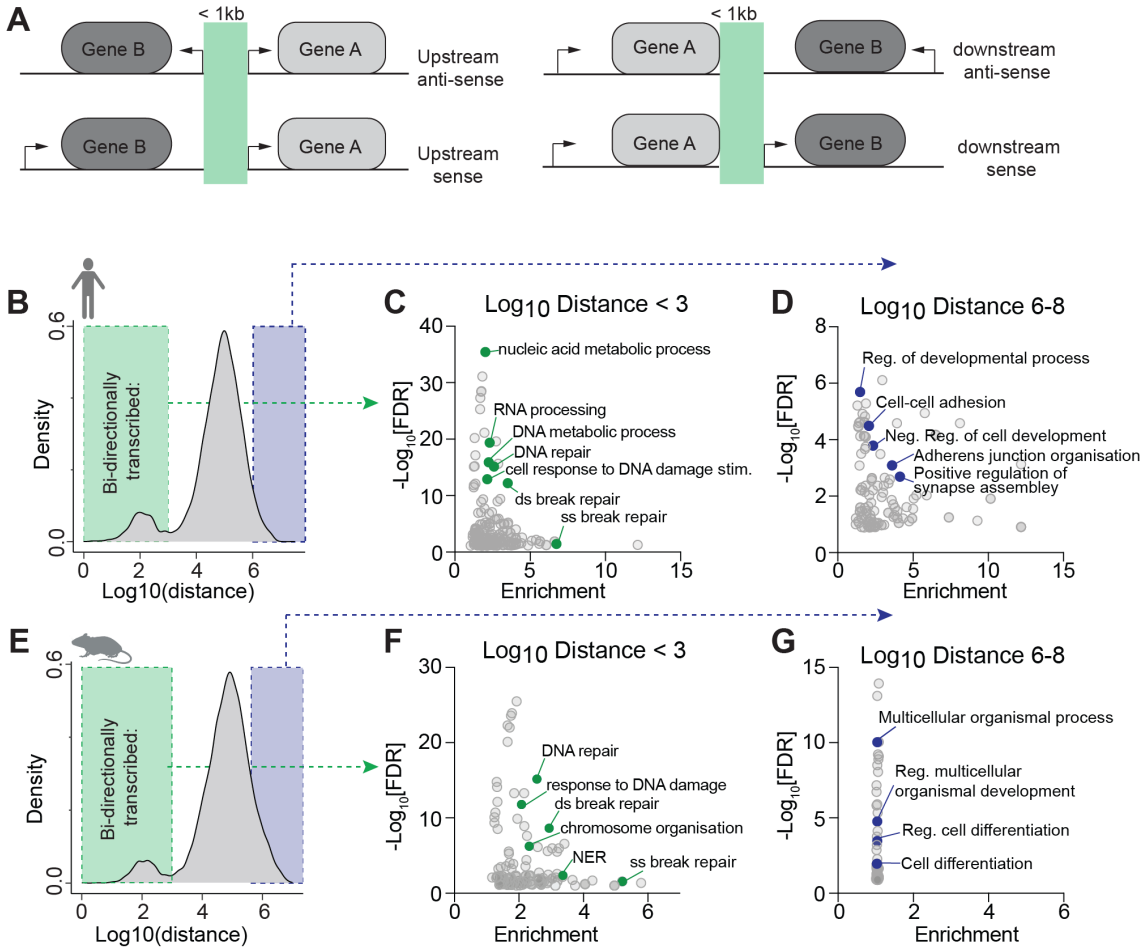

Figure 6.1: **Bi-directionally transcribed genes are conserved between human and mouse** **A** Scheme illustrating how genes can be oriented. **B** Histogram showing density of genes and proximity to a neighbouring gene in the upstream anti-sense direction in the human genome. Genes with a neighbour< 1kb (blue box) and genes with a neighbour> 100kb (red box). **C** Gene ontology analysis on genes with a neighbour within 1kb in the human genome. **D** Gene ontology on genes with a neighbour more than 100kb away in the human genome. **E** Histogram showing the density of genes and proximity to another gene in the mouse genome. Genes with a neighbour< 1kb (blue box) and genes with a neighbour> 100kb (red box).**F** Gene ontology analysis on genes with a neighbour within 1kb in the mouse genome. **G** Gene ontology on genes with a neighbour more than 100kb away in the mouse genome. Analysis with GOrilla (Eden, Lipson, et al., 2007, Eden, Navon, et al., 2009).

6.2.2 *SRSF2 is required for expression of genes organised in clusters*

In order to explore whether genomic organisation correlates with the observed changes to gene expression when SRSF2 was depleted I first used the RNA sequencing data I generated from human cell lines and analysed genes that were differentially expressed in all cell lines (Figure 6.2 A). The frequency of down regulated genes that had a neighbour within 1kb was significantly higher in all orientations (Figure 6.2 B, C). Interestingly in the downstream anti-sense direction, where the 3' ends of two genes are in close proximity, down and up regulated genes were enriched to have a neighbour within 1kb, this orientation may be more prone to collisions and Pol II run-on.

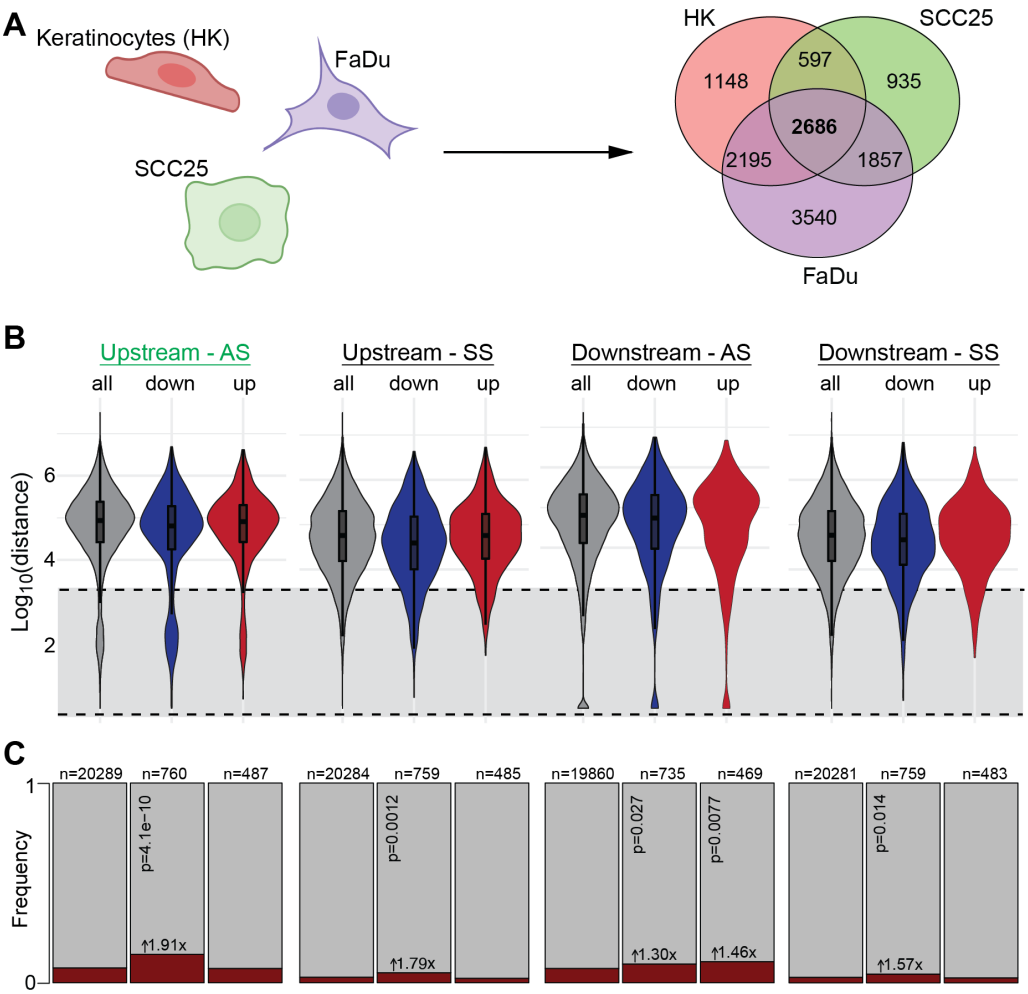

Figure 6.2: **Genes in clusters require SRSF2 for proper gene expression**
A Illustration of cell lines used for analysis and the genes analysed. **B** Violin plots showing enrichment of genes, up (red) or down (blue) regulated after SRSF2 depletion, compared to all (gray), with genes in different orientations; upstream anti-sense (upstream AS), upstream sense (upstream SS), downstream anti-sense (downstream AS), downstream sense (downstream SS). **C** frequency plots, arrow indicates enrichment, n=genes, p= Fisher's exact test.

Down regulated genes organised in the upstream anti-sense direction, which includes bi-directional gene pairs showed the most enrichment, I therefore focused on genes organised in this orientation (Figure 6.3). I first asked how distance and differential expression correlate (Figure 6.3 A-C). Genes with a neighbour within 1kb tended to be downregulated in all cell types. To confirm this I grouped genes by distance. Strinkingly, bi-directionally transcribed genes were downregulated to the greatest extent and the effect appeared to be lost when looking at all other distances (Figure 6.3 D-F). I classified genes as bi-directionally transcribed, or not, which revealed that across cell lines down regulated genes were significantly enriched for being bi-directionally transcribed (Figure 6.3 G). Finally, I asked whether this is conserved using the data generated from mouse back skin after SRSF2 deletion (Figure 6.3 H). Indeed, bi-directionally transcribed genes also showed relative down-regulation, which suggests that SRSF2 acts through a conserved mechanism

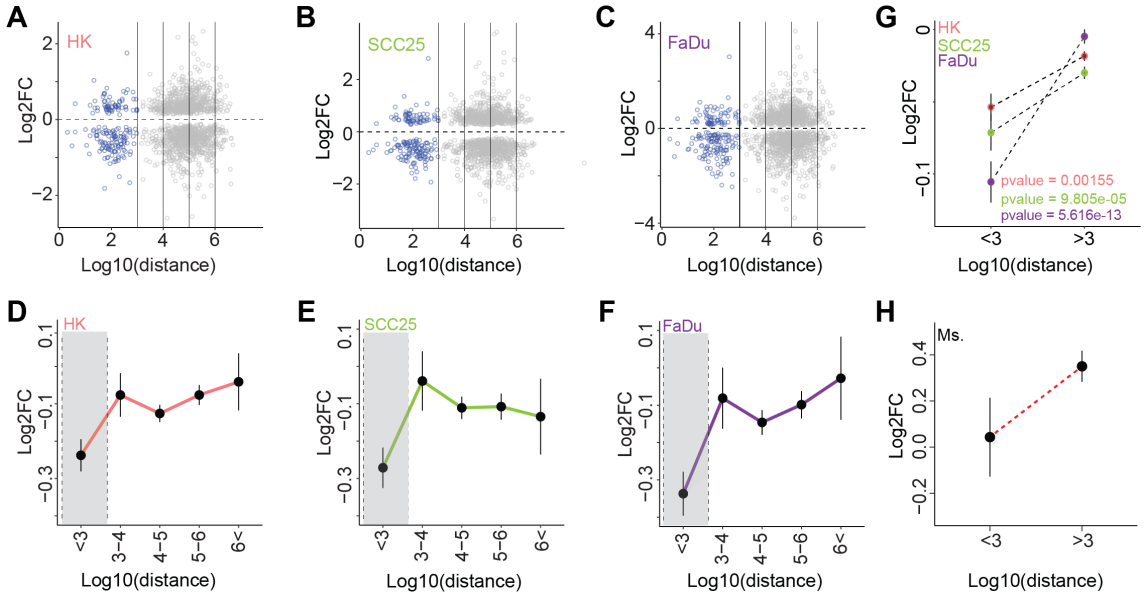

Figure 6.3: **Down regulated genes are bi-directionally transcribed**
(**A-C**) Log_2 fold change (Log2FC) common differentially expressed genes seperated by distance in **A** HK, **B** SCC25 and **C** FaDu cells. (**D-F**) Mean log_2 fold change (Log2FC) of commonly differentially expressed genes grouped by distance to the closest neighbour in **D** HK, **E** SCC25 and **F** FaDu cells, error bars= SEM. **G** Log_2 fold change (Log2FC) common differentially expressed genes less than 1kb or more than 1kb in HK (pink) SCC25 (green) and FaDu (purple) cells.**H** Log_2 fold change (Log2FC) differentially expressed genes less than 1kb or more than 1kb in mouse (Ms.).

to regulate expression of bi-directionally transcribed genes.

In order to ensure that gene organisation and not another variable, such as gene length was driving the changes to gene expression I also analysed whether gene length correlated with changes to gene expression.

I began by looking at how gene length varies between different classes of RNA, which revealed that protein-coding genes show the most variability in length (Figure 6.4 A). Focusing on protein-coding genes that are commonly mis-expressed across cell types I found that gene length does not correlate with changes to gene expression (Figure 6.4 B), but that genes, both shorter and longer than the median gene length are most affected by SRSF2 loss (figure 6.4 C).

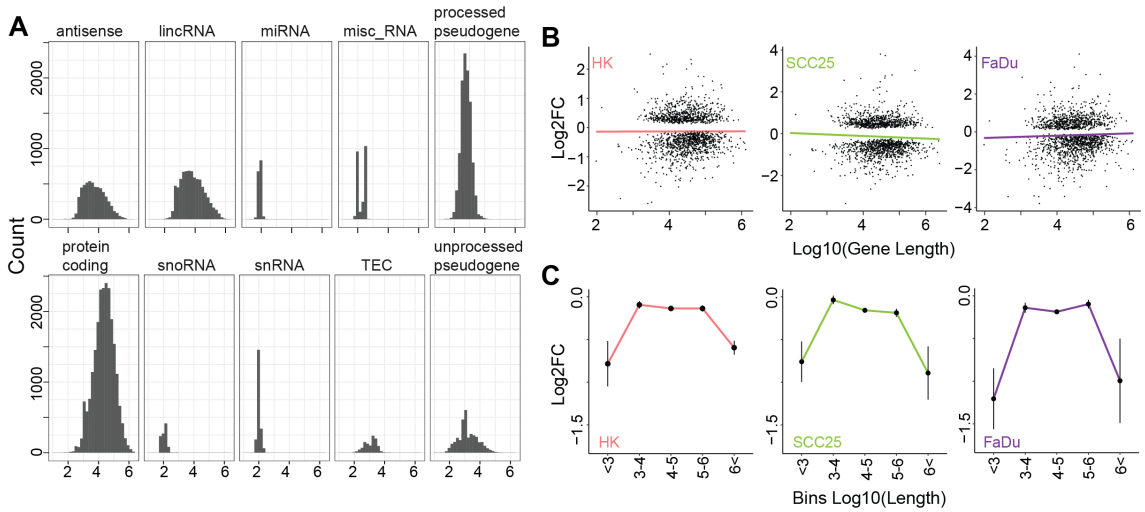

Figure 6.4: **Longer genes are not more susceptible to SRSF2 loss**
A Gene length of different classes of RNA. **B** Log_2 fold change (Log_2FC) of common differentially expressed genes seperated by length in HK (left) SCC25 (middle) and FaDu (right) . **C** Log_2 fold change (Log2FC) common differentially expressed genes binned by length in (HK) SCC25 (middle) and FaDu (right), error bars= SEM.

6.2.3 *SRSF2 depletion leads to increased DNA damage*

SRSF2 depletion causes cell cycle arrest and I have shown that this is in part due to altered transcription of cell cycle genes, which is in line with previously published work (Bapat et al., 2018, Xiao et al., 2007). DNA repair genes were also consistently down regulated after SRSF2 depletion, therefore, I next asked whether DNA damage increased when SRSF2 was depleted (Figure 6.5). Initially this was done by western blotting for γH2AX, which marks double strand breaks, indeed levels were increased when SRSF2 was depleted in all cell lines (Figure 6.5 A,B, C). Interestingly, levels of γH2AX increased across the time course and both SCC lines had increased signal 48 hours after SRSF2 depletion. To better quantify DNA damage that occured after SRSF2 depletion comet assays were performed. Here cells were permeabilised and DNA was denatured before undergoing electrophoresis (Figure 6.5 D). If DNA was fragmented a tail formed and the ratio of the "head" and "tail" area was used as a measure of DNA damage, called an olive tail moment (OTM).

Suprisingly, no increase in DNA damage was seen in HK cells after SRSF2 knock down, while in both cancer cell lines significantly more DNA damage was measured after SRSF2 depletion (Figure 6.5 E). Probing γH2AX levels had shown that there were more double strand breaks when SRSF2 was depleted in all cells. Therefore, that DNA damage was not observed via comet assay is likely to be due to the limitations of the assay, which includes several steps in which cells may be lost and that cells which are too damaged are difficult to quantify.

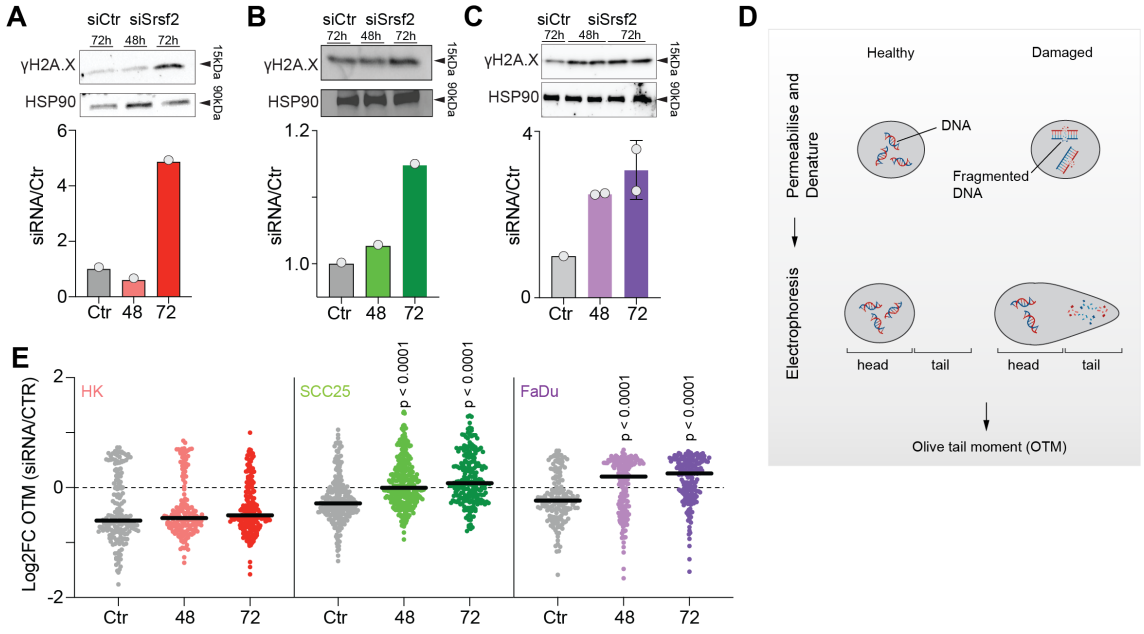

Figure 6.5: **SRSF2 depletion leads to increased DNA damage**
(**A-C**) Western blot of γH2AX staining (upper panel) and quantification (lower panel) in **A** HK, **B** SCC25 and **C** FaDu cells. **D** Scheme showing comet assay procedure and analysis. **E** Log_2 fold change (Log2FC) of relative olive tail moment (OTM) in control (Ctr), and 48 or 72 hours after SRSF2 knock down in HK (left), SCC25 (middle) and FaDu (right). Each point represents one cell, a minimum of 100 cells were counted per condition, line=median, p= unpaired t-test.

6.2.4 *DNA repair is impaired in the absence of SRSF2*

Having established that more DNA damage was present after SRSF2 depletion I also wanted to assay whether this was due to increased occurance of DNA damage, or whether DNA repair was impaired. In order to do this HK cells, either transfected with a scrambled control siRNA or the siRNA pool targetting SRSF2, were irradiated and repair was monitored over the course of an hour (Figure 6.6 A). This revealed that in SRSF2 depleted cells the level of damage was sustained, while in control cells damage was repaired within an hour (Figure 6.6 B).

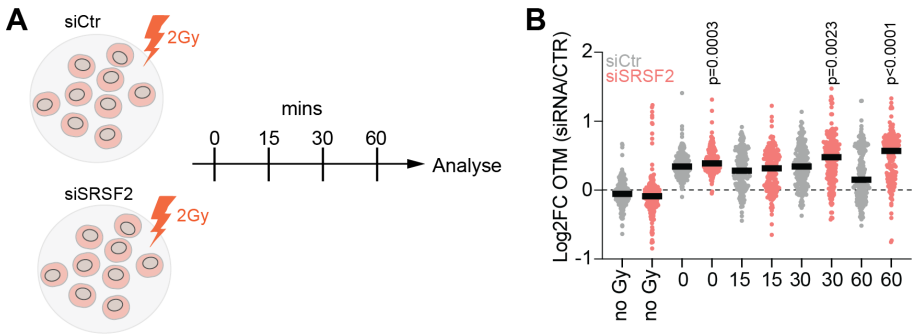

Figure 6.6: **SRSF2 is required to coordinate DNA repair**
A Experimental procedure used to assay DNA damage repair after SRSF2 depletion in keratinocytes (HK). **B** Log_2 fold change (Log2FC) olive tail moment (OTM), control (siCtr) and SRSF2 depleted (siSRSF2) from the same time point were compared, each point represents one cell, a minimum of 100 cells were measured per condition, bar= median, p= unpaired t-test.

6.2.5 *Summary, discussion and future directions*

DNA repair and replication genes are often organised as bi-directionally transcribed gene pairs (Adachi and Lieber, 2002, Trinklein et al., 2004, C. F. Xu, Chambers, and Solomon, 1997) and I have shown that SRSF2 is required to ensure they are properly expressed. Cells depleted for SRSF2 have a higher DNA damage burden, this was particularly prenounced in the SCC cell lines, which also have p53 mutations. This is at least partly due to an inability to repair damage. The cell cycle arrest that occurs after SRSF2 depletion may be exacerbated by the increased DNA damage as it is essential for cells to repair damage in order to maintain the fidelity of the genome and cell viability.

Under homeostatic conditions DNA damage results in cell cycle arrest until the damage is repaired, if the damage cannot be repaired cells undergo apoptosis. These events are in large part mediated by p53, "the guardian of the genome" (Lane, 1992). Previous work has suggested that SRSF2 depletion can be partially rescued by p53 inactivation (Xiao et al., 2007). However, I carried out experiments using primary cells, which have functional p53 and cancer cell lines with p53 mutations and observed largely the same phenotype and the same changes to gene expression in all cell lines. This suggests that the phenotypes observed are not due to a p53 mediated response.

The human genome project found that genes involved in biologically related processes are often found in close proximity to one another, reminiscent of the organisation of bacterial operons (Jacob and Monod, 1961, Britten and Davidson, 1969). Transcription of so called bi-directional gene pairs requires a sophisticated level of regulation to ensure that only the intended genes are transcribed as chromatin architecture and histone marks cannot be used as means of regulation, due to the region having to be in an open conformation to allow Pol II binding and transcription. Mechanisms modulating bi-directionally transcribed gene pairs have been elucidated (Flynn et al., 2011, Leighton, Waterfall, and Lis, 2008, Gilchrist et al., 2012). However, it is likely to vary depending on context, which has made it challenging to study. I hypothesise that genes organised bi-directionally are likely to be particularly vulnerable to perturbations in SRSF2 function due to high levels of activity at these loci.

Replication and transcription are essential cellular processes, however, they can also both cause DNA damage. SRSF2 is required for transcription and dysregulated transcription can lead to increased DNA damage. In conjunction with dysregulated transcription, SRSF2 is required for expression of repair genes. This makes SRSF2 interesting as the processes that go awry when SRSF2 is depleted exacerbate the amount of DNA damage that occurs. This offers an explanation as to why SRSF2 is an essential gene and why it is frequently mutated in cancer (H.-Y. Wang et al., 2001, André et al., 2017).

In the future it will be interesting to focus on specific types of DNA damage and identify where precisely they occur and whether there are specific regions where damage accumulates when SRSF2 is depleted. This could be assessed via exome sequencing following SRSF2 depletion. It is likely that damage accumulates in all transcribed regions and perhaps most at busy hubs of transcription, such as bi-directionally transcribed genes. Additionally, DNA-RNA hybrid immuno-

precipitation (DRIP-chip) sequencing (Chan et al., 2014) could be performed, this would improve understanding of whether specific regions or sequences are particularly susceptible to this type of DNA damage in the absence of SRSF2. I would also investigate whether mutations in specific genes occur more frequently when SRSF2 is depleted in order to address whether this may also be a reason why SRSF2 is often mutated in cancer and why this has been associated with clonal expansion (Kon et al., 2018, E. Kim et al., 2015, Yoshimi et al., 2019). Finally, proximity ligation assays focusing on components of the transcription and replication machinery may also provide further insights into whether these processes are colliding more frequently when SRSF2 is depleted.

Chapter 7

SRSF2 plays a unique role in mediating transcription

Comet assays were performed in collaboratotion with Dr. Ali Bakr.

7.1 *Aims of this chapter*

SRSF2 belongs to the SR protein family, a highly conserved family of proteins. They have been shown to function as a network which regulates itself in a cell type and cell state specific manner (Leclair et al., 2020, Lareau et al., 2007). Studies of the network have focused on regulation via alternative splicing. Whether SR proteins also collaborate to perform wider RNA processing functions remains unknown.

SRSF1 has been shown to be essential for proliferation and maintaining genome stability, however the molecular mechanisms through which it acts have not been fully elucidated. Interestingly, it also binds *RN7SK*, but has not been implicated in direct regulation of transcription (J. Wang, Takagaki, and Manley, 1996, L. Han et al., 1995, Ji et al., 2013). Furthermore, SRSF5 and SRSF2 are the only SR protein which contain a nuclear retention signal, thus rendering them predominantly nuclear (Botti et al., 2017). There is very little known about SRSF5's modes of action, interestingly I found that SRSF5 was consisitenly upregulated after SRSF2 depletion in all epithelial cell lines assayed. Therefore, SRSF2 depletion may lead to dysregulation of a network of SR proteins which are required for proper regulation of transcription.

Finally, splicing events mediated by SRSF2 may have downstream effects which are responsible for the phenotypes I have described in previous chapters. INTS3 has been indentified as a target for splicing mediated by SRSF2 (Yoshimi et al., 2019). The presence of mutant SRSF2 leads to mis-splicing and intron retention which culminates in degradation of the INTS3 transcript and down regulation of the gene. INTS3 belongs to the integrator complex, which has diverse roles in RNA processing and the DNA damage response (Welsh and Gardini, 2023). In order to ascertain whether the loss of INTS3 caused by the presence of mutant SRSF2 is causing the phenotype they observed the authors depleted INTS3 and investigated whether this recapitulated the affects of mutant SRSF2. INTS3 depletion only partly recapitulated what the authers observed when mutant SRSF2 was present (Yoshimi

et al., 2019), suggesting that SRSF2 has broad functions. I also found INTS3 to be mis-spliced and down regulated upon SRSF2 depletion in all cell lines assayed.

Whether the effects that I have associated with SRSF2 loss are unique to SRSF2 and not common functions to SR proteins or whether SRSF2 targets such as INTS3 mediate the phenotypes I have associated with SRSF2 remains unknown. The aims of this chapter were as follows:

1. Establish SRSF1, SRSF5 and INTS3 knock down in epithelial cell lines. To achieve this HK, SCC25 and FaDu cells will be transfected with siPools containing 20 different siRNAs targetting SRSF1, SRSF5 or INTS3. Transfection efficiency will then be assayed by measuring mRNA and protein levels, via qPCR and western blotting.

2. Assay proliferation, viability and DNA damage levels after SRSF1, SRSF5 and INTS3 knock down, to determine whether the effects seen after SRSF2 depletion are due to the collapse of a wider network. This will done via flow cytometry based assays and γH2AX staining and comet assays.

3. Decipher whether SRSF2's role in transcription is unique. This will be done by western blotting for the Ser2 and Ser5 phosphorylated forms of Pol II.

7.2 *Results*

7.2.1 *SR proteins are required for proliferation*

In previous chapters I have shown that SRSF2 is required for cell division and DNA repair and that it acts by regulating transcription. However, SRSF2 acts independently or through a network which collapses is unknown. In order to address this I knocked down two further SR proteins SRSF1 and SRSF5 and INTS3.

I depleted SRSF1, SRSF5 and INTS3 using siRNA pools in HK, SCC25 and FaDu cells for 48 and 72 hours (Figure 7.1 A, B). RNA and protein levels were signifcantly reduced compared to control in all cell lines for all proteins (Figure 7.1 C-E).

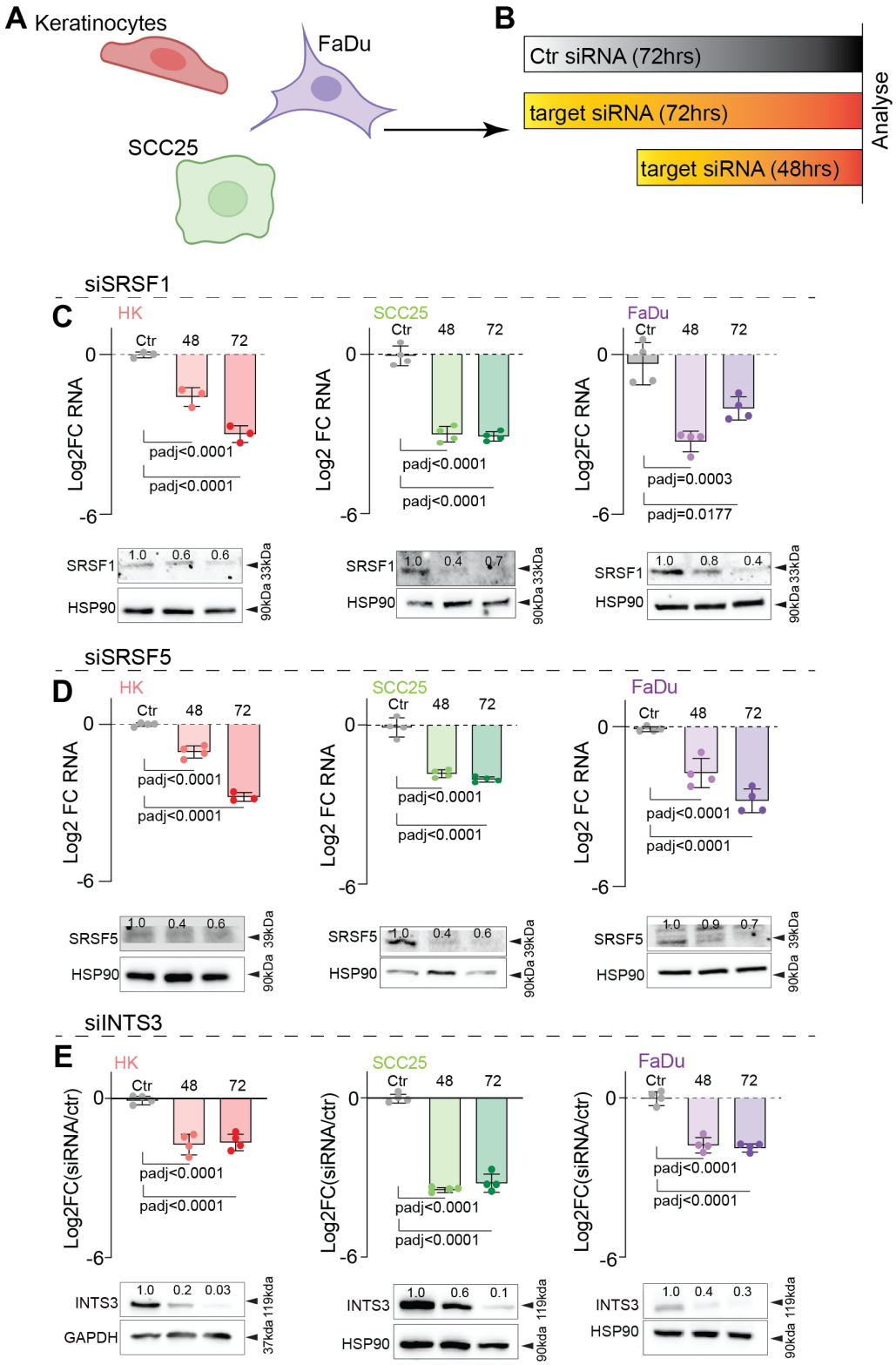

Figure 7.1: Caption on the following page

Figure 7.1: **Establishing knock down of SR protein family members and SRSF2 targets**
A Scheme showing cell lines used **B** and experimental outline for establishing knock down. **C** Log2 fold change (Log2FC) SRSF1 RNA levels (top panel) and protein (bottom panel) in control (Ctr) and after knock down for 48 or 72 hours in HK (left), SCC25 (middle) and FaDu (right). **D** Log2 fold change (Log2FC) SRSF5 RNA levels (top panel) and protein (bottom panel) in control (Ctr) and after knock down for 48 or 72 hours in HK (left), SCC25 (middle) and FaDu (right). **E** Log2 fold change (Log2FC) INTS3 RNA levels (top panel) and protein (bottom panel) in control (Ctr) and after knock down for 48 or 72 hours in HK (left), SCC25 (middle) and FaDu (right). (**C**-**E**) loading control= HSP90/GAPDH. padj= Dunnett's multiple comparisons test

SRSF2 depletion consistently caused cell cycle arrest in epithelial cell lines (Chapter 4), I therefore began by ascertaining whether the same was true when SRSF1, SRSF5 or INTS3 was depleted (Figure 7.2). SRSF1 knock down led to a reduction in the proportion of cells in S phase in all cell lines. In both SCC cell lines the proportion of cells in subG0 was increased after 72 hours, however, this was not the case in HK cells, where cells accumulated in G1/G0 (Figure 7.2 A). Directly assying the proportion of cells proliferating via EdU incorporation, confirmed that SRSF1 depletion led to a signficant reduction in the proportion of cells dividing in all cell lines (Figure 7.2 B). These results are in line with previous reports of SRSF1 being required for proliferation in other cell types (Y. Fu et al., 2013, Wagner and Frye, 2021).

SRSF5 depletion led to a reduction in the proportion of cells in SubG0 and S phase in HK cells, but did not consistently change the proportion of cells in any stage of the cell cycle in either of the SCC cell lines (Figure 7.2 C). Furthermore, assaying proliferation directly confirmed that this was unchanged in alll cell lines following SRSF5 depletion (Figure 7.2 D).

Finally, INTS3 depletion led to a reduction in the proportion of cells in S phase in all cell lines (figure 7.2 E,F). Furthermore, after 72 hours of INTS3 depletion the proportion of cells in SubG0 was also significantly increased in all cell lines.

The consequences of INTS3 knock down most closely resemble those observed after SRSF2 depletion, therefore loss of INTS3 after SRSF2 knock down is likely to contribute to the cell cycle arrest that was observed.

Despite causing cell cycle arrest SRSF2 depletion only led to a minor decrease in cell viability and did not trigger apoptosis. I also investigated whether cell viability was altered after depletion of SRSF1, SRSF5 and INTS3 in HK, SCC25 and FaDu cells (Figure 7.3). Surprisingly, HK cells were the least sensetive to depletion of SRSF1, SRSF5 and INTS3. SRSF1 depletion did not alter viability of HK cells while apoptosis was increased in both SCC25 and FaDu cells (Figure 7.3 A). Following SRSF5 depletion apoptosis was also increased in SCC25 and FaDu cells, but to a lesser extent compared to SRSF1 depletion (Figure 7.3 B). INTS3 knock down led to increased cell death and apoptosis (Figure 7.3 C).

The effects on cell viability after depletion of SRSF1, SRSF5 and INTS3 were more variable than depletion of SRSF2. Interestingly, both cancer cell lines appeared more sensitive to depletion of these proteins and the reduction in cell viability is likely to be causing the cell cycle arrest, which suggests that the mechanism through which they are acting is not the same as the mechanism through which SRSF2 acts.

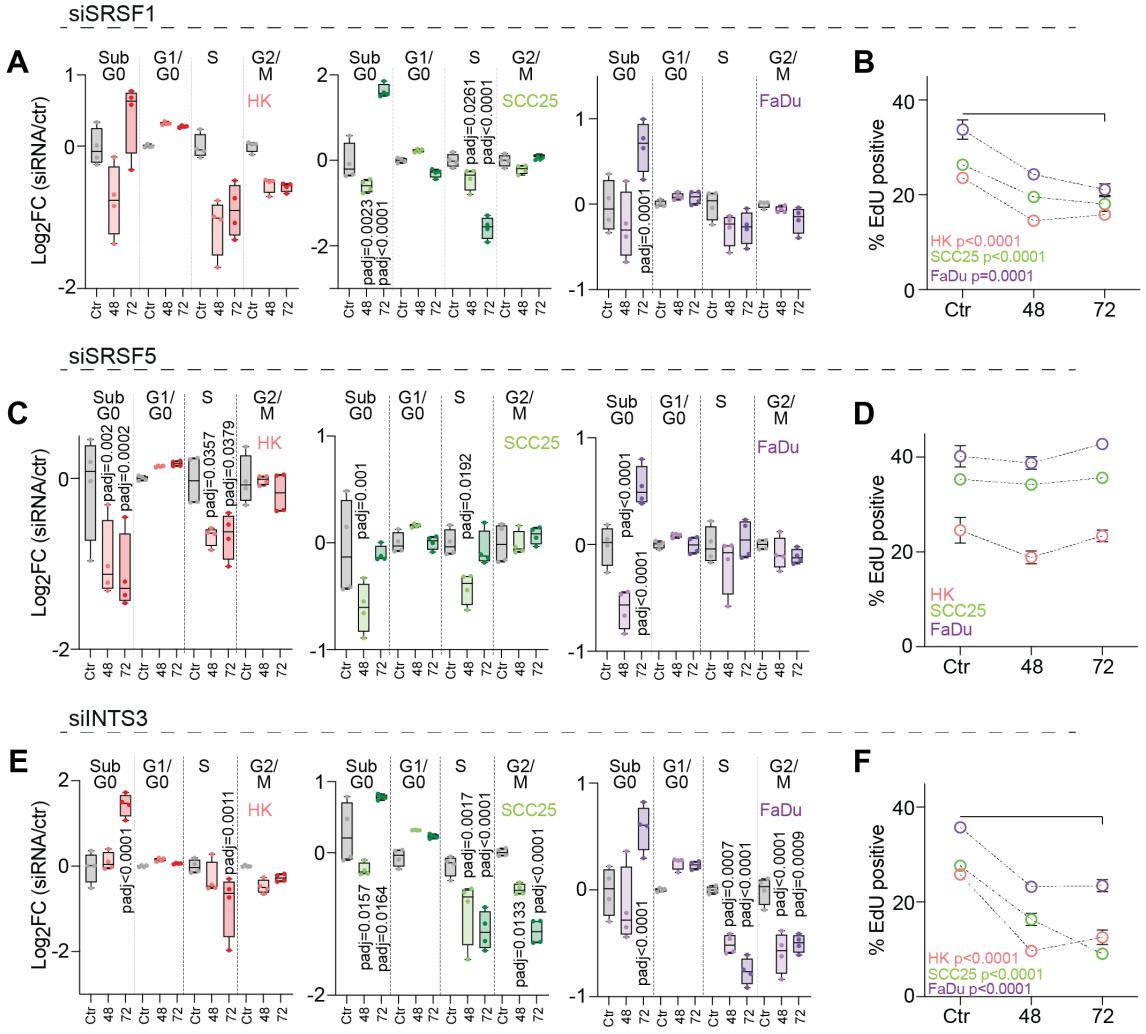

Figure 7.2: **Not all SR proteins are required for proliferation**
A Log2 fold change (Log2FC) of the number of cells in the phases of the cell cycle after SRSF1 depletion for 48 or 72 hours in HK (left), SCC25 (middle) and FaDu (right) cells. **B** Percentage (%) EdU positive cells in HK (pink), SCC25 (green) and FaDu (purple) cells treated with a scrambled siRNA (Ctr) or an sipool targetting SRSF1 for 48 and 72 hours. **C** Log2 fold change (Log2FC) of the number of cells in the phases of the cell cycle after SRSF5 depletion for 48 or 72 hours in HK (left), SCC25 (middle) and FaDu (right) cells. **D** Percentage (%) EdU positive cells in HK (pink), SCC25 (green) and FaDu (purple) cells treated with a scrambled siRNA (Ctr) or an sipool targetting SRSF5 for 48 and 72 hours.**E** Log2 fold change (Log2FC) of the number of cells in the phases of the cell cycle after INTS3 depletion for 48 or 72 hours in HK (left), SCC25 (middle) and FaDu (right) cells. **F** Percentage (%) EdU positive cells in HK (pink), SCC25 (green) and FaDu (purple) cells treated with a scrambled siRNA or an sipool targetting INTS3 for 48 and 72 hours. p= unpaired t-test.

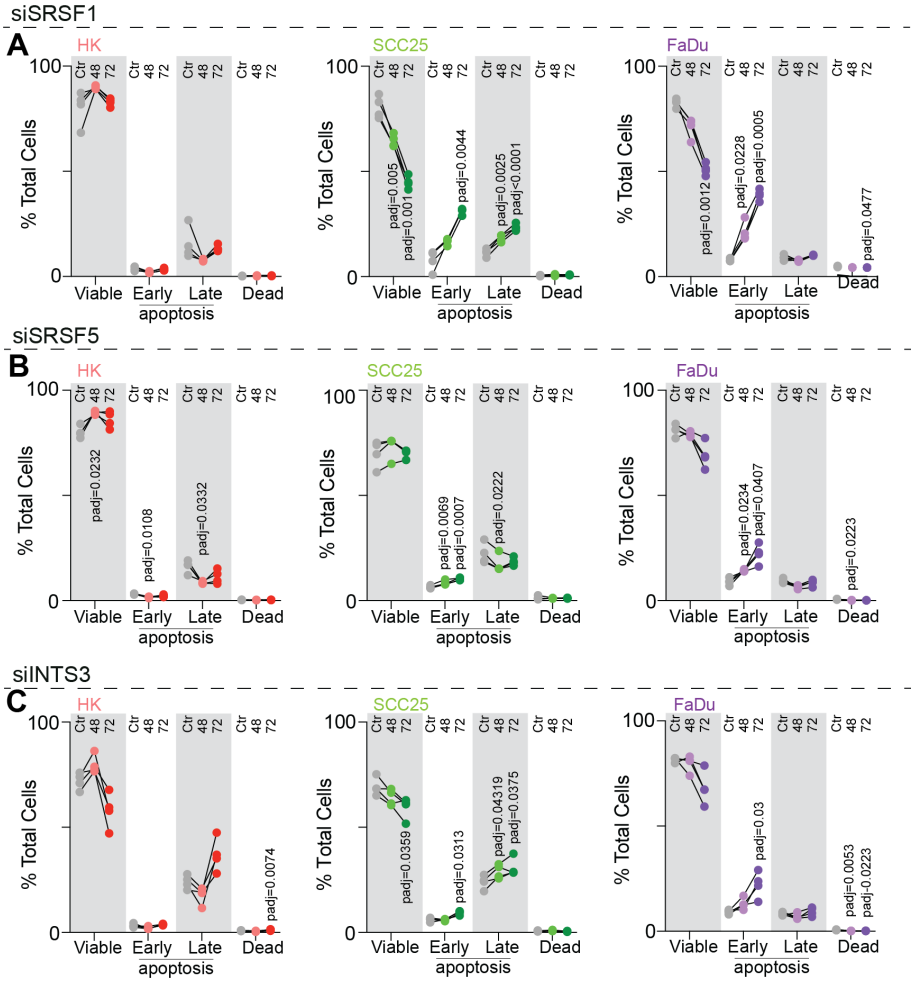

Figure 7.3: **SR proteins are required for viability of cancer cells**
A Percentage (%) viable, apoptotic and dead HK (left), SCC25 (middle) and FaDu (right) cells transfected with control siRNA (Ctr) or an sipool targetting SRSF1 for 48 or 72 hours. **B** Percentage (%) viable, apoptotic and dead HK (left), SCC25 (middle) and FaDu (right) cells transfected with control siRNA (ctr) or an sipool targetting SRSF5 for 48 or 72 hours. **C** Percentage (%) viable, apoptotic and dead HK (left), SCC25 (middle) and FaDu (right) cells transfected with control siRNA (Ctr) or an sipool targetting INTS3 for 48 or 72 hours.padj= Dunnett's multiple comparisons test.

7.2.2 *SRSF2 uniquely leads to DNA damage in epithelial cells*

SRSF2 depletion led to an increase in DNA damage and repair was impaired (Chapter 6), which was at least in part due to improper transcription of repair genes. SRSF1 depletion has also been shown to lead to DNA damage due ot the formation of R-loops (X. Li and Manley, 2005), while, INTS3 is part of the SOSS complex, which is important for sensing single stranded DNA and promoting DNA strand break repair (Y. Jia et al., 2021, Welsh and Gardini, 2023). Whether SRSF5 is important for the DNA damage response remains unknown, however, as the only other SR protein which does not shuttle between the nucleus and the cytoplasm it may play a role in mediating the DNA damage response.

SRSF1, SRSF5 and INTS3 were depleted for up to 72 hours in HK, SCC25 and

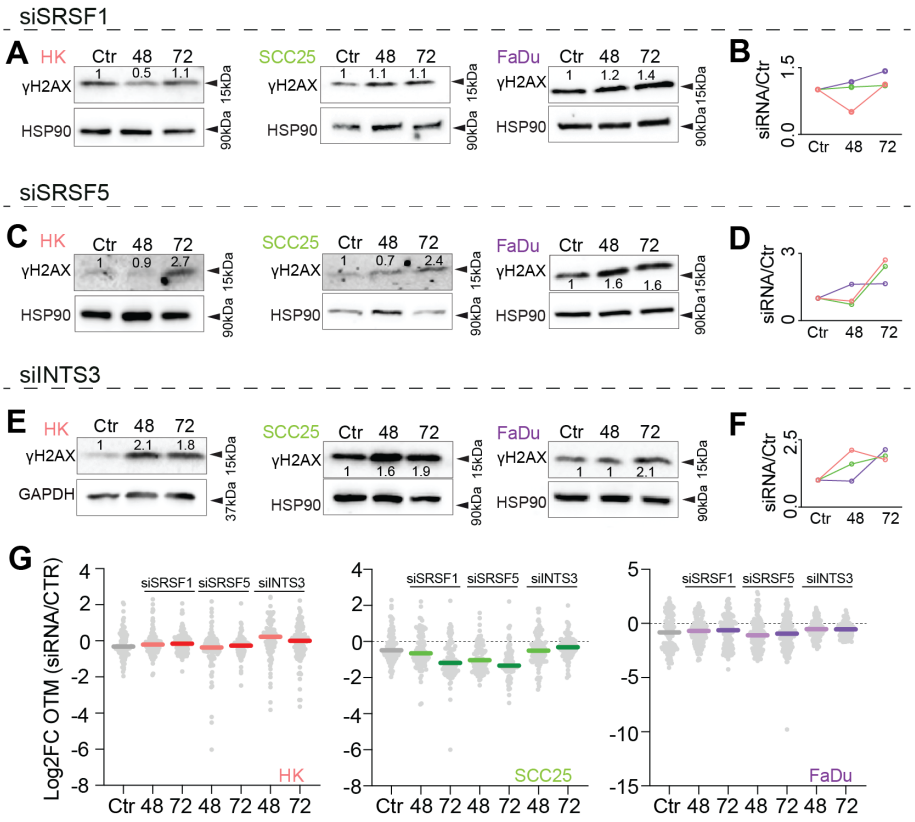

Figure 7.4: **Depletion of select SR proteins and INTS3 does not cause DNA damage**
A Western blot for γH2AX in HK (left) SCC25 (middle) and FaDu (right) in cells treated with a scrambled siRNA (Ctr) or an sipool targetting SRSF1 for 48 or 72 hours. **B** Quantification of relative protein amount in HK (pink), SCC25 (green) and FaDu (purple). **C** Western blot for γH2AX (top) in HK (left) SCC25 (middle) and FaDu (right) in cells treated with a scrambled siRNA (Ctr) or an sipool targetting SRSF5 for 48 or 72 hours. **D** Quantification of relative protein amount in HK (pink), SCC25 (green) and FaDu (purple). **E** Western blot for γH2AX (top) in HK (left) SCC25 (middle) and FaDu (right) in cells treated with a scrambled siRNA (Ctr) or an sipool targetting INTS3 for 48 or 72 hours. **F** Quantification of relative protein amount in HK (pink), SCC25 (green) and FaDu (purple). **G** Log2 fold change (Log2FC) of olive tail moment (OTM) in HK (left), SCC25 (middle) and FaDu (right) cells after knock down of SRSF1 (left), SRSF5 (middle) and INTS3 (right), bar= median. HSP90= loading control.

FaDu cells and DNA damage was probed via western blotting for γH2AX and comet assays (Figure 7.4). SRSF1 depletion did not lead to an increase in γH2AX levels in any of the cell lines tested (Figure 7.4 A, B). Interestingly, SRSF5 and INTS3 depletion led to an increase in γH2AX in all cell lines, however, the effect was milder than that observed after SRSF2 depletion (Figure 7.4 C-F). Comet assays did not show any significant increase in DNA damage in any of the cell lines (7.4 G), which was not the case when SRSF2 was depleted. This suggests that the level of DNA damage caused by SRSF2 depletion is due to SRSF2 having an essential role in mediating DNA repair via transcription.

7.2.3 *Pol II phosphorylation is uniquely mediated by SRSF2*

Finally, SRSF2 is important for regulating transcription, specifically mediating PTEFb function, which is essential for pause release and active transcription (Ji et al., 2013, Lin, Coutinho-Mansfield, et al., 2008). I have shown that SRSF2 is important for this process genome wide and that bi-directionally transcribed genes are particularly vulnerable to SRSF2 loss (Chapters 5 & 6). However, whether this function is unique to SRSF2 remains unknown. I therefore assessed the level of Ser2 phosphorylated Pol II following knock down of SRSF1, SRSF5 or INTS3 (Figure 7.5). SRSF1 depletion led to an increase in Ser2 phosphorylation in HK and FaDu cells, but levels were unchanged in SCC25 cells (Figure 7.5 A-C). SRSF5 depletion did not alter Ser2 phosphorylation levels in HK cells and led to a mild increase in phosphorylation in SCC25 and FaDu cells (Figure 7.5 D-F). Finally, INTS3 depletion led to a reduction in Ser2 phosphorylation levels in HK cells, but this was not the case in SCC25 and FaDu cells, where the levels were unchanged (Figure 7.5 G-I). These results are in line with previous work, which hypothesised that SRSF2 plays a unique role in regulating transcription elongation via PTEFb (Ji et al., 2013).

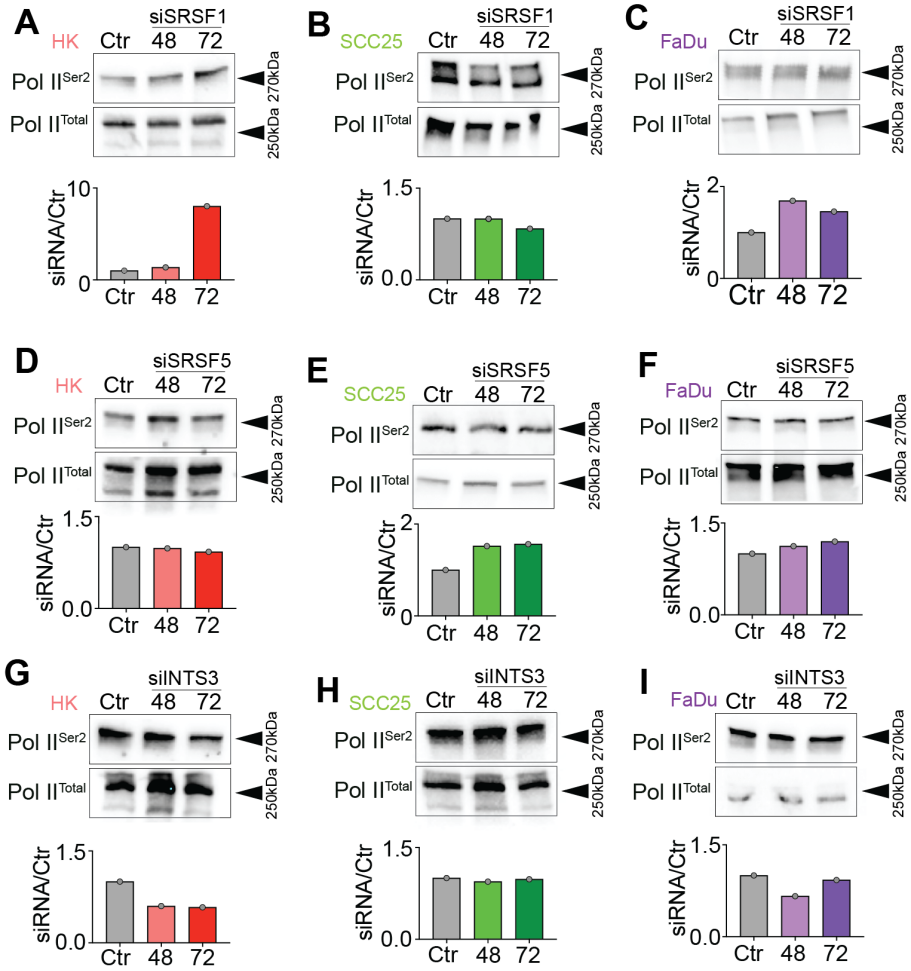

Figure 7.5: **Pol II phosphorylation is exclusively regulated by SRSF2**
(**A-C**) Western blot for total Pol II (Pol II^{Total}) and ser2 phorphorylated Pol II (Pol II^{ser2}) in **A** HK **B** SCC25 and **C** FaDu cells treated with a scrambled siRNA (Ctr) or an sipool targetting SRSF1 for 48 or 72 hours (upper panel) and quantification (lower panel). (**D-F**) Western blot for total Pol II (Pol II^{Total}) in **D** HK, **E** SCC25 and **F** FaDu cells treated with a scrambled siRNA (Ctr) or an sipool targetting SRSF5 for 48 or 72 hours (upper panel) and quantiifcation (lower panel). (**G-H**) Western blot for total Pol II (Pol II^{Total}) in **G** HK, **H** SCC25 and **I** FaDu cells treated with a scrambled siRNA (Ctr) or an siPool targetting INTS3 for 48 or 72 hours (upper panel) and quantiifcation (lower panel.

7.2.4 *Summary, discussion and future directions*

I established knock down of INTS3, a splicing target of SRSF2 and two additional SR proteins; SRSF1 and SRSF5 in order to ascertain whether the functions associated with SRSF2 are unique or whether SRSF2 acts as part of a wider network, which collapses when any single member is lost.

Knock down of SRSF5 did not alter proliferation, while knock down of SRSF1 and INTS3 did lead to a reduction in proliferation. INTS3 depletion additionally lead to an increase in the proportion of cells in SubG0 72 hours after transfection, which was similar to the alterations to the cell cycle that occured after SRSF2 depletion. Interestingly, depletion of SRSF1, SRSF5 and INTS3 had stronger effects on apoptosis in the SCC cell lines than in the primary HK cells. This was not the case when SRSF2 was depleted, where the changes to viability were minor in all cell

lines.

Quantifying DNA damage after depletion of SRSF1, SRSF5 or INTS3 revealed that DNA damage does not accumulate to the same extent as it does following SRSF2 depletion, despite SRSF1 and INTS3 both being implicated in genome stability (X. Li and Manley, 2005, Y. Jia et al., 2021). This may be because these genes do not cause both increased damage and prevent repair simultaneously, as is the case for SRSF2. Finally, I assessed the phosphorylation status of Ser2 of the CTD of Pol II following knock down of SRSF1, SRSF5 and INTS3 and largely observed no changes. INTS3 depletion led to a reduction in Ser2 phosphorylation in HK cells, but this was not consistent in the two cancer cell lines.

Theses findings confirm that SRSF2 has unqiue functions in mediating PTEFb activity and that the absence of SRSF2 has a direct effect on DNA damage burden, which ultimately leads to cell cycle arrest but does not trigger apoptosis. It is interesting to note that SRSF1, SRSF5 and INTS3 depletion had more severe affects on cell viability and that these were particularly strong in the cancer cell lines. In the future it would be interesting to investigate which molecular processes are regulated by these proteins and why they are essential for cancer cell viability. Furthermore, investigating the effects on the SR protein network when SRSF1 or SRSF5 are depleted and comparing the changes with those seen when SRSF2 is depleted could provide further insights into the SR protein network. Finally, it would also be interesting to look at all SR proteins in a systematic way in order to decipher their functions in further contexts and better understand the functions that confer their essential nature.

Chapter 8

Reduced SRSF2 levels allow survival with increased DNA damage

Generation and initial characterisation of the CRISPR-cas9 clones was done with Leonie Arnetzl, a Master's student whom I supervised and part of this work contributed to her Master's thesis (Arnetzl, 2021). RNA sequencing data analysis was assisted by Dr. Anke Heit-Mondrzyk. Comet assays were performed in collaboration with Dr. Ali Bakr. Orthotopic tongue injections were done with Dr. Sylvain Delaunay.

8.1 *Aims of this chapter*

SRSF2 is mutated in approximaltey 30% of MDS cases and 2% of all cancers (Papaemmanuil et al., 2013, Yamaguchi, Shibata, and Handa, 2013, André et al., 2017). The field has largely focused on how a P95H point mutation alters splicing (J. Zhang et al., 2015, E. Kim et al., 2015, Bapat et al., 2018, Thol et al., 2012, Rahman et al., 2020, Kon et al., 2018). This has led to a better understanding of how SRSF2's RNA binding affinity is altered and aberrant splicing patterns have been characterised (Kon et al., 2018, E. Kim et al., 2015). Interestingly, the alterations to splicing are not consistent between mouse and human (E. Kim et al., 2015, Ilagan et al., 2015, Qiu et al., 2016, Shirai et al., 2015, Mupo et al., 2017), which has made understanding how mutant SRSF2 contributes to disease progression difficult to understand.

SRSF2 belongs to a group of splicing factors that are commonly mutated in MDS, these mutations are predominantly heterozygous and mutually exclusive (Dvinge et al., 2016). This highlights the essential nature of these proteins and suggests that their contributions to carcinogenesis may converge on a common mechanism. Indeed mutations in SRSF2 and U2AF1 were both shown to increase genomic instability via increased R-loops which form as a result of increased Pol II stalling (L. Chen et al., 2018).

Despite advances in understanding the role of SRSF2 in carcinogenesis there

are still many open questions. For example, how SRSF2's role in regulating Pol II activity and gene expression contributes to cancer development and progression, furthermore, as mutant SRSF2 is usually observed heterozygous it is unknown whether reducing SRSF2 levels further than 50% is possible in the context of cancer. The aims of this chapter were as follows:

1. Determine the effects of stable SRSF2 reduction on cell survival. This will be done by using CRSISPR-cas9 mediated genome editing to stably reduce SRSF2 expression.

2. Assess essential SRSF2 functions. This will be done by measuring cell cycle, DNA damage and changes to gene expression in clones with stably reduced SRSF2 levels.

3. Assay clones with lower levels of SRSF2 for their ability to form tumours *in vivo*. This will be done by carrying out orthotopic tongue injections and monitoring tumour growth.

8.2 *Results*

8.2.1 *Stable SRSF2 depletion via CRISPR-cas9*

The SRFF2 gene was perturbed using CRISPR-cas9 mediated gene editing (Hsu, Lander, and F. Zhang, 2014) in order to investigate the effects of stable reduction of SRSF2 levels on cell survival. Three guides were designed, two that targeted exon 1 of the SRSF2 gene, (guide 2 and 4) and one that targetted the 3'UTR (guide 5) (Figure 8.1 A, B). Cells were also transfected with an empty vector, which contained the cas9 in order to assess the clonogenicity of the parental cell line used and control for any off target mutations that may be introduced due to the presence of the cas9.

285 cells were sorted from cells transfected with one of the three guides. Analysis of clones that grew out after transfection with guide 2 revealed that no clone with mutations in the SRSF2 gene survived (Figure 8.1 C). Analysis of clones that grew out after transfection with guide 4, which targeted the middle of exon 1, revealed that one clone (#0) with a heterozygous insertion led to a frameshift and premature termination codon and led to a 50% reduction in SRSF2 protein levels survived (Figure 8.1 C, D). Finally, from guide 5, which targeted the 3'UTR, two clones (#1 and #2) with a 40% reduction in SRSF2 levels were viable. Clones #1 and #2 had homozygous insertions, of one and two nucleotides respectively (Figure 8.1 B, C, D).

Clonogenicity of control cells was 13%, while when the SRSF2 gene was targetted clonogenecity was reduced to approximately 1%. This highlights that SRSF2 is essential and reducing levels by more than half is deleterious for cell growth and survival, even in the context of cells originating from a cancer.

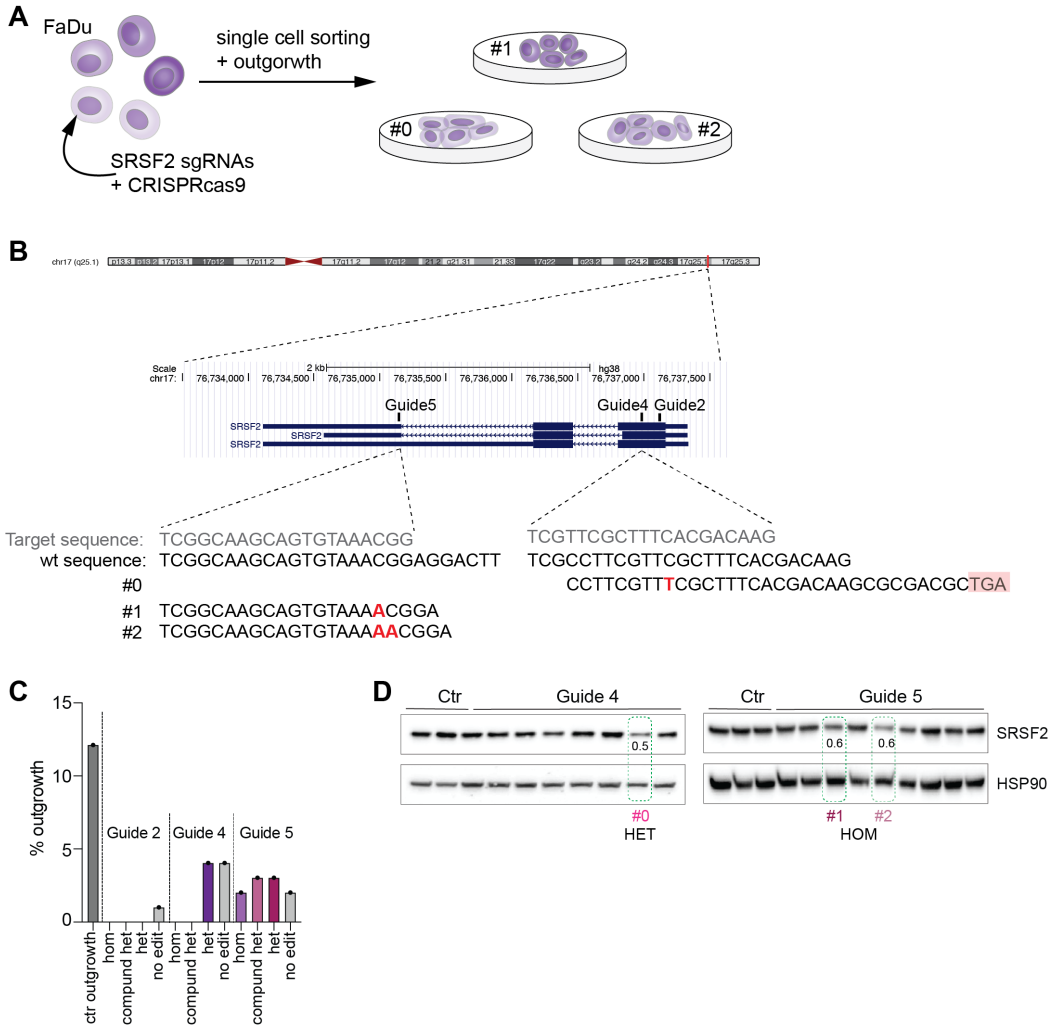

Figure 8.1: **SRSF2 deletion is deleterious for cell survival**
A Cells were transfected with CRISPR-cas9 and a guide targeting different regions of the SRSF2 gene, cells were then single cell sorted.**B** Genome browser view of where guides target the SRSF2 gene and the mutations that lead to lower levels of SRSF2, stop codon and insertions are highlighted in red. **C** Percentage (%) of clonal outgrowth. **D** Western blots highlighting clones with reduced SRSF2 levels, number is protein level relative to control, loading control= HSP90.

8.2.2 *Epithelial cell identity altered in clones with reduced SRSF2 levels*

Surviving clones were characterised in order to elucidate SRSF2's essential functions (Figure 8.2 A). The morphology of clone #0 closely resembled that of the control, while clone #1 and #2 were flatter and squarer. Quantifiying cell size confirmed that cells from clones with reduced SRSF2 levels were larger than those of the control clone (Figure 8.2 B), reminiscent of the increased cell size that occured when functional SRSF2 was lost in the mouse epidermis and suggests that cells may also have cell division defects.

SRSF2 has previously been linked to ECAD splicing in head and neck cancer (Sharma et al., 2011). Indeed all clones showed reduced ECAD expression at both the mRNA and protein levels. (Figure 8.2 C). ECAD is a tumour supressor and a

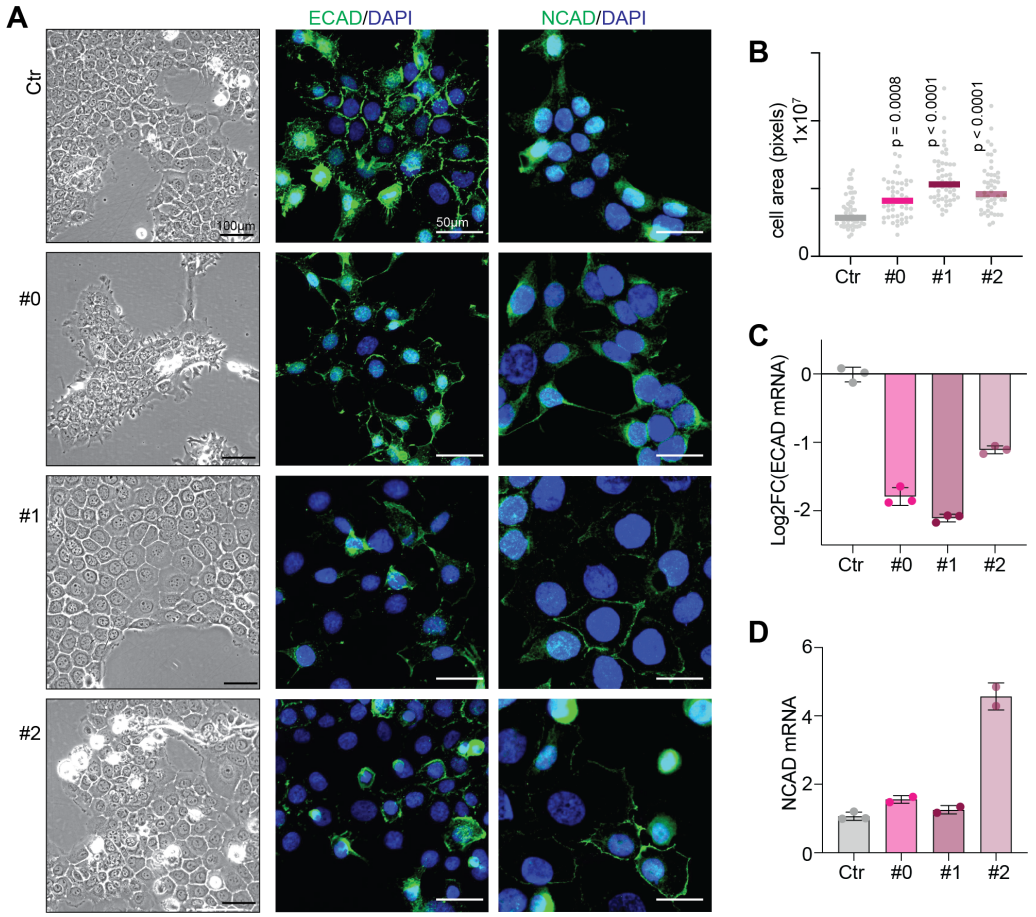

Figure 8.2: **Clones with lower SRSF2 levels are larger**
A Bright field images in a control clone (Ctr) and clones with stably reduced SRSF2 levels, #0. #1 and #2 (left) Immunofluorescence imaging of E-cadherin (ECAD) (middle) and N-Cadherin (NCAD) (right), counter stained with DAPI. Representative images of three technical replicates. **B** Quantification of cell size, each point represents one cell, 100 cells were measured, p= unpaired T-test. (**C, D**) Log_2 fold change (Log2FC) of **C** ECAD and **D** NCAD mRNA levels in clones with reduced SRSF2 levels #0, #1 and #2 relative to Ctr and normalised to 18S, each point represents one RNA extraction and the average of three technical replicates, shown is mean± SD.

reduction in ECAD expression has been linked to epithelial to mesenchymal transition (EMT) and associated with tumour aggressivity as this is an essential step in metastasis (Hanahan and Weinberg, 2000). During EMT ECAD is replaced by N-cadherin (NCAD). Staining for NCAD did not reveal that a switch had taken place, however, when NCAD mRNA levels were assayed clone #2 expressed twice as much as the control clone (Figure 8.2 D). The changes to cell morphology and expression of key epithelial markers suggests that SRSF2 is required for maintaining cell identity.

8.2.3 *Stable reduction of SRSF2 disrupts the cell cycle, but does not perturb cell growth*

SRSF2 levels were not reduced by more than half in any of the surviving clones, this is likely to be because if levels are reduced further cell growth is attenuated,

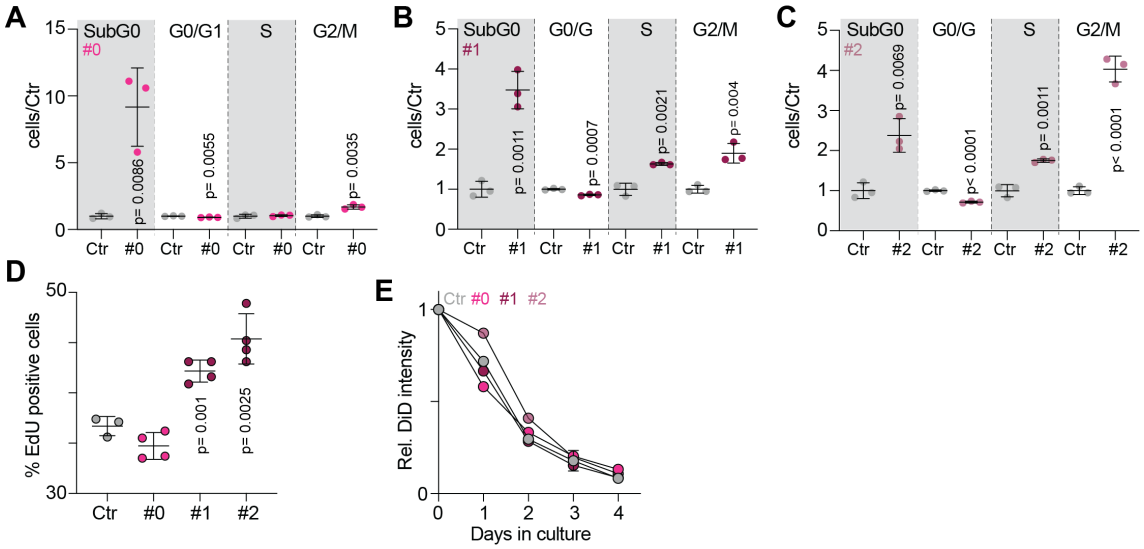

Figure 8.3: **Cell cycle arrest can be overcome when SRSF2 is stably depleted**
A-C Proportion of cells in the phases of the cell cycle, relative to control in clone (**A**) #0,(**B**) #1 and (**C**) #2, each point represents 10,000 events.**D** Percentage (%) of EdU positive cells in a control clone (Ctr), and clones with reduced SRSF2 levels, #0, #1 and #2, p= unpaired t-test. Shown is mean± SD. **E** 1,1'-Dioctadecyl-3,3,3',3'-Tetramethylindodicarboncyanine (DiD) intenstiy across 5 days normalised to day 0, each point is the mean of 10,000 events.

which indicates that a minimum level of SRSF2 is sufficient for cells to divide normally. In order to test this the cell cycle was assayed in clones with reduced SRSF2 levels and a control clone. This revealed that the proportion of cells in SubG0 was increased in all clones with reduced SRSF2 levels, as was the case when SRSF2 was acutely depleted (Figure 8.3 A). Clones #1 and #2 also showed an increase in the proportion of cells in S phase and G2/M. Along with the observed increased cell size this suggests that the clones may need longer to complete the cell cycle, as has been observed in other cell types (Bapat et al., 2018, Xiao et al., 2007). I also assayed the propotion of cells proliferating directly by EdU incorporation and this revealed that clone #1 and #2 showed an increase in the percentage of cells in S phase, in line with the cell cycle analysis (Figure 8.3 B). However, when I assayed division rates I did not observe any significant changes (Figure 8.3 C). Therefore, survivng clones can proliferate at a comparable rate to control with 50% less SRSF2.

The subG0 phase of the cell cycle represents cells with fragmented DNA. This may be due to cells undergoing apoptosis. In order to assess this viability was assayed, this revealed that viability was compromised in clone #0 and #1 (Figure 8.4 A-C). However, cells were not undergoing apoptosis, which was expected as these clones survived.

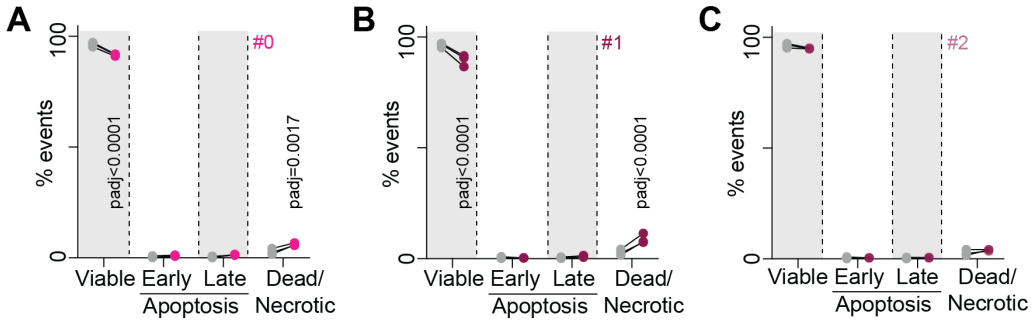

Figure 8.4: **Clones with reduced SRSF2 levels are viable**
(**A-C**) Percentage (%) of viable, early apoptotic, late apoptotic and dead cells in control (gray) and **A** clone #0 , **B** #1 and **C** clone #2, each point represents 10,000 events. padj= two-way ANOVA.

8.2.4 Reduced SRSF2 levels allow cells to tollerate more DNA damage

Clones with reduced SRSF2 levels were able to overcome the cell cycle arrest that was observed after accute SRSF2 depletion, however, all clones showed an increase in the proportion of cells in G2/M. This can occur when DNA repair has to take place, which maybe the case as I previously showed that SRSF2 is required for transcription of DNA repair genes.

To assess whether clones with reduced SRSF2 levels accumulated more DNA damage the amount of γH2AX and R-loop foci were evaluated (Figure 8.5 A). Quantification of nulcei with more than 6 γH2AX foci revealed that the number of double strand breaks was increased in all clones (Figure 8.5 B). Interestingly, only clones #1 and #2 had increased levels of R-loops (Figure 8.5 C). To better quantify the extent of damge in all clones comet assays were performed. All clones had significantly more DNA damage than the control clone (Figure 8.5 D). In order to better understand the types of DNA damage that were occurring the number of unique fusion genes within each clone was determined (Figure 8.5 B). Clone #1 and #2 had significantly more unique fusion genes than the control clone. Fusion genes form when chromosomal rearrangements occur and this can happen as a result of improper resolution of double strand breaks, highlighting how increased DNA damage can be advantageous during cancer development.

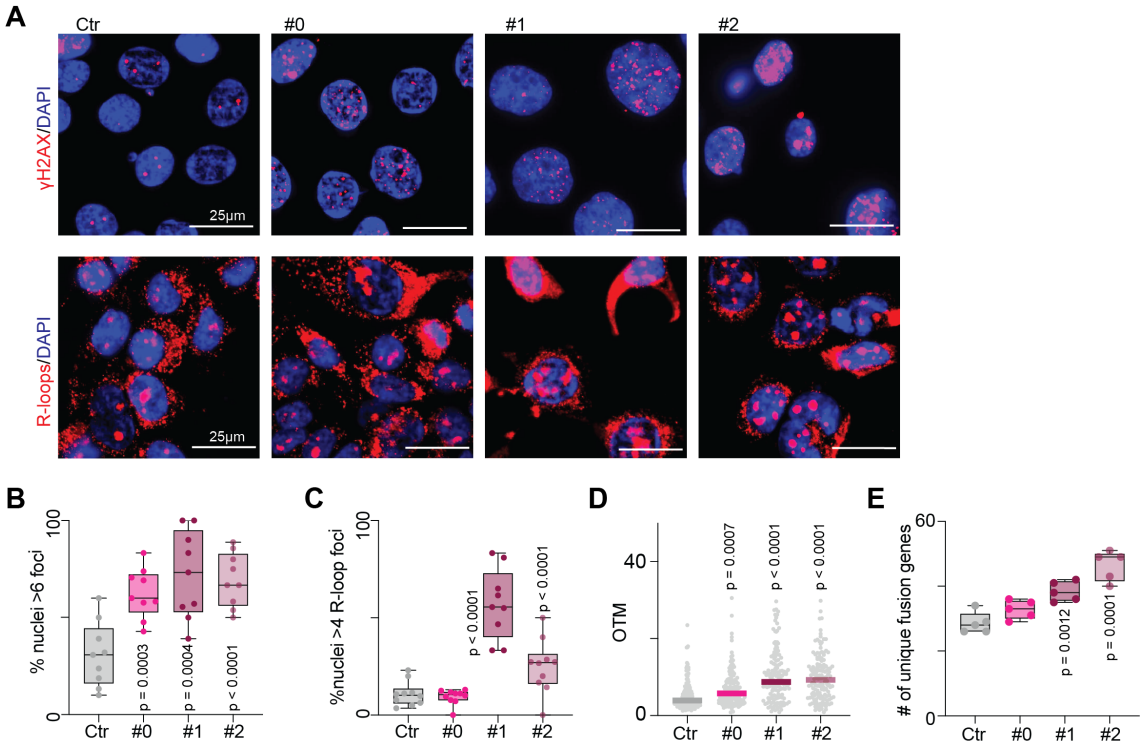

Figure 8.5: **A reduction in SRSF2 levels allows DNA damage to persist**
A Immunofluorescence of γH2AX (top) and R-loops (bottom) staining in a control clone (Ctr) and clones with reduced SRSF2 levels; #0, #1 and #2. **B** Quantification of the percentage (%) of nuclei with more than 6 γH2AX foci in a control clone (Ctr) and clones with reduced SRSF2 levels; #0, #1 and #2. **C** Percentage (%) of nuclei with more than four R-loop foci in a control clone (Ctr) and clones with reduced SRSF2 levels; #0, #1 and #2. **D** Olive tail moment (OTM) from a control clone (Ctr) and clones with stable SRSF2 depletion; #0, #1 and #2, each point is a cell, a minimum of a 100 cells were counted per condition, bar= median. **E** The number of unique fusion genes in a control clone (Ctr) and clones with reduced SRSF2 levels; #0, #1 and #2, each point represents one RNA extraction. (**B, C, E**) Box plots show minimum, first quartile, median, third quartile, and maximum. (**B-E**) p= unpaired t-test

8.2.5 *Reduced SRSF2 levels do not inhibit tumour growth in vivo*

SRSF2 mutations are associated with poor prognosis in AML (Yoshimi et al., 2019). However, how mutant SRSF2 contributes to disease development and progression remains largely unknown. In order to assess whether clones with reduced SRSF2 levels are able to form tumours cells from clones with reduced SRSF2 levels were injected orthotopically into mice and tumour growth was monitored for 28 days (Figure 8.6 A). The number of tumours that developed from clone #0 and #2 was lower than the control and clone #1 (Figure 8.6 B). The cell line used is known to metastasis to the lymph nodes when used for orthotopic injection experiments (Pascual et al., 2017, Delaunay et al., 2022), therefore the number of metastasis was also monitored. All clones metastasised to the lymph nodes, clones #0 and #2 formed fewer metastasis than control and clone #1 (Figure 8.6 C). Although there were differences between clones with reduced SRSF2 levels and the control clone these differences are likely to be accounted for by genetic differences leading to sur-

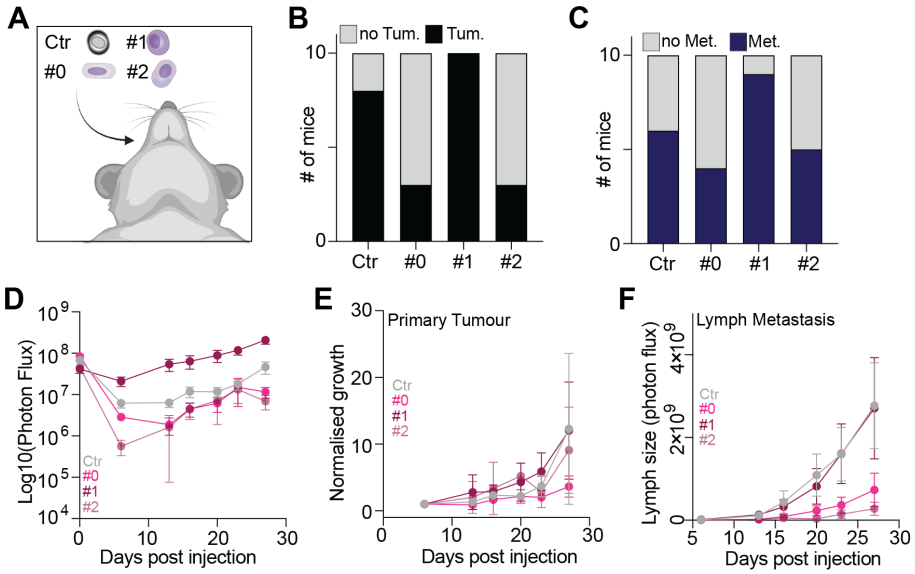

Figure 8.6: **Clones with reduced SRSF2 levels can form tumours in vivo**
A Illustration of orthotopic tongue injections. (**B-C**) Bar graph of the number of mice that developed **B** primary tumours (Tum.) and **C** lymphnode metastasis (Met.) following orthotopic injection with one of the following clones; a control clone (Ctr), or clone #0, #1 or #2. **D** Tumour size, Log_{10} photon flux in control (Ctr; gray), #0 (pink), #1 (dark purple) and #2 (light purple). **E** Tumour growth, realtive to day 6 in control (Ctr; gray), #0 (pink), #1 (dark purple) and #2 (light purple). **F** Lymph node metastasis size (photon flux) in control (Ctr; gray), #0 (pink), #1 (dark purple) and #2 (light purple).Error bars= SEM, 10 mice per condition.

vival advantage. The reduction in the signal between day 0 and day 6 was largest in clones #0 and #2, while clone #1 showed a smaller reduction in signal than the control (Figure 8.6 D). Tumour growth was not significantly different between any of the clones and the control in the primary tumour (Figure 8.6 E). The size of lymph node metastases were larger in the control and clone #1, however, this is likely to also be due to the survival advantage as colonising a new site is a rare event dependent largely on cells surviving the journey to the new site, although there are other possibilities such as the metastatic cells possessing more appropriate homing and trafficking signals.

In conclusion, cells with reduced SRSF2 levels were able to form tumours *in vivo*. This demonstrates that if SRSF2 levels are only reduced by 50% proliferation was not affected. Furthermore, when cell cycle arrest was overcome cells survive with more DNA damage, which may enable faster clonal evolution and therefore worse prognosis, this was illustrated by clone #1, which showed enhanced tumour formation and survival.

8.2.6 *Positive regulation of Pol II allows survival with reduced SRSF2 levels*

Cell division defects have been widely reported in association with alterations to SRSF2 expression levels (Bapat et al., 2018, Xiao et al., 2007, Kon et al., 2018). However, all clones were able to divide at equal rates to control, suggesting that with regard to cell division there is no haploinsufficiency. I wondered whether other SRSF2 functions were more sensitive to SRSF2 dosage.

To explore this Pol II phosphorylation status was examined, as SRSF2 regulates Pol II phosphorylation via PTEFb (Lin, Coutinho-Mansfield, et al., 2008, Ji et al., 2013). This revealed that both the amount of total Pol II and Ser2 phosphorylated Pol II were reduced in all clones (Figure 8.7 A-C). This suggests that transcription is altered in all clones, to better understand these alterations the transcriptomes of all clones were analysed (Figure 8.7 D). Initially, focusing on understanding how they survive with not only less Pol II but also less active Pol II I looked at the expression of positive regulators of Pol II (Figure 8.7 E). This revealed that these were upregulated in all clones, which was not the case after accute SRSF2 depletion, which raises the possibility that cells compensate for the lack of SRSF2 by globally increasing transcription. Previous work in this thesis showed that DNA repair and replication genes were most susceptible to SRSF2 loss due to their organisation as bi-directional gene pairs. I hypothesised that if these genes require more regulation for proper transcription then globally up regulating transcription will not fully compensate for the reduced levels of SRSF2. I therefore looked at the expression of genes with bi-directional promoters compared to all other genes in the three clones, which confirmed that, bi-directionally transcribed genes were more down regulated than other genes (Figure 8.7 F-H).

In order to further understand how the clones survive I performed gene ontology analysis on upregulated genes in all clones (Figure 8.8 A-F). Processes involved in mitosis and G2/M transition were upregulated in all clones. Interestingly, differentiation was upregulated in clones #0 and #2, which may explain why these clones did not form tumours as readily as clone #1 *in vivo*. Interestingly, splicing is only upregulated in clone #2, while transcription is upregulated in all clones. This may further indidcate that SRSF2 plays a unique and essential role in transcription, while its role in splicing can be compensated for by other members of the SR protein family.

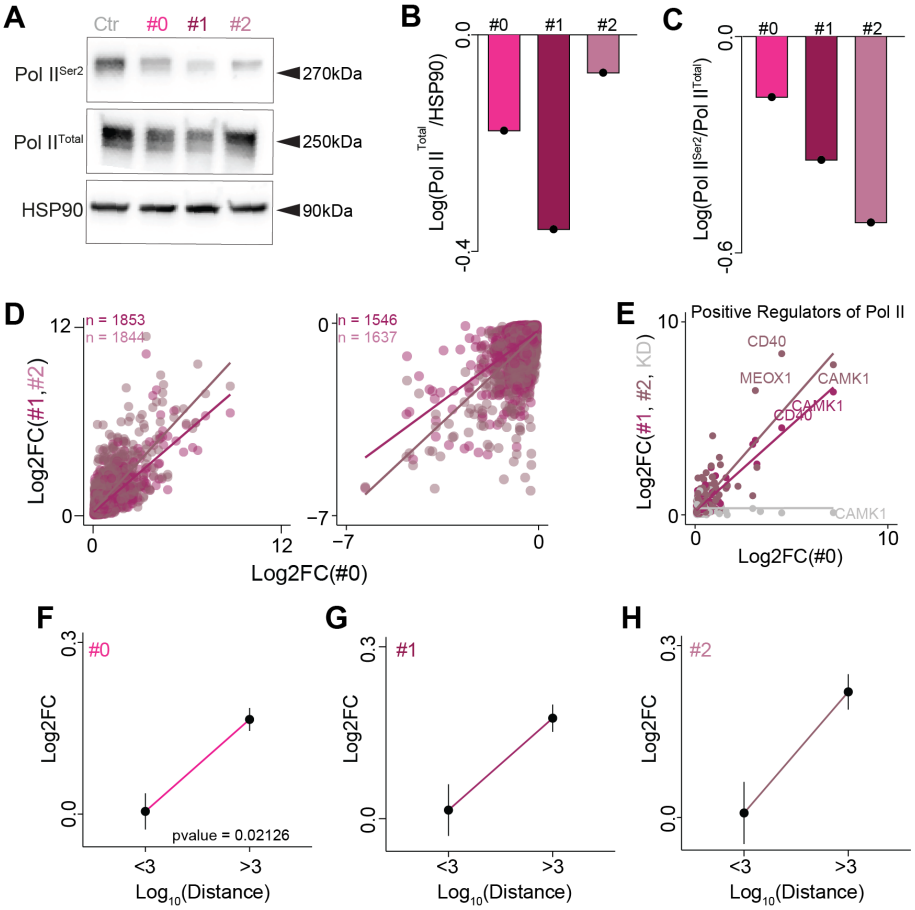

Figure 8.7: **Direct regulation of Pol II allows cells to survive with less SRSF2**
A Western blot of total Pol II (Pol II ^{Total}), Ser2 phosphorylated Pol II (Pol II ^{Ser2}) and house keeper HSP90. **B** Quantification of total Pol II levels relative to HSP90 and normalised to control. **C** Quantification of Ser2 phosphorylated Pol II relative to total Pol II levels and normalised to control. **D** Log_{t2} fold change (Log2FC) of differentially expressed genes from clone #0, #1 (puprle) and #2 (light purple), compared to a control clone, n= genes. **E** Expression of positive regulators of Pol II (GO:0045944) after SRSF2 knock down (KD; gray) and in clones with stable SRSF2 depletion; #0, #1 (purple) and #2 (light purple). (**F**-**H**) Log2FC of commonly differentially expressed genes after stable SRSF2 depletion seperated by distance to the closest neighbour less than 1kb, or more than 1kb in **F** clone #0, **G** clone #1 and **H** clone #2, error bars= SEM, p= Wilcoxen test with continuity correction.

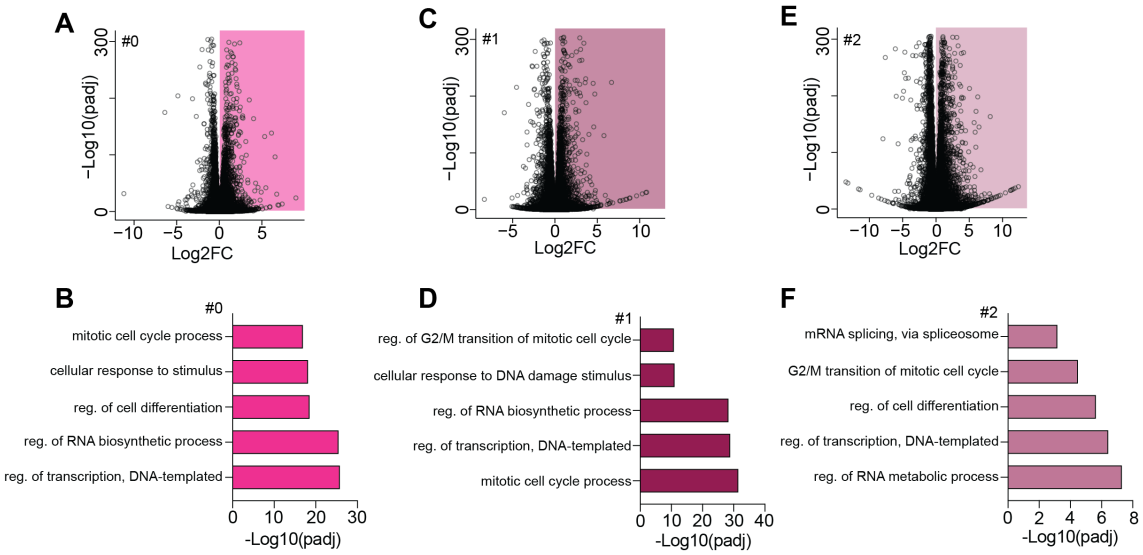

Figure 8.8: **Survival with reduced levels of SRSF2 requires changes to gene expression** **A** Volcano plot of differentially expressed genes in clone #0, significantly upregulated genes are highlighted. **B** Gene ontology analysis on up regulated genes in clone #0, compared to control. **C** Volcano plot of differentially expressed genes in clone #1, significantly up regulated genes are highlighted. **D** Gene ontology analysis on upregulated genes in clone #1, compared to control. **E** Volcano plot of differentially expressed genes in clone #2, significantly up regulated genes are highlighted. **F** Gene ontology analysis on up regulated genes in clone #2, compared to control. Analysis done using EnrichR (Xie et al., 2021, Kuleshov et al., 2016, E. Y. Chen et al., 2013).

8.2.7 *Summary, discussion and future directions*

Three cell lines with stably reduced SRSF2 levels were established in order investigate whether cells are able to compensate for the loss of SRSF2 and if so how they do so. No clones survived with more than 50% reduction in SRSF2 expression, suggesting that levels bellow this are deleterious for viability. Furthermore, this was done using an established SCC cell line with p53 muations, therefore at least 50% SRSF2 levels are required for viability even in the context of p53 deficiency.

Clones with reduced SRSF2 levels all displayed a higher proportion of cells in SubG0 and G2/M. However, division rates were not significantly different to that of a control clone and this remained the case when cells were transplanted *in vivo*. Differential gene expression analysis revealed that genes related to G2/M transition were upregulated. This may simply reflect that a larger proportion of cells in clones with reduced SRSF2 levels were in G2/M, compared to control, which is likely to be due to the elevated levels of DNA damage that was observed in all clones. The G2/M checkpoint is important for preventing the onset of mitosis until DNA damage has been repaired.

Strikingly, total Pol II and Ser2 phosphorylated Pol II levels were reduced in all clones. This demonstrates that SRSF2 is necessary for regulating PTEFb activity. Cells compensated for a reduction in active Pol II by upregulating positive regulators of transcription. Regulating total Pol II levels may be a stress response invoked to reduce DNA damage (Tufegdzic et al., 2020), if there are fewer Pol II molecules the chances of collision with each other or the replication machinery are reduced and so this may limit damage and sustain viability.

In the future explore whether transcription rates are altered in the clones via SLAM sequencing would be interesting. Furthermore, addressing where DNA damage accumulates most readily and if there are some common mutations among all surviving clones which aid survival would help to better understand how SRSF2 mutations contribute to cancer development. Assessing the clones' capacity to repair DNA damage would also be a pressing question for future work. This could be done by treating cells with a mutagen and determining how quickly resulting DNA damage is resolved, this would help to elucidate whether a reduction in functional SRSF2 contributes to relapse. Finally, examing whether more intrinsic damage occurs in the clones as a result of transcription-replication collisions, or whether down regulating Pol II limits these collisions.and the resulting damage would help to better understand how cellls compensate for having reduced SRSF2 levels and improve our understanding of SRSF2's essential functions.

Chapter 9

The role of SRSF2 in carcinogenesis

Chemical carcinogenesis experiments were done in collaboration with Dr. Karin Müller-Decker. RNA sequencing data analysis was assisted by Dr. Anke Heit-Mondrzyk.

9.1 *Aims of this chapter*

Heterozygous point mutations in SRSF2 are frequently observed in various types of blood cancer (Papaemmanuil et al., 2013). However, how mutations in SRSF2 contribute to disease progression remains largely unknown.

Cancer development is associated with the gain of specific characteristics, including uncontrolled proliferation, genome instability and avoidance of cell death (Hanahan and Weinberg, 2000). SRSF2 mutations have been shown to contribute to maintenance of a stem cell like state and therefore clonal expansion via missplicing of EZH2 in AML (E. Kim et al., 2015). Furthermore, mutant SRSF2 has been shown to increase DNA damage via R-loop formation (L. Chen et al., 2018). During carcinogenesis cells stop sensing DNA damage, often through mutations in p53 (Hanahan and Weinberg, 2000, Lane, 1992), which increases the rate of aquisition of mutations allowing cancer to evolve faster.

Chemical carcinogenesis is a well established method for studying cancer intitiation and development (Nassar et al., 2015, Verma and Boutwell, 1980, E. L. Abel et al., 2009, Luch, 2005). In the skin chemical carcinogenesis first leads to benign growths called papillomas, which are caused by cells that have mutations that lead to uncontrolled proliferation, mostly in KRAS and p53 (Nassar et al., 2015). The transition from papilloma to squamous cell carcinoma is associated with increased mutational burden but not the aquisition of specific mutations (Nassar et al., 2015). Here I aimed to ascertain whether mutant SRSF2 contributes to cancer initiaiton. The specific aims of this chapter were as follows:

1. Establish whether heterozygous point mutations in the skin alter carcinogenesis. This was done by inducing mutant SRSF2 followed by a skin carcinogenesis

protocol and monitoring papilloma formation.

2. Probe whether mutant SRSF2 can contribute to skin when it is forced to proliferate. This was done by looking at reporter expression in the skin treated with the promoting factor TPA.

3. Assay the effects of mutant SRSF2 on the transcriptome. This was done by sorting reporter positive cells from back skin and performing total RNA sequencing.

9.2 *Results*

9.2.1 *Investigating whether mutant SRSF2 contributes to cancer initiation*

In order to address whether mutant SRSF2 contributes to skin carcinogenesis I generated conditional mouse models in which the P95H mutation could be induced in the LGR5 positive cells of the hair follicle alone or both LGR5 and KRT14 positive cells of the interfollicular epidermis (Figure 9.1 A). This was achieved by crossing mice bearing a Cre-recombinase with the tamoxifen responsive horomone-binding domain of the estrogen receptor (Cre-ERtam) under the control of the human KRT14 and LGR5 promoters, a tdTomato construct under the Rosa26 locus and an SRSF2 domain consisting of loxP sites flanking exon 1 and 3, which allows expression of a mutant version of SRSF2 (Figure 9.1 B).

The mutation was heterozygous, in order to mimic the conditions that are seen most frequently in patients (Papaemmanuil et al., 2013). The point mutation was induced and a week later the skin was topically treated with the carcinogen 9,10-dimethyl-1,2-benyanthracene (DMBA), promotion with tetradecanoyl-phorbol acetate (TPA) was then started a further week later and carried out three times a week for up to 20 weeks (Figure 9.1 C). Chemically induced carcinogenesis begins with the formation of papillomas between 10 and 20 weeks, which can then progress into carcinomas if the protocol is continued for between 20 and 50 weeks (E. L. Abel et al., 2009). I was interested in whether mutant SRSF2 influences cancer initaition, therefore I monitored papilloma development for up to 20 weeks (Figure 9.1 C).

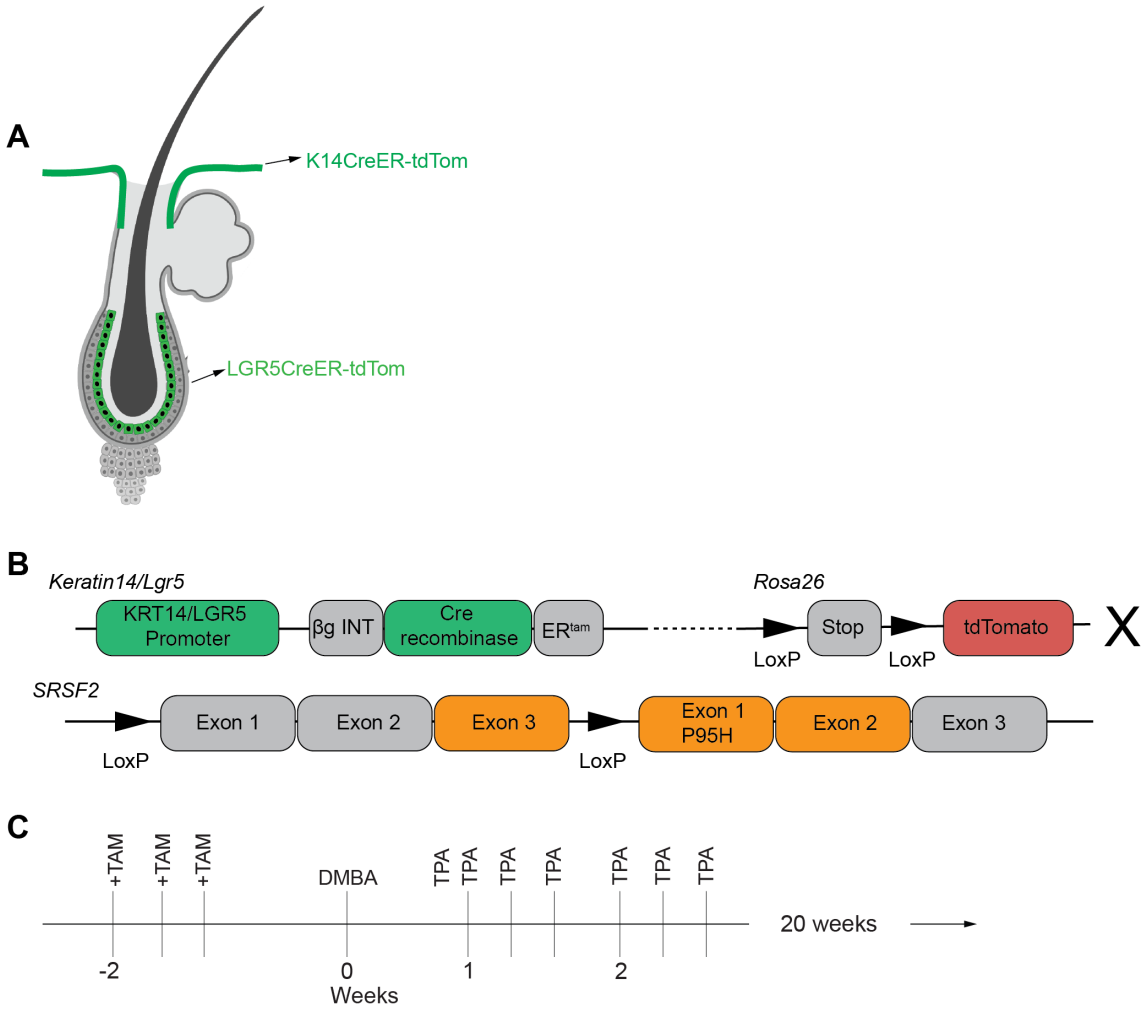

Figure 9.1: **Investigating whether mutant SRSF2 contributes to skin carcinogenesis**
A Scheme of the skin and hair follicle showing the populations where mutant SRSF2 was induced. **B** Scheme showing transgenes to induce mutant SRSF2. **C** Outline of time course with tamoxifen (TAM), DMBA and TPA.
DMBA-9,10-dimmethyl-1-2-benyanthracene, TPA-tetradecanoyl-phorbol acetate.

9.2.2 *Mutant SRSF2 reduces the rate of carcinogenesis*

All control mice developed papillomas within 20 weeks, however, only half the mice with the P95H mutation developed papillomas (Figure9.2 A). Not only was papilloma occurance reduced but the mutant mice which did develop papillomas developed fewer papillomas compared to control (Figure 9.2 B). At 20 weeks control mice all had a minimum of three papillomas, while mice with mutant SRSF2 had a maximum of two papillomas (Figure 9.2 C). Papilloma formation requires uncontrolled proliferation, therefore, the presence of mutant SRSF2 must prevent cells from acquiring this capacity. However, the difference in papilloma formation suggests that cells with mutant SRSF2 are able to survive and contribute to the IFE when induced to proliferate via TPA treatment.

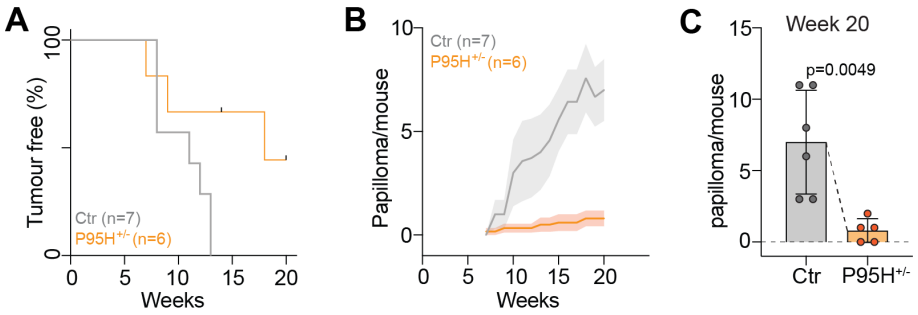

Figure 9.2: **Papilloma formation is inhibited by the presence of mutant SRSF2**
A Percentage (%) of tumour free mice, control (Ctr) or with P95H mutation (P95H$^{+/-}$). **B** Average number of papillomas per mouse in control (Ctr) and SRSF2 mutant (P95H $^{+/-}$), shaded area = SD. **C** Average number of papillomas in control (Ctr) and SRSF2 mutant (P95H $^{+/-}$) mice after 20 weeks of TPA treatment. n= mice, p= unpaired t-test.

9.2.3 *Cells with mutant SRSF2 contribute to the IFE, but do not form tumours*

The extent to which mutant SRSF2 contributed to the IFE and hair follicle, and whether the papillomas which did form in mutant mice originate from SRSF2 mutant cells was not clear. In order to investigate this I looked at the proportion of reporter positive cells in back skin and papillomas (Figure 9.3). I began by quantifying the proportion of RFP in the IFE, this revealed that cells heterozygous for the SRSF2 point mutation were able to contribute to maintenance of the IFE (Figure 9.3 A, B) and the trend suggests that clonal expansion may be enhanced when mutant SRSF2 was present, which would be in line with what has been observed in MDS and AML (Yoshimi et al., 2019, E. Kim et al., 2015).

I next asked whether the papillomas that did grow in mice with mutant SRSF2 were reporter positive (Figure 9.3 C,D). Two papillomas were positive, while one was negative, while in the control four papillomas were positive while two were negative for reporter expression. This shows that when papillomas did grow out from skin with mutant SRSF2 the chance that they originated from reporter positive cells was equivalent, however, initiation was severly impaired. Interestingly, when I looked at epidermal thickness this was increased when mutant SRSF2 was present (Figure 9.3 E), furthermore, the average size of the papillomas that grew was larger in skin with mutant SRSF2 (Figure 9.3 F). These results show that when cell cycle arrest is overcome, in this case via selection by topical treatment with TPA, the presence of mutant SRSF2 can be advantageous and lead to clonal outgrowth. However, the precise barriers that must be overcome and how this occurs remain unclear.

In order to explore the molecular alterations that enable cells to grow out with mutant SRSF2 reporter positive cells were sorted from back skin of control and mutant mice after approximately 20 weeks of TPA treatment (Figure 9.4 A). The number of up and down regulated genes was equally distributed (Figure 9.4 B). However, I focused on the down regulated genes, which revealed that processes involved in differentiation, proliferation, stress and cell death were most affected when mutant SRSF2 was present (Figure 9.4 C). Interestingly genes involved in regulating transcription were also down regulated along with genes involved in the

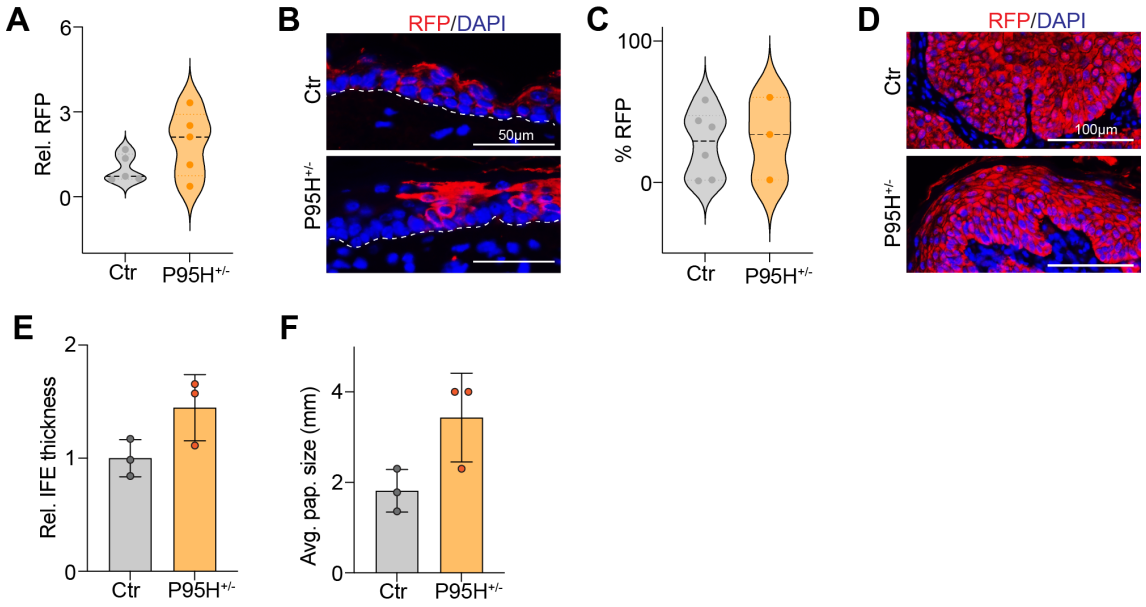

Figure 9.3: **Cells are able to maintain skin when expressing mutant SRSF2**
A Relative (Rel.) proportion of RFP in control (Ctr) and SRSF2 mutant (P95H $^{+/-}$) skin, each point represents one biological replicate and the average of five images. **B** Representative images from control (ctr; upper panel) and SRSF2 mutant (P95H $^{+/-}$; lower panel) epidermis, dotted line represents the basement membrane. **C** Violin plot showing the % of RFP positive cells in control (Ctr) and SRSF2 mutant (P95H $^{+/-}$; orange) papillomas, each point represents one biological replicate and the average of five technical replicates. **D** Representative images from control (Ctr; upper panel) and SRSF2 mutant (P95H $^{+/-}$; lower panel) papillomas. **E** Quantification of epidermal thickness, each point represents a bioloical replicate and the average of three measurements from each of five seperate images. **F** Average papilloma size, each point represents one biological replicate and the average of up to 14 measurements. A minimum of three mice were analysed per condition.

mitogen activated protein kinase (MAPK) cascade, which may be why papilloma formation was significantly less when mutant SRSF2 was present. I also looked at upregulated genes and noted that genes involved in DNA replication were enriched. As I previously showed that SRSF2 was required for proper transcription of DNA replication genes this was a surprise. To explore this further I asked how genes with bi-directional promoters were differentially expressed, the opposite trend was now seen suggesting that forced proliferation enables cells with mutant SRSF2 to overcome the barrier to efficient transcription of bi-directionally transcribed genes (Figure 9.4 D).

To confirm this I looked at double strand breaks in the skin (Figure 9.5 A,B). This revealed that there were not more double strand breaks in mutant skin, when compared to control. I also looked at double strand breaks in papillomas (Figure 9.5 C, D). The number of double strand breaks was higher in papillomas compared to the skin, however, there was not more DNA damage in papillomas from mutant mice, compared to controls. Finally I looked at proliferation in both the skin and papillomas (Figure 9.5 E, F). There was more proliferation in papillomas compared to skin, however, there was no observable differences between control and SRSF2 mutant tissue, which suggested that forced proliferation does lead to selection of cells able to overcome cell cycle arrest, which occured when functional SRSF2 was

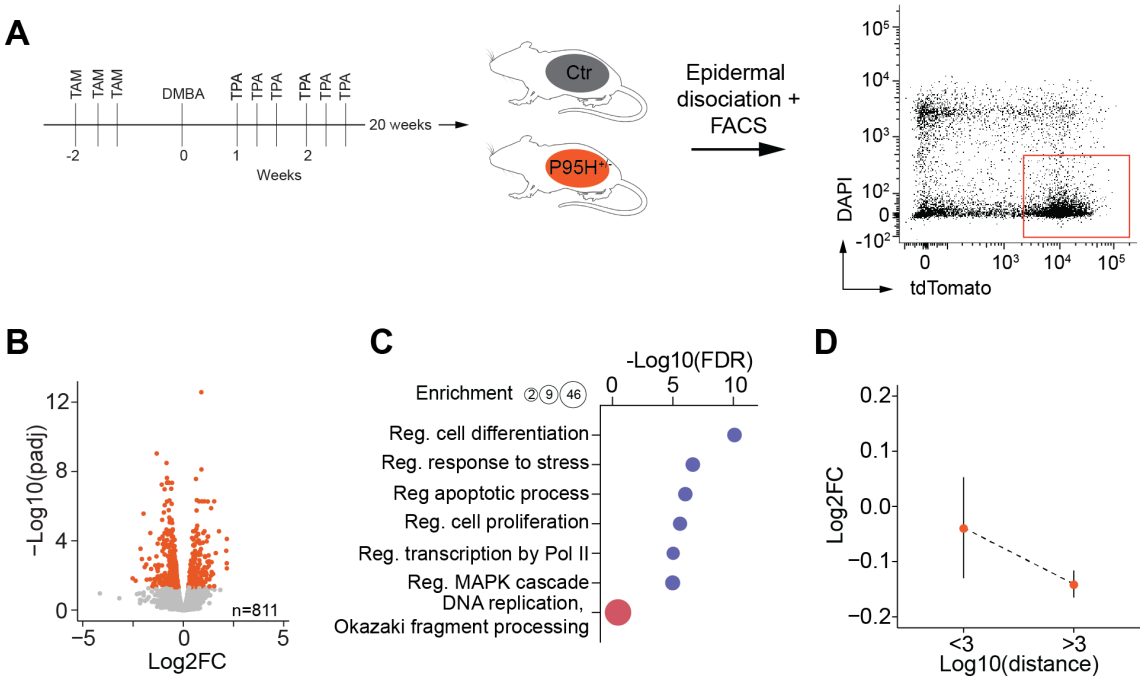

Figure 9.4: **Surviving Cells with mutant SRSF2 rescue transcription of bi-directionally transcribed genes**
A Scheme of experimental procedure for sorting cells from mouse back skin. **B** Volcano plot showing differentially expressed genes in cells with mutant SRSF2, padj<0.05 in orange, n= genes. **C** Gene ontology analysis of down regulated genes, using GOrilla (Eden, Lipson, et al., 2007, Eden, Navon, et al., 2009). **D** Expression of differentially expressed genes padj<0.05 with a neighbour within 1Kb, or further than 1kb away.

lost in skin under homeostatic conditions.

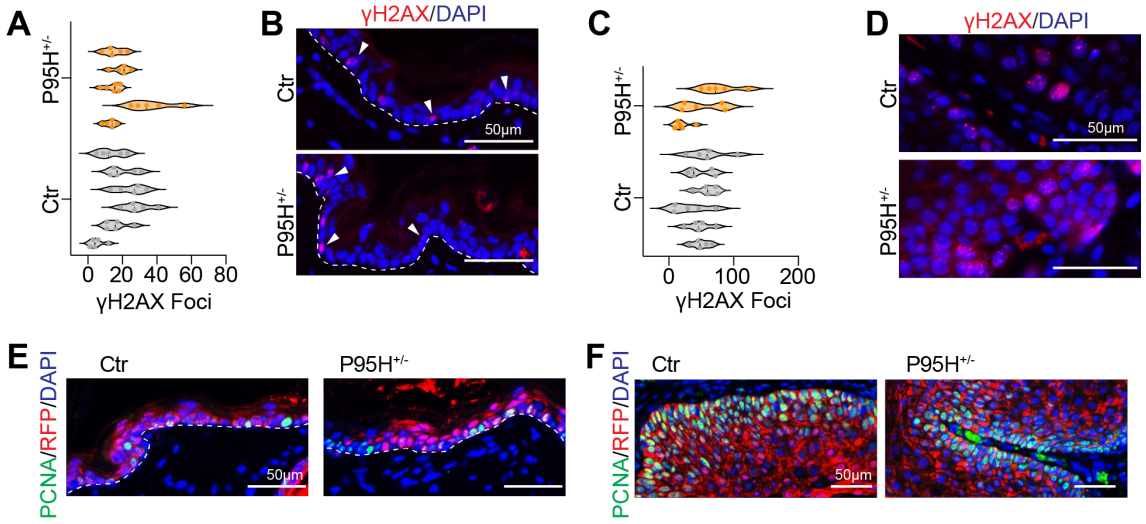

Figure 9.5: **TPA treatment forces cell cycle arrest to be compensated for in cells with mutant SRSF2**
A Quantification of γH2AX poitive cells in control (Ctr) and SRSF2 mutant (P95H $^{+/-}$) skin, each violin represents one biological replicate and each point represents a technical replicate. **B** Representative images from control (ctr; upper panel) and SRSF2 mutant (P95H $^{+/-}$; lower panel) epidermis, dotted line represents the basement membrane and arrows indicate positive nuclei. **C** Quantification of γH2AX positive cells from control (Ctr; upper panel) and SRSF2 mutant (P95H $^{+/-}$; lower panel) papillomas. **D** Representative images of papillomas (left) and epidermis (right) in control (ctr; upper pannel) and SRSF2 mutant (P95H $^{+/-}$; lower panel). **E** Representative images of skin, control (Ctr, upper panel) and SRSF2 mutant (P95H $^{+/-}$; lower panel) stained for PCNA and RFP, the dotted line indicates the basement membrane. **F** Representative images of papillomas, control (Ctr, upper panel) and SRSF2 mutant (P95H $^{+/-}$; lower panel stained for PCNA, RFP the dotted line indicates the basement membrane. A minimum of three mice per condition were anaylsed.

9.2.4 *Summary, discussion and future directions*

To investigate whether the presence of mutant SRFS2 aids cancer initiation an established protocol for probing carcinogenesis was used. This revealed that the heterozygous point mutation in SRSF2 inhibited carcinogenesis, as the number of papillomas was reduced. However, mutant cells were able to contribute to the epidermis, with a slight increase in clonal expansion and an increase in the thickness of the IFE. This finding was in line with my previous findings, presented in Chapter 3 and the literature which showed that the presence of mutant SRSF2 led to attenuated differentiation, and therefore increased clonal expansion (E. Kim et al., 2015). Interestingly, papillomas that did form also tended to be larger, perhaps also suggesting the cells able to escape the proliferative disadvantage that mutant SRSF2 inherently carries were able to grow and evolved faster. Transcriptomic analysis revealed that MAP kinase signalling was altered in SRSF2 mutant cells, along with transcription, differentiation, replication and apoptosis, all key processes that become dysregulated during carcinogenesis and processes that SRSF2 has been shown to be important for regulating. Finally, I also found that bi-directionally transcribed genes are no longer repressed, suggesting that surviving cells have found a mechanism to overcome this.

In the future it would be interesting to look at the carcinogenesis process in skin

with mutant SRSF2 in more detail. In order to do this I would employ Cre-confetti mice (Snippert et al., 2010) in order to look at clone size and then perform single cell multiomics to look at the mutational burden upon DMBA treatment and how the transcriptome is altered immediatly after DMBA treatment and then weekly in order to identify mutations that work synergistically with mutant SRSF2 to enable cells to overcome the cell cycle arrest that usually occurs when mutant SRSF2 is present. Finally, I would also look at splicing and whether isoform usage changes during carcinogenesis via single cell long read sequencing.

Chapter 10

Summary, Discussion and Future Perspectives

10.1 *Summary and Key Findings*

During my PhD I set out to better understand the molecular mechanisms that the SR protein family member SRSF2 acts through, largely focusing on its role in transcription. Furthermore, I investigated the consequences of a point mutation on SRSF2 function and asked whether the point mutant contributes to carcinogenesis. Below is a summary of the key findings I presented in this thesis. I will then discuss these findings and highlight future directions.

The main findings from this thesis are:

- The SRSF2 P95H mutant largely phenocopies SRSF2 knock out, which suggests that this point mutation causes a loss of function.
- Functional SRSF2 is required for homeostasis of the IFE.
- IFE stem and progenitor cells are unable to replicate and differentiate fully when functional SRSF2 is absent.
- SRSF2 depletion in human epidermal cell lines affects both normal and cancer cell viability, but does not trigger apoptosis.
- Acute SRSF2 depletion leads to cell cycle arrest in normal and cancer cells.
- SRSF2 depletion leads to a global reduction in transcription.
- DNA replication and repair genes are most vulnerable to SRSF2 loss due to their organisation as bi-directional gene pairs.
- DNA damage accumulates when SRSF2 is depleted both stably and acutely, this is due to an inability to repair and is likely to be exacerbated by an increase in damage occuring due to dysregulated transcription.
- Stably reducing SRSF2 levels beyond half, via CRISPR-cas9 is deleterious for cell survival.

- Cells are able to replicate with only 50% SRSF2, *in vitro* and *in vivo*.
- PTEFb activitiy cannot be rescued in cells with stably reduced levels of SRSF2, therefore, SRSF2 is essential for mediating PTEFb function in phosphorylating Ser2 of the CTD of Pol II.
- Cells upregulate positive regulators of Pol II to compensate for PTEFb misregulation when SRSF2 levels are stably reduced.
- The presence of heterozygous point mutations in SRSF2 in skin inhibits cancer initiation by chemical carcinogenesis.

10.2 *Discussion*

10.2.1 *Functional SRSF2 is required for skin homeostasis*

SRSF2 is an essential gene and has been shown to be important for homeostasis of several tissues (H.-Y. Wang et al., 2001, Cheng et al., 2016, Kon et al., 2018). Interestingly, SRSF2 is also mutated in cancer, however, how mutations alter protein function and how the mutant protein contributes to carcinogenesis remains largely unknown. To date no study has directly compared the consequences of SRSF2 knock out and mutation on tissue homeostasis. Understanding the molecular mechanisms through which SRSF2 acts has been complicated by it's dual function in both splicing and transcription, processes that are also intricately linked (L. Chen et al., 2018, Moore and Proudfoot, 2009).

My findings showed that the P95H mutation leads to a loss of function and that functional SRSF2 is required for maintaining homeostasis of the IFE. Loss of either wild type or mutant SRSF2 led to a loss of cellularity in the basal layer, but an increase in individual cell size. This is likely to be caused by cells lacking functional SRSF2 not being able to complete cell division and arresting during G2 or mitosis, when all cellular components and DNA have been duplicated. Furthermore, differentiation was impaired, as has been reported in other tissues, including blood, liver and heart (Bapat et al., 2018, Kon et al., 2018, Cheng et al., 2016, Ding et al., 2004). However, why cells are unable to differentiate when functional SRSF2 is absent remains unknown.

SRSF2 has roles in both splicing and transcription, while it has been shown to function promiscuously as a splicing factor (Pandit et al., 2013), its role as a transcriptional activator has been hypothesised to be unique (Ji et al., 2013, Lin, Coutinho-Mansfield, et al., 2008). The alterations to splicing that occur when the P95H mutant form of SRSF2 is present have been studied extensively (Kon et al., 2018, E. Kim et al., 2015, Yoshimi et al., 2019). However, a clear mechanism by which mutant SRSF2 alters splicing of differentiation or proliferation related transcripts across cell types remains unknown. This is in part due to alterations to splicing being inconsistent when comparing human data to mouse models (Mupo et al., 2017, L. Chen et al., 2018). This suggests that SRSF2's role in transcription may be conserved, while this is not the case for its function as a splicing factor.

Examining both the effects of deletion and mutation of SRSF2 in parallel offers a unique perspective, as the effects of the protein being absent can be compared with those seen when the protein is present but the function is changed. The effects on skin homeostasis were strikingly similar when SRSF2 was deleted or mutated, which indicates that a function which is lost in both cases must be largely responsible. Furthermore, the transcriptomic changes that occured as a result of SRSF2 being absent indicated that transcription but not splicing was altered, confirming that SRSF2 is required for transcriptional activation, but not splicing.

Further investigation into the effects of SRSF2 deletion and mutation in parallel will bring further insights into SRSF2 function, and will aid in the unraveling of how changes to transcription alter splicing and how this impacts cell identity and ability to respond to stress.

10.2.2 Skin with mutant SRSF2 is resistant to skin carcinogenesis

In order to ascertain whether mutant SRSF2 contributes to cancer initiation a well established protocol for chemical carcinogenesis was used (Luch, 2005, Nassar et al., 2015). Papilloma formation was reduced in skin with the P95H mutation, suggesting that this hinders growth. However, cells with the P95H mutation were capable of maintaining the tissue. I wondered whether forcing cells to proliferate leads to the selection of cells that are able to compensate for having reduced levels of functional SRSF2. Looking at expression of bi-directionally transcribed genes revealed these genes were no longer repressed, suggesting that the surviving cells were able to rescue this vulnerability. How the cells are able to do this and whether this requires acquisition of a specific set of mutations remains unclear.

10.2.3 *SRSF2 depletion results in cell cycle arrest, irrespective of p53 status*

In both primary human keratinocytes and two SCC cell lines with p53 mutations acute SRSF2 depletion resulted in a decrease in the number of cells in S phase. This decrease occurred after 48 hours of SRSF2 depletion, when the protein is only partially reduced. This suggests that replication requires a certain level of SRSF2 and that this is independent of p53 status. CDC45 was consistenly downregulated when SRSF2 was depleted. It is part of a helicase complex, CDC45-MCM-GINS (CMG), which is important for DNA unwinding and replication initiation (Rzechorzek et al., 2020, Baris et al., 2022, Guillou, Coloma, and Montoya, 2009). Therefore CDC45 may be an important factor through which SRSF2 coordinates DNA replication.

SRSF2 depletion also caused an increase in the proportion of cells in SubG0, which encompasses cells with fragmented DNA and cells undergoing apoptosis. Probing the levels of double strand breaks and fragmented DNA directly confirmed that SRSF2 depletion leads to an increase in DNA damage. Transcription and replication, while essential cellular processes both cause DNA damage. Both processes also go awry when SRSF2 is depleted which may result in increased DNA damage. Furthermore, I showed that DNA repair is also attenuated when SRSF2 is

absent, which would also result in a higher DNA damage burden, thereby amplifying the consequences of SRSF2 loss and resulting in a higher level of DNA damage. Strinkingly, cell viability was only mildy reduced and cells were not apoptotic. This creates a situation where the fidelity of the genome is compromised, but cell viability is maintained, indicating that SRSF2 loss may be advantageous to carcinogenesis, if cell cycle arrest can be overcome.

10.2.4 *SRSF2 has a unique role as a co-transcriptional regulator*

SRSF2 has been shown to bind the *RN7SK* complex and was shown to promote PTEFb dissociation, facilitating productive elongation (Ji et al., 2013, Lin, Coutinho-Mansfield, et al., 2008). These findings led to SRSF2 being termed a transcriptional activator. However, whether SRSF2 is important for this process genome wide or at a specific group of genes was unknown. I confirmed SRSF2's role as a transcriptional activator by looking at Pol II occupancy and the nascent transcriptome after SRSF2 depletion. This revealed that SRSF2 is required for efficient Pol II release into productive elongation genome wide. However, replication and repair genes are most vulnerable to SRSF2 depletion, due to their organisation as bi-directional gene pairs. These are genomic regions of high transcriptional activity, therefore, chromatin architecture and histone modifications cannot be used to modulate Pol II activity. The risk of the Pol II machinery colliding is higher making proper expression of these genes more dependent on higher order regulation, such as that carried out by SRSF2.

Th expression of several SR protein family members was altered in all cell lines, following SRSF2 depletion, however, the dysregulation of the pool varied depending on the cell line. A better understanding of how the SR protein network functions in various contexts and how much redundancy there is between family members remains largely unknown. However, the Ser2 phosphorylated form of Pol II was not affected by depletion of SRSF1 or SRSF5, thus confirming that this role is unique to SRSF2 and that it is not acting as part of a network that breaks down when one member is lost.

All SR proteins are known to function as splicing factors and redundancy between family members has been reported (Pandit et al., 2013, L. Han et al., 1995, Björk et al., 2009). This complicates investigating the consequences of SRSF2 depletion on splicing. Nevertheless, changes to splicing that occur as a result of SRSF2 depletion were examined. The alterations to splicing were extensive, however, these changes did not correlate with alterations to gene expression and there was a very limited number of splicing events that happened in all cell lines. This is likely to be due to the degree of redundancy within the SR proteins in their splicing function and due to the dysregulation of the SR protein network varying depending on the cell line. This is in stark contrast to SRSF2's non-redundant role in regulating PTEFb activity, which is unique and appears to confer it's essential function.

Finally, splicing is largely a co-transcriptional process and transcription rates have been shown to impact spliceosome function (Moore and Proudfoot, 2009), therefore, SRSF2 functions are again likely to compound on the alterations to splic-

ing hindering the elucidation of SRSF2's precise role in splicing.

10.2.5 *DNA damage accumulates due to repair deficits when SRSF2 is depleted*

SRSF2 is an RNA binding protein therefore, observing an increase in DNA damage was unexpected. Transcription can cause DNA damage through the formation of DNA-RNA hybirds, as can conflicts with the replication machinery (Petermann, Lan, and Zou, 2022, Saxena and Zou, 2022). Aberrant transcription and replication may lead to increased DNA damage when SRSF2 is depleted. Nonetheless, repair mechanisms should prevent DNA damage persisting. Transcriptomic analysis revealed that DNA repair and replication genes are most sensitive to SRSF2 depletion. DNA damage has also been shown to accumulate when SRSF2 is mutated (Xiao et al., 2007, L. Chen et al., 2018). I directly assessed the repair capacity of primary human keratinocytes after SRSF2 depletion. This revealed that repair was impaired when SRSF2 was depleted.

I speculate that DNA damage occurs more frequently in actively transcribed regions when SRSF2 is depleted, which would further the deleterious effects of SRSF2 depletion on cell viability. However, this also means that mutations will occur more frequently in regions of the genome which are protein coding, in contrast to mutagens that randomly introduce mutations this may allow cancer to develop faster, under permissive circumstances.

10.2.6 *Stable reduction of SRSF2 allows cells to proliferate with DNA damage*

SRSF2 mutations are mostly heterozygous, which is in line with my findings that the mutation leads to a loss of function as SRSF2 is an essential gene. I was interested in understanding how little SRSF2 cells are able to survive with.

This revealed that even already cancerous cells are not viable if SRSF2 expression is reduced by more than 50% as illustrated by clonogenecity being reduced 10 fold in cells where the SRSF2 gene was edited via CRISPR-cas9. However, clones that did survive with reduced SRSF2 levels divided at rates comparable to control. This suggests that a minimum level of SRSF2 is enough to enable continued cell cycle progression. However, clones had less Ser2 phosphorylated Pol II confirming that SRSF2 is required for coordinating PTEFb activity. Transcription is essential and clones compensate for reduced levels of active Pol II by upregulating positive regulators of transcription, however, bi-directionally transcribed genes were still most susceptible to SRSF2 depletion. In line with this clones also had a higher level of DNA damage compared to control.

If cells are able to tolerate increased DNA damage and continue to proliferate there is an opportunity for faster evolution and more aggressive tumours.

10.3 *Concluding remarks and future perspectives*

Work in this thesis shows that SRSF2 is required for replication, differentiation and DNA damage repair. I explored SRSF2's function as a transcriptional activator, which revealed that it mediates PTEFb activity globally. Bi-directionally transcribed genes are most vulnerable to SRSF2 loss due to their organisation. Genes involved in DNA repair and replication are found to have bi-directional promoters more frequently than other sets of genes (Adachi and Lieber, 2002). Therefore DNA damage accumulates when SRSF2 is depleted which leads to cell cycle arrest. Further investigation is required to fully understand how SRSF2 acts at a molecular level.

This work could be continued in the following ways:

- Determine to what extent SRSF2's role in transcription drives the differentiation defect that was observed when SRSF2 was deleted or mutated *in vivo*. This could be done by measuring transcription rates in differet populations in the IFE and determing whether this is altered when functional SRSF2 is absent.

- Establish whether functional SRSF2 is required in other stem cell populations in the skin, such as LGR5 and KRT19 positive populations in the hair follicle and if so whether similar differentiation and proliferation defects are observed in these compartments.

- Explore how the SR protein network changes across differentiation, by looking at their expression in different cell types in the skin and determine whether dysregulation of this network is a driving factor in the differentiation block that was observed in the IFE.

- Probe where DNA damage occurs, using whole genome sequencing, when SRSF2 is depleted and what types of damage occur, focusing on DNA-RNA hybrids and transcription-replication conflicts. Also ascertain whether mutations in specific proteins are present in cells which are able to compensate for having reduced levels of SRSF2.

- Understand how SRSF2 depletion alters DNA repair processes, are specific repair pathways more affected than others, for example transcription coupled repair, may be more affected than double strand break repair.

- Probe when during carcinogenesis mutations in SRSF2 become advantageous, by inducing the mutation at various times during the carcinogenesis protocol and looking at the consequences on the numbers of papillomas and their size.

These experiments will enhance our understanding of SRSF2 function in both homeostasis and disease. This may then provide an opportunity to target the processes regulated by SRSF2 to improve treatment of diseases such as cancer.

Chapter 11

Appendix

11.1 *List of Abbreviations*

Abbreviation	Description
4OHT	4-hydroxy tamoxifen
4sU	4-thiouridine 5'-monophosphate
5FU	5-fluorouracil
A3SS	alternative 3' splice site
A5SS	alternative 5' splice site
AML	acute Myeloid leukemia
BCC	basal cell carcinoma
CDC25	cell division cycle 25
CDC45	cell division cycle 45
CDK	cycllin dependent kinase
cDNA	complementary DNA
CRISPR	clustered regularly interspaced short palindromic repeats
CTD	C-terminal domain
Ctr	control
DAPI	4',6-diamidino-2-phenylindole
DiD	1,1'-Dioctadecyl-1-3,3,3',3'-Tetramethylindodicarboncyanine
DMBA	9,10-dimethyl-1-2-benyanthracene
DNA	deoxyribonuclease
dNTPs	deoxynucleotide triphosphate
DP	dermal Papilla
DSIF	DRB-sensitivity inducing factor
ECL	enhancer chemiluminescence
ECAD	E-cadherin
ECM	extracellular matrix
EMT	epithelial to mesenchymal transition
ES	exon skipping
EtOH	ethanol
EZH2	enhancer of zeste homolog 2
GAPDH	glyceraldehyde 3-phosphate Dehydrogenase

FBS	fetal bovine Serum
GFP	green fluorescent protein
GO	gene Ontology
Gy	gray
H&E	haematoxylin and Eosin
Het	heterozygous
HEXIM	hexamethylene bisacetamide-inducible
HF	hair follicle
HK	human keratinocytes
hnRNP	heterogeneous nuclear ribonucleoprotein
Hom	homozygous
HR	homologous recombination
HRP	horseradish peroxidase
HSP90	heat shock protein 90
IF	immunofluorescence
IFE	Interfollicular epidermis
IHC	immunohistochemistry
INTS3	integrator complex subunit 3
IRES	internal ribosome entry site
IR	intron retention
ITGα6	Intergrinα6
kb	kilobase
KRT	keratin
KD	knock-down
kDa	kiloDaltons
KI	knock-in
KO	knock-out
LGR5	leucine-rich repeat containing G protein coupled receptor 5
LOR	loricrin
M	Molar
MALAT1	metastasis associated lung adenocarcinoma transcript 1
MCM10	minichromosome maintenance 10
MDS	myelodysplastic Syndrome
EMEM	eagle's minimum essential medium
Met	metastasis
MMR	mismatch repair
mRNA	messenger RNA
MUT	mutant
n	number of observations
NCAD	N-cadherin
ncRNA	noncoding RNA
NELF	negative elongation factor
NER	nilceotide excission repair
NHEJ	non-homogolous end joining
NMD	non-sense mediated decay
OCT	optimal Cutting Temperature Solution

OTM	olive tail moment
p	p-value
PBS	phosphate buffered Saline
PCNA	proliferating-cell-nuclear-antigen
PCR	polymerase chain reaction
PFA	paraformaldehyde
Pol II	RNA polymerase II
PRC2	polycomb repressive 2
PTEFb	positive transcription elongation factor b
qPCR	quantitative polymerase chain reaction
RFP	red Fluorescent protein
RNA	ribonucleic acid
RNAse	ribonuclease
RRM	RNA recognition motif
SAC	spindle assembly checkpoint
SCC	squamous cell carcinoma
SD	standard deviation
SEM	standard error of the mean
Ser	serine
siRNA	small interfering RNA
SLAM	thiol(SH)-linked alkylation for the metabolic
snRNPs	small nuclear ribonuclear protein particles
SRSF	serine arginine rich splicing factor
SG	sebaceous Gland
TAE	tris-acetate-EDTA
TBE	tris-borate-EDTA
TFIIH	transcription factor IIH
Thr	threonine
TPA	tetradecanoyl-phorbol acetate
TGM	transglutaminase
TSS	transcription start site
TTS	transcription termination site
Tum	tumour
Tyr	tyrosine
U	unit
UTR	untranslated region
UV	ultraviolet
V	volt
VTRNA1.1	vault RNA1.1
v/v	volume/volume
W	watt
w/v	weight/volume

Bibliography

Abel, Daan van et al. (2011). "SFRS7-mediated splicing of tau exon 10 is directly regulated by STOX1A in glial cells". In: *PLoS ONE* 6 (7). DOI: 10.1371/journal.pone.0021994.

Abel, Erika L. et al. (2009). "Multi-stage chemical carcinogenesis in mouse skin: Fundamentals and applications". In: *Nature Protocols* 4 (9), pp. 1350–1362. DOI: 10.1038/nprot.2009.120.

Adachi, Noritaka and Michael R. Lieber (2002). "Bidirectional Gene Organization: A Common Architectural Feature of the Human Genome". In: *Cell* 109. DOI: 10.1521/ijgp.2014.64.1.138.

Adelman, Karen and John T. Lis (2012). "Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans." In: *Nature reviews. Genetics* 13 (10), pp. 720–731. DOI: 10.1038/nrg3293.

Aibar, Sara et al. (Oct. 2017). "SCENIC: Single-cell regulatory network inference and clustering". In: *Nature Methods* 14 (11), pp. 1083–1086. DOI: 10.1038/nmeth.4463.

Allshire, Robin C and Hiten D Madhani (2018). "Ten principles of heterochromatin formation and function". In: *Nature Reviews Molecular Cell Biology* 19 (4), pp. 229–244. DOI: 10.1038/nrm.2017.119.

Alonso, L. and Elaine Fuchs (2006). "The hair cycle". In: *Journal of Cell Science* 119 (3), pp. 391–393. DOI: 10.1242/jcs02793.

André, Fabrice et al. (2017). "AACR Project GENIE : Powering Precision Medicine through an International Consortium". In: *Cancer Discovery*. DOI: 10.1158/2159-8290.CD-17-0151.

Anko, Minna Liisa et al. (2012). "The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes". In: *Genome Biology* 13 (3). DOI: 10.1186/gb-2012-13-3-r17.

Aragona, Mariaceleste et al. (2020). *Mechanisms of stretch-mediated skin expansion at single-cell resolution*. Vol. 584. Springer US, pp. 268–273. DOI: 10.1038/s41586-020-2555-7.

Arnetzl, Leonie F. (2021). "Elucidating the role of SRSF2 in head and neck squamous cell carcinoma development". In.

Aslanzadeh, Vahid et al. (2018). "Corrigendum: Transcription rate strongly affects splicing fidelity and cotranscriptionality in budding yeast". In: *Genome research* 28 (4), pp. 203–213. DOI: 10.1101/gr.236265.118.

Atwood, Scott X., Kavita Y. Sarin, et al. (Mar. 2015). "Smoothened Variants Explain the Majority of Drug Resistance in Basal Cell Carcinoma". In: *Cancer Cell* 27 (3), pp. 342–353. DOI: 10.1016/j.ccell.2015.02.002.

Atwood, Scott X., Ramon J. Whitson, and Anthony E. Oro (2014). "Advanced treatment for basal cell carcinomas". In: *Cold Spring Harbor Perspectives in Medicine* 4 (7). DOI: 10.1101/cshperspect.a013581.

Aubol, Brandon E. et al. (2003). "Processive phosphorylation of alternative splicing factor/splicing factor 2". In: *Proceedings of the National Academy of Sciences of the United States of America* 100 (22), pp. 12601–12606. DOI: 10.1073/pnas.1635129100.

Aznarez, Isabel et al. (2018). "Mechanism of Nonsense-Mediated mRNA Decay Stimulation by Splicing Factor SRSF1". In: *Cell Reports* 23 (7), pp. 2186–2198. DOI: 10.1016/j.celrep.2018.04.039.

Bandiera, Roberto et al. (2021). "RN7SK small nuclear RNA controls bidirectional transcription of highly expressed gene pairs in skin". In: *Nature Communications* 12 (1). DOI: 10.1038/s41467-021-26083-4.

Bapat, Aditi et al. (2018). "Myeloid Disease Mutations of Splicing Factor SRSF2 Cause G2-M Arrest and Skewed Differentiation of Human Hematopoietic Stem and Progenitor Cells". In: *Stem Cells* 36 (11), pp. 1663–1675. DOI: 10.1002/stem.2885.

Baris, Yasemin et al. (2022). "Fast and efficient DNA replication with purified human proteins". In: *Nature* 606. DOI: 10.1038/s41586-022-04759-1.

Basu, Souradeep et al. (2022). "Core control principles of the eukaryotic cell cycle". In: *Nature* 607 (July 2022). DOI: 10.1038/s41586-022-04798-8.

Bentley, David L, Christine Guthrie, and Joan Steitz (2005). "Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors". In: *Current Opinion in Cell Biology* 17, pp. 251–256. DOI: 10.1016/j.ceb.2005.04.006.

Bertoli, Cosetta, Jan M. Skotheim, and Robertus A.M. De Bruin (Aug. 2013). "Control of cell cycle transcription during G1 and S phases". In: *Nature Reviews Molecular Cell Biology* 14 (8), pp. 518–528. DOI: 10.1038/nrm3629.

Björk, Petra et al. (2009). "Specific combinations of SR proteins associate with single pre-messenger RNAs in vivo and contribute different functions". In: *Journal of Cell Biology* 184 (4), pp. 555–568. DOI: 10.1083/jcb.200806156.

Blanco, Sandra et al. (2016). "Stem cell function and stress response are controlled by protein synthesis". In: *Nature* 534 (7607), pp. 335–340. DOI: 10.1038/nature18282.

Blanpain, Cedric et al. (2004). "Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche". In: *Cell* 118 (5), pp. 635–648. DOI: 10.1016/j.cell.2004.08.012.

Blanpain, Cédric and Elaine Fuchs (2006). "Epidermal stem cells of the skin". In: *Annual Review of Cell and Developmental Biology* 22, pp. 339–373. DOI: 10.1146/annurev.cellbio.22.010305.104357.

— (2009). "Epidermal homeostasis: A balancing act of stem cells in the skin". In: *Nature Reviews Molecular Cell Biology* 10 (3), pp. 207–217. DOI: 10.1038/nrm2636.

Boija, Ann et al. (Dec. 2018). "Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains". In: *Cell* 175 (7), 1842–1855.e16. DOI: 10.1016/j.cell.2018.10.042.

Botti, Valentina et al. (2017). "Cellular differentiation state modulates the mRNA export activity of SR proteins". In: *Journal of Cell Biology* 216 (7), pp. 1993–2009. DOI: 10.1083/jcb.201610051.

Braastad, Corey D., Mariana Leguia, and Eric A. Hendrickson (2002). "Ku86 autoantigen related protein-1 transcription initiates from a CpG island and is induced by p53 through a nearby p53 response element". In: *Nucleic Acids Research* 30 (8), pp. 1713–1724. DOI: 10.1093/nar/30.8.1713.

Britten, Roy J and Eric H Davidson (1969). "Gene Regulation for Higher Cells : A Theory". In: *Science* (July), pp. 349–357.

Cáceres, Javier F., Gavin R. Screaton, and Adrian R. Krainer (1998). "A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm". In: *Genes and Development* 12 (1), pp. 55–66. DOI: 10.1101/gad.12.1.55.

Cahill, Daniel P. et al. (1998). "Mutations of mitotic checkpoint genes in human cancers". In: *Nature*.

Califice, Sophie et al. (2012). "A single ancient origin for prototypical Serine/arginine-rich splicing factors". In: *Plant Physiology* 158 (2), pp. 546–560. DOI: 10.1104/pp.111.189019.

Castro, Ignacio Pérez de, Guillermo de Cárcer, and Marcos Malumbres (May 2007). *A census of mitotic cancer genes: New insights into tumor cell biology and cancer therapy*. DOI: 10.1093/carcin/bgm019.

Cazalla, Demian et al. (2002). "Nuclear Export and Retention Signals in the RS Domain of SR Proteins". In: *Molecular and Cellular Biology* 22 (19), pp. 6871–6882. DOI: 10.1128/mcb.22.19.6871-6882.2002.

Centeno, Patricia P., Valeria Pavet, and Richard Marais (June 2023). "The journey from melanocytes to melanoma". In: *Nature Reviews Cancer* 23 (6), pp. 372–390. DOI: 10.1038/s41568-023-00565-7.

Chan, Yujia A. et al. (2014). "Genome-Wide Profiling of Yeast DNA:RNA Hybrid Prone Sites with DRIP-Chip". In: *PLoS Genetics* 10 (4). DOI: 10.1371/journal.pgen.1004288.

Chen, Edward Y et al. (2013). "Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool". In: *BMC Bioinformatics*.

Chen, Fei Xavier, Edwin R. Smith, and Ali Shilatifard (2018). "Born to run: Control of transcription elongation by RNA polymerase II". In: *Nature Reviews Molecular Cell Biology* 19 (7), pp. 464–478. DOI: 10.1038/s41580-018-0010-5.

Chen, Liang et al. (2018). "The Augmented R-Loop Is a Unifying Mechanism for Myelodysplastic Syndromes Induced by High-Risk Splicing Factor Mutations". In: *Molecular Cell* 69 (3), 412–425.e6. DOI: 10.1016/j.molcel.2017.12.029.

Chen, Mo and James L. Manley (2009). "Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches". In: *Nature Reviews Molecular Cell Biology* 10 (11), pp. 741–754. DOI: 10.1038/nrm2777.

Cheng, Yuanming et al. (2016). "Liver-Specific Deletion of SRSF2 Caused Acute Liver Failure and Early Death in Mice". In: *Molecular and Cellular Biology* 36 (11), pp. 1628–1638. DOI: 10.1128/mcb.01071-15.

Cockburn, Katie et al. (2022). "Gradual differentiation uncoupled from cell cycle exit generates heterogeneity in the epidermal stem cell layer". In: *Nature Cell Biology* 24 (December). DOI: 10.1038/s41556-022-01021-8.

Cohen, Erez et al. (2022). "Revisiting the significance of keratin expression in complex epithelia". In: *Journal of cell science* 135 (20). DOI: 10.1242/jcs.260594.

Connelly, Margery A. et al. (1998). "The promoters for human DNA-PK(cs) (PRKDC) and MCM4: Divergently transcribed genes located at chromosome 8 band q11". In: *Genomics* 47 (1), pp. 71–83. DOI: 10.1006/geno.1997.5076.

Cowper, Alison E. et al. (2001). "Serine-Arginine (SR) Protein-like Factors That Antagonize Authentic SR Proteins and Regulate Alternative Splicing". In: *Journal of Biological Chemistry* 276 (52), pp. 48908–48914. DOI: 10.1074/jbc.M103967200.

Crncec, Adrijana and Helfrid Hochegger (Oct. 2019). "Triggering mitosis". In: *FEBS Letters* 593 (20), pp. 2868–2888. DOI: 10.1002/1873-3468.13635.

Cui, Mingxue et al. (2008). "Genes involved in pre-mRNA 3prime-end formation and transcription termination revealed by a lin-15 operon Muv suppressor screen". In: *PNAS* 105 (43), pp. 16665–16670. DOI: 10.1073/pnas.0807104105.

Delaunay, Sylvain et al. (July 2022). "Mitochondrial RNA modifications shape metabolic plasticity in metastasis". In: *Nature* 607 (7919), pp. 593–603. DOI: 10.1038/s41586-022-04898-5.

Ding, Jian Hua et al. (2004). "Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart". In: *EMBO Journal* 23 (4), pp. 885–896. DOI: 10.1038/sj.emboj.7600054.

Dvinge, Heidi et al. (2016). "RNA splicing factors as oncoproteins and tumour suppressors". In: *Nature Reviews Cancer*. DOI: 10.1038/nrc.2016.51.

Earnshaw, William C. and Ann F. Pluta (1994). "Mitosis". In: *BioEssays* 16 (9), pp. 639–643. DOI: 10.1002/bies.950160908.

Eden, Eran, Doron Lipson, et al. (2007). "Discovering motifs in ranked lists of DNA sequences". In: *PLoS Computational Biology* 3 (3), pp. 0508–0522. DOI: 10.1371/journal.pcbi.0030039.

Eden, Eran, Roy Navon, et al. (Feb. 2009). "GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists". In: *BMC Bioinformatics* 10. DOI: 10.1186/1471-2105-10-48.

Edmond, Valerie et al. (2013). "A new function of the splicing factor SRSF2 in the control of E2F1-mediated cell cycle progression in neuroendocrine lung tumors". In: *Cell Cycle*, pp. 1267–1278.

Fernández-Nogales, Marta et al. (2014). "Huntington's disease is a four-repeat tauopathy with tau nuclear rods". In: *Nature Medicine* 20 (8), pp. 881–885. DOI: 10.1038/nm.3617.

Flynn, Ryan A. et al. (2011). "Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome". In: *Proceedings of the National Academy of Sciences of the United States of America* 108 (26), pp. 10460–10465. DOI: 10.1073/pnas.1106630108.

Fu, X. D. and T. Maniatis (1992). "The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3prime splice site". In: *Proceedings of the National Academy of Sciences of the United States of America* 89 (5), pp. 1725–1729. DOI: 10.1073/pnas.89.5.1725.

Fu, Yu et al. (May 2013). "SRSF1 and SRSF9 RNA binding proteins promote Wnt signalling-mediated tumorigenesis by enhancing β-catenin biosynthesis". In: *EMBO Molecular Medicine* 5 (5), pp. 737–750. DOI: 10.1002/emmm.201202218.

Galgóczy, Petra, André Rosenthal, and Matthias Platzer (2001). "Human-mouse comparative sequence analysis of the NEMO gene reveals an alternative promoter within the neighboring G6PD gene". In: *Gene* 271 (1), pp. 93–98. DOI: 10.1016/S0378-1119(01)00492-9.

Gao, Lei et al. (Jan. 2007). "SR protein 9G8 modulates splicing of tau exon 10 via its proximal downstream intron, a clustering region for frontotemporal dementia mutations". In: *Molecular and Cellular Neuroscience* 34 (1), pp. 48–58. DOI: 10.1016/j.mcn.2006.10.004.

Ge, Hui and James L. Manley (1990). "A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro". In: *Cell* 62 (1), pp. 25–34. DOI: 10.1016/0092-8674(90)90236-8.

Ge, Yejing et al. (2017). "Stem Cell Lineage Infidelity Drives Wound Repair and Cancer". In: *Cell* 169 (4), 636–650.e14. DOI: 10.1016/j.cell.2017.03.042.

Gilchrist, Daniel A. et al. (2012). "Regulating the regulators: The pervasive effects of Pol II pausing on stimulus-responsive gene networks". In: *Genes and Development* 26 (9), pp. 933–944. DOI: 10.1101/gad.187781.112.

Gkatza, Nikoletta Athanasia (2017). "The role of RNA modifications in cell homeostasis and in response to oxidative stress". In: (October).

Gout, Stephanie et al. (2012). "Abnormal Expression of the Pre-mRNA Splicing Regulators SRSF1, SRSF2, SRPK1 and SRPK2 in Non Small Cell Lung Carcinoma". In: *PLoS ONE* 7 (10). DOI: 10.1371/journal.pone.0046539.

Graveley, Brenton R and Tom Maniatis (1998). "Arginine/Serine-Rich Domains of SR Proteins Can Function as Activators of Pre-mRNA Splicing". In: *Molecular Cell* 1, pp. 765–771.

Gudikote, Jayanthi P. et al. (2005). "RNA splicing promotes translation and RNA surveillance". In: *Nature Structural and Molecular Biology* 12 (9), pp. 801–809. DOI: 10.1038/nsmb980.

Guillou, Emmanuelle, Javier Coloma, and Guillermo Montoya (2009). "The human GINS complex associates with Cdc45 and MCM and is essential for DNA replication". In: *Nucleic Acids Research* 37 (7), pp. 2087–2095. DOI: 10.1093/nar/gkp065.

Guo, Yang Eric et al. (Aug. 2019). "Pol II phosphorylation regulates a switch between transcriptional and splicing condensates". In: *Nature* 572 (7770), pp. 543–548. DOI: 10.1038/s41586-019-1464-0.

Haferlach, T. et al. (2014). "Landscape of genetic lesions in 944 patients with myelodysplastic syndromes". In: *Leukemia* 28 (2), pp. 241–247. DOI: 10.1038/leu.2013.336.

Han, Joonhee et al. (2011). "SR Proteins Induce Alternative Exon Skipping through Their Activities on the Flanking Constitutive Exons". In: *Molecular and Cellular Biology* 31 (4), pp. 793–802. DOI: 10.1128/mcb.01117-10.

Han, Lu et al. (1995). "Overexpression of the SR proteins ASF/SF2 and SC35 influences alternative splicing in vivo in diverse ways". In: *Cold Spring Harbor Laboratory Press*, pp. 1–17.

Hanahan, Douglas and Robert A. Weinberg (2000). "The Hallmarks of Cancer". In: *Cell* 100. DOI: 10.1107/S2059798322003928.

Heidemann, Martin et al. (2013). "Biochimica et Biophysica Acta Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription". In: *Biochimica et Biophysica Acta* 1829 (1), pp. 55–62. DOI: 10.1016/j.bbagrm.2012.08.013.

Hentze, Matthias W. and Elisa Izaurralde (2013). "Making sense of nonsense". In: *Nature Structural and Molecular Biology* 20 (6), pp. 651–653. DOI: 10.1038/nsmb.2601.

Herzog, Veronika A. et al. (2017). "Thiol-linked alkylation of RNA to assess expression dynamics". In: *Nature Methods* 14 (12), pp. 1198–1204. DOI: 10.1038/nmeth.4435.

Hir, Hervé Le et al. (2001). "The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay". In: *EMBO Journal* 20 (17), pp. 4987–4997. DOI: 10.1093/emboj/20.17.4987.

Hnisz, Denes et al. (Mar. 2017). *A Phase Separation Model for Transcriptional Control*. DOI: 10.1016/j.cell.2017.02.007.

Hochegger, Helfrid, Shunichi Takeda, and Tim Hunt (2008). "Cyclin dependent kinases and cell cycle transitions: does one fit all?" In: *Nature Reviews Molecular Cell Biology*, pp. 910–916.

Howard, Jonathan M. and Jeremy R. Sanford (2015). "The RNAissance family: SR proteins as multifaceted regulators of gene expression". In: *Wiley Interdisciplinary Reviews: RNA* 6 (1), pp. 93–110. DOI: 10.1002/wrna.1260.

Hsu, Patrick D., Eric S. Lander, and Feng Zhang (2014). "Development and applications of CRISPR-Cas9 for genome engineering". In: *Cell* 157 (6), pp. 1262–1278. DOI: 10.1016/j.cell.2014.05.010.

Huang, Lulu et al. (2011). "RNA Homeostasis Governed by Cell Type-Specific and Branched Feedback Loops Acting on NMD". In: *Molecular Cell* 43 (6), pp. 950–961. DOI: 10.1016/j.molcel.2011.06.031.

Ilagan, Janine O. et al. (2015). "U2AF1 mutations alter splice site recognition in hematological malignancies". In: *Genome Research* 25 (1), pp. 14–26. DOI: 10.1101/gr.181016.114.

Jackson, Stephen P. and Jiri Bartek (Oct. 2009). "The DNA-damage response in human biology and disease". In: *Nature* 461 (7267), pp. 1071–1078. DOI: 10.1038/nature08467.

Jacob, Franis and Jacques Monod (1961). "Genetic Regulatory Mechanisms in the Synthesis of Proteins". In: *J. Mol. Biol.* DOI: 10.1016/S0022-2836(61)80072-7.

Jaks, Viljar et al. (2008). "Lgr5 marks cycling, yet long-lived, hair follicle stem cells". In: *Nature Genetics* 40 (11), pp. 1291–1299. DOI: 10.1038/ng.239.

Janes, Sam M. and Fiona M. Watt (2006). "New roles for integrins in squamous-cell carcinoma". In: *Nature Reviews Cancer* 6 (3), pp. 175–183. DOI: 10.1038/nrc1817.

Jensen, Kim B., Ryan R. Driskell, and Fiona M. Watt (2010). "Assaying proliferation and differentiation capacity of stem cells using disaggregated adult mouse epidermis". In: *Nature Protocols* 5 (5), pp. 898–911. DOI: 10.1038/nprot.2010.39.

Jensen, Mads A., John E. Wilkinson, and Adrian R. Krainer (2014). "Splicing factor SRSF6 promotes hyperplasia of sensitized skin". In: *Nature Structural and Molecular Biology* 21 (2), pp. 189–197. DOI: 10.1038/nsmb.2756.

Jensen, Uffe B., Sally Lowell, and Fiona M. Watt (1999). "The spatial relationship between stem cells and their progeny in the basal layer of human epidermis: A new view based on whole-mount labelling and lineage analysis". In: *Development* 126 (11), pp. 2409–2418.

Ji, Xiong et al. (2013). "SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase". In: *Cell* 153 (4), pp. 855–868. DOI: 10.1016/j.cell.2013.04.028.

Jia, Rong et al. (2010). "SRp20 is a proto-oncogene critical for cell proliferation and tumor induction and maintenance". In: *International Journal of Biological Sciences* 6 (7), pp. 806–826. DOI: 10.7150/ijbs.6.806.

Jia, Yu et al. (Dec. 2021). "Crystal structure of the INTS3/INTS6 complex reveals the functional importance of INTS3 dimerization in DSB repair". In: *Cell Discovery* 7 (1). DOI: 10.1038/s41421-021-00283-0.

Joost, Simon et al. (Sept. 2016). "Single-Cell Transcriptomics Reveals that Differentiation and Spatial Signatures Shape Epidermal and Hair Follicle Heterogeneity". In: *Cell Systems* 3 (3), 221–237.e9. DOI: 10.1016/j.cels.2016.08.010.

Jumaa, Hassan, Grace Wei, and Peter J. Nielsen (1999). "Blastocyst formation is blocked in mouse embryos lacking the splicing factor SRp20". In: *Current Biology* 9 (16), pp. 899–902. DOI: 10.1016/S0960-9822(99)80394-7.

Karni, Rotem et al. (2007). "The gene encoding the splicing factor SF2/ASF is a proto-oncogene". In: *Nature Structural and Molecular Biology* 14 (3), pp. 185–193. DOI: 10.1038/nsmb1209.

Kim, Eunhee et al. (2015). "SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition". In: *Cancer Cell* 27 (5), pp. 617–630. DOI: 10.1016/j.ccell.2015.04.006.

Kim, J. et al. (2014). "Splicing factor SRSF3 represses the translation of programmed cell death 4 mRNA by associating with the 5 prime-UTR region". In: *Cell Death and Differentiation* 21 (3), pp. 481–490. DOI: 10.1038/cdd.2013.171.

Kim, V. N., N. Kataoka, and G. Dreyfuss (2001). "Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex". In: *Science* 293 (5536), pp. 1832–1836. DOI: 10.1126/science.1062829.

Koenigs, Vanessa et al. (2020). "SRSF7 maintains its homeostasis through the expression of Split-ORFs and nuclear body assembly". In: *Nature Structural and Molecular Biology* 27 (3), pp. 260–273. DOI: 10.1038/s41594-020-0385-9.

Kon, Ayana et al. (2018). "Physiological Srsf2 P95H expression causes impaired hematopoietic stem cell functions and aberrant RNA splicing in mice". In: *Blood* 131 (6), pp. 621–635. DOI: 10.1182/blood-2017-01-762393.

Kovaka, Sam et al. (Dec. 2019). "Transcriptome assembly from long-read RNA-seq alignments with StringTie2". In: *Genome Biology* 20 (1). DOI: 10.1186/s13059-019-1910-1.

Krainer, Adrian R., Greg C. Conway, and Diane Kozak (1990). "Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells". In: *Genes and Development* 4 (7), pp. 1158–1171. DOI: 10.1101/gad.4.7.1158.

Kuleshov, Maxim V. et al. (July 2016). "Enrichr: a comprehensive gene set enrichment analysis web server 2016 update". In: *Nucleic Acids Research* 44 (1), W90–W97. DOI: 10.1093/nar/gkw377.

Lane, David P. (1992). "guardian of the genome". In: *Nature*.

Lans, Hannes et al. (2019). "The DNA damage response to transcription stress". In: *Nature Reviews Molecular Cell Biology* 20 (12), pp. 766–784. DOI: 10.1038/s41580-019-0169-4.

Lareau, Liana F. et al. (2007). "Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements". In: *Nature* 446 (7138), pp. 926–929. DOI: 10.1038/nature05676.

Leclair, Nathan K. et al. (2020). "Poison Exon Splicing Regulates a Coordinated Network of SR Protein Expression during Differentiation and Tumorigenesis". In: *Molecular Cell* 80 (4), 648–665.e9. DOI: 10.1016/j.molcel.2020.10.019.

Leighton, J. Core, Joshua J. Waterfall, and John T. Lis (2008). "Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Inititation at Human Promoters". In: *Science* 322.

Li, Jun et al. (Dec. 2022). "Spatially resolved proteomic map shows that extracellular matrix regulates epidermal growth". In: *Nature Communications* 13 (1). DOI: 10.1038/s41467-022-31659-9.

Li, Xialu and James L. Manley (2005). "Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability". In: *Cell* 122 (3), pp. 365–378. DOI: 10.1016/j.cell.2005.06.008.

Li, Yang I. et al. (2018). "Annotation-free quantification of RNA splicing using LeafCutter". In: *Nature Genetics* 50 (1), pp. 151–158. DOI: 10.1038/s41588-017-0004-9.

Lin, Shengrong, Gabriela Coutinho-Mansfield, et al. (2008). "The splicing factor SC35 has an active role in transcriptional elongation". In: *Nature Structural and Molecular Biology* 15 (8), pp. 819–826. DOI: 10.1038/nsmb.1461.

Lin, Shengrong, Ran Xiao, et al. (2005). "Dephosphorylation-dependent sorting of SR splicing factors during mRNP maturation". In: *Molecular Cell* 20 (3), pp. 413–425. DOI: 10.1016/j.molcel.2005.09.015.

Liu, Lijun et al. (2019). "The cell cycle in stem cell proliferation, pluripotency and differentiation". In: *Nature Cell Biology* 21 (9), pp. 1060–1067. DOI: 10.1038/s41556-019-0384-4.

Long, Jennifer C. and Javier F. Caceres (2009). "The SR protein family of splicing factors: Master regulators of gene expression". In: *Biochemical Journal* 417 (1), pp. 15–27. DOI: 10.1042/BJ20081501.

Longman, Dása, Iain L. Johnstone, and Javier F. Cáceres (2000). "Functional characterization of SR and SR-related genes in Caenorhabditis elegans". In: *EMBO Journal* 19 (7), pp. 1625–1637. DOI: 10.1093/emboj/19.7.1625.

Lu, Yu et al. (July 2014). "Alternative splicing of MBD2 supports self-renewal in human pluripotent stem cells". In: *Cell Stem Cell* 15 (1), pp. 92–101. DOI: 10.1016/j.stem.2014.04.002.

Luch, Andreas (2005). "Nature and nurture - Lessons from chemical carcinogenesis". In: *Nature Reviews Cancer* 5 (2), pp. 113–125. DOI: 10.1038/nrc1546.

Lykke-Andersen, Jens, Mei Di Shu, and Joan A. Steitz (2000). "Human Upf proteins target an mRNA for nonsense-mediated decay when downstream of a termination codon". In: *Cell* 103 (7), pp. 1121–1131. DOI: 10.1016/S0092-8674(00)00214-2.

Mahat, Dig B. et al. (2016). "Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation". In: *Molecular Cell* 62 (1), pp. 63–78. DOI: 10.1016/j.molcel.2016.02.025.

Manley, James L. and Adrian R. Krainer (2010). "A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins)". In: *Genes and Development* 22 (9), pp. 2926–2929. DOI: 10.1105/tpc.110.078352.

Maquat, Lynne E., Woan Yuh Tarn, and Olaf Isken (2010). "The pioneer round of translation: Features and functions". In: *Cell* 142 (3), pp. 368–374. DOI: 10.1016/j.cell.2010.07.022.

Masai, Hisao et al. (2010). "Eukaryotic chromosome DNA replication: Where, when, and how?" In: *Annual Review of Biochemistry* 79, pp. 89–130. DOI: 10.1146/annurev.biochem.052308.103205.

Matthews, Helen K., Cosetta Bertoli, and Robertus A.M. de Bruin (2022). "Cell cycle control in cancer". In: *Nature Reviews Molecular Cell Biology* 23 (1), pp. 74–88. DOI: 10.1038/s41580-021-00404-3.

Meers, Michael P. et al. (2019). "Improved CUT&RUN chromatin profiling tools". In: *eLife* 8, pp. 1–16. DOI: 10.7554/eLife.46314.

Meselson, M. and F. Stahl (1958). "The replication of DNA in Escherichia coli". In: *PNAS*, pp. 671–682.

Moll, Roland, Markus Divo, and Lutz Langbein (2008). "The human keratins: Biology and pathology". In: *Histochemistry and Cell Biology* 129 (6), pp. 705–733. DOI: 10.1007/s00418-008-0435-6.

Moore, Melissa J. and Nick J. Proudfoot (2009). "Pre-mRNA Processing Reaches Back to Transcription and Ahead to Translation". In: *Cell* 136 (4), pp. 688–700. DOI: 10.1016/j.cell.2009.02.001.

Morgan, D. O. (2007). "The Cell Cycle: Principles of Control". In: *Yale Journal of Biology and Medicine*.

Mueller-Roever, Sven et al. (2001). "A Comprehensive Guide for the Accurate Classification of Murine Hair Follicles in Distinct Hair Cycle Stages". In: *The Journal of Investigative Dermatology*, pp. 3–15. DOI: 10.1046/j.0022-202x.2001.01377.x.

Mupo, A. et al. (Mar. 2017). "Hemopoietic-specific Sf3b1-K700E knock-in mice display the splicing defect seen in human MDS but develop anemia without ring sideroblasts". In: *Leukemia* 31 (3), pp. 720–727. DOI: 10.1038/leu.2016.251.

Murray, Andrew W (1992). "Creative blocks: cell cycle checkpoints and feedback controls". In: *Nature*, pp. 599–604.

Narla, Goutham et al. (2005). "A Germline DNA Polymorphism Enhances Alternative Splicing of the KLF6 Tumor Suppressor Gene and Is Associated with Increased Prostate Cancer Risk". In: *Cancer Research*.

Nassar, Dany et al. (July 2015). "Genomic landscape of carcinogen-induced and genetically induced mouse skin squamous cell carcinoma". In: *Nature Medicine* 21 (8), pp. 946–954. DOI: 10.1038/nm.3878.

Navarro, J Manuel, José Casatorres, and José L Jorcano (1995). "Elements Controlling the Expression and Induction of the Skin Hyperproliferation-associated Keratin K6*". In: *The Journal of Biological Chemistry*.

Neumann, Tobias et al. (2019). "Quantification of experimentally induced nucleotide conversions in high-throughput sequencing datasets". In: *BMC Bioinformatics* 20 (1), pp. 1–16. DOI: 10.1186/s12859-019-2849-7.

Nojima, Takayuki et al. (2015). "Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing". In: *Cell* 161 (3), pp. 526–540. DOI: 10.1016/j.cell.2015.03.027.

Nouspikel, Thierry P., Nevila Hyka-Nouspikel, and Philip C. Hanawalt (Dec. 2006). "Transcription Domain-Associated Repair in Human Cells". In: *Molecular and Cellular Biology* 26 (23), pp. 8722–8730. DOI: 10.1128/mcb.01263-06.

Nowak, Jonathan A. et al. (2008). "Hair follicle stem cells are specified and function in early skin morphogenesis". In: *Cell Stem Cell* 3 (1), pp. 33–43. DOI: 10.1016/j.stem.2008.05.009.

Pandit, Shatakshi et al. (2013). "Genome-wide Analysis Reveals SR Protein Cooperation and Competition in Regulated Splicing". In: *Molecular Cell* 50 (2), pp. 223–235. DOI: 10.1016/j.molcel.2013.03.001.

Papaemmanuil, Elli et al. (2013). "Clinical and biological implications of driver mutations in myelodysplastic syndromes". In: *Blood* 122 (22), pp. 3616–3627. DOI: 10.1182/blood-2013-08-518886.

Pascual, Gloria et al. (2017). "Targeting metastasis-initiating cells through the fatty acid receptor CD36". In: *Nature*. DOI: 10.1038/nature20791.

Pertea, Mihaela, Daehwan Kim, et al. (Sept. 2016). "Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown". In: *Nature Protocols* 11 (9), pp. 1650–1667. DOI: 10.1038/nprot.2016.095.

Pertea, Mihaela, Geo M. Pertea, et al. (2015). "StringTie enables improved reconstruction of a transcriptome from RNA-seq reads". In: *Nature Biotechnology* 33 (3), pp. 290–295. DOI: 10.1038/nbt.3122.

Peterlin, B Matija and David H Price (2006). "Controlling the Elongation Phase of Transcription with P-TEFb". In: *Molecular Cell* 2, pp. 297–305. DOI: 10.1016/j.molcel.2006.06.014.

Petermann, Eva, Li Lan, and Lee Zou (2022). "Sources, resolution and physiological relevance of R-loops and RNA–DNA hybrids". In: *Nature Reviews Molecular Cell Biology* 23 (8), pp. 521–540. DOI: 10.1038/s41580-022-00474-x.

Phatnani, Hemali P. and Arno L. Greenleaf (2006). "Phosphorylation and functions of the RNA polymerase II CTD". In: *Genes and development*, pp. 2922–2936. DOI: 10.1101/gad.1477006.important.

Philip, Ewels et al. (Mar. 2020). "The nf-core framework for community-curated bioinformatics pipelines". In: *Nature Biotechnology* 38 (3), pp. 272–276. DOI: 10.1038/s41587-020-0446-y.

Platzer, Matthias et al. (1997). "Ataxia-telangiectasia locus: Sequence analysis of 184 kb of human genomic DNA containing the entire ATM gene". In: *Genome Research* 7 (6), pp. 592–605. DOI: 10.1101/gr.7.6.592.

Poonian, M S, W W Mccomas, and A L Nussbaum (1977). "Chemical Synthesis of two deoxyribododecanucleotides for the attachment of restriction termini to an artificial minigene". In: *Gene* 1, p. 357.

Qiu, Jinsong et al. (2016). "Distinct splicing signatures affect converged pathways in myelodysplastic syndrome patients carrying mutations in different splicing regulators". In: *Cold Spring Harbor Laboratory Press* 22 (10), pp. 1535–1549. DOI: 10.1261/rna.056101.116.

Rahman, Mohammad Alinoor et al. (2020). "Recurrent SRSF2 mutations in MDS affect both splicing and NMD". In: *Genes and Development* 34 (5), pp. 413–427. DOI: 10.1101/gad.332270.119.

Ratnadiwakara, Madara et al. (2018). "SRSF3 promotes pluripotency through nanog mRNA export and coordination of the pluripotency gene expression program". In: *eLife* 7, pp. 1–28. DOI: 10.7554/eLife.37419.

Reed, Robin and Ed Hurt (2002). "A conserved mRNA export machinery coupled to pre-mRNA splicing". In: *Cell* 108 (4), pp. 523–531. DOI: 10.1016/S0092-8674(02)00627-X.

Rheinwald, James G and Howard Green (1975). "Serial Cultivation of Strains of Human Epidermal Keratinocytes: the Formation of Keratinizing Colonies from Single Cells". In: *Cell* 6, pp. 1–344.

Richardson, Dale N. et al. (2011). "Comparative analysis of serine/arginine-rich proteins across 27 eukaryotes: Insights into sub-family classification and extent of alternative splicing". In: *PLoS ONE* 6 (9). DOI: 10.1371/journal.pone.0024542.

Ronen, Amiram and Barry W. Glickman (2001). "Human DNA repair genes". In: *Environmental and Molecular Mutagenesis*. DOI: 10.1126/science.1056154.

Rué, Pau and Alfonso Martinez Arias (2015). "Cell dynamics and gene expression control in tissue homeostasis and development". In: *Mol Syst Biol* 11. DOI: 10.15252/msb.

Rzechorzek, Neil J et al. (2020). "CryoEM structures of human CMG – ATPgammaS – DNA and CMG – AND-1 complexes". In: *Nucleic Acids Research* 48 (12), pp. 6980–6995. DOI: 10.1093/nar/gkaa429.

Sajini, Abdulrahim A. et al. (2019). "Loss of 5-methylcytosine alters the biogenesis of vault-derived small RNAs to coordinate epidermal differentiation". In: *Nature Communications* 10 (1). DOI: 10.1038/s41467-019-10020-7.

Salamonsen, L. A. et al. (2001). "Genes involved in implantation". In: *Reproduction, Fertility and Development* 13 (1), pp. 41–49. DOI: 10.1071/RD00046.

Sanchez-Danes, Adriana and Cedric Blanpain (Sept. 2018). "Deciphering the cells of origin of squamous cell carcinomas". In: *Nature Reviews Cancer* 18 (9), pp. 549–561. DOI: 10.1038/s41568-018-0024-5.

Sanford, J. R., J. Ellis, and J. F. Càceres (2005). "Multiple roles of arginine/serine-rich splicing factors in RNA processing". In: *Biochemical Society Transactions* 33 (3), pp. 443–446. DOI: 10.1042/BST0330443.

Sanford, Jeremy R et al. (2004). "A novel role for shuttling SR proteins in mRNA translation". In: *Genes and Development*, pp. 755–768. DOI: 10.1101/gad.286404.nucleocytoplasmic.

Sapra, Aparna K. et al. (2009). "SR Protein Family Members Display Diverse Activities in the Formation of Nascent and Mature mRNPs In Vivo". In: *Molecular Cell* 34 (2), pp. 179–190. DOI: 10.1016/j.molcel.2009.02.031.

Sarkar, Sinjini et al. (June 2021). *Mitotic checkpoint defects: en route to cancer and drug resistance*. DOI: 10.1007/s10577-020-09646-x.

Saxena, Sneha and Lee Zou (2022). "Hallmarks of DNA replication stress". In: *Molecular Cell* 82 (12), pp. 2298–2314. DOI: 10.1016/j.molcel.2022.05.004.

Schmidt-Ullrich, Ruth and Ralf Paus (2005). "Molecular principles of hair follicle induction and morphogenesis". In: *BioEssays* 27 (3), pp. 247–261. DOI: 10.1002/bies.20184.

Schneider, Robert and Rudolf Grosschedl (2007). "Dynamics and interplay of nuclear architecture , genome organization , and gene expression". In: *Genes and Development*, pp. 3027–3043. DOI: 10.1101/gad.1604607.The.

Sharma, Sanjai et al. (2011). "Exon 11 skipping of E-cadherin RNA downregulates its expression in head and neck cancer cells". In: *Molecular Cancer Therapeutics* 10 (9), pp. 1751–1759. DOI: 10.1158/1535-7163.MCT-11-0248.

Shirai, Cara Lunn et al. (2015). "Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo". In: *Cancer Cell* 27 (5), pp. 631–643. DOI: 10.1016/j.ccell.2015.04.008.

Siegel, Rebecca L. et al. (Jan. 2023). "Cancer statistics, 2023". In: *CA: A Cancer Journal for Clinicians* 73 (1), pp. 17–48. DOI: 10.3322/caac.21763.

Singh, Guramrit et al. (2012). "The Cellular EJC Interactome Reveals Higher-Order mRNP Structure and an EJC-SR Protein Nexus". In: *Cell* 151 (4), pp. 915–916. DOI: 10.1016/j.cell.2012.10.032.

Skene, Peter J., Jorja G. Henikoff, and Steven Henikoff (2018). "Targeted in situ genome-wide profiling with high efficiency for low cell numbers". In: *Nature Protocols* 13 (5), pp. 1006–1019. DOI: 10.1038/nprot.2018.015.

Snippert, Hugo J. et al. (2010). "Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells". In: *Cell* 143 (1), pp. 134–144. DOI: 10.1016/j.cell.2010.09.016.

Sureau, A. et al. (2001). "SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs". In: *EMBO Journal* 20 (7), pp. 1785–1796. DOI: 10.1093/emboj/20.7.1785.

Tessarz, Peter and Tony Kouzarides (2014). "Histone core modifications regulating nucleosome structure and dynamics". In: *Nature Reviews Molecular Cell Biology* 15. DOI: 10.1038/nrm3890.

Thol, Felicitas et al. (2012). "Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes". In: *Blood* 119 (15), pp. 3578–3584. DOI: 10.1182/blood-2011-12-399337.

Trinklein, Nathan D. et al. (2004). "An abundance of bidirectional promoters in the human genome". In: *Genome Research* 14 (1), pp. 62–66. DOI: 10.1101/gr.1982804.

Tripathi, Vidisha et al. (2012). "SRSF1 regulates the assembly of pre-mRNA processing factors in nuclear speckles". In: *Molecular Biology of the Cell* 23 (18), pp. 3694–3706. DOI: 10.1091/mbc.E12-03-0206.

Tubbs, Anthony and André Nussenzweig (2017). "Endogenous DNA Damage as a Source of Genomic Instability in Cancer". In: *Cell* 168 (4), pp. 644–656. DOI: 10.1016/j.cell.2017.01.002.

Tufegdzic, Ana et al. (2020). "Regulation of the RNAPII Pool Is Integral to the DNA Damage Response". In: *Cell*, pp. 1245–1261. DOI: 10.1016/j.cell.2020.02.009.

Tumbar, Tudorita et al. (2004). "Defining the Epithelial Stem Cell Niche in Skin". In: *Science* 303 (January), pp. 359–363.

Twyffels, Laure, Cyril Gueydan, and Certain Sr (2011). "Shuttling SR proteins : more than splicing factors". In: *The FEBS Journal* 278, pp. 3246–3255. DOI: 10.1111/j.1742-4658.2011.08274.x.

Vaquero-Garcia, Jorge et al. (2016). "A new view of transcriptome complexity and regulation through the lens of local splicing variations". In: *eLife* 5 (FEBRUARY2016), pp. 1–30. DOI: 10.7554/eLife.11752.

Verma, Ajit K and R K Boutwell (1980). "Effects of dose and duration of treatment with the tumor-promoting agent, 12-O-tetradecanoylphorbol-13-acetate on mouse skin carcinogenesis". In: *Carcinogenesis* 1.

Vermeulen, Katrien, Dirk R Van Bockstaele, and Zwi N Berneman (2003). "The cell cycle : a review of regulation , deregulation and therapeutic targets in cancer". In: *Cell Prolif.*, pp. 131–149.

Vervoort, Stephin J. et al. (2022). "Targeting transcription cycles in cancer". In: *Nature Reviews Cancer* 22 (1), pp. 5–24. DOI: 10.1038/s41568-021-00411-8.

Vitale, Ilio et al. (2017). "DNA Damage in Stem Cells". In: *Molecular Cell* 66 (3), pp. 306–319. DOI: 10.1016/j.molcel.2017.04.006.

Wagner, Rebecca E. and Michaela Frye (2021). "Noncanonical functions of the serine-arginine-rich splicing factor (SR) family of proteins in development and disease". In: *BioEssays* 43 (4), pp. 1–11. DOI: 10.1002/bies.202000242.

Wakano, Clay et al. (July 2012). "The dual lives of bidirectional promoters". In: *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1819 (7), pp. 688–693. DOI: 10.1016/j.bbagrm.2012.02.006.

Wang, Eric T et al. (2008). "Alternative isoform regulation in human tissue transcriptomes". In: *Nature*. DOI: 10.1038/nature07509.

Wang, Huan-You et al. (2001). "SC35 Plays a Role in T Cell Development and Alternative Splicing of CD45". In: *Molecular Cell* 7, pp. 331–342.

Wang, Jin, Yoshio Takagaki, and James L. Manley (1996). "Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability". In: *Genes and Development* 10 (20), pp. 2588–2599. DOI: 10.1101/gad.10.20.2588.

Wang, Shuxiong et al. (2020). "Single cell transcriptomics of human epidermis identifies basal stem cell transition states". In: *Nature Communications* 11 (1). DOI: 10.1038/s41467-020-18075-7.

Wang, Yingzi et al. (Apr. 2005). "Tau exons 2 and 10, which are misregulated in neurodegenerative diseases, are partly regulated by silencers which bind a SRp30c·SRp55 complex that either recruits or antagonizes htra2β1". In: *Journal of Biological Chemistry* 280 (14), pp. 14230–14239. DOI: 10.1074/jbc.M413846200.

Welsh, Sarah A. and Alessandro Gardini (Mar. 2023). "Genomic regulation of transcription and RNA processing by the multitasking Integrator complex". In: *Nature Reviews Molecular Cell Biology* 24 (3), pp. 204–220. DOI: 10.1038/s41580-022-00534-2.

Wheeler, Emily C. et al. (Mar. 2022). "Integrative RNA-omics Discovers GNAS Alternative Splicing as a Phenotypic Driver of Splicing Factor–Mutant Neoplasms". In: *Cancer Discovery* 12 (3), pp. 836–855. DOI: 10.1158/2159-8290.CD-21-0508.

Whitson, Ramon J. et al. (Mar. 2018). "Noncanonical hedgehog pathway activation through SRF-MKL1 promotes drug resistance in basal cell carcinomas". In: *Nature Medicine* 24 (3), pp. 271–281. DOI: 10.1038/nm.4476.

Williams, Lucy H. et al. (2015). "Pausing of RNA Polymerase II Regulates Mammalian Developmental Potential through Control of Signaling Networks". In: *Molecular Cell* 58 (2), pp. 311–322. DOI: 10.1016/j.molcel.2015.02.003.

Wojcik, Sonja M., Donnie S. Bundman, and Dennis R. Roop (July 2000). "Delayed Wound Healing in Keratin 6a Knockout Mice". In: *Molecular and Cellular Biology* 20 (14), pp. 5248–5255. DOI: 10.1128/mcb.20.14.5248-5255.2000.

Wood, Richard D. et al. (2001). "Human DNA repair genes". In: *Science* 291. DOI: 10.1016/j.mrfmmm.2005.03.007.

Wu, Shang Ju et al. (2013). "Clinical implications of U2AF1 mutation in patients with myelodysplastic syndrome and its stability during disease progression". In: *American Journal of Hematology* 88 (11), pp. 3106–3111. DOI: 10.1002/ajh.23541.

Xiao, Ran et al. (2007). "Splicing Regulator SC35 Is Essential for Genomic Stability and Cell Proliferation during Mammalian Organogenesis". In: *Molecular and Cellular Biology* 27 (15), pp. 5393–5402. DOI: 10.1128/mcb.00288-07.

Xie, Zhuorui et al. (Mar. 2021). "Gene Set Knowledge Discovery with Enrichr". In: *Current Protocols* 1 (3). DOI: 10.1002/cpz1.90.

Xu, Chun Fang, Julie A. Chambers, and Ellen Solomon (1997). "Complex regulation of the BRCA1 gene". In: *Journal of Biological Chemistry* 272 (34), pp. 20994–20997. DOI: 10.1074/jbc.272.34.20994.

Xu, Liting et al. (Dec. 2018). "Phosphorylation of serine/arginine-rich splicing factor 1 at tyrosine 19 promotes cell proliferation in pediatric acute lymphoblastic leukemia". In: *Cancer Science* 109 (12), pp. 3805–3815. DOI: 10.1111/cas.13834.

Ya, Chieh Hsu and Elaine Fuchs (2022). "Building and Maintaining the Skin". In: *Cold Spring Harbor Perspectives in Biology* 14 (7). DOI: 10.1101/cshperspect.a040840.

Yamaguchi, Yuki, Hirotaka Shibata, and Hiroshi Handa (2013). "Transcription elongation factors DSIF and NELF: Promoter-proximal pausing and beyond". In: *Biochimica et Biophysica Acta* 1829 (1), pp. 98–104. DOI: 10.1016/j.bbagrm.2012.11.007.

Yoshida, Kenichi et al. (2011). "Frequent pathway mutations of splicing machinery in myelodysplasia". In: *Nature* 478 (7367), pp. 64–69. DOI: 10.1038/nature10496.

Yoshimi, Akihide et al. (2019). *Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis*. Vol. 574. Springer US, pp. 273–277. DOI: 10.1038/s41586-019-1618-0.

Yu, Qingming, Jun Guo, and Jianhua Zhou (July 2004). "A minimal length between tau exon 10 and 11 is required for correct splicing of exon 10". In: *Journal of Neurochemistry* 90 (1), pp. 164–172. DOI: 10.1111/j.1471-4159.2004.02477.x.

Zaborowska, Justyna, Sylvain Egloff, and Shona Murphy (2016). "The pol II CTD : new twists in the tail". In: *Nature Structural and Molecular Biology* 23 (9). DOI: 10.1038/nsmb.3285.

Zahler, A. M. et al. (1992). "SR proteins: A conserved family of pre-mRNA splicing factors". In: *Genes and Development* 6 (5), pp. 837–847. DOI: 10.1101/gad.6.5.837.

Zappa, Francesca et al. (2019). "The TRAPP complex mediates secretion arrest induced by stress granule assembly". In: *The EMBO Journal* (September), pp. 1–24. DOI: 10.15252/embj.2019101704.

Zhang, Jian et al. (2015). "Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities". In: *Proceedings of the National Academy of Sciences of the United States of America* 112 (34), E4726–E4734. DOI: 10.1073/pnas.1514105112.

Zhong, Xiang Yang et al. (2009). "SR Proteins in Vertical Integration of Gene Expression from Transcription to RNA Processing to Translation". In: *Molecular Cell* 35 (1), pp. 1–10. DOI: 10.1016/j.molcel.2009.06.016.

Zinani, Oriana Q.H., Kemal Keseroğlu, and Ertuğrul M. Özbudak (Jan. 2022). "Regulatory mechanisms ensuring coordinated expression of functionally related genes". In: *Trends in Genetics* 38 (1), pp. 73–81. DOI: 10.1016/j.tig.2021.07.008.