The role of isoform switching and neoantigen formation in the response to radiation

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas Term, 2023

Majd Abdulghani

Balliol College, University of Oxford

Abstract

This thesis investigates the role of isoform switching in response to ionising radiation (IR) and the modulation of this process by SRSF1, a proto-oncogenic splicing factor. Utilising deep RNA-sequencing of B-cell lines from ten healthy individuals, the study reveals extensive IR-induced isoform switching across the transcriptome, leading to potentially shorter transcripts that influence DNA damage response, apoptosis, and cell cycle arrest. Intriguingly, nearly half of the genes exhibiting isoform-level changes showed no differential expression at the gene level, highlighting the importance of isoform-specific analysis in understanding cellular responses to IR.

The RNA-binding protein SRSF1 is identified as a mediator of IR-induced isoform switching. Loss of SRSF1 expression, which is a common response to IR across various cell types, enhances radiosensitivity in cell lines and in cancer patients.

Moreover, the thesis explores the combined effect of SRSF1 knockdown and IR on triple-negative breast cancer cells, revealing an altered antigenic landscape with 86 putative neoantigens, and therefore offering insights into novel targeted immunotherapies. The findings propose SRSF1 as a prognostic marker for radiotherapy efficacy in the short-term, and present a foundation for future therapeutic approaches targeting SRSF1 in cancer treatment.

Acknowledgements

"On the Earth are signs for those who seek certainty. And in your own

selves, will you not then see?"

- Qur'an Translated Meaning [51:20]

I thank Allah first and foremost for allowing me to see some of His signs in our selves, and for the unearned blessing of being able to write this, physically and mentally. I pray that He accept it from me and amplify its benefits for me and for the world.

Second, it is strange to be writing and submitting a thesis, going about life business-as-usual, when genocide and ethnic cleansing are actively taking place in Palestine. I pray to Allah, and implore every person reading this to call for an immediate and lasting ceasefire and for the liberation of the colonized Palestinian people.

Now, this may be the last acknowledgements section I write in my life, so please bear with me its length, as I attempt to properly acknowledge everyone who directly or indirectly made this thesis possible (though I can never thank any of them enough—or thank them all). I don't think I will get another chance to say what I can say here.

My fantastic supervisor, Prof. Tim Humphrey. Tim, all your students probably tell you this, but I have to say it too: I hit the jackpot with you. I'm the envy of many of my friends for having you as a supervisor. The example I always give is how you start almost every meeting with, "how can I help you today?" which summarizes how you see your role as my supervisor: not exploitative, not "tough-loving", not even kindly observing, but encouraging, supporting, enabling, and empowering. You have taught me what it means to lead with

compassion, to not take life too seriously, and to never let failure be a setback. I can't thank you enough for making my DPhil journey as fulfilling as I hoped it would be.

Prof. Shona Murphy, my secondary supervisor. Many other DPhil students I meet say they never talk to their secondary supervisor. I always tell them they don't know what they're missing, although I suspect just talking to their secondary supervisor won't make much of a difference—it has to be you, Shona! Your cheerfulness every time we talk, your investment in my success from the beginning, and your practical solutions to all my lab woes have made a huge difference in my DPhil. I end every meeting with you energized, no matter how dejected I was feeling before it. Thank you.

Prof. Vivian G. Cheung, to whom I owe the honour of having Tim as a supervisor. This was only one of the many ways you proactively and extraordinarily supported me. From our first encounter, where you reached out to me after hearing my radio interview in 2016, to the unwavering support during the challenges of my Master's, the pandemic, and then motherhood, your mentorship has been a limitless supply of encouragement and inspiration. Your belief in me, especially when I was faced with the decision on whether to apply for the Rhodes, was a pivotal force in my success. As a minority in Western academia, your support has been life-changing. You embody the true essence of mentorship, showing me that it's about proactive involvement, understanding, and advocating for mentees. This has not only shaped my career, but also my own approach to mentorship. I'm striving hard to live up to the person you believe me to be, hoping that I do "change the world one day."

Thank you to Mr. Muhammad Alagil and to the Rhodes Trust, especially Mary Eaton, the registrar, and Domefafa Atiso, the selection and outreach manager. Mr. Alagil, your generous endowment has changed my life in more ways than you can imagine. The fact that

you extend your mentorship and generosity beyond it is extraordinary. I promise to live the rest of my life carrying the honour and responsibility of being the first Saudi Rhodes Scholar with pride and humility, and to work hard to have it be the starting point, rather than the biggest achievement. Mary, it's easy to understand why crying on your proverbial couch is a rite of passage for Rhodes Scholars, me no exception. From the first call with you in November, 2018, your thoughtful questions, gentle demeanour and genuine compassion were unmatched. Thank you for your kindness and support when the pressure got just a little bit too much, between motherhood and a DPhil. Fafa, thank you for being on my side from the start of the Rhodes chapter, for showing me the ropes before I even came to Oxford, for driving our family home from the hospital that one wonderful April day, and for always lending a listening ear. I treasured your and your family's warmth in the cold of Oxford.

Thank you to the Saudi Arabian Cultural Bureau for unbelievably generous yearly grants, which gave me the intellectual and financial freedoms to perform all the experiments in this thesis, and all the ones that never made it in. I'm so proud to be Saudi, and I can't wait to give back to my country. Thank you especially to Prof. Amal Fatani, the Saudi Cultural Attaché to the UK, for your relentless support for Saudi students. For believing in me, highlighting my achievements, and always answering my messages, even though you must receive hundreds a day. You are an absolute inspiration to all Saudis with your dedication to duty and humble spirit.

Thank you to Dr. Nicola Ternette, who I met in 2019 to talk about the project that is now the fourth chapter of my thesis. It took four years, but you never stopped supporting and encouraging me to get it done. Thank you for generously allowing me to use your lab

Thank you to postdocs in the Ternette Lab, Dr. Sam Palmer and Dr. Robert Parker. Sam, your thoughtful questioning of all my techniques during our weekly meetings helped me better understand what I was doing and make sure I'm running the appropriate analyses. Thank you for your generous help with running and understanding RNA-seq analyses. Rob, thank you for running my samples through the mass spec and helping me understand the output, for patiently answering all my questions, even up to your last day in the lab.

Thank you, Drs. Ejung Moon, Abhay Singh, Salem Almujari, and Su Phyu for kindly sharing cell lines whenver requested, no questions asked. Thank you, Prof. Mark Hill, for generously offering your tissue culture hood, incubator, and other equipment and consumables for my experiments; the last stage of my DPhil went much more smoothly because of your generosity. Thank you, Dr. Aaron Jeffries from the University of Exeter's Sequencing Facility. You have gone above and beyond to help me with the long-read sequencing, from figuring out the best strategy, to processing the data.

On to personal acknowledgements:

My mother, Prof. Heba Kurdi, this is for you, more than anyone else. Following in your footsteps has been one of the greatestest honours of my life (second only to being Muslim), and if I am even a quarter of the woman you are, I would count myself extraordinary (but I'm not). My ultimate objective in life is to please you and my father, after Allah, and make you proud. I hope this thesis brings you some joy and pride. You are my ultimate source of inspiration. Your unrelenting hard work and focus have taught me even more than your encouraging words have, although whenever I am stuck, I do remember what you always say, ‟علي أن أسعى، وليس علي إدراك النجاح.” I have never questioned, like other women and girls, "how do I balance work and family?" because I grew up with you as a mother, and you

always made it look so easy: raising five children with love, while climbing the academic ladder from BSc all the way to Professorship, and while patenting and publishing left and right. I am incredibly proud to be your daughter, and I thank Allah for it every single day.

My father, Dr. Yousef Abdulghani, thank you for being my rock, for being a voice of reason when nothing else makes sense, and for the pride you show in the little things I do and accomplish. In contrast to my mother's fast-paced approach to life and work, you have taught me to slow down and smell the roses, to see the beauty in little things, and to take nothing for granted. To embrace spontaneity and take a break when everything seems overwhelming. You also taught me that it's never too late to change, never too late to admit mistakes and to do better. Above all, thank you for teaching me the most important things in life to know: who my Lord is, who my Prophet is, what my religion is, and what my book is. When I answer these questions in the grave, *in Sha Allah*, I will think of you.

To my brothers. Majed, for 29+ years of making me laugh. Mohannad, for 29+ years of being there for me, no matter the time. Moath, for 25+ years of having something to teach me and something to learn from me. And Mohammad, for 22+ of being up for a good time, even when you started med school. It's because I have you four that I pushed for my daughter to have brothers (albeit not biological).

Thank you to my second family, the Arifs. Prof. Ibrahim Arif, thank you for your wisdom and support before, during, and after my PhD, and for encouraging my career trajectory and always helping where you can. Mama Faridah, thank you for raising the boy who became my incredible husband, and for watching Samar so I can write this thesis. I'm so lucky I get to consider you my second mtother. Waad, thank you for being the example that inspires me to work hard on this PhD and to stay focused, while taking all the breaks I

need. Thank you for being our only family in the UK; sharing Eid with you was one of the highlights of our years there. Farah, thank you for being excited with me about anything and everything I'm excited about. And thank you both for being the best (and only!) aunts for Samar.

Thank you Amna Alghamdi, for being the best second mom I can think of for my daughter, and for giving me peace of mind to write the thesis while you cared for her and let her bond with her brothers.

I always wished I had an older sister growing up. At some point, I came to accept the reality of the situation, but then Allah gave me three older sisters in the form of best friends. I only regret that none of them have ever met (yet?). Balsam: it's been 13 years of enduring and amazing friendship, despite over half of them being long-distance. We don't talk or text all the time, but we will always talk like no time has passed at all. Your clarity of purpose inspires me. You push me to carve a better life for myself professionally and personally, and in many ways, you mentor me in both. Thank you for sticking with me all this time. Bekah: out of all my friends, you are the most unexpected, in the best way possible. Thank you for inviting me to Girls in Science with you, even though you probably didn't want to at the time. Going through graduate school, pregnancy, and motherhood together is more than I could have hoped to go through with any biological sister. We haven't actually seen each other in person in over four years, but it feels like just months? I never imagined our friendship would stay this strong after this long apart, knowing we will probably never even live in the same continent again. Thank you for making it work. Thank you for always being there, through all my motherhood and grad school ups and downs—no one gets me like you do (not even Anmar). I can't wait to see you again, whenever that is. Last but certainly not

least, Asma: I never saw you coming, yet the joy you brought to me and my family is unimaginable. Thank you for suddenly giving me someone I can talk to about anything and everything any time—things I would not be able to tell anyone except you. Most of all, thank you for truly letting me feel what it's like to have an older sister in the house: sneaking into your bedroom late at night just to talk, silently communicating with looks from across the room, helping plan gatherings and being there from their start to when they end late at night, eating the leftovers and laughing about the chaos that we just went through. I can't wait to meet Prince Ali ;).

My two Saudi friends with whom I started the journey of the DPhil in Oncology: Norah Alrishedan and Ftoon Aljarbou. I pray that Allah eases the rest of your journey and multiply your reward. You both helped me through thick and thin, you were family in a strange and foreign place, and you selflessly gave everything you could when I needed help. Norah, I'm so proud of the work we've done together on our podcast, and I'm so grateful to you for actually making it happen. Thank you for always listening to my complaints of the DPhil, and for offering to help however you can. Ftoon, my first motherhood friend: I can't tell you how inspired I am by your strength. No matter how hard life hits you, you get up again and try even harder. You're the strongest person I ever met. You're an amazing mother, and Lamar and Fahad are so lucky to have you, as am I to call you my friend.

Dineo Serame, thank you for being the most surprising, yet one of the closest, friendships I made in Oxford. How many people can say they have a friend who wrote a poem about them? I can, thanks to your beautiful words :"). Thank you for in turn bringing out the poet in me (albeit temporarily), for uplifting me, and seeing in me something I don't. I can't wait to say I know the president of Botswana!

To my family of Saudi Rhodes Scholars: Sami, for connecting with me from that first interview in 2018, for endless supply of painkillers postpartum, for being Samar's first friend, for baked goods (which I always make better versions of, but it's nice to not have to make them myself), and for trusting me; Ghadah, thank you for persisting in your love-hate friendship with Samar, and for making me feel normal at a time when I really didn't—you have no idea how much your friendship helped even when I didn't think I needed help; Omar, thank you for always being up for a coffee or a meal out, never cancelling, always showing up (except this one time); Mneera, thank you for educating me about very wrong things going on in the world and locally, you've opened my eyes; Ahmed, thank you for your immediate فزعات, I always know I can count on you; Jood, thank you for being a force for justice, you've inspired me more than you know to stand up for myself and others; Mohammed, thank you for all your outreach efforts, for standing up with me on every stage, and going to the ones I couldn't go to. For the Scholars I sadly haven't overlapped with: Ahmed Alsulami, Tarfa Alshubayli, Haifa Alarrafah, Mohammed Alwelyee, Yousef Bukhari, and all future ones: thank you for being part of this family, I'm rooting for you.

Laiba and Humaid. Few are good couple friends, and even fewer are ones that let you feel so comfortable and unjudged. From board game nights to family-friendly lunches, you have made space for us and made us feel as normal as possible in a pandemic and after our lives turned upside down with Samar's birth. Laiba, your friendship through it all was a pillar I knew I could lean on (and heavily did), and that meant the world during my PhD. I can't wait to see how far you'll go, and I count myself extremely lucky that I get to call you my friend.

Thank you, Dr. Walaa Qattan, for your friendship from our first year in college, for your tips and feedback on my manuscript and posters, and for selflessly helping me navigate the strange job market in Saudi. And thank you, Dr. Waad Albawardi, for the strength you gave me when I couldn't imagine going any further.

Samar/Summer, my daughter. Although you might one day read the above, mentioning how challenging I found it to have and raise you during my DPhil, I want you to know you are the best decision I ever made, and one of the greatest blessings from Allah. You're only 2.5, but you make me smile and lift my spirits like no one else can, even when I'm at my lowest. I treasure that I get to hold you, that I get to receive the immense pure love that you pour into my life. Thank you for grounding me and reminding me what matters when I feel lost. You will always be my proudest achievement.

Last but not least, my husband Anmar, my best friend. Thank you for leading our family with respect, kindness, and love. Thank you for not caring what people thought or said about the fact that you became a stay-at-home dad for two years, while I worked on my DPhil, when you could have been an Assistant Professor. Thank you for never allowing me to feel guilty about it, and for fiercely loving me every day. Thank you for showing me that 'not all men,' is actually true. Thank you for helping me emotionally, mentally, physically, nutritionally (with your amazing cooking), and even with my DPhil—troubleshooting my code and even my experiments, although you barely understood them. I'll never forget that you helped me figure out it's trypsin. I can't believe how blessed I am to share my life with you. Thank you.

Finally, thank you to everyone who offered a smile, an encouraging word, who sat for coffee, or asked about my research in interest. You make the world better.

Contents

Statement of Contribution

The project that extends from Chapter 2 to Chapter 3.3.2 is a project I picked up from the Vivian G. Cheung Lab at the University of Michigan. When I started working on it, it had the bare bones (the differential gene expression and isoform switching data and a few preliminary figures), but I took it from there and made it my own during my DPhil. Therefore, although I did not contribute to the generation of all the data, all the Figures and Tables are my own work, except where otherwise indicated in the text of the legends. In addition, in the main text, where I say 'I' or 'we' when describing the results, indicates whether I performed the analysis or not, respectively.

I confirm that permission to reproduce any figures or tables was obtained from co-authors and that, where reference to the work of others was made, appropriate credit has been given.

List of Abbreviations

ACN: Acetonitrile

ANOVA: Analysis of Variance

ARC: Advanced Research Computing

ATM: Ataxia Telangiectasia Mutated

BLC: B-cell Lymphoma

BMRC: Biomedical Research Computing

BRCA: Breast Cancer

BWT: Burrows-Wheeler Transform

CAR: Chimeric Antigen Receptor

CEPH: Centre d'Etude du Polymorphisme Humaine

CPC2: Coding Predictor Calculator 2

DDR: DNA Damage Response

DMEM: Dulbecco's Modified Eagle's Medium

DSB: Double-strand Break

EDTA: Ethylenediaminetetraacetic acid

EJC: Exon Junction Complex

ERAAP: Endoplasmic Reticulum Aminopeptidase Associated with Antigen Processing

FBS: Fetal Bovine Serum

FDR: False Discovery Rate

FMLRC: FM-index Long Read Corrector

GEO: Gene Expression Omnibus

GO: Gene Ontology

HLA: Human Leukocyte Antigen

hnRNP: Heterogeneous Nuclear Ribonucleoprotein

HPLC: High-Performance Liquid Chromatography

ICD: Immunogenic Cell Death

IDR: Intrinsically Disordered Region

IFN: Interferon

IR: Ionising Radiation

KD: Knockdown

LC-MS^2: Liquid Chromatography-Tandem Mass Spectrometry

LDF: Linear Discriminative Function

LGG: Lower-Grade Glioma

MHC: Major Histocompatibility Complex

MISO: Mixture of Isoforms

NCBI: National Centre for Biotechnology Information

NDLB: Non-Denaturing Lysis Buffer

NF-κB: Nuclear Factor Kappa B

NMD: Nonsense-Mediated Decay

PABP: Poly(A)-Binding Protein

PBS: Phosphate-Buffered Saline

PCV: Percent Coefficient of Variation

PEV: Percent Explained Variance

PFI: Progression-Free Interval

PLC: Peptide Loading Complex

PMAD: Percent Median Absolute Deviation

PTC: Premature Termination Codon

PSI: Percent Spliced In

PVDF: Polyvinylidene Fluoride

RBP: RNA-Binding Protein

RNA-IP: RNA-Immunoprecipitation

RNAPII: RNA Polymerase II

RNA-seq: RNA sequencing

RPKM: Reads Per Kilobase of transcript per Million mapped reads

RRM: RNA Recognition Motif

RUV: Remove Unwanted Variation

SEM: Standard Error of the Mean

SOM: Self-Organizing Map

SRSF: Serine-arginine-rich Splicing Factor

TAP: Transporter Associated with Antigen Presentation

T2T: Telomere-to-Telomere

TCGA: The Cancer Genome Atlas

TFA: Trifluoroacetic Acid

TNBC: Triple-Negative Breast Cancer

VSN: Variance Stabilization Normalization

WebGestalt: WEB-based GEne SeT AnaLysis Toolkit

Chapter 1   Introduction

1.1 The Hallmarks of Cancer

Cancer is a multifaceted disease and one of the world's most pressing issues. At the heart of understanding it are the hallmarks of cancer, which were introduced in 2000 by Hanahan and Weinberg as a framework to systematically understand this disease (Hanahan and Weinberg 2000). This distilled the heterogeneity of cancer into a manageable set of underlying characteristics, which describe the essential traits that cancer cells acquire during their development, offering a clearer understanding of cancer's biological basis.

The original hallmarks outlined six fundamental properties in cancerous cells: self-sufficiency in growth signals, insensitivity to anti-growth signals, evading apoptosis, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis (Hanahan and Weinberg 2000). Additionally, they introduced genome instability as an enabling characteristic, a mechanism that facilitates the acquisition of the hallmarks by premalignant cells.

In 2011, the framework was updated to include two new hallmarks: deregulating cellular energetics and avoiding immune destruction, along with a new enabling characteristic, tumour-promoting inflammation (Hanahan and Weinberg 2011). This update reflected the significant advances in cancer research that transpired between 2000 and 2011.

Most recently, in 2022, Hanahan further expanded this framework (Hanahan 2022). This update introduced new hallmarks and enabling characteristics to accommodate emerging research and understanding in the field. The added hallmarks are unlocking phenotypic

plasticity, nonmutational epigenetic reprogramming, polymorphic microbiomes, and the inclusion of senescent cells as a significant component of the tumour microenvironment (Figure 1.1). These additions underscore the evolving nature of cancer research, emphasizing the adaptive and dynamic characteristics of cancer cells and their interactions with the surrounding environment.
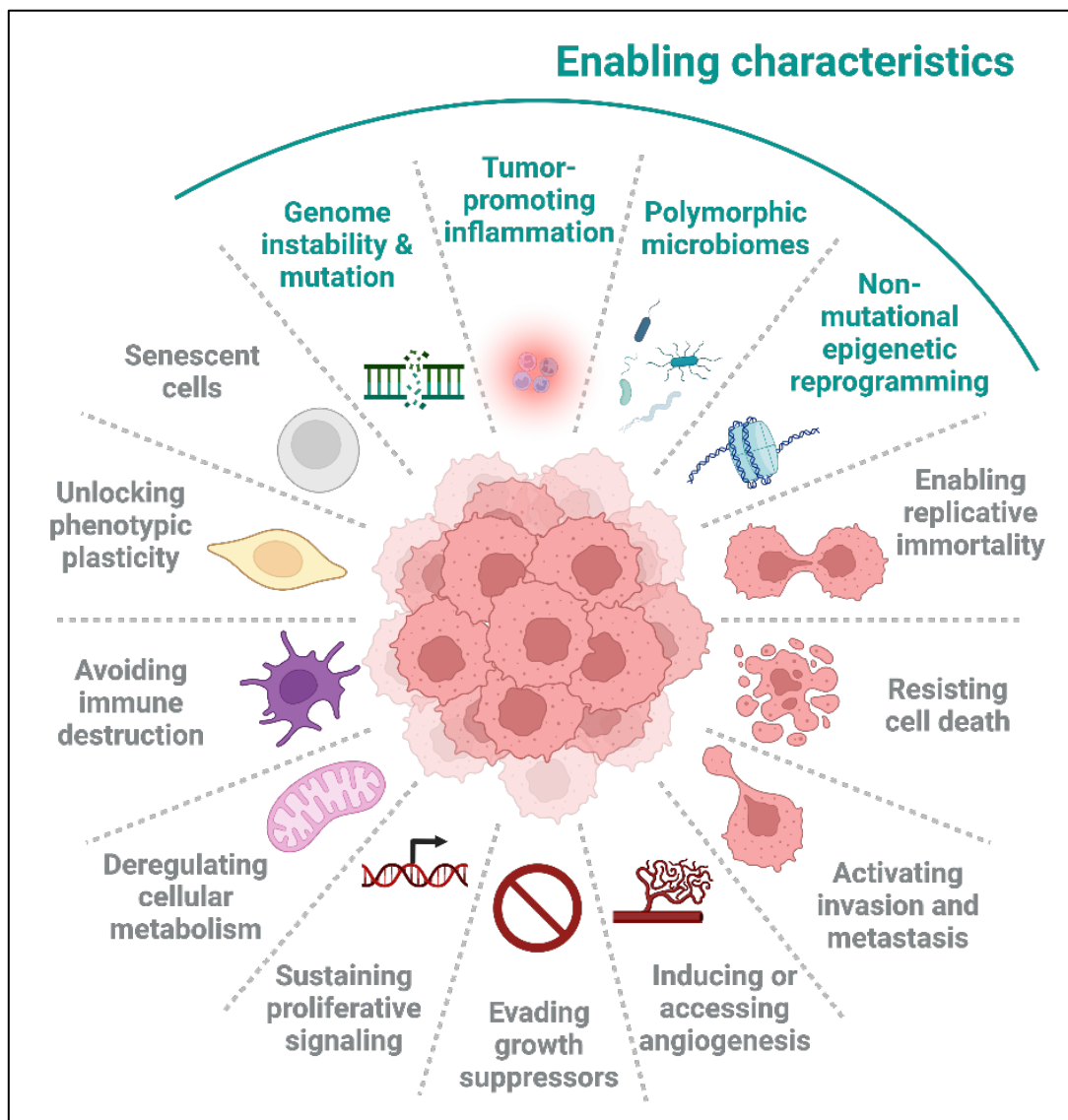
Figure 1.1 The hallmarks of cancer. Created with BioRender.com, adapted from Erin Marshall.

In this thesis, my work on understanding the impact of radiation on normal and cancerous cells frequently overlaps with a few of these hallmarks/enabling characterestics, the most relevant of which to my work are: resisting cell death, avoiding immune destruction, and genome instability and mutation.

As our understanding of cancer improves, these hallmarks provide a valuable conceptual framework. They not only encapsulate the current state of knowledge but also guide future research directions, offering insights into therapeutic avenues.

1.2 Radiotherapy

Since the first patient treatment in 1896 (Thariat et al. 2013), radiotherapy has been a mainstay of cancer treatment and mangement. As the accuracy and precision of imaging and treatment delivery techniques advanced (e.g. intensity-modulated radiation therapy), use of radiotherapy in cancer has expanded. Approximately 60% of patients in the US receive curative radiotherapy at some point in their treatment (Schaue and McBride 2015).

Radiotherapy works by delivering high-energy ionising radiation (IR; e.g. X-rays, α-rays, γ-rays, protons, and heavy ions) to malignant cells, which damages the DNA directly and indirectly, causing mutations, genomic instability, and apoptosis (Alizadeh et al. 2015). Direct DNA damage occurs when radiation directly hits the DNA molecule, while indirect effects are mostly the result of water radiolysis products, which generate reactive oxygen species that oxidize DNA and attack the sugar-phosphate backbone (Juan et al. 2021). The resulting genomic lesions include single-strand breaks, double-strand breaks (DSBs), DNA-DNA and DNA-protein crosslinks, abasic sites or other chemical modifications, and multiply damaged DNA sites (Hill and Ullrich 2019).

These lesions, of which DSBs are the most toxic, trigger damage sensors in the cell to activate the repair machinery. The MRE11/RAD50/NBS1 (MRN) complex binds to DSBs, and recruits the ATM kinase, which gets activated and phosphorylates the Histone 2 variant, H2AX. The production of phosphorylated H2AX (γH2AX) is a key step in DNA damage signalling, and is amplified by MDC1, leading to the recruitment of 53BP1 and BRCA1, among other DNA Damage Response (DDR) proteins. The activation of ATM also leads to the phosphorylation of CHK2, p53 and CDC25, triggering cell cycle arrest in the G1/S or G2/M checkpoints, although the majority of cells are arrested at G2, regardless of dose (Maity et al. 1994; Santivasi and Xia 2014; Vignard et al. 2013). This allows time for the repair of DNA damage before cell cycle progression. Non-homologous end joining is the major repair pathway after IR-induced damage, while homologous recombination occurs mainly after replication fork collapse in S-phase (Vignard et al. 2013).

If DNA damage is not repaired, some cells proceed with the cell cycle anyway, in a phenomenon termed checkpoint adaptation. This was first observed in yeast (Sandell and Zakian 1993), but is now known to occur in mammalian cells as well (Syljuåsen et al. 2006). Regardless, the fate of most cells with unrepaired DNA, especially through DSBs, is cell death; even if cells undergo a few divisions with unrepaired DSBs, most eventually die due to mitotic catastrophe (Dodson et al. 2007; Huber et al. 2015).

A variety of IR-induced cell death mechanisms exist: apoptosis, necrosis, and autophagy (Cohen–Jonathan et al. 1999; Palumbo and Comincini 2013; Roninson et al. 2001), as well as other mechanisms (Dodson et al. 2007; Lei et al. 2020; Smith et al. 2021; Wang et al. 2018a; Zhu et al. 2021). Choice of cell death mechanism depends on many factors, including

radiation dose, with higher doses (≥ 8 Gy) generally triggering necrotic death, and lower doses (≤ 5 Gy) generally triggering apoptotic death (Zhu et al. 2021).

A fundemental framework in radiobiology is the 5Rs (Withers 1975; Steel et al. 1989): repair, repopulation, redistribution, reoxygenation, and radiosensitivity, which determine the tissue response to multiple IR doses. Recently, a 6th R was proposed: reactivation of anti-tumour immune response (Boustani et al. 2019). The Rs are summarized in Figure 1.2, and I describe the ones relevant to this thesis in more detail.
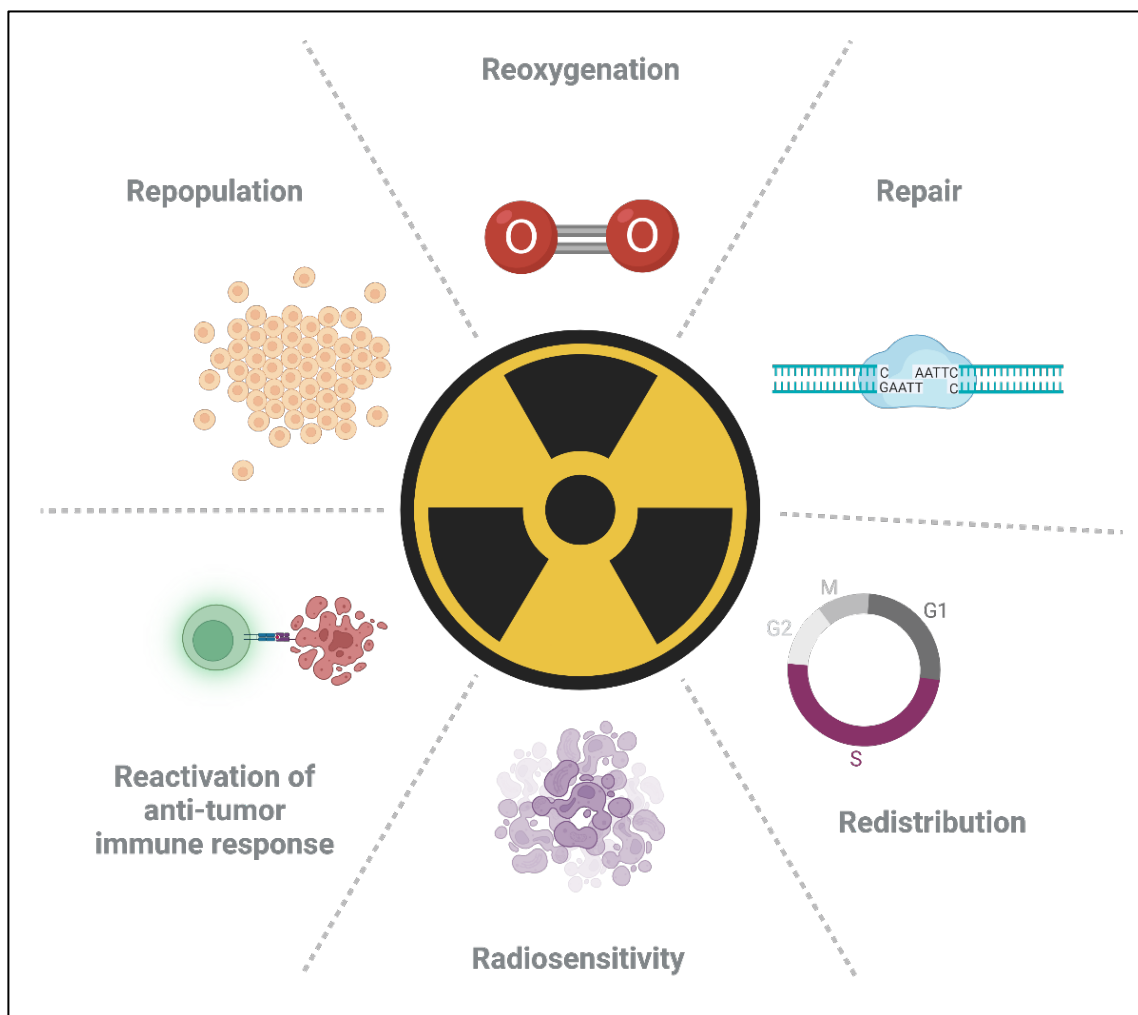
Figure 1.2 The 6 Rs of Radiotherapy. Created with BioRender.com.

1.2.1 Radiosensitivity

Individual response to radiation exposure is highly variable and difficult to predict. One patient may experience extreme adverse side-effects, such as increased cell death and higher incidence of secondary cancers, while another patient exposed to a similar dose may tolerate it well. This was first observed in patients with rare disease mutations that impact DNA repair pathways, such as ataxia-telangiectasia patients with mutations in the kinase ATM (Lavin and Khanna 1999).

However, not all individuals who experience sensitivity to IR are carriers of these mutations. Moreover, even in one individual, different tumours will have varying sensitivities. Radioresistance is typically associated with melanoma and glioblastoma, and with *KRAS* and *EGFR* mutations (Schaue and McBride 2015). Yet generally, mechanisms underlying individual response to radiation remain poorly understood. In order to minimize risks of accidental exposure and improve the efficacy of radiotherapy, a basic understanding of the mechanisms underlying cellular response to IR, and the factors determining radiosensitivity, is essential.

1.2.2 Reactivation of Anti-tumour Immune Response

The relationship between the immune response and radiotherapy is multifaceted. On one hand, some immunosuppressive effects have been attributed to IR (Rückert et al. 2021). On the other hand, radiotherapy can dramatically promote antitumour immunity through immunogenic cell death (ICD), in a type I interferon-dependent manner (Burnette et al. 2011; Deng et al. 2014). ICD is the release of tumour-associated or -specific antigens from dying cancer cells into the tumour microenvironment, effectively transforming the death of tumour cells into an *in situ* vaccine, which can have systemic benefits (Formenti and Demaria 2009;

Golden and Apetoh 2015). ICD also involves release of other immunostimulatory molecules, such as chemokines, calreticulin, and damage-associated molecular patterns (DAMPs), to name a few (Gameiro et al. 2014; Liao et al. 2020). Another way by which IR promotes antitumour immune response is by remodeling the tumour vasculature to increase immune cell infiltration (Jarosz-Biej et al. 2019).

In the context of this thesis, the role of radiotherapy in promoting production and release of cancer-specific antigens (i.e. neoantigens) is especially of interest. It has been known for nearly two decades that radiotherapy upregulates Major Histocompatibility Complex class I (MHC-I) on the cell surface and that it increases the antigen repertoire, therefore promoting antitumour T-cell response (Reits et al. 2006).

However, the focus has mostly been on neoantigens derived from mutated sites in the genome, which can generate peptides that are not present in normal cells (Formenti et al. 2018; Lhuillier et al. 2021; Ho et al. 2023). Yet, it is quite obvious that DNA mutations are not the only way to generate neoantigens. Alternative splicing occurs in about 95% of human genes (Pan et al. 2008; Wang et al. 2008), and can lead to the production of neoantigens, due to intron retention (Dong et al. 2021; Smart et al. 2018), for example, among other splicing aberrations. Indeed, aberrant alternative splicing has been proposed as one of the hallmarks of cancer (Ladomery 2013; Oltean and Bates 2014; El Marabti and Younis 2018; Zhang et al. 2019b; Wang and Lee 2018). Still, it has generally been overlooked as a mechanism for neoantigen formation, even though this has largely limited the use of immunotherapy to cancers with a high mutational burden, such as melanoma and non-small cell lung cancer (Alexandrov et al. 2013; Sha et al. 2020).

1.3 Antigen Presentation on MHC-I

MHCs (human leukocyte antigens [HLAs] in humans) present antigens to T cells to allow them to distinguish between self versus non-self, or harmless versus pathogenic, cells. Two classes of MHCs exist: class I and class II, and they differ in the cells that express them and the type of antigens they present. MHC-I is present on all nucleated cells and mainly presents products of cytosolic protein degradation by proteasomes to CD8+ T cells. On the other hand, MHC-II is on the surface of professional antigen-presenting cells, such as macrophages and B cells, and mainly presents products of lysosomal or endosomal proteolysis of cell-surface or extracellular proteins to CD4+ T cells (Rock et al. 2016). Here, I will focus on MHC-I, because the work in this thesis that pertains to antigen presentation is not carried out in professional antigen-presenting cells.

When a protein is destined for degradation, for example by getting ubiquitylated, the proteasome breaks it down into smaller peptides. While most of these are rapidly degraded by peptidases in the cytoplasm, some are instead translocated to the endoplasmic reticulum by transporter associated with antigen processing (TAP) (Reits et al. 2003). TAP accepts peptides that mostly conform to MHC-I binding specificities: 8-10 amino acids long and with specific anchor residues required for binding to MHC-I's peptide-binding groove (Pishesha et al. 2022; Nguyen et al. 2021). The required anchor residues differ significantly by HLA-I allele, of which > 26,000 exist in six genes (Barker et al. 2023). This significant polymorphism allows HLA-I molecules to present a wide array of peptides.

Once a peptide is translocated to the endoplasmic reticulum by TAP, it may be trimmed further by the endoplasmic reticulum aminopeptidase associated with antigen processing (ERAAP), after which it enters the peptide loading complex (PLC), which consists of TAP,

tapasin, MHC-I, ERp57, and calreticulin. The PLC loads peptides onto the MHC-I molecule. When a peptide with sufficient affinity binds to MHC-I, it is released from the endoplasmic reticulum and transported to the cell membrane to present the bound peptide on the surface (Neefjes et al. 2011).

In conditions of disease, such as cancer or microbial infections, the antigen repertoire expands significantly through several mechanisms. For example, in the context of cancer, DNA mutations or aberrations in splicing can generate cancer-specific antigens (neoantigens). In addition, activation of IFNγ alters the catalytic subunits of the proteasome to subunits that preferentially generate peptides more likely to bind MHC-I. This altered IFNγ-induced proteasome is called the immunoproteosome (Pishesha et al. 2022).

1.4 mRNA Splicing

In 1977, Phillip Sharp and Richard Roberts published their groups' independent discoveries of pre-mRNA splicing (Berget et al. 1977; Chow et al. 1977). This groundbreaking discovery has transformed our understanding of eukaryotic gene expression and proteome diversity. Pre-mRNA splicing is a co-transcriptional process where introns, which are generally non-protein-coding, are removed (spliced out) from the transcript, while exons are connected, to form the mature mRNA transcript. This process is divided into two categories: 1) constitutive splicing, where the same exons and/or introns are always removed from the mature transcript, regardless of cell type or environmental conditions; and 2) alternative splicing, where some exons are selectively included in the mature transcript, depending on cell type or environmental conditions. Alternative splicing can also affect introns, as some introns are occasionally retained in the mature transcript, and are referred to as retained introns.

Splicing serves several functions. For example, regulating gene expression (Brinster et al. 1988; Ding and Elowitz 2019; Jacob and Smith 2017; Shaul 2017), producing non-coding RNAs (Rearick et al. 2011; Westholm and Lai 2011), and diversifying the proteome (Liu et al. 2017), as different exons are spliced in or out of a transcript, producing new isoforms from the same gene, sometimes with different functions. Less than 50% of protein-protein interactions are preserved between a given pair of isoforms from the same gene (Yang et al. 2016). About 97% of protein coding genes contain at least one intron (Grzybowska 2013), and ~95% of those are expressed as more than one isoform, i.e. alternatively spliced (Pan et al. 2008; Wang et al. 2008). Each of the ~20,000 protein coding genes in the human genome has an average of ~6 isoforms, ~4 of which are protein-coding (The ENCODE Project Consortium 2012).

Mechanistically, splicing involves both *cis* and *trans* factors. *Trans* factors consist mainly of small nuclear ribonucleoproteins (snRNPs) and splicing factor proteins, for example heterogeneous nuclear ribonucleoproteins (hnRNPs) and SR (serine-arginine-rich) proteins. On the other hand, *cis* elements include exonic and intronic splicing enhancers and silencers, which are sequences recognized by specific splicing factors to enhance or inhibit splicing within an intron or exon (Figure 1.3).

Constitutive exons, which are always included in the mature mRNA, differ from alternatively-spliced exons, which may or may not be included in the mature mRNA, in terms of exon length (alternatively spliced elements are shorter), reading frame preservation (alternatively spliced exon length is usually a multiple of 3), splice site strength (splice sites of constitutive exons have a stronger sequence conservation to the consensus), among others (Zheng et al. 2005). Both constitutive and alternative splicing are carried out by a huge and

dynamic ribonucleoprotein machine called the spliceosome, made up of five core snRNPs. Over 300 cofactors can also associate with the spliceosome, and those aid sequence recognition and ATP hydrolysis, among other functions (Chen and Moore 2015).

Splicing mostly occurs co-transcriptionally, when the spliceosome binds to splice sites at the exon-intron junction, following the "GU-AG" rule to differentiate exons from introns. Specifically, it recognizes a 5' splice site with the sequence GU at the start of an intron, a 3' splice site with the sequence AG at the end of the intron, a branch point sequence upstream of the 3' splice site, with conserved adenine, and a polypyrimidine (PY) tract between the branchpoint and the 3' splice site (Kornblihtt et al. 2013; Zhang et al. 2021b).

Defects in constitutive and alternative splicing contribute to, as well as cause, many diseases, including cancer (Jiang and Chen 2021; Montes et al. 2019; Oltean and Bates 2014; Zhang et al. 2021b).
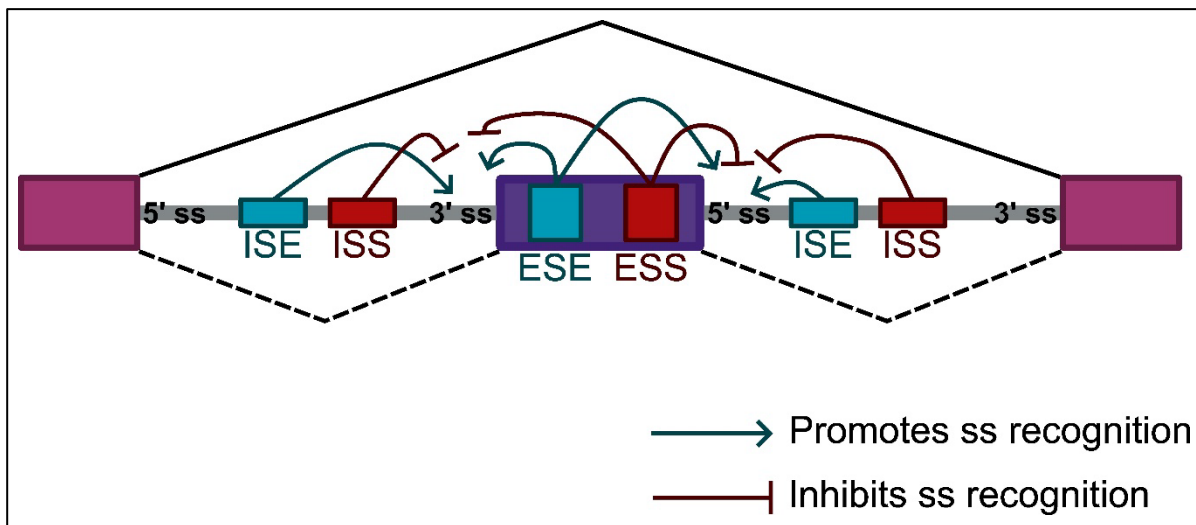
Figure 1.3 *Cis* elements that affect mRNA splicing include intronic and exonic splicing enhancers and silencers. *Trans* factors bind to these sequences to promote or inhibit splicing. ss, splice site; ISE, intronic splicing enhancer; ISS, intronic splicing silencer; ESE, exonic splicing enhancer; ESS, exonic splicing silencer.

1.4.1 Mechanism of Alternative mRNA splicing

As mentioned above, the majority of both alternative and constitutive splicing in humans is carried out by the spliceosome (Figure 1.4). This process starts when the snRNA within U1 snRNP base-pairs with the nascent mRNA's 5' splice site, which is at the start (5' end) of an exon-intron junction. The 5' splice site marks the end of the exon with the bases AG, and the start of an intron with the consensus sequence GURAGU (where R is a purine) (Marasco and Kornblihtt 2022). Concomitantly, Splicing Factor 1 (SF1) binds to the branch point sequence, closer to the 3' end of the intron (consensus sequence YNYURAY, where Y is a pyrimidine, and A is the branch point adenosine) and the U2 auxiliary factor (U2AF) binds to the polypyrimidine (PY) tract, located within the last 40 bases of the intron, and to the 3' splice site (consensus sequence AG). The binding of these three components (U1, SF1, and U2AF) of the spliceosome to the nascent mRNA forms the Early (E) complex (Shenasa and Bentley 2023). At this stage, the intron loops to bring the 5' and 3' splice sites within proximity, likely due to association of the 5' splice site with the transcription machinery during transcription (Leader et al. 2021). Replacement of SF1 with U2 at the branch point, and removal of U2AF, form the A complex. Next, the pre-catalytic B complex forms when the U4/U5/U6 tri-snRNP binds to U1 and the 5' splice site. To begin catalysis, U1 and U4 are destabilized. The major snRNPs at this stage are U6 at the 5' splice site, U2 at the branch point, and U5 stabilizing the interaction between the two (Shenasa and Bentley 2023). This complex allows for the first transesterification reaction to occur. Here, the hydroxyl (OH) group of the branch point adenosine attacks the phosphate of the first guanosine in the intron at the 5' splice site, releasing the upstream exon (which is kept in the vicinity of the reaction by the snRNPs) and generating an intron lariat attached to the

downstream exon. This is the C complex (Rogalska et al. 2022; Kretova et al. 2023). At this stage, the second transesterification reaction occurs between the free 3' OH of the upstream exon and the 5' phosphate group of the guanosine at the 3' splice site, creating a phosphodiester bond between the exons and releasing the intron lariat (Rogalska et al. 2022).
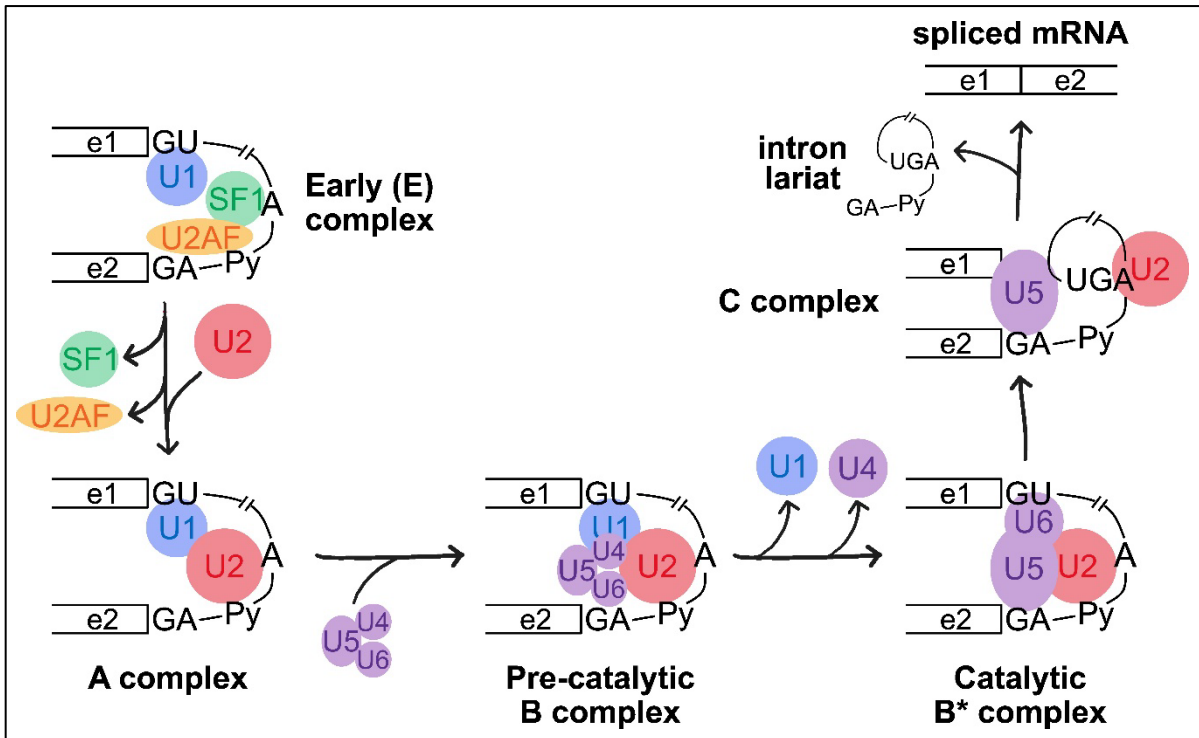
Figure 1.4 Mechanism of mRNA splicing. e1, exon 1; e2, exon 2; GU, 5' splice site; A, branch point adenosine; Py, polypyrimidine tract; AG, 3' splice site; SF1, splicing factor 1; U2AF, U2 auxiliary factor.

1.4.2 Types of Alternative Splicing

The degeneracy of splice site consensus sequences, and the presence of intronic and exonic splicing enhancers and silencers, create room for alternative splicing to occur, generating multiple transcripts from the same gene, depending on the optimality of the splice site and the binding of *trans* factors to the splicing enhancers and silencers (Figure 1.3).

Common types of alternative splicing are listed in Figure 1.5. Cassette exons, where an exon that is flanked by two other exons may be skipped or included in the mature transcript, are the most common type of alternative splicing (Dvinge and Bradley 2015; Choi et al. 2023). Alternative first exons are cases where two or more exons may be used as the first exon in the mature transcript. This is also sometimes called alternative promoters and, although it is frequently considered as such, may not strictly be a type of alternative splicing because it can occur without any splicing, simply by using an alternative promoter to start transcription. Similarly, alternative last exon events, also sometimes called alternative polyadenylation sites, may not involve any splicing, as transcription may terminate at an earlier exon, eliminating the need to splice out the more downstream one (Zhang et al. 2021a).

Alternative 5' or 3' splice sites are cases where, within the same exon, an upstream or downstream splice site may be used, resulting in a shorter or longer version of the same exon. Splice site selection depends on various factors, such as the closeness of the splice site in question to the consensus splice site sequence. Furthermore, splice site selection can also be influenced by the rate of transcription, as proposed by the kinetic coupling model, where the "window of opportunity" for recognizing upstream splice sites before downstream ones are synthesized is increased or decreased, depending on RNA pol II elongation rate (Kornblihtt 2007).

Finally, intron retention, where an intron is not spliced out of the mature mRNA (Grabski et al. 2021), occurs in nearly 90% of human multi-exon protein-coding genes (Middleton et al. 2017). Intron retention serves many functions, such as detaining the transcript in the nucleus for delayed splicing, downregulating gene expression by promoting regulated unproductive splicing and translation (RUST) and nonsense-mediated decay, and even generating novel proteins (Grabski et al. 2021).
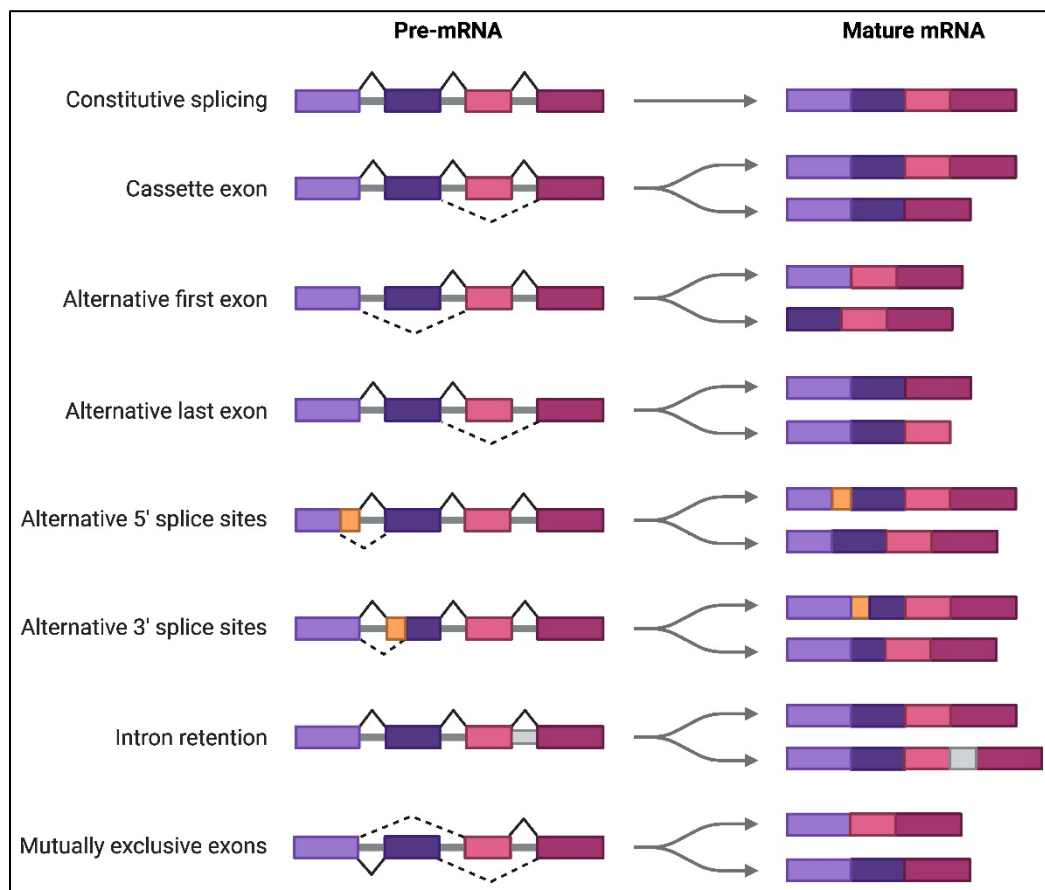
Figure 1.5 Types of mRNA splicing. Gray lines indicate introns, colored boxes indicate exons, solid versus dashed black lines indicate splicing possibilities, as shown to the right of the arrows. Adapted from "mRNA Splicing Types", by BioRender.com (2024). Retrieved from https://app.biorender.com/biorender-templates.

In this thesis, I differentiate between alternative splicing and isoform switching. Alternative splicing is a mechanism by which different isoforms are produced from the same

gene, as shown in Figure 1.5. On the other hand, isoform switching refers specifically to the change in expression of those isoforms across conditions. This may be due to alterations in transcription, alternative splicing, or in transcript stability.

1.4.3 Alternative Splicing and Nonsense-mediated Decay

During splicing, a multiprotein complex called the exon junction complex (EJC) is deposited ~24 nucleotides upstream of exon-exon junctions. This serves as a quality control mechanism to ensure the transcript is suitable for translation. If the ribosome encounters a stop codon upstream of an EJC during the first (pioneer) round of translation takes place, this signals that the stop codon is premature, and that the transcript is defective. Consequently, nonsense-mediated decay (NMD) machinery, which includes RNA nucleases, are recruited rapidly to degrade the transcript (García-Moreno and Romão 2020). Another mechanism to trigger NMD is when the termination codon encountered is distal to the termination promoting factor poly(A)-binding protein (PABP), which indicates that this termination codon is premature (Karousis and Mühlemann 2019). This can occur through splicing when splicing defects, or even physiological splicing, introduce a frameshift in the transcript that leads to the presence of a premature termination codon (PTC) (da Costa et al. 2017).

NMD, partly through its link with splicing, therefore serves as a quality control mechanism during translation, but also functions to regulate gene expression. Around a third of alternatively-spliced transcripts contain PTCs (Lewis et al. 2003). Cassette exons with PTCs to induce NMD, and therefore downregulate expression, are called poison exons. Those exons are prevalent in the serine-arginine (SR)-rich family of splicing factors (SRSFs), in addition to other RNA-binding proteins (RBPs), allowing them to autoregulate

their levels during differentiation by splicing their own transcripts to include or exclude the poison exon (Leclair et al. 2020). Interestingly, two recent studies found that poison exons exhibited anti-tumour effects (Thomas et al. 2020; Leclair et al. 2020).

In fact, the link between NMD and malignancy is well-known (da Costa et al. 2017; Tan et al. 2022; Nagar et al. 2023). However, whether it is beneficial or detrimental is context-dependent. In some cases, NMD serves to downregulate tumour suppressor genes, but not oncogenes, and it also removes aberrant transcripts that could have stimulated the anti-tumour immune response by forming neoantigens. In other cases, NMD is detrimental to tumours by degrading tumourigenic transcripts (Tan et al. 2022; Nagar et al. 2023).

1.4.4 RNA-binding Proteins

RBPs are a large class of proteins that serve diverse functions related to RNA. RBPs constitute at least ~8% of protein-coding genes, as there are over 1500 validated RBPs (Gerstberger et al. 2014), but some estimates bring them up to 4000 (Zhao et al. 2022). Some RBPs contain well-establish RNA-binding domains, but many others do not (Corley et al. 2020). An RBP can serve one or more of the following functions, as well as non-RNA-related ones: RNA editing, modification, splicing, polyadenylation, transport, translation, and degradation (Zhao et al. 2022).

As mentioned in the previous section, many RBPs serve as splicing factors to regulate their own and other transcripts' splicing, making them key players in proteome landscape and diversity. The two major families of splicing factors are the SRSF and the heterogenous nuclear ribonucleoprotein (hnRNP) families (Dvinge et al. 2016; Van Nostrand et al. 2020). SRSFs are canonically considered splicing enhancers, meaning they promote splice site

recognition and therefore exon inclusion. On the other hand, hnRNPs are canonically silencers, promoting exon exclusion. However, studies have shown that splicing factors deviate from their canonical roles, acting as both silencers and enhancers in a context-dependent manner (Fu and Ares 2014).

Splicing factors also serve functions independent of splicing (Wagner and Frye 2021). For example, SRSF1 and SRSF2 contribute to transcription by serving as subunits in the 7SK snRNP complex, which facilitates release of RNA Pol II from promoter-proximal pausing (Ji et al. 2013). In addition, several splicing factors interact with the EJC, promoting its binding to mRNA and therefore enhancing NMD in PTC-containing transcripts (Singh et al. 2012).

Interestingly, splicing factors are strongly associated with malignancy, and many of them, such as SRSF1, are considered oncoproteins (Dvinge et al. 2016). In this thesis, I focus on SRSF1 because I identified an important role for it in DNA-damage-associated isoform switching.

Serine-Arginine (SR)-rich Splicing Factor 1 (SRSF1)

First discovered by Adrian R. Krainer and Tom Maniatis (Krainer and Maniatis 1985), SRSF1 is the flagship member of the SR family of proteins. Like many of them, it contains an N-terminal RNA recognition motif (RRM) and a C-terminal SR-rich region, which facilitates protein-protein interactions. In between the two domains, SRSF1 also has a pseudo-RRM. The two RRMs are required for efficient binding to and splicing of RNA (Zuo and Manley 1993).

SRSF1 contributes to both constitutive and alternative splicing (Xiao and Manley 1998). Although canonically it binds to exonic splicing enhancers to promote splicing, SRSF1 also

functions as a splicing repressor (Anczuków et al. 2015; Du et al. 2021; Mayeda et al. 1993; Erkelenz et al. 2013; Pandit et al. 2013), depending on several factors, such as transcript length (Mayeda et al. 1993) and its precise binding location within the transcript (Anczuków et al. 2015; Erkelenz et al. 2013).

SRSF1 binds to RNA in a sequence- or structure-specific manner. Two structural motifs have been identified for SRSF1, one with a symmetrical internal loop and a 4-nucleotide hairpin loop, and another with a 7-nucleotide hairpin loop (Adinolfi et al. 2019). In terms of sequence-specificity, SRSF1-binding sequence motifs have been investigated extensively (Adinolfi et al. 2019; Feng et al. 2019; Anczuków et al. 2015; Barbagallo et al. 2021; Du et al. 2021; Pandit et al. 2013; Sanford et al. 2009; Ray et al. 2013; Tacke and Manley 1995). Although not identical, all identified SRSF1 motifs are purine-rich, and almost all of them specifically contain GGA, which alone is sufficient for SRSF1 binding (Feng et al. 2019).

Though it is classified as a splicing factor, SRSF1, like other splicing factors, has other functions, such as facilitating NMD, nuclear export, translation, and genome stability. Many of these functions are facilitated by posttranslational modifications, primarily phosphorylation and methylation of the serine and arginine residues, respectively, in the SR domain (Das and Krainer 2014). In terms of NMD, SRSF1 enhances it in an EJC-independent manner by recruiting UPF1, a core NMD protein, to the mRNA (Aznarez et al. 2018), and its splicing activity promotes its effect on NMD (Aznarez et al. 2018; Zhang and Krainer 2004).

In terms of genome stability, which has direct implications in cancer, SRSF1 loss leads to the accumulation of R-loops, three-stranded structures that occur when the nascent RNA binds to the template DNA strand, causing damage in the exposed non-template DNA strand (Crossley et al. 2019). Overexpression of RNase H1, an enzyme that facilitates R-loop

resolution, suppressed the cell cycle arrest and genome instability caused by SRSF1 depletion but only delayed cell death (Li and Manley 2005).

As mentioned in 1.4.4, SRSF1 is a proto-oncogene: it is frequently overexpressed in several cancers; its overexpression transforms rodent fibroblasts, which go on to form sarcomas; and finally, restoring its expression reverses the transformed state (Karni et al. 2007). Consistent with this, SRSF1 has been shown to inhibit apoptosis *in vitro* in a p53-independent manner (Das et al. 2012). Furthermore, increased SRSF1 expression is correlated with increased glioma grade and poor patient survival, and SRSF1 promotes proliferation, survival, and invasion of glioma cell lines (Zhou et al. 2018). Similarly, SRSF1 increases the aggressiveness of breast cancer by promoting migration, proliferation, and cell cycle progression, while inhibiting apoptosis (Anczuków et al. 2015; Du et al. 2021).

1.4.5 Alternative Splicing in Tumourigenesis

Alternative splicing is infamously aberrant in cancer cells (Bonnal et al. 2020; Oltean and Bates 2014; Schwenzer et al. 2021) and contributes to each of the hallmarks of cancer (Hanahan 2022). Aberrations result from a myriad of defects, from mutations in splice sites and splicing enhancers/suppressors to overexpression of splicing factors. Furthermore, driver mutations in splicing factors are frequent in myeloid neoplasms and lymphoproliferative disorders. For example mutations in splicing factors, such as SRSF2 and SF3B1, are present in 45-85% of myeloid neoplasms with features of myelodysplasia (Yoshida et al. 2011). Another study identified putative driver mutations in 119 splicing factors across 33 cancer types, including lung adenocarcinoma and bladder carcinoma (Seiler et al. 2018).

A review by Yoshimi and Abdel-Wahab highlighted the "therapeutic window" available for targeting spliceosome-mutant/dysregulated cancers, including those that upregulate SRSF1 (Yoshimi and Abdel-Wahab 2017). Spliceosome-targeting is an exciting area of research, with the development of antisense oligonucleotides to inhibit specific splicing, and drugs that target core, as well as accessory splicing machinery to partially inhibit splicing nonspecifically.

For example, H3B-8800, a small-molecule inhibitor of the SF3b complex, which is an essential spliceosome component, showed promise in early clinical trials (Stanley and Abdel-Wahab 2022). Another example is E7820, which degrades the accessory splicing factor RBM39. This drug is currently in Phase II clinical trials for refractory myeloid neoplasms with splicing factor mutations (NCT05024994).

4.1.1.1 *Alternative splicing in avoiding immune destruction*

One of the ways cancer cells avoid recognition by the immune system is by expressing Human Leucocyte Antigen (HLA)-G. This is expressed in placental cells and maternal bloodstream during pregnancy to suppress rejection of the fetus by the immune system (Xu et al. 2020). Cancer has co-opted this mechanism of evading immune destruction. HLA-G has several isoforms, some of which are functionally superior in inhibiting immunocompetent cells (Martín-Villa et al. 2022). In addition to HLA-G, mutations in several splicing factors increase NF-κB signaling, dependence on STAT1/Type I interferon signaling, and cytokine production (Sciarrillo et al. 2020).

4.1.1.2 *Alternative splicing in genome instability and mutation*

Mutations or alterations in the expression of splicing factors or their upstream regulators can alter splicing transcriptome-wide (Tam and Stirling 2019; Agrawal et al. 2018), thus

affecting genes that maintain genomic integrity directly or indirectly. In addition, splicing factors themselves directly maintain genomic integrity in several ways, which can be disrupted in cancer.

Splicing factors, such as SR splicing factors 1-3 (SRSF1-3), prevent nascent RNA binding to the template DNA, restricting formation of R-loops, which contribute to genomic instability (Li and Manley 2005; Paulsen et al. 2009; Naro et al. 2015). Indeed, inhibiting splicing with pladienolide B promotes R-loop-induced DNA damage (Wan et al. 2015). Therefore, aberrations in splicing or mutations in splicing factors in malignant cells contribute to genomic instability by promoting R-loop formation.

In addition, splicing factors such as hnRNP U and hnRNP G facilitate faithful mitosis (Sundaramoorthy et al. 2014; Somma et al. 2020), which is crtitical to chromosomal integrity; protect telomeres (Flynn et al. 2011; Zhang et al. 2006); and participate in DNA repair. Furthermore, some DDR proteins function in splicing, such as BRCA1, which forms a complex with splicing machinery to regulate the splicing of a number of DDR genes (Savage et al. 2014).

4.1.1.3 *Alternative splicing in resisting cell death*

Many regulators of apoptosis exist as multiple isoforms with opposite functions (Schwerk and Schulze-Osthoff 2005; Wu et al. 2003). After DNA damage, the shorter, pro-apoptotic isoform of B-cell Lymphoma (BCL)-X is expressed at a higher level relative to the longer anti-apoptotic isoform (BCL-XL), as signaling pathways are triggered to alter the activity of splicing repressors (Shkreta et al. 2011). In the context of cancer, BCL-XL is usually overexpressed and contributes to treatment resistance (Keitel et al. 2014; Wang et al. 2014; Amundson et al. 2000; de Jong et al. 2018).

Caspases, crucial apoptosis regulators, also exist in antagonistic isoforms. In renal cancer, SRSF2 is reduced, promoting the the expression of the anti-apoptotic isoforms of caspases 8 and 9 (caspase-8L and caspase-9b), while decreasing, or interfering with, the pro-apoptotic isoforms (caspase-8a and caspase-9a) (Kędzierska et al. 2016).

Even at the level of cell surface receptors, the death receptor FAS is alternatively spliced to either the full-length isoform or to soluble FAS (sFAS), missing the transmembrane domain. sFAS is upregulated in cervical cancer and melanoma (Contreras-Ochoa et al. 2022; Owen-Schaub 2001).

P53 itself is alternatively spliced into at least 12 isoforms exhibiting various specialized functions (Joruiz and Bourdon 2016). For example, while full-length p53 promotes apoptosis, P53β promotes sensescence (Chen et al. 2021). Work in the Kastan lab has demonstrated the importance of P53β in IR-induced cellular senescence, mediated by ATM-dependent SMG1 inhibition (Chen et al. 2017a; McCann et al. 2023).

This list is not exhaustive by any means, neither to resiting cell death, nor to the other hallmarks discussed. Alternative splicing is integral to every aspect of the hallmarks of cancer, to the extent that it was proposed as another enabling characterestic (Ladomery 2013).

1.4.6 Alternative Splicing and Neoantigen Formation

Seminal work from the Abdel-Wahab and Bradley labs demonstrated that the splicing-modulating drugs indisulam or MS-023 enhanced the efficacy of immune checkpoint blockade in a mouse xenograft model by producing splicing-derived immunogenic neoantigens (Lu et al. 2021). Another study found that the number of peptides derived from

neojunctions (exon-exon junctions exclusive to tumours) is nearly triple those derived from single-nucleotide variants (Kahles et al. 2018).

One of the mechanisms attributed to alternative splicing-derived neoantigen formation is intron retention, which is common in cancer (Dvinge and Bradley 2015; Jung et al. 2015). mRNAs with retained introns are normally degraded via nonsense-mediated decay (NMD) (Grabski et al. 2021; Monteuuis et al. 2019), but this is dependent on the pioneer round of translation to detect the premature termination codons that trigger NMD. Therefore, despite the retained-intron-containing transcript not being translated to a full-length protein, peptides derived from it can be generated and be a source of neoantigens (Apcher et al. 2011; Smart et al. 2018).

1.4.7 Alternative Splicing in Response to Ionising Radiation

While the transcriptional response to IR is well-characterized (Jen and Cheung 2003; Kis et al. 2006; Rieger and Chu 2004; Rødningen et al. 2005; Smirnov et al. 2009, 2012), our understanding of the underlying isoforms contributing to the transcriptional IR response is lacking, even though the relationship between splicing and the DNA damage response (DDR) is increasingly appreciated, as expanded on in 4.1.1.2 and reviewed in (Tam and Stirling 2019).

One of the earliest studies of alternative splicing response to IR in humans focussed on Fibroblast Growth Factor 2 (FGF2). The authors identified one isoform of FGF2, the 24 kDa form, as key to FGF2-mediated radioresistance in HeLa cells, increasing G2 arrest, while three other tested isoforms did not influence radiosensitivity (Cohen-Jonathan et al. 1997). They then went on to show that in 24 kDa FGF2-transfected cells, an alternatively spliced

variant of nucleophosmin 2 (NPM2) is overexpressed and plays a role in 24 kDa FGF2-mediated radioresistance (Dalenc et al. 2002).

Around the same time, another group was investigating RAD17 alternative splicing in response to IR. After verifying the expresion of 4 isoforms at the protein level, they found that the 73 kDa isoform increases $\geq$ 5-fold after IR, but they did not identify the mechanism or implications of this increased isoform expression (Chen et al. 2001).

In contrast, more research has gone into investigating Clusterin (CLUs) proteins' role in the response to IR (Miyake et al. 2001; Gleave et al. 2001; Yang et al. 2000; Leskov et al. 2003; Criswell et al. 2003) (reviewed in (Klokov et al. 2004)). CLU exists in two main isoforms: a secretory glycoprotein (sCLU), which promotes radioresistance, and a nuclear one (nCLU), which has the opposite effect. After IR, sCLU is downregulated in a P53-dependent manner, allowing cells to die (Criswell et al. 2003).

Sprung et al. were the first to investigate the relationship between alternative splicing and IR on a genome-wide scale and identified IR-specific isoform switching (i.e. differential isoform usage) over different doses, times, and cell lines (Sprung et al. 2011). Other studies since have confirmed the relevance of isoform switching to the IR response (Macaeva et al. 2016; Quintens et al. 2015; Forrester et al. 2012; Chen et al. 2017a), but all used exon arrays, which are limited by exon annotation and thus may overlook alternative splicing involving un-annotated exons. Wahba et al. sequenced total and polysome-bound RNA to characterize isoform switching 1 hour post exposure to 2 Gy IR in a glioblastoma stem-like cell line (Wahba et al. 2017). They found that not only is isoform switching induced in response to IR, but translation also favors specific isoforms.

Recently, publicly available next-generation RNA-seq datasets have started to be utilized to further elucidate the interplay between IR and alternative splicing. Through this, it was found that alternative splicing of protein arginine methyltransferase 5 (PRMT5), itself a known regulator of many splicing factors' activities, is also implicated in the IR response. *PRMT5-ISO5* is radiosentitizing in hepatocellular carcinoma patients, and its levels are positively regulated by HNRNPH1 and negatively regulated by SRSF3 (Wen et al. 2022).

Despite these efforts, there remains a large gap in our understanding of how isoform switching modulates the early and late IR response in normal human cells.

1.5 Thesis Aims and Objectives

In this thesis, my goal is to bridge the gap in our understanding of the role of isoform switching in modulating the IR response in human cells. I have three aims to achieve. I list them below, broken down by specific objectives.

Aim 1: Investigate the genome-wide RNA isoform switching response to ionising radiation (Chapter 2)

Objectives:

- Identify differentially expressed genes and isoforms in response to IR in B-cell lines derived from healthy individuals.
- Characterize the temporal dynamics and types of alternative splicing events induced by IR.
- Determine the impact of IR-induced isoform switching on transcript length and protein function.

Aim 2: Elucidate the role of the splicing factor SRSF1 in mediating the isoform switching response to IR and its implications for radio- and immunotherapy resistance (Chapter 3).

Objectives:

- Identify IR-responsive RBPs and assess their binding potential to alternatively-spliced isoforms *in silico*.

- Validate SRSF1 as a potential mediator of IR-responsive isoform switching through RNA-immunoprecipitation and sequencing.

- Investigate the role of SRSF1 in the IR response using *in vitro* knockdown and overexpression experiments.

- Assess the impact of SRSF1 expression on patient outcomes after radiotherapy and immunotherapy using data from The Cancer Genome Atlas.

Aim 3: Evaluate the effects of SRSF1 knockdown and IR on gene expression, isoform switching, and neoantigen generation in triple-negative breast cancer cells (Chapter 4).Objectives:

- Analyse gene- and isoform-level expression changes in response to SRSF1 knockdown and IR in a triple-negative breast cancer cell line (MDA-MB-231).

- Assess the impact of SRSF1 knockdown and IR on transcript length and protein function in MDA-MB-231 cells.

- Profile the immunopeptidome of MDA-MB-231 cells under different conditions to identify differentially presented peptides and potential neoantigens.

- Integrate transcriptome and immunopeptidome data to investigate the relationship between gene expression and antigen presentation and to identify antigens generated from unique transcripts.

Chapter 2 Genome-wide ionising radiation-induced isoform switching

2.1 Introduction

Although the response to ionising radiation is well-studied at the gene-level, far fewer studies account for individual isoform expression. Existing studies utilise 2-3 cell lines, at most, and none investigate the consequences of IR-induced isoform switching on protein structure across the transcriptome. In this chapter, I used deep RNA-sequencing data from B-cell lines derived from 10 healthy, unrelated individuals at three timepoints (before IR, and 2 and 6 hours after IR) to investigate the impact of IR on isoform switching and the mediators of this response.

2.2 Materials and Methods

2.2.1 Cell Culture

Ten human B-cell lines (GM07000, GM10838, GM10839, GM11992, GM11993, GM11994, GM12716, GM12717, GM12877, GM12878) from unrelated members of the Centre d'Etude du Polymorphisme Humaine (CEPH) families (Coriell Cell Repositories) (Table 2.1) were used in this study. These cells were chosen because this study is a continuation on previous studies from the Cheung Lab at the University of Michigan that investigated the impact of ionising radiation at different time points and doses on normal human cells (Jen and Cheung 2003, 2005; Smirnov et al. 2009; Nayak et al. 2014). B cells

can be extracted from individuals easily in large amounts, so are a common human sample. These cells, which were all immortalised the same way, are valuable for understanding normal response to radiation and variability in radiation sensitivity among individuals, because they represent 10 healthy, unrelated individuals from 10 families in Utah, USA, with well-documented pedigrees. The cells were cultured in RPMI 1620 medium supplemented with 15% heat-inactivated fetal bovine serum (FBS), 1% Penicillin-Streptomycin, and 1% L-Glutamine. The culture and harvesting of these cells were done by others at the Cheung Lab in the University of Michigan.

Table 2.1 B-cell line information, acquired from Cellosaurs.org (Bairoch 2018) and the Coriell Institute for Medical Research catalog (catalog.coriell.org), both accessed in March 2024.

Cell line	Transformant	Sex	Age (years)	Race	Ethnicity	Country of origin
GM07000	Epstein-Barr Virus	Female	66	White	Utah/Mormon	USA
GM10838	Epstein-Barr Virus	Male	Unspecified	White	Utah/Mormon	USA
GM10839	Epstein-Barr Virus	Female	Unspecified	White	Utah/Mormon	USA
GM11992	Epstein-Barr Virus	Male	Unspecified	White	Utah/Mormon	USA
GM11993	Epstein-Barr Virus	Female	Unspecified	White	Utah/Mormon	USA

GM11994	Epstein-Barr Virus	Male	Unspecified	White	Utah/Mormon	USA
GM12716	Epstein-Barr Virus	Male	Unspecified	White	Utah/Mormon	USA
GM12717	Epstein-Barr Virus	Female	Unspecified	White	Utah/Mormon	USA
GM12877	Epstein-Barr Virus	Male	Unspecified	White	Utah/Mormon	USA
GM12878	Epstein-Barr Virus	Female	Unspecified	White	Utah/Mormon	USA

2.2.2 Irradiation

B-cells were plated at a density of 0.5×10^6 cells/mL 18 hours prior to irradiation and exposed to 10 Gy ionising radiation using a ^{137}Caesium irradiator. Protein and RNA were isolated from three conditions (no irradiation, 2hr post-irradiation, and 6hr post-irradiation). This was done by others at the Cheung Lab. In the context of understanding the impact of ionising radiation on DNA damage repair and cell cycle checkpoint activation (i.e. the first steps to making sure cells respond properly to IR), these timepoints are well-suited. After all, DNA repair takes place within minutes to hours, and the cell cycle needs to be halted in this time frame to allow for repair, but cells must also prepare for the possibility of apoptosis. Previous studies by the Cheung Lab and others have demonstrated dramatic changes in DNA repair, stress response, and cell cycle genes within these time frames (Jen and Cheung 2003; Khodarev et al. 2001; Jen and Cheung 2005; Smirnov et al. 2009; Nayak et al. 2014). When

we shifted to investigating apoptosis, we used later timepoints: 24, and 48 hours later, as that is when cells begin to apoptose if DNA is unrepaired.

2.2.3 RNA Sequencing

RNA was extracted by others in the Cheung Lab using RNeasy micro-kit (QIAGEN) from the 10 cell lines in the same conditions. cDNA was synthesized using random hexamer primers by TaqMan Reverse Transcription Reagent kit (Applied Biosystems). RNA-seq libraries were prepared using Illumina TruSeq Stranded Total RNA Library Prep kit (Illumina, #20020596) and sequenced on HiSeq 2500. I did not contribute to this.

The sequencing data was processed by Joshua Burdick and Zhengwei Zhou in the Cheung Lab. An average of 60 million unique 100-nt reads per sample was generated. Sequencing reads were pre-processed to remove adapters using FASTX-Clipper (http://hannonlab.cshl.edu/fastx_toolkit/), low-quality sequences., and reads shorter than 35nt. Reads were aligned to human reference (hg18) using GSNAP (version 2013-10-28) (Wu and Nacu 2010) using the following parameters: mismatches % [(read length + 2)/12-2]; mapping score R 20; soft-clipping on (-trim-mismatch-score=-3). Read counts from each sample were normalized to the total number of mapped reads.

2.2.4 Gene Expression Analysis

Gene expression levels were calculated using the Remove Unwanted Variation (RUV) method (Risso et al. 2014). Genes with RPK >8 in at least 9/30 samples were defined as "expressed," yielding 13,936 expressed genes. Differences in gene expression between each condition were evaluated by Analysis of Variance (ANOVA) with repeated measures and controlled using FDR (Benjamini and Hochberg 1995) in R. Genes were considered IR-

responsive if they had a corrected p of < 0.01. We identified 5,618 IR-responsive genes. This was done prior to my arrival at the Cheung Lab by others.

2.2.5 Detection and Analysis of Alternatively-spliced Transcripts

Alternatively-spliced transcripts were detected from a reference set of transcripts (Wang et al. 2008). The Mixture of Isoforms (MISO) algorithm version 0.5.2 (Katz et al. 2010) was run with default settings to calculate the Percent Spliced In (PSI) for exons/introns detected in each cell line and condition (no irradiation, 2hr post-irradiation, 6hr post-irradiation). PSI is the number of reads that align to a particular exon or intron, divided by the total number of reads for that gene. Isoform switching between conditions was evaluated by ANOVA with repeated measures and controlled using FDR (Benjamini and Hochberg 1995) in R. Transcripts were considered IR-responsive if they had a corrected p of < 0.05. In total, there were 1,881 IR-responsive transcripts. This at the Cheung Lab by others.

2.2.6 Gene Ontology Analysis

I performed Gene Ontology analyses using WEB-based GEne SeT AnaLysis Toolkit (WebGestalt) (Liao et al. 2019) with the following parameters: Method of Interest: Over-Representation Analysis (ORA), Functional Database: Gene Ontology (Biological Process or Biological Process noRedundant), Reference Set: genome protein-coding. The FDR threshold was set to 0.05.

2.2.7 Identifying Potentially Novel IR-responsive Genes

To uncover IR-responsive genes missed in prior studies due to exclusive isoform-level differential expression, I fine-tuned the stringency criteria. For isoform-level expression, I used an FDR threshold of 0.01 and required a minimum 10% change in PSI after IR. This identified 641 significant events (from 498 genes). For gene-level expression, I increased

the FDR threshold to 0.05, resulting in 7,784 differentially expressed genes. Among these, 335 genes were differentially expressed at both levels, and 163 genes were exclusively differentially expressed at the isoform-level.

To determine novelty, I conducted comprehensive literature searches for each gene, searching on Google and Google Scholar for gene symbol and aliases in conjunction with "ionising radiation". Known IR-linked genes or previously identified differentially expressed genes in response to IR were excluded. Through this analysis, I confidently identified 37 genes that had not been previously linked to the IR response, suggesting they are novel IR-responsive genes (Appendix Table 7.1).

2.2.8 Prediction of Functional Consequences of Alternative Splicing

To identify functional protein domains encoded within IR-responsive cassette exons, I converted the coordinates of cassette exons to the GRCh37/hg19 assembly using the UCSC LiftOver tool (Kent et al. 2002), then retrieved the amino acid sequences encoded by cassette exons, using R package ensmebldb (Rainer et al. 2019). For alternative first and last exon events, I inputted the UCSC ID associated with each isoform into the UCSC table browser to return the peptide sequence of the isoform. The following settings were used to retrieve the peptide sequence: assembly: NCBI36/hg18; group: Genes and Gene Predictions; track: UCSC genes; table: knownGenePep; output format: all fields from the selected table. I then ran the peptide sequences from cassette exon, alternative first exon, and alternative last exon events on Linux InterProScan software, version 5 in standalone mode (Jones et al. 2014). I filtered hits to protein domains, coils, and predicted disorder regions. I used literature searches to investigate the consequences of functional domain loss shown in Figure 2.2D.

2.2.9 Motif Identification and *de novo* Motif Discovery

We used a list of human RNA-binding proteins (RBPs) with known motifs (Ray et al. 2013). An RBP was considered IR-responsive if the gene expression level had a corrected p < 0.05 after ANOVA (see Gene Expression Analysis) and a fold change of ±1.2 at any condition. In total, 35 RBPs met this criterion. To elucidate the regulatory function of these RBPs in irradiated cells, their binding motifs were identified within IR-responsive cassette exons. Specifically, MEME-suite version 4.11.3 (Bailey et al. 2009) was used to scan (single-strand only) the entire nucleotide sequence of each of the 645 IR-responsive cassette exons for binding motifs of IR-responsive RBPs with a p-value cut-off of 0.001. Only the binding motifs listed in Ray et al. (Ray et al. 2013) were used for scanning. A cassette exon was considered a putative target of an RBP if it contained >1 of the RBP's binding motif(s).

MEME-suite version 5.0.5 (Bailey et al. 2009) was also used to identify RNA binding motifs *de novo*. Oligonucleotides of length k=8 nucleotides were enumerated from the 645 IR-responsive cassette exons by searching their entire nucleotide sequence (single-strand only). In total, two motifs were enriched. The locations of these de novo-enriched motifs within IR-responsive cassette exons were identified using FIMO (Grant et al. 2011). This was all done by Joshua Burdick from the Cheung Lab.

2.2.10 Self-Organizing Map (SOM) Clustering

Transcript isoforms with similar temporal patterns of splicing were identified using SOM clustering as described by Tamayo et al. (Tamayo et al. 1999). In brief, Z-scores were calculated from the PSI of each IR-responsive alternatively-spliced event across all 10 samples and 3 conditions. For each type of alternative splicing, SOMs were generated by GenePattern (Reich et al. 2006) using a cluster range of 1-6. The iteration that produced the

highest number of unique clusters was used (Figure 2.4B). This was done by others in the Cheung Lab.

2.2.11 Statistical Analysis

Experimental results from replicates are shown as mean ± standard error of the mean (SEM). Analyses were carried out in R, Excel, and GraphPad Prism. I used the R package VennDiagram (https://cran.r-project.org/package=VennDiagram) to create Figure 2.1B. I used the Rank-Sum test in Figure 2.2B to test whether the difference in length between included or spliced out introns and exons is statistically significant. In Figure 2.2A and C, a two-tailed test of proportions was carried out to test if observed proportions were higher than expected by chance. PY tract density was calculated by counting the number of pyrimidines (C and T) in the last 40 bases of each intron and dividing the number by 40 (the length of the segment). GC content was calculated by counting the number of Gs and Cs and dividing by the length of the intron/exon. The Mann-Whitney U test was employed to determine if the differences in length, GC content, and PY tract densities across exons and introns is statistically significant.

2.3 Results

2.3.1 Radiation-induced Changes in Gene Expression

In this study, we used B-cell lines derived from 10 healthy, unrelated individuals to study normal cellular response to ionising radiation. We sequenced the cDNA of these cells at three time points: immediately before exposure to IR (baseline), 2 hours after IR, and 6 hours after IR (Figure 2.1A), generating about 60 million uniquely aligned reads per sample.

To identify the radiation-responsive genes, we carried out analyses of variance (ANOVA) and identified 5,618 differentially expressed genes (false discovery rate (FDR) < 1%) at 2 and/or 6 hours after IR. Overall, more genes were downregulated than upregulated (Table 2.2). As expected, upregulated genes play a role in DNA repair (e.g. *MDM2*, *RAD51*), inflammation (e.g. *TNF*, *CEBPB*), and apoptosis (e.g. *BAX*, *TP53BP2*), while genes in pathways related to the cell cycle (e.g. *CDK4, CDK10*) or synthesis of biomolecules (e.g. *FDPS*, *NSDHL*), were downregulated (Table 2.2). These results are consistent with previous findings that, upon radiation exposure, human cells induce DNA damage response pathways, while shutting down other cellular processes (Jen and Cheung 2003; Tusher et al. 2001; Wang 2014; Zhou and Elledge 2000).
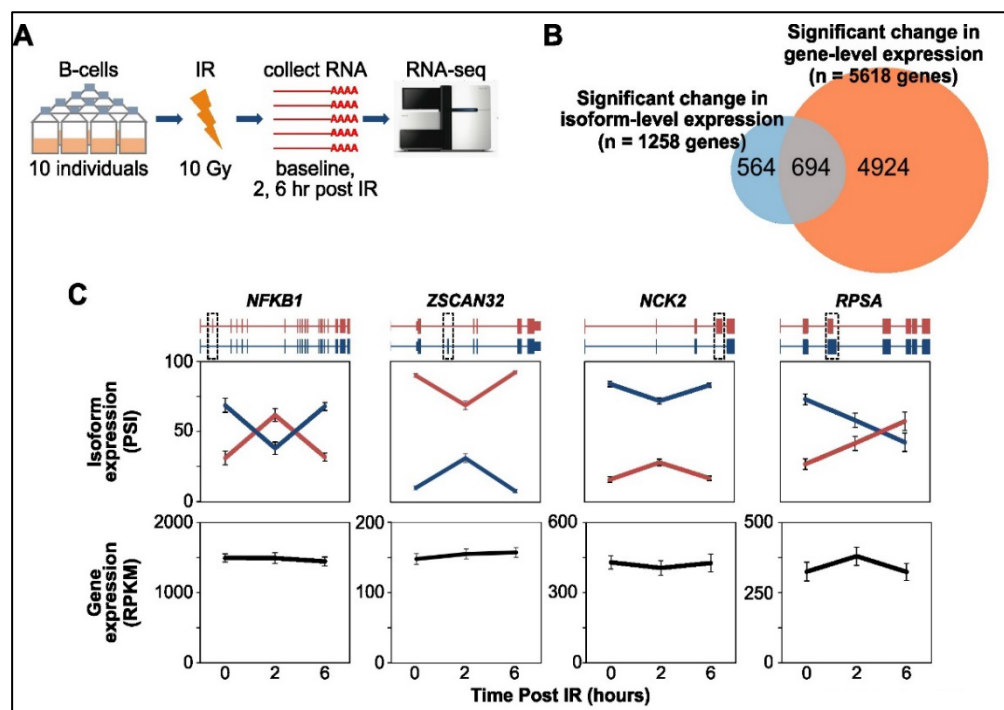
Figure 2.1 Gene-level expression of IR-responsive genes. **A)** Experimental design. The experiment was designed and data generated before my arrival. The figure is partly made by Niema Razavian (NR), alum of the Cheung Lab at University of Michigan and partly by me (he sourced the pictures and I assembled the figure and text). **B)** Overlap between genes exhibiting gene-level or isoform-level differential expression in response to IR. **C)** Examples

of genes exhibiting IR-induced isoform switching without a significant change in gene-level expression. Red and blue correspond to the expression of the same-color isoform schematic shown above each panel. Error bars represent standard error of the mean from 10 biological replicates. Apart from *ZSCAN32*, the panels in this figure were made by NR. IR, ionising radiation; PSI, percent spliced in; RPKM, reads per kilobase of transcript per million reads mapped.

Table 2.2 Gene-level response to ionising radiation. The RNA sequencing (RNA-seq) data was generated and processed, including differential expression analysis, before my arrival in the Cheung lab. I used it to create this table.

Time (hr)	Induced by irradiation			Repressed by irradiation		
	# genes	Range % (median %)	Examples (related pathway)	# genes	Range % (median %)	Examples (related pathway)
0 to 2	1,733	5 – 834 (25)	*MDM2, CHEK1* (DNA repair)	1,950	-3 – -76 (-23)	*CDK4, CKD10* (cell division) *IDI1, IDI2* (cholesterol biosynthesis)
0 to 6	1,554	5 – 2621 (32)	*BCL2L1, TP53BP2* (apoptosis) *TNF, FOXO1* (stress response)	2,715	-4 – -92 (-22)	
2 to 6	1,954	5 – 724 (37)		2,760	-5 – -82 (-22)	

2.3.2 Ionising Radiation Leads to Global Isoform Switching and Potentially Shorter Transcripts

As we examined the sequencing results, we noted that in addition to changes in gene expression levels, there were also alternative splicing events. To quantify them systematically, using the same cDNA sequencing data, we identified alternatively-spliced isoforms and quantified their expression levels using the Mixture of Isoforms (MISO) algorithm (Katz et al. 2010), which assigns a Percent Spliced In (PSI) value to each exon or intron. PSI is the number of reads that align to a particular exon or intron, divided by the total number of reads for that gene. We identified 1,881 alternatively-spliced isoforms (from

1,258 genes) that were significantly (FDR < 5%) induced or repressed at 2 and/or 6 hours after IR.

Among the 1,258 genes with radiation-responsive change in isoform expression, almost half (565/1,258) exhibited no gene-level differential expression when the expression was averaged across isoforms (Figure 2.1B); these would have been missed if we focused only on total expression levels. Figure 2.1C shows four genes that showed no changes in total expression levels but significant isoform-level response to IR. For example, isoform 1 of *NFKB1*, which promotes apoptosis, increases in expression post-IR, while isoform 2, which is anti-apoptotic, decreases.

Although *NFKB1* is known to be IR-responsive, through an extensive literature search and stringent filtering criteria (see Materials and Methods), I identified at least 37 genes that, to the best of our knowledge, have not been associated directly with the IR response before (Appendix Table 7.1). These genes encode transcription regulators including six zinc finger proteins (ZCCHC10, ZFAND4, ZNF266, ZNF821 and ZSCAN32). In *ZSCAN32*, an exon encoding the C-terminal portion of the SCAN/oligomerization domain is spliced out, potentially reducing interactions with other proteins (Williams et al. 1999).

The most common alternative splicing events in response to IR involved cassette exons (34% of all isoform-switching events, of which 43% are exon inclusion events, and 57% are exclusion events). Next were alternative last exon events (24% of all events), intron retention (19% of all events), and alternative first exon events (11% of all events). These four types encompassed 1,669 of the 1,881 isoform-switching events (Figure 2.2A).
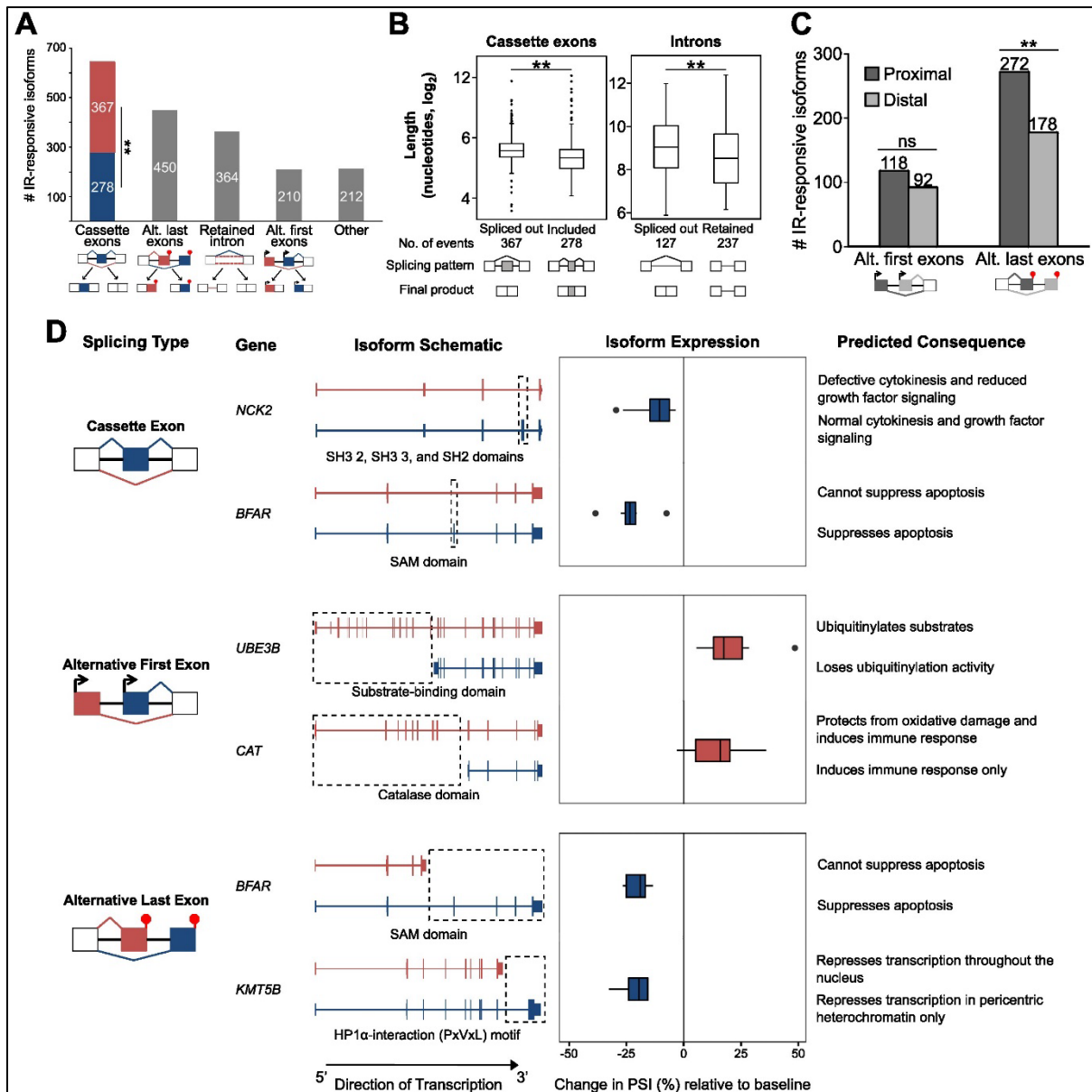
Figure 2.2 IR-induced isoform switching likely results in shorter transcripts, affecting function. A) The most common alternative splicing types observed in response to IR. The bar for cassette exons is divided into exon inclusion (blue) and exclusion (red) events. A schematic clarifying each alternative splicing type is shown. B) Length ($\log_2$-transformed nucleotide number) of IR-responsive cassette exons and introns, along with a schematic clarifying each outcome. C) Frequency of proximal versus distal IR-responsive alternative first or last exon usage, with a schematic clarifying relative positions of exons. NR made the schematics that simplify the splicing type. Otherwise, the figure and analysis are my own work. D) Examples of protein domains affected by alternative splicing in cassette, alternative

first, and alternative last exon events (n = 10) and the anticipated functional consequences of the resulting isoforms. The change in Percent Spliced In (PSI) is shown for the regions highlighted in dashed lines for each gene to reflect the relative abundance of the region. The examples shown are the ones with the most dramatic change at 2 or 6 hours after IR that also have data available on the potential consequence of the isoform switch, based on literature searches. All PSI changes are based on 2 hours relative to baseline, except for *CAT*, which is based on 6 hours relative to baseline. The predicted consequence of each isoform (with and without the region in dashed lines) is shown on the right side of the plot). NR made the schematics that simplify the splicing type. Otherwise, the figure and analysis are my own work. IR, ionising radiation; ns, not significant at $p = 0.05$; PSI, percent spliced in; ** $p < 0.01$.

Next, I characterized the sequence features surrounding the alternatively spliced cassette exons, including length, guanine-cytosine (GC) content and PY tract density of adjacent introns (Figure 2.3). Although there was no significant difference in PY tract density of introns adjacent to IR-responsive cassette exons compared to those adjacent to non-IR-responsive cassette exons (Figure 3.3A), I found that the introns adjacent to IR-responsive cassette exons had higher GC content than corresponding regions of non-IR-responsive transcripts (median 31% versus 27% for upstream introns and 31% vs 25% for downstream introns; $p < 0.0001$) (Figure 3.3B). Similarly, the exons downstream of the IR-responsive cassette exons also had significantly higher GC content (median 47% versus 45%; $p < 0.0001$).
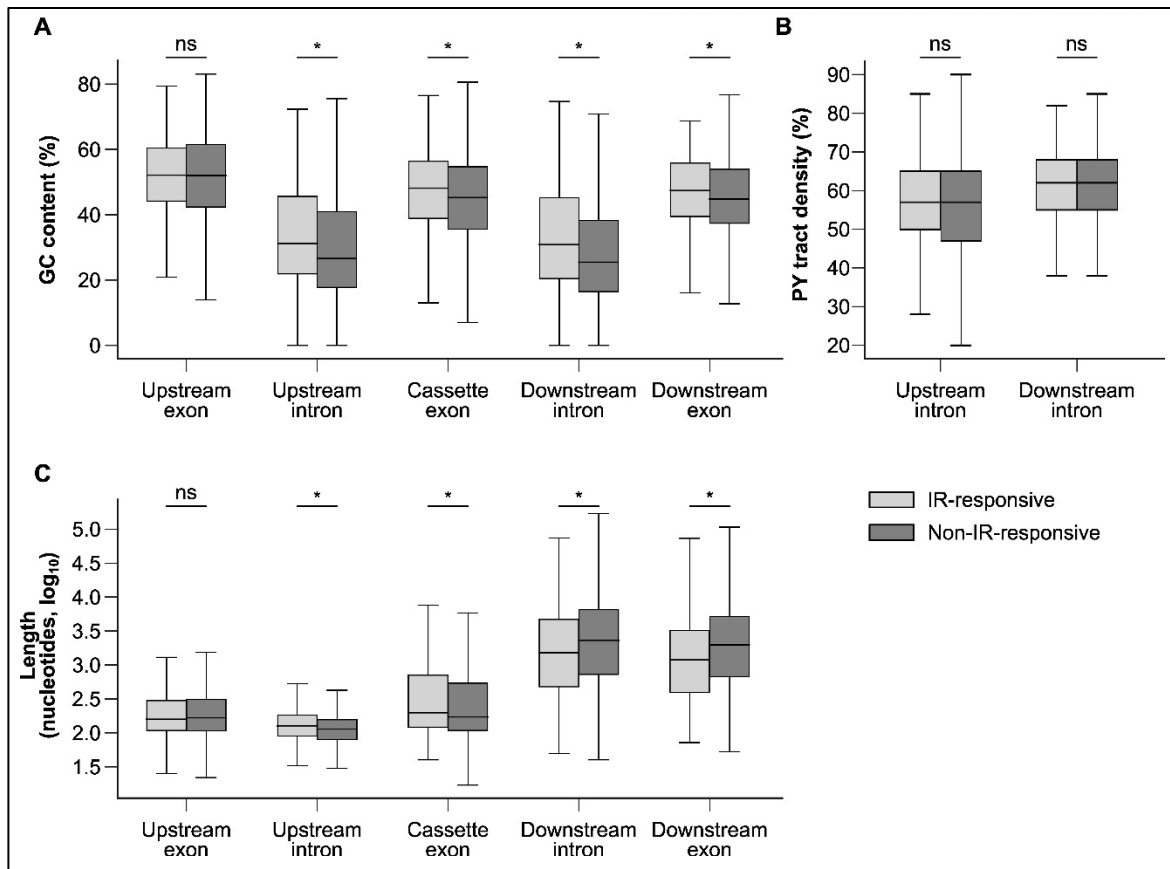
Figure 2.3 Basic sequence characteristics of all IR-responsive cassette exons (regardless of timepoint; n = 645) compared to non-IR-responsive cassette exons (n = 22,488) in terms of the **A)** PY tract density (for adjacent introns only), **B)** GC content, and **C)** length, for them, as well as for their upstream and downstream exons and introns. * p < 0.0001 (Mann-Whitney U); ns, not significant; IR, ionising radiation; PY, polypyrimidine.

Previous studies have suggested that stress influences transcript, exon, or intron lengths (Boutz et al. 2015; Braunschweig et al. 2014; Chang et al. 2015; Sadek et al. 2019; Fontana et al. 2017), but this has not been simultaneously studied across different alternative splicing patterns. I asked how ionising radiation affects transcript, exon, and intron lengths, across the four most common types of alternative splicing (cassette exons, alternative last and first exons, and retained introns). I found that IR-responsive cassette exons are typically longer than their non-IR-responsive counterparts, with medians of 126 and 113 nucleotides

respectively (p < 0.0001) (Figure 2.3C). However, a majority (57%) of these cassette exons are spliced out in response to IR (p < 0.001) (Figure 2.2A). The ones retained are on average 40 nucleotides shorter than those that are spliced out (p < 0.001) (Figure 2.2B). A similar trend is observed with intron retention: IR-responsive retained introns (237/364) are, on average 105 nucleotides shorter than those that are excluded (127/364; Figure 2.2B). Overall, this data suggests that exons and introns retained after IR exposure are shorter.

I also examined the transcripts by proximal versus -distal first or last exon usage. While there was no significant difference in the usage of proximal or distal alternative first exons, (Figure 2.2C), 60% (272/450) of alternative last exon events used the more proximal last exons (p < 0.00001) (Figure 2.2C), indicating shorter transcripts. One of those events occurred in the *Bifunctional Apoptosis Regulator* (*BFAR*), which normally suppresses BAX-induced cell death by binding to BCL-2 and BCL-X(L) via its SAM domain (Zhang et al. 2000). In response to IR, a proximal last exon is used, which leads to the loss of the downstream exons that code for the SAM domain (Figure 2.2D). This likely abolishes the anti-apoptotic function of BFAR. Another example is *KMT5B*. The canonical isoform of this lysine methylase deposits the H4K20me3 mark to repress transcription in pericentric heterochromatin only (Tsang et al. 2010), but in the irradiated cells, a shorter isoform is expressed more abundantly (Figure 2.2D), and it codes for a truncated protein that trimethylates lysine 20 on H4 throughout the nucleus (Tsang et al. 2010), potentially contributing to IR-induced global transcriptional repression. Taken together, these data suggest that the response to ionising radiation is carried out mostly through shorter isoforms.

2.3.3 Isoform Switching is an Early Response to IR

Next, we examined the temporal dynamics of alternative splicing in response to IR. We grouped the IR-responsive transcripts by their expression patterns using an unsupervised neural network. Results from self-organizing maps show three main temporal patterns of alternatively-spliced isoforms (Figure 2.4A and B): early but transient response (e.g. *NFKB1* and *BFAR*), which occurs at 2 hours but is back to baseline levels by 6 hours after IR; early and sustained response (e.g. *TCF7* and *RAPGEF6*), which occurs at 2 hours and maintains similar levels at 6 hours; and late response (e.g. *HNRNPDL* and *CAT*), which only occurs at 6 hours after IR. Most (88% of 1,669) isoform switching occurred within two hours after irradiation, but those isoforms were induced only transiently, as their expression levels had returned to baseline by six hours (Clusters 1 and 2 in Figure 2.4B). I performed functional enrichment analysis using Gene Ontology, which revealed that these early-transient isoforms mount the initial transcriptional response to radiation by regulating apoptosis and gene expression (e.g. *BAX* and *RBM14* isoforms, respectively), responding to DNA damage (e.g. *POLE* isoform), and arresting the cell cycle (e.g. *WEE1* isoform) (Figure 2.4C). While fewer isoforms changed at, or were sustained for, six hours, some were identified and they are shown as Clusters 3-6 in Figure 2.4B. These transcripts are enriched for pathways related to DNA damage (e.g. *MDM2* isoform), cell cycle regulation (e.g. *WEE1* isoform), and negative regulation of gene expression (e.g. *HDAC8* isoform), indicating these aspects of the IR response persist to 6 hours after IR (Figure 2.4C).
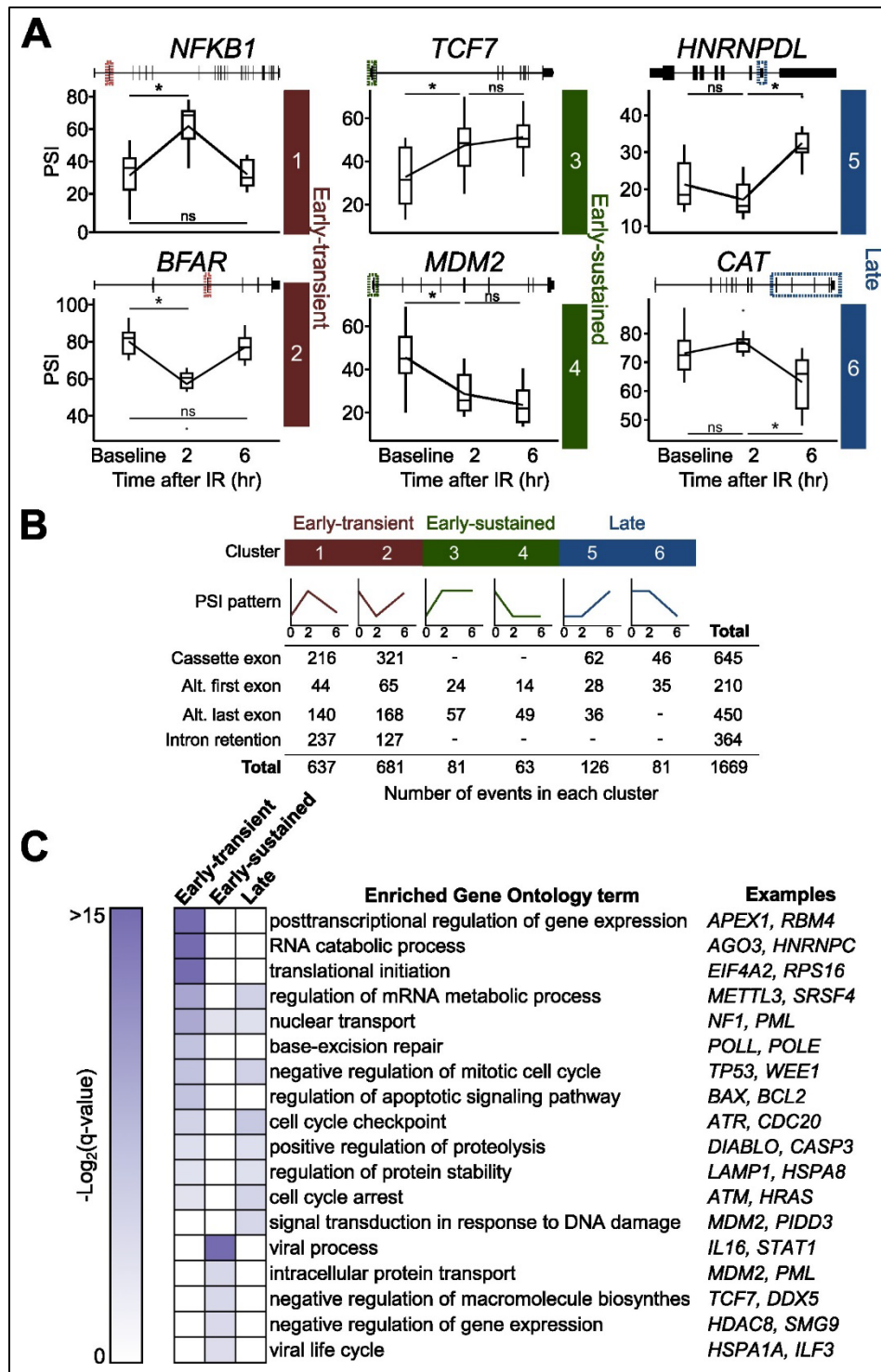
Figure 2.4 IR-induced isoform switching occurs early, affecting stress response genes. **A)** Examples of the six kinetic clusters of IR-responsive isoform switching. Each box plot represents the PSI of the affected exons (highlighted by dotted lines) in the respective genes. **B)** Kinetic clusters of the most common IR-induced alternative splicing types. Each cluster

is depicted by a representative graph with PSI on the y-axis and time points (Baseline, 2hr post-IR, and 6hr post-IR) on the x-axis. **C)** Heatmap of some of the most significant Gene Ontology Biological Process terms associated with isoform-switching genes in the three broad categories of temporal response to IR (early-transient, early-sustained, and late), along with example genes. Color gradient represents the significance based on Benjamini–Hochberg-corrected p-value. All non-white cells are $q < 0.05$. PSI, percent spliced in; ns, not significant; * $p < 0.05$.

2.3.4 Alternatively-spliced IR-responsive Isoforms Have Different Functions

To more closely investigate the consequences of alternative splicing on protein function, I asked what protein domains were affected after IR exposure in the isoform-switching events. Using InterProScan (Jones et al. 2014), I identified known domains within the IR-responsive isoforms that exhibited cassette exon, alternative last exon, or alternative first exon splicing events. I found 496 IR-responsive domains, within 345 isoforms of 265 genes. Significantly more domain-encoding exons are spliced out (281/496) than induced (215/496; binomial test $p < 0.01$), and six examples are detailed in Figure 2.2D, including the aforementioned *KMT5B* and *BFAR*. In addition to the shorter isoform of *BFAR* that results due to proximal alternative last exon usage, I identified a cassette exon in *BFAR*, which codes for the second half of the SAM domain (Zhang et al. 2000). In response to IR, the SAM domain is spliced out in the B cells of all 10 individuals, resulting in an isoform of *BFAR* that most likely cannot suppress apoptosis. Another example is *NCK2*, where exon 3, which encodes the second and third SH3 domains and most of the SH2 domain, is skipped. The SH2 domain is critical for cytokinesis (Jacquet et al. 2018) and the two SH3 domains are essential for growth factor signaling, as well as other signaling pathways (Liu et al. 2006; Tu et al. 1998). Similarly, usage of an alternative first exon in the catalase *CAT*, results in

an increase of the full-length isoform, which protect from IR-induced oxidative stress by neutralizing reactive oxygen species. The clustering was performed before my arrival at the Cheung Lab. NR made the PSI pattern icons. I generated the rest of the figure. **C)** Enriched Gene Ontology terms in genes that exhibited early-transient, early-sustained, and late isoform switching, with example genes in the enriched term on the right. IR, ionising radiation; PSI, Percent Spliced In.

2.4 Discussion

While the transcriptional response to IR is well-characterized, the precise contribution of isoform switching to the DNA damage response remains unclear. In the present study, we analyzed data from cell lines derived from 10 healthy individuals to uncover isoform switching as an important feature of the cellular response to IR. We find that IR induces a high degree of isoform switching, leading to potentially shorter transcripts that modulate the DNA damage response by preferentially facilitating apoptosis and cell cycle arrest. Together, our findings underscore the importance of isoform-level expression changes in coordinating the cellular response to IR.

While numerous studies have established a connection between shorter transcripts and stress response, they have primarily focused on either entire gene lengths or specific alternative splicing events, rather than across multiple alternative splicing types simultaneously (Boutz et al. 2015; Chang et al. 2015; Sadek et al. 2019; Fontana et al. 2017; McKay et al. 2004; Williamson et al. 2017). This study addresses this knowledge gap by revealing that IR-induced stress results in the generation of likely shorter transcripts, which arise through increased cassette exon exclusion (Figure 2.2A), shorter retained introns/exons compared to those that are spliced out (Figure 2.2B), and preferential utilization of proximal

last exons over distal last exons (Figure 2.2C). Although transcript length cannot be determined definitively without long-read sequencing, this analysis offers valuable insights into the diverse isoform-switching events involved in the cellular response to IR.

Previous work from the Sharp lab and others demonstrated that DNA damage-associated intron retention events predominantly involve shorter genes and shorter introns (Boutz et al. 2015; McKay et al. 2004). My findings corroborate these observations, as I report increased intron retention and shorter retained introns in response to IR, suggesting a shared DNA damage response mechanism. This may be attributed to the increased probability of longer introns incurring damage and, consequently, not being transcribed.

Fontana et al. (Fontana et al. 2017) reported an increase in proximal last exon usage in response to oxidative stress, due to the depletion of the SWI/SNF protein SMARCA2, which typically inhibits 3'-end processing and transcript cleavage at proximal last exons. IR is known to trigger oxidative stress by generating reactive oxygen species (Azzam et al. 2012; Buonanno et al. 2011; Maier et al. 2016). Correspondingly, I also observed *SMARCA2* depletion following IR exposure, indicating its potential involvement in IR-induced proximal last exon usage. In addition, UV irradiation led to shorter isoforms due to increased usage of proximal alternative last exons, which was associated with a significant slowdown in RNAPII elongation (Williamson et al. 2017). IR-induced damage is also implicated in slowing down RNAPII in an ATM-dependent manner (Shanbhag et al. 2010), which may contribute to the generation of shorter transcripts. This contrasts with UV-induced damage, which causes ATM-independent RNAPII stalling (Muñoz et al. 2009). In addition to favouring suboptimal splice sites, RNAPII stalling is associated with increased exon

skipping (Dujardin et al. 2014), consistent with our finding that most IR-responsive cassette exons are spliced out.

Our analysis emphasizes the need to incorporate isoform-level expression analysis into gene expression studies, an aspect that has been largely overlooked. Nearly half of the IR-responsive genes that exhibited isoform switching in our data were not differentially expressed at the gene level. Such missed "hits" leave a largely untapped pool of precision oncology targets (Robinson et al. 2019). Beyond missing hits, analysis of isoform-level expression reveals a much more nuanced response to stress. In this respect, a gene's function is no longer reduced to what its canonical isoform does, but rather the potentially unique functions of each isoform are considered. I identified the protein domains encoded by differentially expressed exons, and found nearly 500 domains, within 265 IR-responsive genes, many of which affect functionally-relevant protein domains (i.e. domains involved in cell division/survival and DNA repair). This finding is undoubtedly conservative, given that domain identification relies on database annotation. This highlights the need for a more comprehensive understanding of the effects of isoform switching on protein function and posttranslational modifications.

In line with this, our investigation of isoform switching in the B cells of 10 individuals showed a clear trend favoring isoforms that promote a stress response after IR. For example, isoforms that facilitate transcriptional repression and DNA repair are induced (e.g. shorter isoform of *KMT5B*, and alternative first exon isoform of *DDX55*, respectively), while those for cell cycle progression and survival are repressed (e.g. longer isoforms of *NCK2* and *BFAR*, respectively) (Figure 2.2D). This further underscores the importance of considering isoform-level expression analysis to better understand the functional implications of gene

regulation under stress conditions, such as IR exposure. However, my analyses are all in B cells, which means my observations could be a cell-specific response. Further interrogations in other cell types are necessary for broader conclusions.

Chapter 3 SRSF1 mediates radio- and immunotherapy resistance

3.1 Introduction

In the previous chapter, I found nearly 2000 IR-induced isoform switching events. I found that they occur early (within 2 hours) after IR and that they lose and gain exons that encode domains which facilitate the response to IR. But what is mediating those isoform switching or alternative splicing events in response to IR? In this Chapter, I sought to identify the RNA binding protein(s) that regulate(s) IR-induced isoform switching, and characterize their role in IR response.

3.2 Materials and Methods

3.2.1 Cell Culture

Ten human B-cell lines (GM11993, GM11994, GM07000, GM12878, GM12921, and GM12922) from unrelated members of the Centre d'Etude du Polymorphisme Humaine (CEPH) families (Coriell Cell Repositories) were cultured in RPMI 1620 medium supplemented with 15% heat-inactivated fetal bovine serum (FBS), 1% Penicillin-Streptomycin, and 1% L-Glutamine. The culture and harvesting of these cells were done by others at the Cheung Lab in the University of Michigan.

HEK293T cells were used in transfections as a model for non-malignant cells because of the ease of transfecting them compared to the B cell lines. MDA-MB-231 triple-negative breast cancer cells were used as a model for malignant cells for the following reasons: 1) breast cancer is the most common cause of cancer in women globally (Howard and Olopade

2021), 2) breast cancer is treated with radiation, 3) triple-negative breast cancer is the most aggressive type of breast cancer (Howard and Olopade 2021), 4) MDA-MB-231 cells are the most common triple-negative breast cancer model used in research laboratories (Wagner 2022).

HEK293T and MDA-MB-231 cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM; Thermo Fisher Scientific, #41966029) supplemented with 10% FBS (Sigma-Aldrich, #F9665-500ML) and 1% Penicillin-Streptomycin (Sigma-Aldrich, #P4333-100ML). During transfections, the media used contained the same supplements, except the antibiotics. The cells were passaged every 2-3 days, at around 75% confluency.

3.2.2 Protein Isolation and Western Blot

Protein was obtained by lysing cells in RIPA buffer (NEB) for B cells or a non-denaturing lysis buffer (NDLB; 20 mM Tris-HCl, 137 mM NaCl, 10% glycerol, 1% NP-40, 2 mM EDTA) for HEK293T and MDA-MB-231 cells. I did not extract, quantify, or blot protein from the B cells (this was done by others in the Cheung Lab), but I did from HEK239T and MDA-MB-231 cells. Buffers contained freshly-added protease/phosphatase inhibitors. Samples were sonicated, centrifuged at maximum speed at 4ºC to pellet insoluble material, then the supernatant was transferred to new tubes. Protein concentration was measured using Qubit BR Protein Assay (ThermoFisher). Next, 14-18 μg of total protein lysate was mixed with 4x Laemmli Sample Buffer (BioRad) and reducing agent (ThermoFisher), boiled at 95ºC for 10 minutes, then loaded into 4-20% Bis-Tris SDS-PAGE gels (BioRad). Protein was then transferred onto polyvinylidene fluoride (PVDF) membranes, blocked 5% milk for one hour and incubated with primary antibody overnight. Primary antibodies used are: anti-SRSF1 (Abcam #38017, 1:1000), anti-GAPDH (Santa

Cruz Biotechnology #25778, 1:200), anti-α-Tubulin (Santa Cruz Biotechnology #8035, 1:1000), anti-apoptosis marker cocktail (Abcam #136812; 1:250). After three 1x PBST (phosphate-buffered saline with 0.1% Tween-20) washes, membranes were incubated in respective HRP-conjugated secondary antibodies for one hour. Bands were visualized using ECL Western Blotting Substrate (ThermoFisher).

The blot in Figure 3.2B is based on the following individuals who were sequenced in Chapter 2: GM11993, GM11994, GM07000, and GM12878, as well as the following two individuals who were not sequenced in this study: GM12921 and GM12922. I quantified the blot in Figure 3.2B using ImageJ region-of-interest method (Davarinejad 2015) and analysed it using GraphPad Prism with a two-way ANOVA, followed by Tukey's multiple comparisons test.

I quantified the blots in Figure 3.5C, Figure 3.7A, and Figure 3.7D with Image Lab (BioRad). The statistical test conducted in Figure 3.5C is a paired two-sample t-test done in Excel (Microsoft).

3.2.3 Reverse Protein-RNA Immunoprecipitation

I did not perform the RNA-IP or process its results; this was done by others in the Cheung Lab. The cells were cultured as in 2.2.1. Reverse immunoprecipitation was carried out using Magna RNA-Binding Protein Immunoprecipitation Kit (Millipore) following the manufacturer's protocol. For each immunoprecipitation reaction, $2x10^7$ cells from GM11994 and GM12878 lines were harvested in three aforementioned conditions. Cells were lysed in 100 μL lysis buffer from the kit (proprietary components) with protease and RNase inhibitors. Ten μg of anti-SRSF1 antibody (Bethyl, #A302-052A) and negative control rabbit IgG (Millipore, #12-370) were conjugated to Magnetic Protein A/G beads.

One-hundred microliters of cell lysate was added into 900 μL Immunoprecipitation Buffer with RNase inhibitor and incubated with 50 μL beads-antibody complex overnight at 4°C. Bead-bound immunoprecipitates were then washed and incubated with proteinase K in the presence of 1% SDS for 30 min at 55°C. RNA was then extracted from supernatants using phenol:chloroform:isoamyl alcohol and precipitated using ethanol. Precipitated RNA was digested by DNase I (Ambion). RNA-seq libraries for SRSF1-IP and input RNA were prepared, preprocessed, and aligned as described above. Enrichment of transcripts in the immunoprecipitate was analyzed using Cufflinks v2.2.1 (Trapnell et al. 2010), considering only transcripts that overlapped with the 645 IR-responsive cassette exons.

All transcripts included for enrichment analysis had FPKM >1 in at least one condition per individual in both SRSF1-IP and RNA input samples. I defined a transcript as an SRSF1 target if it 1) had a fold enrichment of ≥2 (SRSF1-IP/input) in at least one condition per individual and 2) had a de novo or known SRSF1 binding motif. The latter condition was added to increase stringency, as native RNA-IP cannot accurately determine the location of protein binding on the transcript and therefore makes it difficult to distinguish between different events in the same gene. With the added motif constraint, I can narrow down the events likely to be associated with SRSF1.

3.2.4 Gene Ontology Analysis

I performed Gene Ontology analyses using WebGestalt (Liao et al. 2019) with the following parameters: Method of Interest: Over-Representation Analysis (ORA), Functional Database: Gene Ontology (Biological Process or Biological Process noRedundant), Reference Set: genome protein-coding. The FDR threshold was set to 0.05.

3.2.5 Motif Identification and *de novo* Motif Discovery

I did not perform this analysis. It was all done in collaboration with Joshua Burdick from the Cheung Lab. We used a list of human RBPs with known motifs (Ray et al. 2013). An RBP was considered IR-responsive if the gene expression level had a corrected $p < 0.05$ after ANOVA (see 2.2.4) and a fold change of ± 1.2 at any condition. In total, 35 RBPs met this criterion. To elucidate the regulatory function of these RBPs in irradiated cells, their binding motifs were identified within IR-responsive cassette exons. Specifically, MEME-suite version 4.11.3 (Bailey et al. 2009) was used to scan (single-strand only) the entire nucleotide sequence of each of the 645 IR-responsive cassette exons for binding motifs of IR-responsive RBPs with a p-value cutoff of 0.001. Only the binding motifs listed in Ray et al. (Ray et al. 2013) from 80 RBPs (appendix Table 7.2) were used for scanning. A cassette exon was considered a putative target of an RBP if it contained >1 of the RBP's binding motif(s).

MEME-suite version 5.0.5 (Bailey et al. 2009) was also used to identify RNA binding motifs *de novo*. Oligonucleotides of length k=8 nt were enumerated from the 645 IR-responsive cassette exons by searching their entire nucleotide sequence (single-strand only). In total, two motifs were enriched (Table 3.3). The locations of these *de novo*-enriched motifs within IR-responsive cassette exons were identified using FIMO (Grant et al. 2011).

3.2.6 SRSF1 Knockdown and Overexpression

I used the following conditions for the knockdown in 6-well plates, as recommended by the (Thermo Fisher Scientific, #13778030) reverse transfection protocol: 500µl OptiMEM (ThermoFisher, #11058021) per well, 2.5ml antibiotic-free medium (see Cell Culture) per well, 20nM siRNA per well, 5µl Lipofectamine RNAiMAX reagent/well. I carried out

overexpression using jetPRIME transfection reagent (Polyplus, #101000015), as per the manufacturer's protocol and the following conditions in a 6-well plate: 200µl of jetPRIME buffer, 500ng DNA, and 4µl jetPRIME reagent, in a total of 3ml media per well.

For cell cycle analysis (Figure 3.5A): I plated 200,000 HEK293T cells on 6-well plates and reverse-transfected using Lipofectamine RNAiMAX (Invitrogen) with an siRNA oligo targeting SRSF1 (Dharmacon, A-018672-13-0005; hereby referred to as siSRSF1-1), a pool of siRNA oligos targeting SRSF1 (Dharmacon, E-018672-00-0005; hereby referred to as siSRSF1-2), or a pool of non-targeting siRNAs as a control (Dharmacon, D-001910-10-50; hereby referred to as Control siRNA). After transfection, I left the cells to grow for two days, and then either irradiated at 10 Gy or not. I collected the cells by trypsinization 48-hours after IR. The cells were fixed with ice-cold 70% ethanol and frozen at -20ºC until propidium iodide staining (See 3.2.8). The experiment was repeated for a total of four independent replicates.

For resazurin viability assay (Figure 3.5B): I plated 7,500 HEK293T cells on four 96-well black plates (one for each dose) with a clear bottom and reverse-transfected using RNAiMAX with siSRSF1-1, siSRSF1-2, or Control siRNA. Two days after transfection, I irradiated the cells at increasing doses or did not irradiate them. Resazurin was added to the cells 48-hours after IR exposure (see Viability Assay). The following conditions were used for the knockdown in 96-well plates, as recommended by the Lipofectamine RNAiMAX reverse transfection protocol: 20µl OptiMEM, 100µl antibiotic-free medium (see Cell Culture), 20nM siRNA, 0.2µl RNAiMAX reagent/well.

For apoptosis western blots (Figure 3.5 and Figure 3.7D): I plated 250,000 HEK293T or MDA-MB-231 cells on 6-well plates and reverse-transfected using Lipofectamine

RNAiMAX with siSRSF1-1 or Control siRNA. After transfection, cells were left to grow for three days, and then either irradiated at 10 Gy or not. I collected the cells by scraping at 24- and 48-hours post-irradiation or mock irradiation. I pelleted and froze the cells at -80ºC until protein extraction. I repeated the experiments for a total of three biological replicates.

For the complementation experiment (overexpression following knockdown) (Figure 3.7A), the SRSF1 plasmid (OriGene, RC201636) contains only the SRSF1 ORF, while the siRNAs target the 3' UTRs of SRSF1. Therefore, there was no possibility of the siRNA downregulating the overexpressed protein. I seeded the cells at a density of 200,000 cells per well, and reverse-transfected with siSRSF1-2 or Control siRNA as above. The next day, I transfected the cells with an empty plasmid (OriGene, PS100001), or a plasmid containing the coding sequence of SRSF1 (OriGene, RC201636) as per the conditions specified above. Twenty-four-hours later, I irradiated one plate of the transfected cells at 10Gy, and not the other. I collected the cells 48-hours after irradiation to assess apoptosis, knockdown, and overexpression using western blotting.

For cell cycle analysis (Figure 3.5A), I irradiated HEK293T cells at 10Gy, or did not irradiate, two days after reverse-transfection (see SRSF1 Knockdown). Two days later, I collected the supernatant media in the wells, trypsinized the cells, and subsequently collected them into the same tubes as respective supernatants. Next, the cells were centrifuged at 200 g for 3 minutes and the supernatant decanted. The cells were then fixed with 70% ethanol then centrifuged at 2400 g for 10 minutes. I washed the pellet with PBS and centrifuged again. I decanted the supernatant and resuspended the pellet in 300µl PBS. Next, 5µl of 10mg/ml RNase A and 2µl of 25% Triton X-100 were added to each sample and incubated at 37ºC for 30 minutes. Finally, I added 12.5µl of 1mg/ml propidium iodide (Sigma-Aldrich,

P4864) to each sample, and read on CytoFlex (Beckman Coulter). At least 10,000 events were read for each sample.

3.2.7 Viability Assay

Cells were plated and transfected as above. When ready for reading, I added alamarBlue HS (Invitrogen, #A50100) at a final concentration of 1/10 to each well, incubated at 37ºC for 2-4 hours, then read on a fluorescent plate reader (POLARstar Omega) using 544 nm as the excitation wavelength and 590 nm as the emission wavelength.

I had done the viability assay with varying replicates per condition: Control and siSRSF1-2 at 0, 6, and 10 Gy were repeated four independent times. At 3 Gy, they were repeated three independent times. siSRSF1-1 was repeated three independent times for 0, 6, and 10 Gy, and two independent times for 3 Gy.

3.2.8 Flow Cytometry-based Assessment of Major Histocompatibility Complex (MHC) Class I Levels on the Cell Surface

I plated 250,000 MDA-MB-231 cells on 6-well plates and reverse-transfected using Lipofectamine RNAiMAX as described in 3.2.1 with siSRSF1-1, siSRSF1-2, or Control siRNA. I let the cells grow for two days, and then either irradiated at 10 Gy or not, changing the media of all wells to 2 ml of antibiotic-containing media before IR or mock IR (leaving the plate on the bench for the same length of time as the irradiation). One well in the mock IR plate is the positive control, Interferon γ (IFNγ). To this well, I added IFNγ solution to a final concentration of 2 ng/µl. Forty-eight hours later, I aspirated media from the wells, washed them gently with PBS, added 500 µl of PBS and collected the cells in 1.5 ml tubes by scraping. I repeated the collection step with another 500 µl of PBS. I then pelleted the

cells and discarded the supernatant. Next, I resuspended the pellet in 310 µl of 5% FBS (in PBS) and moved 150 µl of the cells to a 96-well plate for easier handling, discarding the rest, except for 10 µl of each condition, which were combined in one well to make up the unstained cells gating control. I centrifuged the plate at 1,800 rounds per minute (rpm) for 3 minutes at 4°C, and in the meantime prepared the anti-MHC-I antibody (Biolegend, #311410), diluting it 1:500 in 5% FBS. Once the centrifugation was completed, I decanted the supernatant from the plate and added 50 µl of the diluted antibody. Next, I incubated for 20-30 minutes at room temperature in the dark, then added 100 µl of 5% FBS and pelleted the cells by centrifugation as above. After decanting the supernatant, I resuspended the cells in 4% paraformaldehyde to fix them and stored them at 4°C for 3-7 days before running the samples on the flow cytometry machine CytoFlex (Beckman Coulter). I repeated the experiments for a total of three biological replicates. The data was analyzed in FlowJo (Becton Dickinson) and Floreada (available at https://floreada.io; accessed on 26/03/2024) and the gating strategy is shown in Figure 3.1. I used Excel to carry out the two-sample Student's t-test in Figure 3.8C.
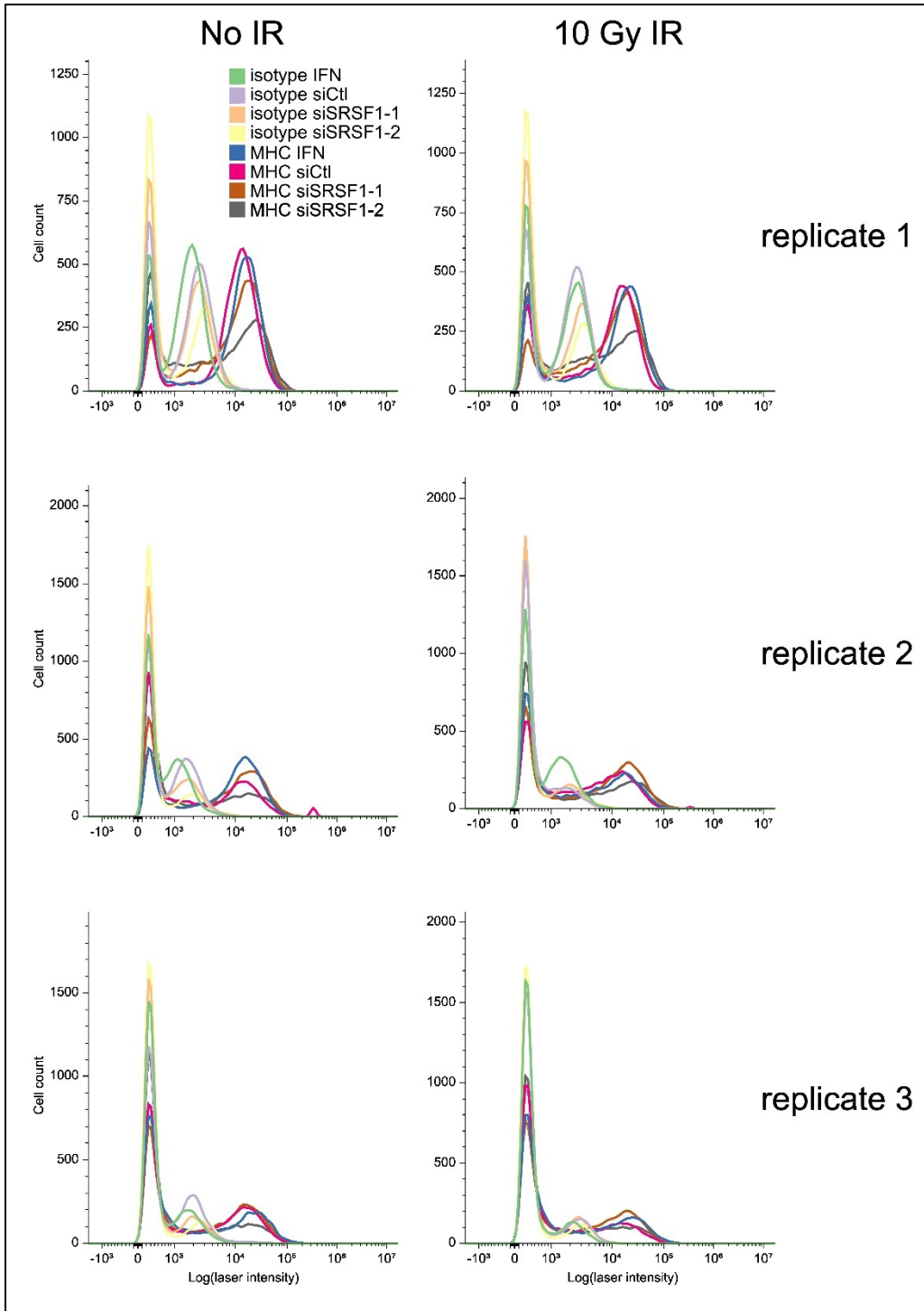
Figure 3.1 Gating strategy used in flow cytometry analysis of MHC-I on the cell surface. IR, ionising radiation; IFN, interferon; siCtl, non-targeting-siRNA-transfected cells; siSRSF1-1, single siRNA targeting SRSF1; siSRSF1-2, siRNA pool targeting SRSF1; MHC, major histocompatibility complex

3.2.9 RNA Extraction

I reverse-transfected 200,000 HEK293T cells as described in 3.2.6. Roughly 36 hours after transfection, I changed the media of all the wells to media with antibiotics, then irradiated on plate at 10 Gy and kept the other unirradiated. I collected the cells 48 hours later by trypsinization. The RNA was extracted using RNeasy Mini Kit (QIAGEN, #74104) with vacuum manifold according to the manufacturer's instructions. The primary modifications involved variations in buffer volumes and additional washing steps, as suggested by user discussions on Research Gate. The procedure was conducted at ambient laboratory temperature (20–25°C). The vacuum manifold was ventilated between loading steps to ensure uniform conditions for each sample, primarily achieved by removing the notch on an unused sample hole. The lids of the RNeasy spin columns were kept open during vacuum application, and the QIAvac 24 Plus vacuum manifold was set up following the handbook instructions. Additionally, RNaseZap was applied to all benchtop equipment, including pipettes and the vacuum manifold, to prevent RNA degradation.

Cell pellets were resuspended in 350 µl Buffer RLT and thoroughly mixed by vortexing. The lysate was then transferred into a QIAshredder spin column placed in a 2 ml collection tube and centrifuged for 2 minutes at full speed. An equal volume of 70% ethanol was added to the homogenized lysate, followed by thorough mixing. Subsequently, 700 µl of the sample from this step, including any precipitates formed, was transferred to each RNeasy spin column on the vacuum manifold, and vacuum was applied.

This was followed by several washing steps. Firstly, 700 µl Buffer RW1 was added to each column, an increase from the standard 350 µl, to ensure complete removal of Buffer RLT. A DNase/RDD mixture (10 µl DNase + 70 µl Buffer RDD per sample) was prepared

and applied to each column, followed by a 15-minute incubation at room temperature. The columns were then washed sequentially with 350 µl and then 700 µl Buffer RW1, followed by two washes with 500 µl Buffer RPE. An additional step not in the original protocol involved a further wash with 500 µl Buffer RPE and a 1-minute incubation to remove all salts attached to the membrane. The vacuum was applied after each addition until the transfer was complete, and the vacuum manifold was ventilated afterwards.

Finally, the RNeasy spin columns were transferred to new 2 ml collection tubes for a 1-minute centrifugation at full speed. This was followed by placing each column in a new 1.5 ml collection tube, onto which 30 µl RNAse-free water was carefully added directly to the spin column membrane. After a 5-minute incubation at room temperature, the tubes were centrifuged for 2 minutes at full speed to elute the RNA. A total of 5 biological replicates were collected.

The RNA concentration was measured using NanoDrop One (Thermo Fisher Scientific) and its integrity assessed with Qubit RNA IQ assay (Thermo Fisher Scientific, #Q33222) on the Qubit 4 (Thermo Fisher Scientific). The remining eluted RNA was stored in -80ºC.

3.2.10 RNA Sequencing and Data Processing

Total RNA was shipped to Novogene in dry ice for Illumina sequencing on Novaseq 6000, paired-end 150 bases. At Novogene, mRNA was purified from total RNA using poly-dT oligo-attached magnetic beads. After fragmentation, the first strand cDNA was synthesized using random hexamer primers followed by the second strand cDNA synthesis. The library was ready after end repair, A-tailing, adapter ligation, size selection, amplification, and purification. The library was checked with Qubit and real-time PCR for quantification and bioanalyzer for size distribution detection. Quantified libraries were

pooled and sequenced on Illumina platforms, according to effective library concentration and data amount. The original fluorescence images obtained from high throughput sequencing platforms are transformed to short reads by base calling. These short reads are recorded in FASTQ format, which contains base information (reads) and corresponding sequencing quality information.

The sequenced reads (raw reads) often contain low quality reads and adapters, which will affect the analysis quality. Therefore, it is necessary to filter the raw reads to get the clean reads. The filtering process is as follows: 1) Remove reads containing adapters. 2) Remove reads containing N > 10% (N represents bases that cannot be determined). 3) Remove reads containing > 50% low quality (Q score ≤ 5) bases. The cleaned reads were sent to me in FASTQ files, and I processed the data from there.

I processed the data on the University of Oxford's Biomedical Research Computing (BMRC) and Advanced Research Computing (ARC) clusters using Salmon v1.10.0 (Patro et al. 2017), mapping to the human genome telomere-to-telomere (T2T)-CHM13v2.0, acquired from National Center for Biotechnology Information (NCBI). First, I created the index, starting with preparing the metadata. Salmon indexing requires the names of the chromosomes, which I extracted with the following command:

```
grep "^>" < chm13v2.0.fa | cut -d " " -f 1 > decoys.txt
sed -i.bak -e 's/>//g' decoys.txt
```

Next, I converted the GFF3 file of the genome to FASTA format:

```
gffread    -w    transcripts.fa    -g    chm13v2.0.fa
chm13v2.0_RefSeq_Liftoff_v5.1.gff3
```

The next step was to construct a "gentrome" fasta file, where the genome and the transcriptome are concatenated. The gentrome is the entire human genome, and I used it as a decoy sequence, as suggested in Salmon's documentation.

I created the index as follows:

```
salmon index -t gentrome.fa -d decoys.txt --keepDuplicates
-p 12 -i salmon_index
```

Finally, I ran Salmon in mapping-based mode with the following command:

```
salmon quant -i salmon_index -l ISR --seqBias --gcBias -1
${sample}_1.clean.fq.gz"  -2  "${sample}_2.clean.fq.gz"  -o
"${sample}_salmon_output" -p 20
```

The data is available on GEO under accession number GSE242550 with the access token cxypcwyqnzkfrop. To assess the overlap between this dataset and the B-cell data, I used IsofromSwitchAnalyzeR subsetSwitchAnalyzeRlist() to extract the isoforms in the HEK cells that were significantly (FDR < 0.05) switching expression between siCtl 0Gy and siSRSF1 0Gy, and defined those as SRF1 targets. I compared this list to the list of SRSF1 targets defined in 2.2.9 for the B cells.

3.2.11 Isoform Switching and Alternative Splicing Analysis

After running Salmon, I used the IsoformSwitchAnalyzeR R package (Vitting-Seerup and Sandelin 2017, 2019) to analyze alternative splicing and isoform switching. The following functions specifically were used: I used importIsoformExpression(), followed by imporRdata() function to import the data from Salmon (Soneson et al. 2016; Robinson and Oshlack 2010). Next, I used isoformSwitchTestDEXSeq() to test for isoform switches

(Ritchie et al. 2015; Anders et al. 2012). Finally, I used analyzeAlternativeSplicing() to annotate alternative splicing events (Vitting-Seerup et al. 2014).

3.2.12 Statistical Analysis and Kaplan-Meier Plots

Due to the varying number of replicates, I carried out the statistical analysis for the viability assay in Figure 3.5B after consulting with Oxford Statistical Consultancy: two-way ANOVA was used, followed by Dunnett's Multiple Comparisons Test, comparing knockdown samples to non-knockdown sample at each dose. I performed the analysis on raw values, and extrapolated the p-values to the percentages shown in Figure 3.5B.

The data in Figure 3.3 and Figure 3.4 were obtained from the Gene Expression Omnibus (GEO). The data in Figure 3.3 is all RNA-seq data, while Figure 3.4 is all microarray data. The accession numbers, citations, and statistical tests performed are in the legend for convenience. The data was taken as is without further processing, except to plot and perform the statistical tests.

I downloaded TCGA data for survival analyses (Figure 3.7C) from Xena Browser (Goldman et al. 2019). Enric Domingo from Oxford's Department of Oncology obtained additional clinical data from Liu et al (Liu et al. 2018): gender, age, stage, grade, histological type and progression-free interval (PFI). PFI is defined as the period from diagnosis to first occurrence of a new tumour event which includes disease progression, locoregional recurrence, distant metastasis, new primary tumour or death with tumour. Patients alive without these events or dead without tumour were censored. I filtered the samples to include only primary tumours, patients who received radiotherapy, and patients with complete survival information, leaving a total of 272 patients for LGG and 540 for BRCA. We used the package survminer (version 0.4.6, https://CRAN.R-project.org/package=survminer) to

statistically determine the optimal expression level cutoff for dividing patients into "low SRSF1 expression" and "high SRSF1 expression" categories and plot the date. The package survival (version 2.38, https://CRAN.R-project.org/package=survival) was used to fit the Cox model. Univariate analyses results are shown in Table 3.1. Based on the significant (p < 0.05) variables in univariate analyses, I fit a multivariate analysis of *SRSF1* expression and stage for breast cancer and it better fit the data, compared to a model with stage alone (Chi-squared p-value = 0.02). For lower-grade glioma, I fit a multivariate analysis of *SRSF1* expression, age, histological grade, and histological type (oligoastrocytoma and oligodendroglioma) and it better fit the data, compared to the same model without *SRSF1* expression (Chi-squared p-value = 0.03).

Table 3.1 Univariate analyses results per dataset

Dataset	Variables	Number of samples	Hazard ratio	P-value	97.5% confidence interval
Breast Cancer	*SRSF1* expression	540	1.83	0.013	1.13 - 2.95
	stage	530	2.81	4E-08	1.94 - 4.06
	age	540	1	0.970	0.98 - 1.02
	gender	540	2.43	0.378	0.34 - 17.56
	histological type	540	1.39	0.276	0.77 - 2.51
Lower-grade Glioma	*SRSF1* expression	272	2	0.003	1.27 - 3.15
	Oligoastrocytoma	272	0.46	0.003	0.28 - 0.77
	Oligodendroglioma	272	0.45	0.001	0.28 - 0.72
	histological grade	272	1.57	0.041	1.02 - 2.43
	age	272	1.03	5E-04	1.01 - 1.04
	gender	272	0.87	0.486	0.59 - 1.28

For the KM plots in Figure 3.8B and C, I used the Kaplan-Meier plotter tool (Győrffy 2023; Kovács et al. 2023) with the following parameters: Affy ID: SRSF1, Survival: PFS, Auto select best cutoff: checked, Follow up threshold: all, Censor at threshold: checked, [then either Anti-PD-1 treatment: all anti-PD-1, Anti-PD-L1 treatment: not used; or Anti-

PD-1 treatment: not used, Anti-PD-L1 treatment: all Anti-PD-L1], Anti-CTLA-4, treatment: not used, Sample acquisition: all, Tumour type: all, Gender: all. Tumour types included in this pan-cancer analysis are bladder cancer (n=73), esophageal adenocarcinoma (n=103), glioblastoma (n=28), hepatocellular carcinoma (n=22), head and neck squamous cell carcinoma (n=5), melanoma (n=423), non-small-cell lung cancer (n=43), urothelial cancer (n=348).

3.3 Results

3.3.1 SRSF1 Mediates Isoform Switching in Irradiated Cells

Having examined the IR-responsive isoforms, we turned to identify the proteins that mediate the alternative splicing. We focused on genes that encode RNA-binding proteins (RBPs) (Ray et al. 2013), and performed ANOVA on their expression levels in the 10 individuals before and two time points after IR to identify candidate regulators. We found 35 IR-responsive RBPs (FDR < 5% and fold change ±1.2) (Table 3.2) that include members of splicing factor families, such as the SR-rich splicing factor family (*SRSF1*, *SRSF2*, *SRSF7*). The expression of most (28/35) of these RBPs decreases in the B cells following radiation. For example, SRSF1 expression levels are significantly ($p < 0.05$) repressed at both the RNA and the protein levels (Figure 3.2A and B).
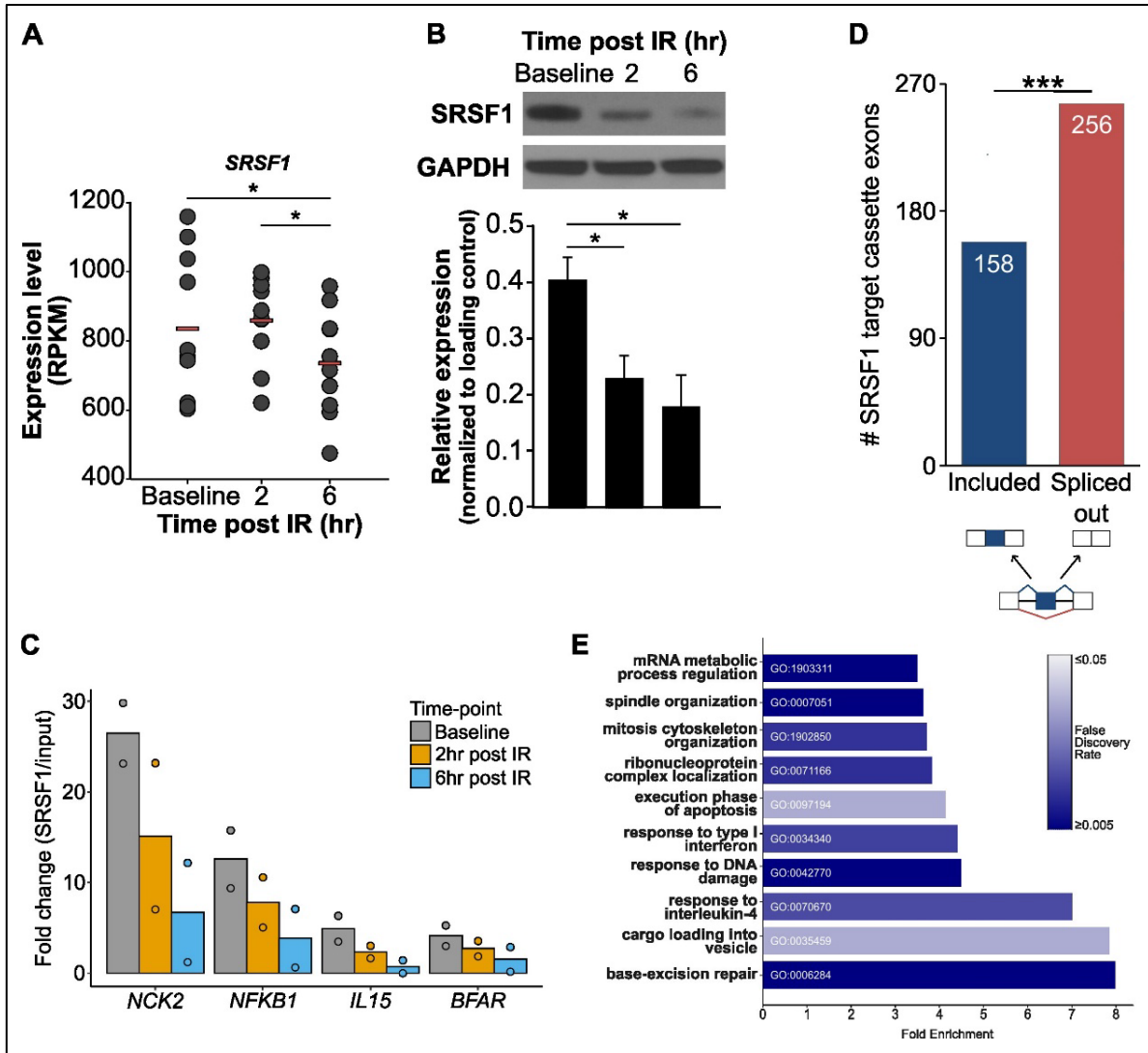
Figure 3.2 SRSF1 loss mediates IR-responsive isoform switching. **A)** RNA expression of SRSF1 at baseline (immediately before IR), and 2 and 6 hours after IR. Each dot represents a sample from one individual (n = 10). The mean expression level is displayed as a red bar. NR performed the analysis, while I created the figure. **B)** Protein expression of SRSF1 at the same timepoints as A. Top: representative western blot of SRSF1 and GAPDH expression. Bottom: quantification of SRSF1 protein expression from 6 biological replicates. For each condition, expression of SRSF1 relative to GAPDH is shown. Error bars represents standard error of the mean. The western blots were completed before my arrival in the Cheung Lab. However, I analyzed and plotted the data. **C)** Mean fold enrichment of SRSF1-RNA-immunoprecipitation over input for example genes across time in response to IR. Each gene contains a cassette exon that is alternatively spliced in response to IR. Individual circles

represent the fold enrichment of the two biological replicates. The RNA immunoprecipitation and sequencing was carried out and processed before my arrival in the Cheung Lab. I plotted the data of my chosen genes of interest. **D)** The outcome of splicing for SRSF1 target cassette exons. **E)** Fold enrichment of relevant Gene Ontology Biological Process terms enrichment in IR-responsive SRSF1 target genes. Red line indicates Benjamini-Hochberg-corrected p-value = 0.05. RPKM, reads per kilobase of transcript per million reads mapped; IR, ionising radiation; * $p < 0.05$; *** $p < 0.001$.

Table 3.2 Expression levels of 35 IR-responsive RBPs and their number of putative target exons. The RNA-seq data was generated and processed, including differential expression analysis, before my arrival in the Cheung lab. I used it to create this table.

RBP gene	Average expression (RPKM) of RBP			ANOVA p-value	# IR-responsive CE with RBP motif
	Baseline	**2hr post IR**	**6hr post IR**		
SRSF1	837.94	861.36	737.32	1.71E-02	358
SRSF2	2722.09	3185.38	2279.40	4.59E-03	183
PCBP2	4131.13	4410.29	3842.44	9.94E-03	173
PCBP1	7982.84	7051.80	5992.70	3.26E-03	144
HNRNPK	4778.22	5138.24	4134.01	4.04E-03	139
HNRNPL	5073.45	4972.22	3787.44	3.31E-03	129
CELF6	14.31	11.11	16.08	7.60E-03	126
RBM6	919.48	763.70	711.90	1.07E-03	123
PTBP1	3532.94	4156.44	3250.68	1.18E-02	115
SART3	922.27	774.60	875.81	9.31E-03	112
SNRPA	1992.52	1994.88	1700.43	1.79E-02	109
G3BP2	1300.21	1585.28	1294.12	4.02E-03	107
HNRNPLL	373.46	362.17	279.19	7.04E-03	103
SFPQ	4111.10	5155.28	3884.70	3.78E-03	101
MBNL1	1132.06	1083.69	951.08	1.48E-02	93
ZC3H10	116.01	72.77	93.49	3.95E-03	91
RBM4	250.13	294.34	245.85	1.00E-02	81
MATR3	1869.45	2124.21	1744.21	5.48E-03	79
HNRNPA2B1	11513.86	11318.96	8771.10	4.11E-03	78
IGF2BP3	410.86	374.84	319.95	4.30E-03	77
ZNF638	510.14	371.23	520.59	2.31E-03	72
CPEB2	56.72	68.58	63.13	1.19E-02	70
ZC3H14	375.21	335.36	332.24	1.25E-02	67

PPRC1	574.50	699.43	317.46	1.84E-03	65
RBM28	462.50	428.69	383.38	1.23E-02	61
TIA1	457.49	420.08	470.89	9.49E-03	61
CPEB4	210.88	313.46	270.00	5.75E-03	59
U2AF2	2528.97	2373.39	2143.97	1.72E-02	56
RBM8A	396.87	423.70	334.88	2.21E-03	47
RBM41	102.66	74.02	139.61	2.75E-04	44
SRSF7	2169.83	2329.04	1712.11	3.01E-03	43
RBM3	2268.48	2747.73	2425.64	7.23E-03	39
QKI	31.58	28.25	26.28	1.10E-02	36
TUT1	410.35	319.20	387.06	8.61E-03	32
DAZAP1	2094.83	2066.82	1678.20	1.23E-02	29

RBP, RNA-Binding Protein; RPKM, Reads Per Kilobase of transcript per Million reads mapped; IR, Ionising Radiation; ANOVA, Analysis of Variance; CE, Cassette Exons; [a] Benjamini-Hochberg-adjusted.

To ask if these IR-responsive RBPs could mediate alternative splicing in response to IR, we looked for the binding sites for these RBPs in the IR-responsive cassette exons. We identified binding sites for the 35 IR-responsive RBPs in 602 of the 645 IR-responsive cassette exons (Table 3.2). The RBP with the largest number of putative targets is SRSF1: over half (358/645) of the IR-responsive cassette exons have SRSF1 binding sites. The second splicing factor with the most putative targets was SRSF2 (183/645 cassette exons). These findings relied on known RBP binding motifs to identify regulatory sequences within the cassette exons. To extend the analysis to unknown potential regulatory sequences, we also performed *de novo* motif enrichment in the IR-responsive exons. This analysis revealed two GA-rich motifs: GRAAAT and GAGGAGGA, where "R" represents G or A, as the most enriched sequences (Table 3.3). Respectively, each motif was found in 414 (64%) and 488 (76%) of the 645 IR-responsive cassette exons (Table 3.3). These enriched motifs resemble known SRSF1 binding motifs (e.g. GGAGGA, GRAGGA), which are GA-rich (Anczuków et al. 2015; Ray et al. 2013; Zhou et al. 2018; Maslon et al. 2014; Feng et al.

2019). They are also largely found in the same exons as known SRSF1 motifs: of the 414 exons with GRAAAT or GAGGAGGA motifs, 247 (~60%) and 333 (~70%), respectively, have sequences that correspond to the known SRSF1 motifs. Additional experiments are required to verify these motifs, but these findings suggest that SRSF1 plays a role in IR-induced isoform switching of cassette exons.

Table 3.3 *De novo*-enriched motifs within ionising radiation-responsive cassette exons. The enrichment was carried out by Joshua Burdick of the Cheung Lab, but I generated this table based on its results.

De novo-enriched motif	Sequence logo	Number of IR-responsive cassette exons with the *de novo*-enriched motif
GRAAAT		414
GAGGAGGA		488

IR, ionising radiation

To validate that SRSF1 indeed mediates IR-responsive alternative splicing, we carried out RNA-immunoprecipitation (RNA-IP) with an antibody against SRSF1, followed by sequencing. The results showed that 458 (71%) of the 645 IR-responsive cassette exon isoforms are bound by SRSF1 (Figure 3.2C). Among these 458 cassette exons, 414 of them have the SRSF1 binding motif. As noted above, SRSF1 decreases in response to IR. As expected, our data suggest that, over the time course of radiation exposure, as SRSF1 expression decreases, its relative binding to its targets tends to decrease as well. (Figure 3.2C). After IR, 62% of SRSF1 target exons are spliced out after IR (p < 0.0001; Figure

3.2D). This suggests that SRSF1 mediates the inclusion of this subset, and its loss contributes to the generation of shorter IR-responsive isoforms. These 414 exons come from 346 genes, which are enriched for GO terms related to apoptosis, cell cycle arrest, and the DNA damage response (Figure 3.2E). Notably, among SRSF1 targets are the aforementioned *NFKB1*, *ZSCAN32* (Figure 2.1C), along with *NCK2* isoform that leads to defective cytokinesis, and the *BFAR* isoform that cannot suppress BAX-induced cell death (Figure 2.4D). Together, these results suggest that normal human B cells downregulate SRSF1 as a mechanism of coping with IR-induced DNA damage.

3.3.2 SRSF1 is Associated with Radioresistance *in vitro* and in Cancer Patients

Next, we asked whether SRSF1 depletion is part of the general response to IR and can be generalized beyond B cells. To answer that, I first leveraged data in the NCBI Gene Expression Omnibus (GEO) (Barrett et al. 2013; Edgar et al. 2002) and found that, in various human cells, including cardiomyocytes, lung and aortic endothelial cells (Chopra et al. 2022), macrophages, as well as in a glioblastoma cell line (Zhang et al. 2019a), the expression level of *SRSF1*, as in B cells, decreased after IR (Figure 3.3). Although SRSF1 expression level decreases across the different cell types, its expression pattern across time post IR varies.
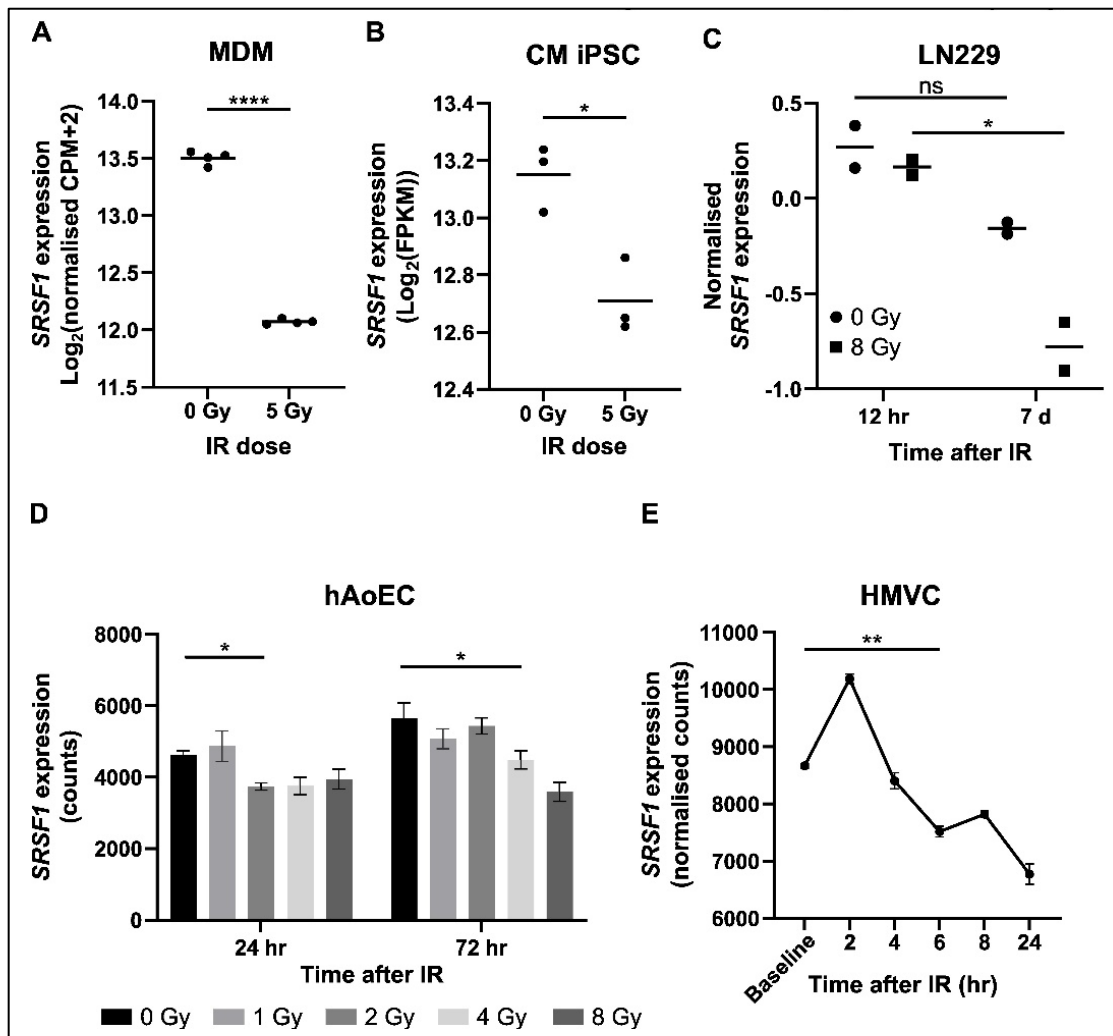
Figure 3.3 SRSF1 is depleted after IR in several human cell lines. **A)** *SRSF1* expression (quantile-normalized log₂(Counts Per Million (CPM) +2)) in primary human monocyte-derived macrophages (MDM) exposed to 5 Gy γ-irradiation, or unirradiated (RNA-seq data; GEO accession GSE145577). No information was provided on the timepoint of collection after IR exposure. **** $p < 0.0001$ (two-tailed Welch's t-test). Each filled circle is a biological replicate and the horizontal line represents the mean. **B)** *SRSF1* expression (log₂(fragments per kilobase million (FPKM))) in induced Pluripotent Stem Cell-derived cardiomyocytes (iPSC CM) 48 hours after 5 Gy X-radiation or sham irradiation (RNA-seq data; GEO accession GSE107685) (Becker et al. 2018). * $p = 0.01$ (two-tailed Student's t-test). Each filled circle is a biological replicate and the horizontal line represents the mean. **C)** Normalized *SRSF1* expression in glioblastoma cell line LN229 12 hours or 7 days after exposure to 8 Gy or mock-irradiation (RNA-seq data; GEO accession GSE121422). Each filled circle is a biological replicate and the horizontal line represents the mean. **D)** *SRSF1*

expression (counts) in human aortic endothelial cells (hAoEC) 24 and 72 hours after exposure to the indicated IR doses (RNA-seq data; GEO accession GSE202119). Each filled circle is a biological replicate and the horizontal line represents the mean. **E)** *SRSF1* expression (normalized counts (DESeq2 size factor method)) in human lung microvascular endothelial cells (HMVC) right before irradiation (baseline) and 2, 4, 6, 8, and 24 hours post 10 Gy X-irradiation (RNA-seq data; GEO accession GSE179810). ** $p = 0.003$ (two-tailed Welch's t-test). Each filled circle represents the mean, and error bars represent the standard error of the mean from three biological replicates.

To ask if SRSF1 depletion is IR-specific or a general response to DNA double-strand breaks (DSBs), I extended the analysis to other agents that induce DSBs. Leveraging the expression data on NCBI Gene Expression Omnibus, I found that, in different human cell types (namely, breast cancer [MCF7], renal [HEK293T], and myeloid leukemia [U937] cells), *SRSF1* levels decrease after treatment with other DSB-inducing agents, such as doxorubicin, etoposide, and camptothecin (Figure 3.4). Altogether, these data support the hypothesis that SRSF1 downregulation is a prevalent mechanism employed by cells to facilitate the response to DSBs.
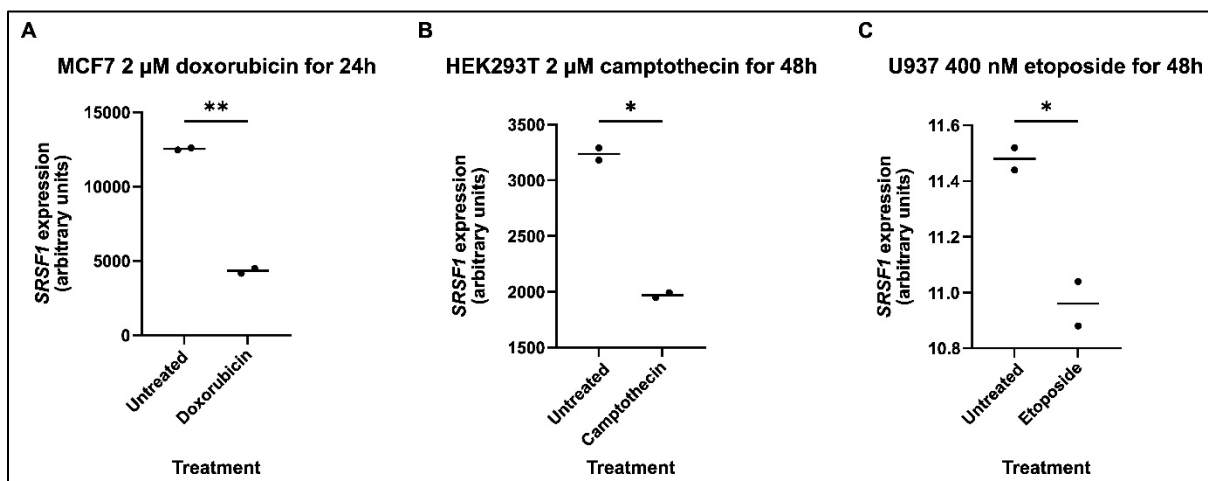
Figure 3.4 *SRSF1* is depleted in response to other double-strand break-inducing agents, namely **A)** doxorubicin (microarray data; GEO dataset GDS4936), **B)** camptothecin (microarray data; GEO dataset GDS1453), and **C)** etoposide (microarray data; GEO dataset

GDS5809). * p < 0.05, ** p < 0.01 (two-tailed paired t-test). The gene expression units are 'arbitrary' as described in GEO Profiles Help page (www.ncbi.nlm.nih.gov/geo/info/profiles.html; accessed 26/03/2024).

To further assess the biological role of SRSF1 in the radiation response, I knocked down SRSF1 by RNA interference in HEK293T cells and assessed the cell cycle profile and viability after IR (Figure 3.5A and B). I used HEK293T cells here instead of B cells because the latter are notoriously challenging to transfect. SRSF1 loss exacerbated IR-induced cell cycle arrest at G2 (p < 0.05; Figure 3.5A) and led to a significant reduction in cell viability at all tested IR doses (p < 0.001; Figure 3.5B). I also assessed apoptosis levels using cleaved PARP as a marker. Upon irradiation, cells with lower SRSF1 had significantly (p < 0.05) higher apoptosis levels compared to controls (Figure 3.5C and Figure 3.6).
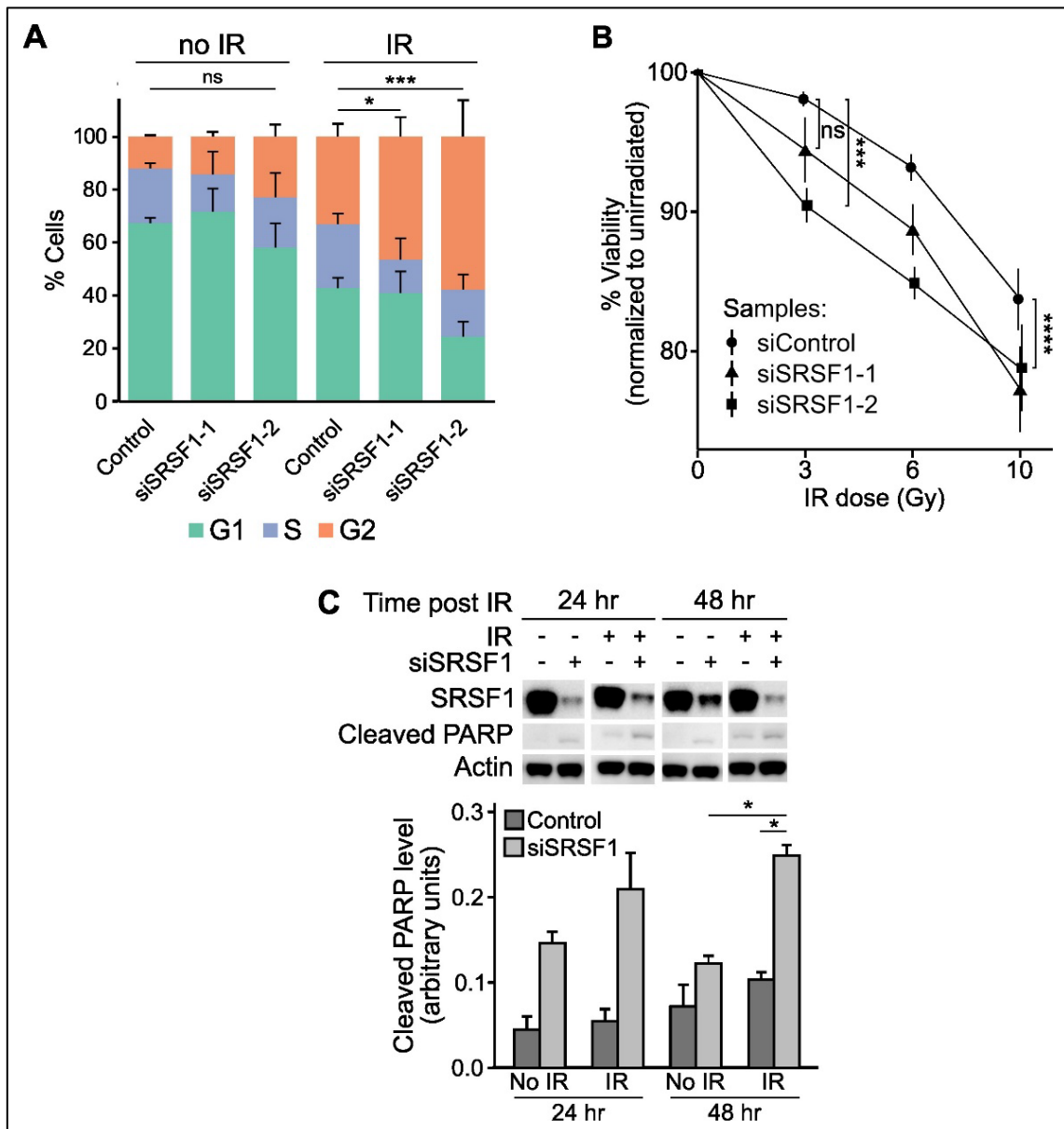
Figure 3.5 SRSF1 loss promotes G2 arrest and cell death in response to radiation. **A)** Cell cycle profile (based on propidium iodide staining) of HEK293T cells transfected with non-targeting control siRNA, an siRNA against SRSF1 (siSRSF1-1), or a pool of siRNAs against SRSF1 (siSRSF1-2), and then either irradiated at 10 Gy (IR) or not (no IR) 48 hours prior to collection. Error bars represent standard error of the mean from four biological replicates. **B)** Resazurin viability assay of HEK293T cells transfected with control siRNA, siSRSF1-1, or siSRSF1-2. The cells were exposed to increasing IR doses and resazurin was added 48 hours later to measure viability. Error bars represent standard error of the mean from 2-4 biological replicates. **C)** Top: representative western blot of HEK293T cells

transfected with a control siRNA (-) or siSRSF1-1 (+), and then either irradiated at 10 Gy or not, 24 and 48 hours prior to collection. Bottom: quantification of cleaved PARP from three biological replicates. Error bars represent standard error of the mean from three biological replicates. IR, ionising radiation; ns, not significant; * $p < 0.05$; *** $p < 0.001$; **** $p < 0.0001$. The 48-hour timepoint was used in A and B to allow time for at least one cell cycle, and because this timepoint had a stronger cleaved PARP signal in C (which was done before A and B).
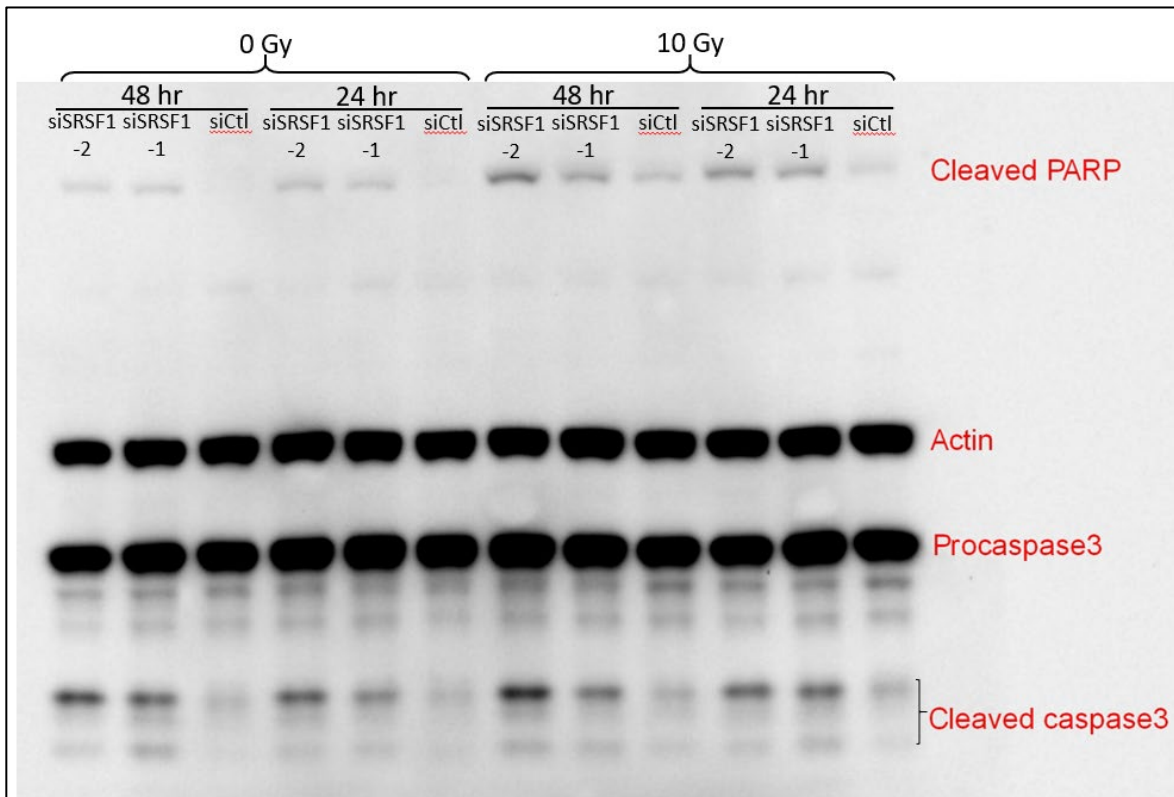
Figure 3.6 Full image of the western blot shown in Figure 3.5C. This blot originally also contained cleaved caspase-3 bands, an early apoptosis marker, showing the same trend as cleaved PARP. Because it is an early apoptosis marker, I did not use it in subsequent replicates, however, and focused instead on cleaved PARP.

Crucially, I then overexpressed SRSF1, which rescued the cell death (Figure 3.7A), thus establishing that the decrease in SRSF1 level promotes cell death by apoptosis. I sequenced cDNA from SRSF1 knockdown and irradiated cells (with respective controls), and found that, despite the inherent cellular differences between B cells and HEK293T cells, over 20%

(77/346; hypergeometric p-value < 0.001) of the IR-responsive SRSF1 target genes identified in B cells also undergo isoform switching in HEK293T cells post-SRSF1 knockdown (Figure 3.7B). These include crucial apoptosis genes, e.g. *MDM2*, *BAX*, *CASP3*; cell cycle genes, e.g. *PRC1*, *CDC16*, *CDK11B*, and DNA repair genes, e.g. *POLL*, *INO80E*, *SETMAR*. This observation not only validates these targets, but also reinforces the pivotal role of SRSF1 in response to IR.
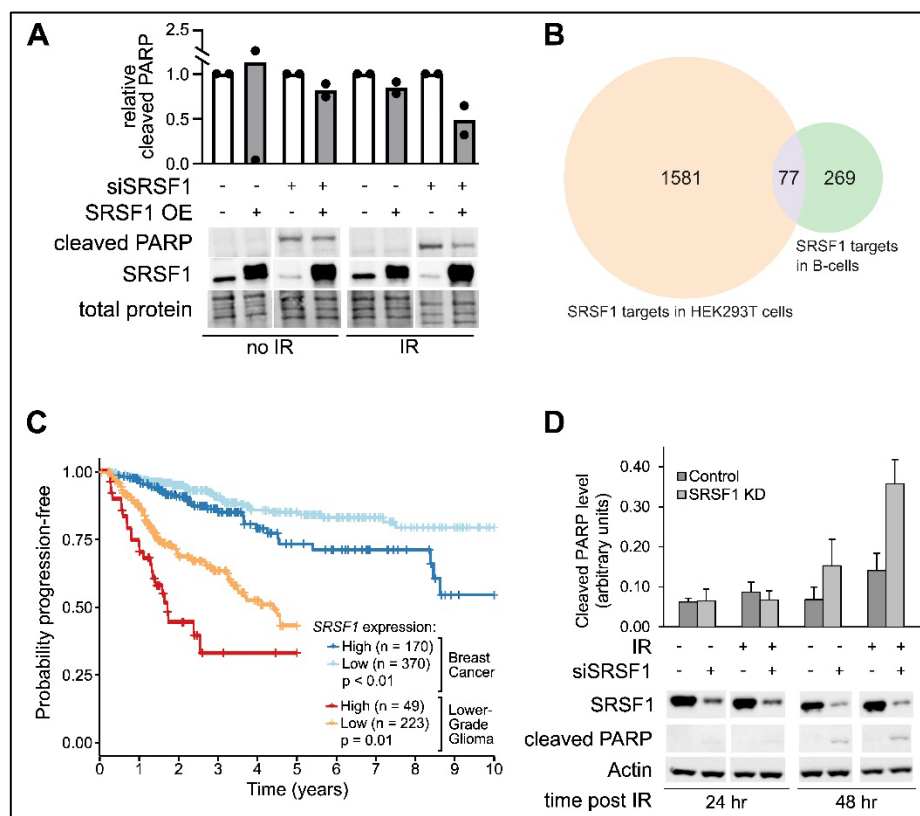
Figure 3.7 SRSF1 is associated with radiosensitivity *in vitro* and in cancer patients. **A)** Western blot (bottom) and quantification (top) of cleaved PARP levels in HEK293T cells transfected with non-targeting control or siSRSF1, as well as either an empty plasmid, or a plasmid containing the coding sequence of SRSF1. Cells were exposed to 10 Gy ionising radiation (IR) 48 hours after seeding, and collected 48 hours after irradiation. Each condition is normalized to the respective empty plasmid-transfected counterpart. The quantification is based on two biological replicates. **B)** Overlap of SRSF1 target genes in HEK293T cells after SRSF1 knockdown and B cells after 10 Gy IR. **C)** Kaplan-Meier plots depicting

progression-free survival of lower-grade glioma and breast cancer patients treated with radiotherapy over the period of 5 years for lower-grade glioma and 10 for breast cancer, stratified by their SRSF1 expression level. Enric Domingo, from the Department of Oncology at Oxford, and I collaborated on this. I accessed the data from The Cancer Genome Atlas (TCGA) and sent it to him, and I wrote the R code to analyse the results. He retrieved additional data on risk factors and adapted my R code to include them. He sent me the results and I plotted the figure. **D)** Western blot (bottom) and quantification (top) of cleaved PARP levels in MDA-MB-231 cells transfected with non-targeting control or siSRSF1. Cells were exposed to 10 Gy ionising radiation (IR) 48 hours after seeding, and collected 48 hours after irradiation. Error bars represent standard error of the mean from three biological replicates.

IonisingIonisingTo determine if this trend can be generalized, I turned to data from The Cancer Genome Atlas (https://www.cancer.gov/tcga). If low SRSF1 confers radiosensitivity, we would expect cancer cells that overexpress SRSF1 to be more radioresistant and therefore patients to have a worse outcome after radiotherapy. I checked large B-cell lymphoma patients, since this study began with sequencing data from B cells. However, there were not enough patients in the cohort to draw reliable conclusions (48 total, only 5 of which received radiotherapy). I broadened my search to cancers that overexpress SRSF1, namely lung, breast, colon, small intestine, thyroid, liver, pancreas, kidney and gliomas (Zhou et al. 2018; Karni et al. 2007; Zheng et al. 2020). I focused on studies that include at least 200 radiotherapy-treated patients, which left only the breast cancer and lower-grade glioma cohorts. With help from Enric Domingo, a postdoc in the Department of Oncology at the University of Oxford, we classified 540 breast cancer and 272 lower-grade glioma patients who received radiotherapy by *SRSF1* expression levels in their tumours. Multivariate analysis of these cohorts (see Materials and Methods section 3.2.12 for details) demonstrated that high *SRSF1* expression results in a shorter progression-free interval, accounting for established prognostic clinical and pathological factors in lower-

grade glioma (HR = 1.71 [95% CI: 1.08 – 2.70]) and breast cancer (HR = 1.74 [95% CI: 1.07 – 2.85]) patients. Patients with lower-grade gliomas that express lower *SRSF1* had a significantly higher chance of staying progression-free for five years, compared to those with high *SRSF1* levels (43% vs. 33%; log-rank p < 0.01) (Figure 3.7C). This trend is also seen in breast cancer patients with lower *SRSF1*: significantly more of them remained progression-free for 10 years (80% vs. 55%; log-rank p = 0.01), compared to those with higher *SRSF1* expression (Figure 3.7C).

In line with the patient data, I observed higher levels of apoptosis in the triple-negative breast cancer cell line, MDA-MB-231, when I knocked down SRSF1 and irradiated the cells, compared to unirradiated cells with normal SRSF1 levels (Figure 3.7D). Altogether, our data suggest that SRSF1 loss modulates isoform switching in response to IR, and this is associated with radiosensitivity.

3.3.3 *SRSF1* is Associated with Immunotherapy Resistance

Examining the top Gene Ontology terms associated with SRSF1 targets, it is notable that there are terms related to immune modulation, such as 'response to interleukin-4' and 'response to type I interferon' (Figure 3.2E). I wanted to investigate SRSF1's impact on immune modulation further, so I started with a simple experiment to assess HLA-A expression on the surface of SRSF1 knockdown and irradiated (10 Gy) cells. Using the aggressive and generally radioresistant triple-negative breast cancer cell line MDA-MB-231 as a model, I found the SRSF1 knockdown, with an individual siRNA (siSRSF1-1) or a pool (siSRSF1-2), significantly (p < 0.05) increased HLA-A expression on the cells surface to levels similar to, or higher, than the positive control, IFNγ (Figure 3.8A). Although radiation has been suggested to also increase HLA-A expression, I did not find that to be the case in

this cell line, as although HLA-A expression slightly increased in non-targeting-siRNA-transfected and siRNA2-transfected cells, the increase was not statistically significant (Figure 3.8A).

Given the finding that SRSF1 depletion increases HLA-A expression, I wondered whether this extends to an impact on immunotherapy. Thus, I analyzed data across cancers for patients who received anti-PD1 or anti-PDL1 immunotherapy drugs and found that lower *SRSF1* expression predicted much longer survival after treatment with either agent (Figure 3.8B and C). Specifically, cancer patients with low *SRSF1* levels who were treated with anti-PD1 had ~35% chance of surviving to 5 years, while those with high *SRSF1* levels were unlikely to even survive to 1.5 years. With anti-PDL1 treatment, patients with low *SRSF1* had a 40% chance of surviving to 3.3 years, but those with high *SRSF1* were unlikely to survive beyond 2.5 years. Altogether, these findings indicate that SRSF1 may influence the effectiveness of immunotherapies, potentially serving as a valuable biomarker for patient response after further validation.
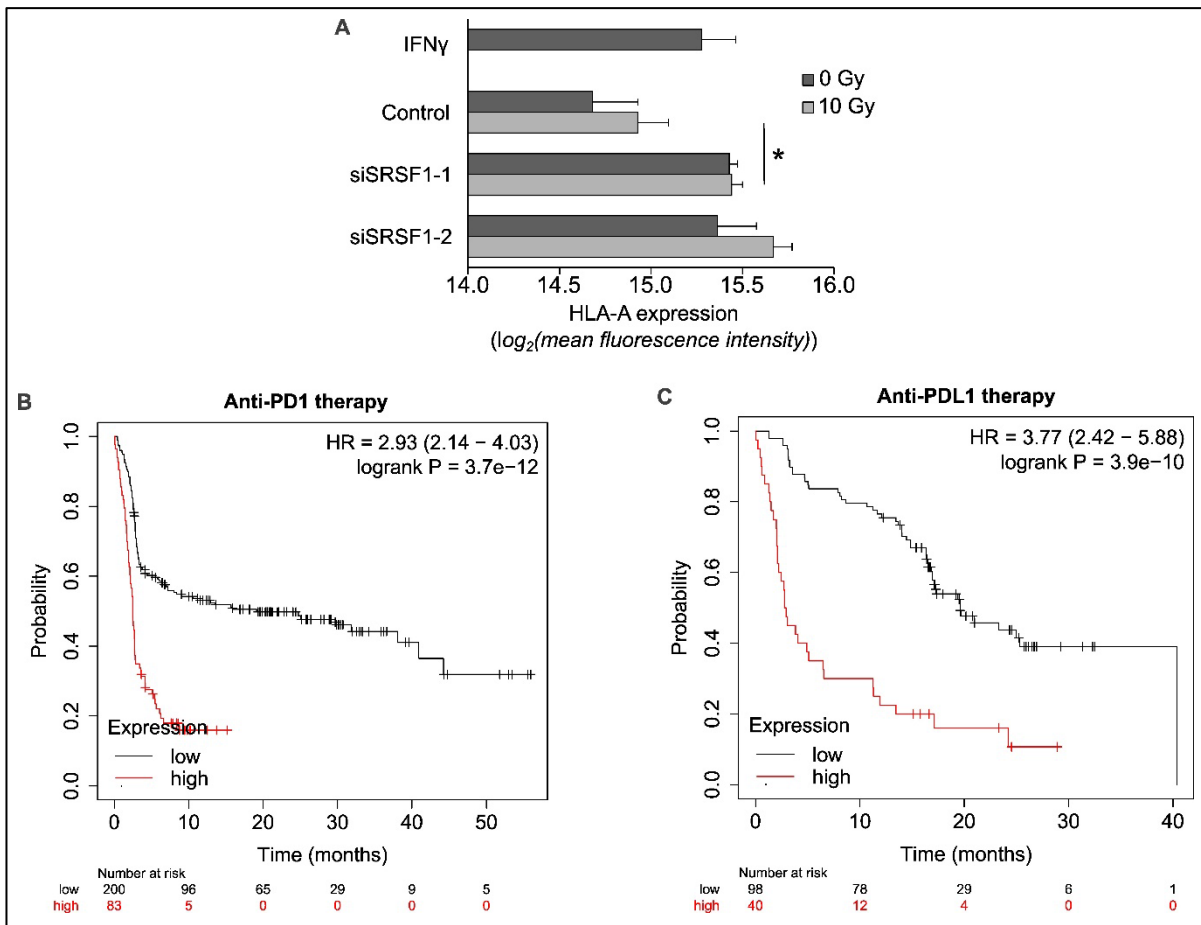
Figure 3.8 *SRSF1* level modulates the response to immunotherapy. **A)** Flow cytometry measurement of human leukocyte antigen A (HLA-A) level on the cell surface of MDA-MB-231 cells 48 hours after 10 Gy ionising radiation or mock irradiation and with or without SRSF1 knockdown with a single siRNA (siSRSF1-1) or a pool of siRNAs (siSRSF1-2). IFNγ is a positive control for HLA-A induction. **B)** and **C)** Kaplan-Meier plots depicting overall survival of pan-cancer patients treated with anti-PD1 (B) or anti-PDL1 (C) immunotherapy, stratified by their *SRSF1* expression level. * $p < 0.05$.

3.4 Discussion

Through a combination of computational and experimental analyses, we identified the splicing factor and proto-oncogene SRSF1 as a potential mediator of IR-induced isoform switching, specifically in the case of cassette exons. However, there are three important caveats to how we arrived at SRSF1: firstly, the list of 80 RBPs we started with (appendix

Table 7.2) is significantly limited compared to the number of known RBPs today (over 1,500 (Gerstberger et al. 2014; Zhao et al. 2022)). Although at the time this project started in the Cheung Lab, many more than 80 RBPs were known to exist, the Ray et al paper was state-of-the-art in characterizing the motifs of the select 80, which is why this list was used. Secondly, although our initial selection of potential RBPs was based on their expression changes post-IR, posttranslational modifications can influence splicing activity without modifying RBP levels (Matsuoka et al. 2007). Thirdly, while several RBPs had enriched binding sequences in IR-responsive cassette exons in the subsequent *in silico* analysis, we chose to focus on SRSF1 due to its highest number of target sequences.

Nonetheless, the evidence indicating that SRSF1 is potentially a key modulator or IR-induced isoform switching are as follows: first, the majority of IR-responsive cassette exons contained known SRSF1 binding motifs, more than any other splicing factor. Second, the two significantly enriched *de novo* sequences largely resemble known SRSF1 motifs and are frequently found within the same cassette exons as known SRSF1 motifs, which is important since splicing factors typically bind more than one motif in the same target gene (Jolma et al. 2020; Sohrabi-Jahromi and Söding 2021). While this is compelling, however, it is important to note that other RNA-binding proteins, such as SRSF9, have a similar motif to SRSF1 (Feng et al. 2019) and that SRSF1 also has structural motifs, which we did not consider (Adinolfi et al. 2019). Third, the majority (64%; 346/544) of genes with IR-responsive cassette exons were also physically associated with SRSF1 in RNA-IP. Finally, SRSF1 target exons are within genes involved in cell survival and division, and exons that promote these functions are removed after IR (e.g. *BFAR* and *NCK2*, respectively).

First identified by Adrian R. Krainer and Tom Maniatis (Krainer and Maniatis 1985), SRSF1 has since been shown to contribute to both constitutive and alternative splicing (Xiao and Manley 1998). It also functions both as a splicing activator or repressor (Anczuków et al. 2015; Du et al. 2021; Mayeda et al. 1993; Erkelenz et al. 2013; Pandit et al. 2013), depending on several factors, such as transcript length (Mayeda et al. 1993) and where it binds on the transcript (Anczuków et al. 2015; Erkelenz et al. 2013). We expanded this to the context of IR: most (62%) SRSF1 target exons were skipped after IR-induced SRSF1 depletion, but many were instead included.

While SRSF1 is mainly known as a splicing factor, it has other functions, such as facilitating nonsense-mediated decay (NMD), nuclear export, translation, and genome stability (Das and Krainer 2014). Isoform switching observed in previous studies was found to be mainly the result of SRSF1's role in alternative splicing specifically, as opposed to NMD (Sheng et al. 2018; Du et al. 2021, 2021; Chen et al. 2017b; Leva et al. 2012; Anczuków et al. 2012). However, NMD is also likely to contribute, as SRSF1's splicing activity promotes its effect on NMD (Aznarez et al. 2018; Zhang and Krainer 2004). Therefore, it will be important to determine which of these functions, splicing and/or NMD, contributes to SRSF1-mediated isoform switching in the context of IR.

Mechanistically, after IR, SRSF1 expression is repressed at the RNA and protein levels. While our design does not include mock-irradiated samples at matched timepoints, publicly-available human cDNA-seq datasets with matched timepoints of IR or mock IR demonstrate that *SRSF1* reduction is associated specifically with IR and not time. This reduction is also not limited to B cells. At least five other human cell lines corroborate IR-induced *SRSF1* depletion as a general phenomenon. Furthermore, SRSF1 depletion appears to be a common

mechanism in response to DSBs in general, as demonstrated by three DSB-inducing chemotherapeutic agents, which depleted *SRSF1*.

The mechanism by which SRSF1 decreases in response to IR will require further study. SRSF1 is regulated at many levels. In this case, SRSF1 protein levels drop significantly 2 hours after IR (Figure 3.2B), but not RNA levels, which are only significantly decreased by 6 hours after IR (Figure 3.2A). This indicates translational and posttranslational control occurring first, followed by transcriptional and posttranscriptional regulation, to downregulate SRSF1 levels. SRSF1 transcription is induced by the oncoprotein MYC (Gonçalves and Jordan 2015). In our data, *MYC* levels are decreased by 6 hours post IR, which could be behind the concomitant drop in *SRSF1* levels. SRSF1 can also autoregulate its expression by splicing an intron in its 3' UTR to produce an NMD isoform (Gonçalves and Jordan 2015). However, this is unlikely to be the cause of SRSF1 decrease in this case, as I did not detect any significant splicing changes in SRSF1.

Knocking down SRSF1 led to increased radiosensitivity in normal and breast cancer cells. While SRSF1 loss in B cells is correlated with an increase in the expression of isoforms that allow cells to respond appropriately to radiation exposure (e.g. the *NCK2* isoform that halts cell division and the isoform of *BFAR* that promotes apoptosis), I cannot exclude the possibility that the increased radiosensitivity is at least partially due to the accumulation of R-loops, which occur when the transcribed RNA binds to the template DNA strand, which can lead to increased levels of DNA damage (Crossley et al. 2019), potentially triggering cell death. In this respect, SRSF1 loss leads to the accumulation of R-loops. Overexpression of RNase H1, an enzyme that facilitates R-loop resolution, suppressed the cell cycle arrest and genome instability caused by SRSF1 depletion but only delayed cell death (Li and

Manley 2005). This indicates that other functions of SRSF1, including potentially its role in isoform switching, contribute to the increased cell death associated with the loss of SRSF1. With that in mind, SRSF1 is known to promote many anti-apoptotic splice variants in breast cancer and acute myeloid leukemia, reducing cell death (Kędzierska and Piekiełko-Witkowska 2017).

Patients with tumours that fail to repress *SRSF1* in response to IR, and therefore continue to divide and survive, have lower chances of remaining progression-free, compared to those with tumours that repress *SRSF1* (33% vs. 43% chance of remaining progression-free, respectively, in lower-grade glioma patients). Consistent with our findings, SRSF1 has been shown to inhibit apoptosis in vitro in a p53-independent manner (Das et al. 2012). Furthermore, Zhou et al. (Zhou et al. 2018) found that increased SRSF1 expression is correlated with increased glioma grade and poor patient survival, and that SRSF1 promoted proliferation, survival, and invasion of glioma cell lines. Similarly, SRSF1 increases the aggressiveness of breast cancer by promoting migration, proliferation, and cell cycle progression, while inhibiting apoptosis (Anczuków et al. 2015; Du et al. 2021). Here, I find that SRSF1 may confer radioresistance in normal and cancerous cells, and in breast cancer and lower-grade glioma patients.

A review by Yoshimi and Abdel-Wahab highlighted the "therapeutic window" available for targeting spliceosome-mutant/dysregulated cancers, including those that upregulate SRSF1 (Yoshimi and Abdel-Wahab 2017). Spliceosome-targeting is an exciting area of research, with the development of antisense oligonucleotides to inhibit specific splicing, and drugs that target core, as well as accessory splicing machinery to partially inhibit splicing nonspecifically.

For example, H3B-8800, a small-molecule inhibitor of the SF3b complex, which is an essential spliceosome component, showed promise in early clinical trials (Stanley and Abdel-Wahab 2022). Another example is E7820, which degrades the accessory splicing factor RBM39. This drug is currently in Phase II clinical trials for refractory myeloid neoplasms with splicing factor mutations (NCT05024994).

The increase of HLA-A expression upon SRSF1 knockdown provides a potential mechanism by which radiotherapy and immunotherapy responses may be improved in low-*SRSF1* tumours: an upregulation of HLA-A means more antigens can be presented on the cell surface. Given that SRSF1 depletion in the tumour can lead to aberrant splicing and therefore novel proteins/peptides, these neoantigens (antigens unique to the tumour cells), which may also be the result of SRSF1-depletion-induced genome instability, will be presented on the increased HLA-A molecules, allowing T cells to selectively kill the tumour cells. SRSF1 alters the splicing of PD-1 in T cells to generate the full-length, transmembrane-domain-containing isoform (Wahid et al. 2023), which promotes T-cell death. On the other hand, inhibiting SRSF1-mediated inclusion of this domain results in the production of the soluble form of PD-1 (Wahid et al. 2023), which inhibits PD-1 signaling and triggers antitumour immunity (Xiao et al. 2007). In addition, SRSF1 has been described as a "master switch" for gene expression in the immune system, regulating splicing, stability, transport of cytokine mRNAs and transcription of CD3ζ and IL-2 (Paz et al. 2020). Therefore, in patients with low SRSF1, a combination of increased 1) cell cycle-inhibiting and cell death-promoting isoforms, 2) immune-activating isoforms, 3) aberrant splicing-derived neoantigens, 4) genome instability-derived neoantigens, 5) HLA-A expression, and 6) soluble PD-1, is likely behind their responsiveness to immuno- and radiotherapy.

With additional validations, *SRSF1* levels may be used as a prognostic marker for the efficacy of radiotherapy in breast cancer and lower-grade glioma patients and for anti-PD1/PDL1 therapies. These validations include experiments that are randomised and blinded on more specimens, randomised clinical trials with stringent statistical controls, followed by demonstrations of clinical validity and utility, as well as the development of a test for SRSF1 levels to be used in the clinic (Ou et al. 2021). Furthermore, my findings raise the possibility of developing therapeutics that selectively target SRSF1, its kinase SRPK1 (Tzelepis et al. 2018), or its methyl-transferase PRMT5 (Zheng et al. 2023), in cancer cells overexpressing SRSF1, including breast cancer (Du et al. 2021), lung cancer (Jiang et al. 2016), glioma (Zhou et al. 2018), and others (Karni et al. 2007), while sparing normal cells, for example using bispecific antibodies.

Chapter 4 SRSF1 depletion and ionising radiation generate neoantigens in triple negative breast cancer

4.1 Introduction

Triple-negative breast cancer (TNBC) presents a significant clinical challenge due to its aggressive nature and limited treatment options (Yin et al. 2020). Recent advances in cancer immunology and genetics have expanded avenues for therapeutic intervention, particularly through targeting tumour-specific antigens (neoantigens). This chapter delves into the complex interplay between SRSF1 depletion and ionising radiation (IR) in TNBC, exploring how these factors influence gene and isoform expression, leading to the potential generation of neoantigens.

SRSF1, a proto-oncogenic splicing factor, plays a pivotal role in RNA processing and has been implicated in various cancers, including breast cancer. Building on my findings in previous chapters, I wanted to investigate the potential of SRSF1 depletion, in conjunction with IR, on the neoantigen landscape in TNBC cells. In this chapter, I present a detailed analysis of gene and isoform-level changes in TNBC cells following SRSF1 knockdown and IR. The focus is on understanding how these changes contribute to the generation of neoantigens derived from aberrantly spliced transcripts, which can be critical targets for immunotherapeutic strategies.

Since I am interested in alternatively-spliced transcripts, short-read cDNA-sequencing, such as that typically employed via Illumina, is not ideal, as transcripts are chopped to a maximum of 150-base long chunks, which leads to a significant loss of structural information, especially since many isoforms are highly similar, so it would be nearly impossible in many cases to confidently assign a 150-base piece to the correct isoform. Long-read sequencing on the other hand, such as that employed by PacBio or Oxford Nanopore Tech, preserves the entire transcript length, thereby removing any bias in read assignment. However, long-read sequencing suffers from a high base-level error rate (relative to short-read sequencing). Therefore, to acquire a deep understanding of both transcript sequence and structure, I performed both long- and short-read cDNA sequencing on my samples. This, along with immunopeptidomics profiling, allowed me to investigate the impact of ionising radiation and/or SRSF1 knockdown on RNA sequence and structure, as well as unique antigens derived from alternatively-spliced transcripts.

The findings highlight the importance of considering isoform-level information in genomic studies, as changes at this level can have significant implications for the generation of novel antigens. Additionally, the study sheds light on the effect of combining SRSF1 knockdown and IR on altering the TNBC antigenic repertoire, providing insights into how these treatments might be harnessed to enhance the immune response against TNBC.

This chapter contributes to the broader understanding of the molecular mechanisms underlying TNBC and offers a foundation for the development of more effective, targeted immunotherapies.

4.2 Materials and Methods

4.2.1 Cell Culture, SRSF1 Knockdown, and Irradiation

MDA-MB-231 cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM; Thermo Fisher Scientific, #41966029) supplemented with 10% FBS (Sigma-Aldrich, #F9665-500ML) and 1% Penicillin-Streptomycin (Sigma-Aldrich, #P4333-100ML). The cells were cultured in T175 flasks (Corning, #431080) and expanded until two flasks were about 75% confluent. At this stage, the cells were trypsinized and counted.

In the meantime, 9.1 ml of OptiMEM (ThermoFisher, #11058021) to each of four T175 flasks for four conditions: siCtl 0Gy, siSRSF1 0Gy, siCtl 10Gy, and siSRSF1 10Gy. Next, I added 6 μl of 100 μM siRNA, followed by 91 μl of Lipofectamine RNAiMAX (Thermo Fisher Scientific, #13778030). I incubated the flasks for 10 minutes at room temperature, then added 4.5×10^6 cells in antibiotic-free media to each flask for a total of 30 ml of media per flask. This was done in the late afternoon. Around 36 hours later, I irradiated the 10 Gy flasks using a ^{137}caesium source (Gamma-Service Medical GmbH), after replacing their media with 20 ml of antibiotic-containing media, at 10 Gy. The 0 Gy flasks were mock irradiated by leaving them out of the radiation chamber for the same length of time. Forty-eight hours later, I washed the cells with PBS and collected by scraping. One ml of the cells was reserved in a 1.5 ml tube for transcriptomics, while the other was reserved in another 1.5 ml tube for immunopeptidomics. Cells in both tubes were pelleted at stored at -80ºC. This was repeated for a total of three biological replicates before RNA or immunopeptidome extraction.

4.2.2 RNA Extraction

RNA extraction was done exactly as in 3.2.9. Extracted RNA was split into 15 µl for short-read sequencing with Novogene Illumina platform and 15 µl for long-read sequencing in collaboration with Dr. Aaron Jeffries from the University of Exeter.

4.2.3 RNA Sequencing and Data Processing

4.1.1.4 Short reads

The short-read RNA sequencing was done exactly as in 3.2.10, up to obtaining the cleaned fastq files from Novogene, because I needed to use a pipeline that integrates with the long-read data, as explained below.

4.1.1.5 Long reads

The RNA was sent to Aaron Jeffries at the University of Exeter's sequencing facility after a collaboration agreement was signed. From his side, Aaron constructed the sequencing library according to Nanopore protocol, with cDNA and PCR amplification steps. The kit used was SQK-PCB111.24 (Oxford Nanopore Technologies), and the 12 samples were barcoded and ran 5 times (5 flow cells, 12 samples per flow cell) on PromethION (Oxford Nanopore Technologies). Flow cell model R4.9.1 (Oxford Nanopore Technologies) were used with Chemistry Kit 10. The following run settings were used: Run length, 72 hours; Active channel selection, On; Pore scan frequency, 1.5 hours; Reserved pores, On; Minimum read length, 200 bp; Read splitting, Off; Basecalling, Super-accurate basecalling 450 bps; Modified basecalling, Off; Trim barcodes, Off, Mid-read barcode filtering, Off. The following software versions were used: MinKNOW, 22.12.5; MinKNOW Core 5.4.3; Guppy (used for basecalling), 6.4.6.

When I received the FASTQ files, I proceeded with filtering and orienting them for quality control. I used Pychopper (https://github.com/epi2me-labs/pychopper) with the following command:

```
pychopper -k PCS111 -r ${report_output} -u
${unclassified_output} -w ${rescued_output} ${input_file}
${full_length_output} -t 48
```

Next, I corrected the Pychopped long reads with the short reads to improve per-base accuracy using FM-index Long Read Corrector (FMLRC) v0.1.8 (Mak et al. 2023; Wang et al. 2018b). First, I build a Burrows-Wheeler Transform (BWT) file from the cleaned FASTQ short-read files, as required by FMLRC, using msbwt2-build. Next, I used that to correct the long reads as follows:

```
fmlrc2 "$BWT_PATH" "$input_file" "$output_file" -t 24
```

Here, BWT_PATH was the path to the BWT file built from the short reads, and input_file was the long-read fastq files. This generated a list of corrected gzipped fasta files for each sample. I then aligned those reads to the transcriptome of T2T-CHM13v2.0 (downloaded from https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_009914755.1/) with the following command:

```
minimap2 -ax splice rna.fna $file -uf --secondary=no -t 24
> $out_sam_file
```

Finally, I ran Salmon in alignment-based mode, with the following command:

```
salmon quant -l SF --ont -t rna.fna -a ${sample}_mapped.sam
-o ${sample}_salmon_output -p 20
```

4.2.4 Gene-level Differential Expression Analysis

I imported and merged the Salmon output files into R with the tximport package (Soneson et al. 2016). Using the makeTxDbFromGFF() command from the GenomicFeatures package (Lawrence et al. 2013), I constructed a transcriptome annotation object (TxDb) from the genome GTF. I then performed the differential gene expression analysis using the EdgeR package (Chen et al. 2016; McCarthy et al. 2012; Robinson et al. 2010), with the glmQLFit() and glmQLFTest() commands to fit a quasi-likelihood negative binomial generalized log-linear model to the data and conduct gene-wise statistical tests, respectively. The latter command was used three times to perform the following pairwise comparisons: siCtl 10Gy to siCtl 0Gy, siSRSF1 0Gy to siCtl 0Gy, and siSRSF1 10Gy to siCtl 0Gy.

4.2.5 Isoform Switching and Alternative Splicing Analysis

I also used the Salmon output as input to the IsoformSwitchAnalyzeR R package (Vitting-Seerup and Sandelin 2017, 2019) to analyze alternative splicing and isoform switching. The following functions specifically were used: I used importIsoformExpression(), followed by imporRdata() function to import the data from Salmon (Soneson et al. 2016; Robinson and Oshlack 2010). I next used isoformSwitchAnalysisPart1() to filter out lowly-expressed genes (gene expression cutoff = 1) and isoforms (isoform fraction cutoff = 0.01), test for isoform switches with DEXSeq based on dIF $\geq \pm 0.1$ and False Discovery Rate < 0.05 (Ritchie et al. 2015; Anders et al. 2012), analyze novel isoforms (Vitting-Seerup et al. 2014), and extract corresponding nucleotide and amino acid sequences into FASTA format.

4.1.1.6 Predicting coding potential

I used the Coding Predictor Calculator 2 (CPC2) (Kang et al. 2017) to predict coding potential of the transcripts, using the nucleotide FASTA file outputted by IsoformSwitchAnalyzeR. Spcifically I used standalone CPC2 v1.0.1 and ran it with the following command:

```
CPC2.py  -i  isoformSwitchAnalyzeR_isoform_nt.fasta  -o
cpc2_out
```

4.1.1.7 Predicting intrinsically disordered regions

I used IUPred3 (Erdős et al. 2021) to predict intrinsically disordered regions within the amino acid sequence of the switching isoforms. To run, I used this command:

```
python3  iupred3.py  -a  -s  medium  $temp_file  long  >>
$output_file
```

4.1.1.8 Predicting signal peptides

To predict signal peptides and their cleavage sites, I used the webserver of SignalP 5.0 (Almagro Armenteros et al. 2019). I chopped the amino acid fasta file to 5000-sequence chunks and used them as input to the web tool. I set Organism group to Eukarya and Output format to Short output (no figures).

4.1.1.9 Predicting protein domains

To identify protein domains within switching isoforms, I used standalone Pfam (downloaded from https://ftp.ebi.ac.uk/pub/databases/Pfam/Tools/) with the following command on the amino acid fasta file:

```
pfam_scan.pl -fasta isoformSwitchAnalyzeR_isoform_AA.fasta
-dir hmmscan-pfam
```

4.1.1.10 Integrating predictions with RNA-seq data

To integrate results from the above tools, I went back to IsformSwitchAnalyzeR and ran isoformSwitchAnalysisPart2(), pointing to the results directories from each of the above tools. This command also incorporates analysing alternative splicing (Vitting-Seerup et al. 2014).

I used extractSwitchOverlap() to plot Figure 4.3A, extractConsequenceEnrichment() to plot Figure 4.3C and ggplot2 (Wickham 2016) to plot Figure 4.1A.

4.1.1.11 Determining transcript length

I used the transcriptLengths() command from the GenomicFeatures R package (Lawrence et al. 2013) to obtain all transcript lengths. I used Graphpad Prism to create the violin plot in Figure 4.3B and perform the rank-sum test to check statistical significance.

4.2.6 Immunopeptidome Extraction

I performed this with Hala Estephan from the Department of Oncology, and the work was carried out in the Nicola Ternette Lab at the University of Oxford. Columns were first cleaned by washing with 10% acetic acid, then with PBS until the pH was neutral. Protein A-Sepharose beads (GE Healthcare) were resuspended with gentle agitation, then 1 ml per sample (12 samples total) was taken into a 15 ml tube (Falcon), topped up with PBS, and centrifuged for 1 minute at 100 g to pellet the beads. The supernatant was discarded. Three mg of pan-HLA-I antibody (clone W6/32) purified from hybridoma cells (ATCC HB-95)

was then added to the bead pellet with PBS and rotated gently at 11 rpm in an orbital rotator for 1.5 hours at 4°C.

Next, the antibody-conjugated beads were washed in a column format. The PBS was allowed to flow through, then borate buffer (50 ml of solution A [0.1 M boric acid, 0.1 M potassium chloride] and 4 ml of 0.1 M NaOH, then filled up to 100 ml with High-Performance Liquid Chromatography [HPLC]-grade water) was added to fill up the column (10x the column volume) at room temperature, and the buffer was allowed to flow through.

The column was then equilibrated in fresh cross-linking solution (40 mM dimethyl pimelimidate dihydrochloride [Sigma, # D8388] in borate buffer; pH 9), added to 10x the column volume. Most of the liquid was allowed to flow through, but then topped up and incubated for 30 minutes at room temperature. The reaction was stopped by adding ice-cold 0.2 M Tris buffer (pH 8.0), and the column was equilibrated with 50 mM Tris (pH 8.0) and stored at 4°C.

The next day, the cell pellets (approximately 6.4×10^6 cells per sample) were defrosted on ice, then lysed with around 1 ml of ice-cold lysis buffer (1% IGEPAL 630, 150 mM NaCl, 50 mM Tris, pH 8.0, supplemented with 2 tablets of protease inhibitor cocktail (Roche) and 1 tablet of PhosStop phosphatase inhibitor (Roche) in 10 ml of lysis buffer). The cells were pipetted gently up and down to obtain a homogenous lysate. The lysates were incubated at 4°C in an orbital rotator to solubilize MHCs. Next, they were centrifuged at 4°C for 15 minutes at 500 g to remove nuclei (pellet) and then at > 45,000 g for 45 minutes at 4°C to pellet other insoluble material. Next, the lysates were diluted in the same volume and added to the antibody-cross-linked beads. The mixture was next incubated overnight with gentle rotation in an orbital rotator at 4°C.

The next morning, clean columns were incubated with 10% acetic acid for 20 minutes, then washed with PBS until the pH was increased to neutral. The antibody-cross-linked beads were loaded into the columns and the liquid allowed to flow through. A series of washes at ≥ 10x the column volume followed. First, wash buffers W1 (50 mM Tris, pH 8; 150 mM NaCl; 5 mM ethylenediaminetetraacetic acid [EDTA]), followed by W2 (50 mM Tris, pH 8; 150 mM NaCl), were used to remove detergents. Next, the beads were washed with wash buffer W3 (50 mM Tris, pH 8; 450 mM NaCl) to remove non-specifically bound material. Finally wash buffer W4 (50 mM Tris, pH 8) was used to remove salts.

The HLA-peptide complexes were eluted into protein LoBind tubes (Eppendorf, # 022431102) using 5x the column volume of 10% acetic acid, then dried in a centrifuge concentrator (Eppendorf). The dried material was resuspended in 750 μl of HPLC-grade water, sonicated in a water bath for 5 minutes without heat, then rotated for 10 minutes.

Next, we filtered peptides below 5 kDa in a column format. Clean columns in 2 ml collection tubes were washed with 500 μl of 10% acetic acid and centrifuged at 12,000 g until all the liquid has flown into the collection tube and discarded. We transferred ≤ 400 μl at a time of the samples to the columns and centrifuged for 30 minutes at 4ºC at 13,000 g, collecting flow-through in 2 ml Protein LoBind tubes. To ensure all peptides were collected, 20% acetonitrile was added to the original 2 ml tubes that the samples were dried in and also filtered in the column. This flow-through was combined with the previous one, and the combination was filtered through a fresh set of columns. The samples were then dried once more overnight in a centrifuge concentrator.

The next morning, we resuspended the samples with 50 μl of loading buffer (0.1% trifluoroacetic acid [TFA], 1% acetonitrile [ACN] in HPLC-grade water), sonicated for 5 minutes in water bath, then vortexed for 10 minutes at room temperature.

C18 spin tips (Pierce, #84850) were positioned in their adaptors and placed in 2 ml Protein LoBind tubes, then 20 μl of 0.1% TFA/80% ACN in water was added and the tubes centrifuged at 1200 g for 1 minute. The tips were then equilibrated by adding 20 μl of 0.1% TFA and centrifuged at 1200 g for 1 minute. Next, we transferred the tips to new LoBind tubes and added each sample to a tip, then centrifuged at 1200 g for 2-3 minutes until all the liquid had flown through. The tips were then washed twice by adding 20 μl of 0.1% TFA and centrifuging at 1200 g for 1 minute. Next, we transferred the tips to new LoBind tubes and eluted the samples twice by adding 20 μl of 0.1% TFA in 80% ACN and centrifuged for 1 minute at 1200 g. Finally, the samples were dried in a concentrator centrifuge, reconstituted in 20 μL of loading buffer, and subjected to liquid chromatography-tandem mass spectrometry analysis (LC-MS^2).

4.2.7 Liquid Chromatography - Tandem Mass Spectrometry (LC-MS^2)

The following was all done by Robert Parker from the Ternette Lab at the University of Oxford. For LC-MS^2, an Ultimate 3000 RSLCnano System paired with a PepMap C18 column (2 μm particle size, 75 μm x 50 cm) from Thermo Scientific was used, connected to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Scientific). A 60-minute linear gradient ranging from 3% to 25% acetonitrile in 5% dimethyl sulfoxide/0.1% formic acid was applied at a flow rate of 250 nL/min for peptide elution. An Easy-Spray Source was used to introduce peptide ions to the mass spectrometer at 2000 V. Full MS detection was performed at a resolution of 120,000 within a 300–1500 m/z range, with precursors

selected using TopSpeed ion selection over a 2-second cycle time and a quadrupole isolation width of 1.2 atomic mass units. MS^2 acquisition was set at a resolution of 30,000, with high-energy collisional dissociation energy tailored for peptides based on their charges.

A nanoElute system (Bruker Daltonics) coupled online to a Trapped Ion Mobility Spectrometry Time-of-Flight (timsTOF) SCP (Bruker Daltonics) was utilized. Samples were loaded onto an Aurora C18 column (25 cm x 75 μm, 1.7 μm, IonOpticks, Australia) using 0.1% formic acid. Peptide separation occurred over 66 minutes using a linear gradient of acetonitrile in acetic acid at 50°C and a flow rate of 150 nL/min. Electrospray ionization conditions employed a CaptiveSpray source (Bruker Daltonics) with specific temperature and voltage settings. Mass spectrometric analysis was conducted in a data-dependent PASEF mode, with ion mobility $1/K0 = 1.7 - 0.7$ Vs cm^2 and mass-to-charge $100 - 1,700$. Ion accumulation and ramp time were set to 166 milliseconds. Optimized collision energies were applied 70 eV at $1/K0 = 1.7$ Vs cm^{-2}; 40 eV at $1/K0 = 1.34$ Vs cm^{-2}; 40 eV at $1/K0 = 1.1$ VS Vs cm^{-2}; 30 eV at $1/K0 = 1.06$ Vs cm^{-2}; 20 eV at $1/K0 = 0.7$ Vs cm^{-2}.

$LC-MS^2$ datasets were analyzed using Peaks v10.6 software. The precursor mass tolerance was set to 20 ppm and the fragment mass tolerance to 0.05 Da. For mass spectrometric analysis, spectral sequence annotation was controlled using a decoy database search build-in with the Peaks software (Bioinformatics Solutions). The probability scores of each spectral annotation was calculated using a linear discriminative function (LDF), which was converted into a P-value that represents the probability of the annotation to be a non-random assignment. For example, a given score cut-off of $-10\log_{10}(P) = 15$ is equivalent to a p value of $p = 0.032$, and $10\log_{10}(P) = 20$ is equivalent to a p value of $p = 0.01$.

4.2.8 Immunopeptidomics Data Processing

Using the FragPipe output, 'combined_modified_peptide.tsv,' which contained 20,664 peptides, I first normalized the peptide intensities using NormalyzerDE (Willforss et al. 2019). This allowed me to test different popular normalization approaches and select the best one for the data. In this case, I selected Variance Stabilization Normalization (VSN) because it reduced intragroup Percent Coefficient of Variation (PCV), Percent Median Absolute Deviation (PMAD), and Percent Explained Variance (PEV). This meant that within each group, the data points varied less in relation to their mean (PCV), were more consistently close to their median value (PMAD), and had less variance attributed to random or unexplained factors (PEV). The NormalyzerDE report also allowed to me to identify the third biological replicate of siSRSF1 0Gy as an outlier (Figure 4.4A and B), so I excluded it from further analyses.

I then used Perseus to import the VSN-normalized intensity data, which I filtered to include only peptides that are quantified (not NA or 0) in at least 2 biological replicates in at least one condition. This left 10,816 peptides.

To test the viability of imputing the missing values, I tested multiple imputation parameters from the 'Replace missing values from normal distribution' imputation tool in Perseus. Each time I tested a certain combination of parameters, I checked the histogram of the data to visually confirm that it still conformed to a normal distribution. The combination that worked best was using width = 0.8 and down shift = 1.5. However, although the histogram looked neatly normally distributed, the inter-replicate correlation was severely reduced, as shown in Figure 4.4C. Therefore, I decided to proceed without imputation.

I next performed a one-way ANOVA on the data to identify differentially presented peptides. To do this in Perseus, I selected 'Multiple-sample tests' from the 'Tests' menu, and used the following default parameters: Grouping, treatment; Test, ANOVA; S0, 0; Use for truncation, Permutation-based FDR; FDR, 0.05; Report q-value, checked; Number of randomizations, 250; Preserve grouping in randomizations, <None>; Log10, checked; Suffix, empty; Write residuals, unchecked. This identified four differentially presented peptides, shown in Figure 4.5A. Which I then imported into Graphpad Prism and performed a mixed effects ANOVA on, followed by Šídák's multiple comparisons test, as recommended by the software, based on the missing values and distribution of the data.

4.2.9 Determining Potential Neoantigenicity and Novelty

To determine which antigens are potentially neoantigens (cancer-specific), and, further, which of those potential neoantigens is novel (not previously discovered in studies), I manually inserted each peptide into caAtlas, which is a database of antigens in normal and cancer cells/tissues (Yi et al. 2021) and HLA Ligand Atlas (Marcu et al. 2021), an antigen database specifically for normal cells/tissues. If, based on looking up the peptide in both databases, it was not found in normal samples, I considered it a potential neoantigen. If the peptide was also not found in any cancer samples, I considered it potentially novel.

4.2.10 Integrating the Transcriptome and Immunopeptidome

The purpose of this analysis was to identify unique (i.e. only present in one condition and absent in others) sequences from upregulated isoforms in the NRA-seq data and cross-check whether they resulted in MHC I-presented peptides in the immunopeptidomics data.

First, by using IsofromSwitchAnalyzeR in 4.2.5, I was able to extract a table that has all upregulated isoforms (dIF ≥ ±0.1 and False Discovery Rate < 0.05) across different conditions and the genes they correspond to, as well as a table that has all expressed isoforms (gene expression cutoff = 1, isoform fraction cutoff = 0.01) and the genes they correspond to. By definition, the second file includes all the data in the first file, but I kept them separate to mark the upregulated isoforms.

Next, I needed the amino acid sequences for all the expressed isoforms, so I extracted them using extractSequence() in IsoformSwitchAnalyzeR with the following arguments: onlySwitchingGenes = FALSE, extractNTseq = FALSE, extractAAseq = TRUE, removeShortAAseq = FALSE, removeLongAAseq = FALSE, alsoSplitFastaFile = FALSE, removeORFwithStop=FALSE, writeToFile = TRUE, which produced a expressed_isoforms.fasta file with the translated amino acid sequences of the isoforms.

I then wrote a Python script to create three new FASTA files, one per condition, for isoforms upregulated in each condition: siSRSF1 0Gy, siSRSF1 10Gy, and both (compared to siCtl 0Gy). For each gene associated with an upregulated isoform (provided the upregulated isoform has an amino acid sequence in sequence in expressed_isoforms.fasta), the script identified all isoforms of that gene within the expressed isoform dataset, cross-referencing the tables extracted from IsoformSwitchAnalyzeR. It then extracted their amino acid sequences from the expressed_isoforms.fasta. All this information was outputted into a CSV file (one per FASTA file) with the following columns: isoform_id_up (the upregulated isoform), gene_id (the gene it corresponds to), isoform_id_all (expressed isoforms of that gene, each in a row), sequence (amino acid sequence of the isoform in isoform_id_all). In addition, the script produced a file based on the last two columns of the CSV file, where the

headers are the isoform in isoform_id_all and the sequences are those associated with them in the final column.

Afterwards, I wrote a shell script that ran the multiple sequence alignment tool CLUSTAL-OMEGA (Sievers et al. 2011) iteratively for each gene containing an upregulated isoform with the other expressed isoforms in Clustal format. The script was run three times, once for each condition. For each condition, it produced one Clustal file per gene with the alignment of the upregulated isoform against the other expressed isoforms. The next task was to automate going through each Clustal file to identify any sequence(s) within the upregulated isoforms that are unique to that isoform and not present in any other expressed isoform of the same gene.

To do this, I used Python, defining a unique sequence as a stretch of 5 or more consecutive amino acids that are only found or are only deleted in the upregulated isoform and not any other expressed isoform, and attached the flanking 15 amino acids up- and downstream of this stretch. These unique sequences (with the flanking regions) were constructed into one FASTA file, keeping track of isoforms upregulated in siSRSF1 0Gy, in siSRSF1 10Gy, or in both, compared to siCtl 0Gy. Finally, this FASTA file was concatenated with the human protein database FASTA and the database search was run on the concatenated file, allowing peptides that match these unique sequences to be identified.

4.2.11 MHC Motif Analysis

To generate sequence logos for SRSF1 depletion, all peptides exclusively presented in siSRSF1 0Gy, siSRSF1 10Gy, and in both at the same time, were used as input for MHC Motif Decon 1.0 (Kaabinejadian et al. 2022), which was run with default parameters after

specifying the HLA alleles expressed in MDA-MB-231 cells (HLA-A02:17, HLA-A02:01, HLA-B40:02, HLA-B41:01, HLA-C02:02, HLA-C17:01). To generate the "typical" sequence logos, all 10,816 peptides quantified in at least 2 biological replicates of at least 1 condition were used as input for MHC Motif Decon 1.0 with the same parameters as above.

4.2.12 Gene Ontology Analysis

I carried out functional enrichment analysis with Gene Ontology terms using WebGestalt (Liao et al. 2019), with the following options: Method of Interest: Over-Representation Analysis (ORA), Functional Database: Gene Ontology (Biological Process or Biological Process noRedundant), and a list of all expressed genes as a reference set. For each analysis, the Weighted set cover was used to assess the most significantly enriched terms while maximizing gene coverage. The FDR threshold was set to 0.05.

4.3 Results

4.3.1 Isoform-level Changes are Not Necessarily Captured at the Gene-level

To understand the effects of radiation and SRSF1 knockdown on triple-negative breast cancer (TNBC), I sequenced cDNA from MDA-MB-231 cells (TNBC model) 48 hours after 10 Gy IR or no IR and with or without SRSF1 knockdown (n = 3 biological replicates per condition), and analysed gene- and isoform-level expression changes across conditions. Interestingly, only 69 genes were captured at both the gene and isoform levels (Figure 4.1A). Moreover, I found no correlation at all between gene-level differential expression (which sums up the expression of all isoforms for a gene) and isoform-level differential expression (i.e. the occurrence of isoform switching between isoforms of the gene) (Figure 4.1B). These

results demonstrate the importance of incorporating isoform-level information in any RNA/cDNA sequencing experiment.
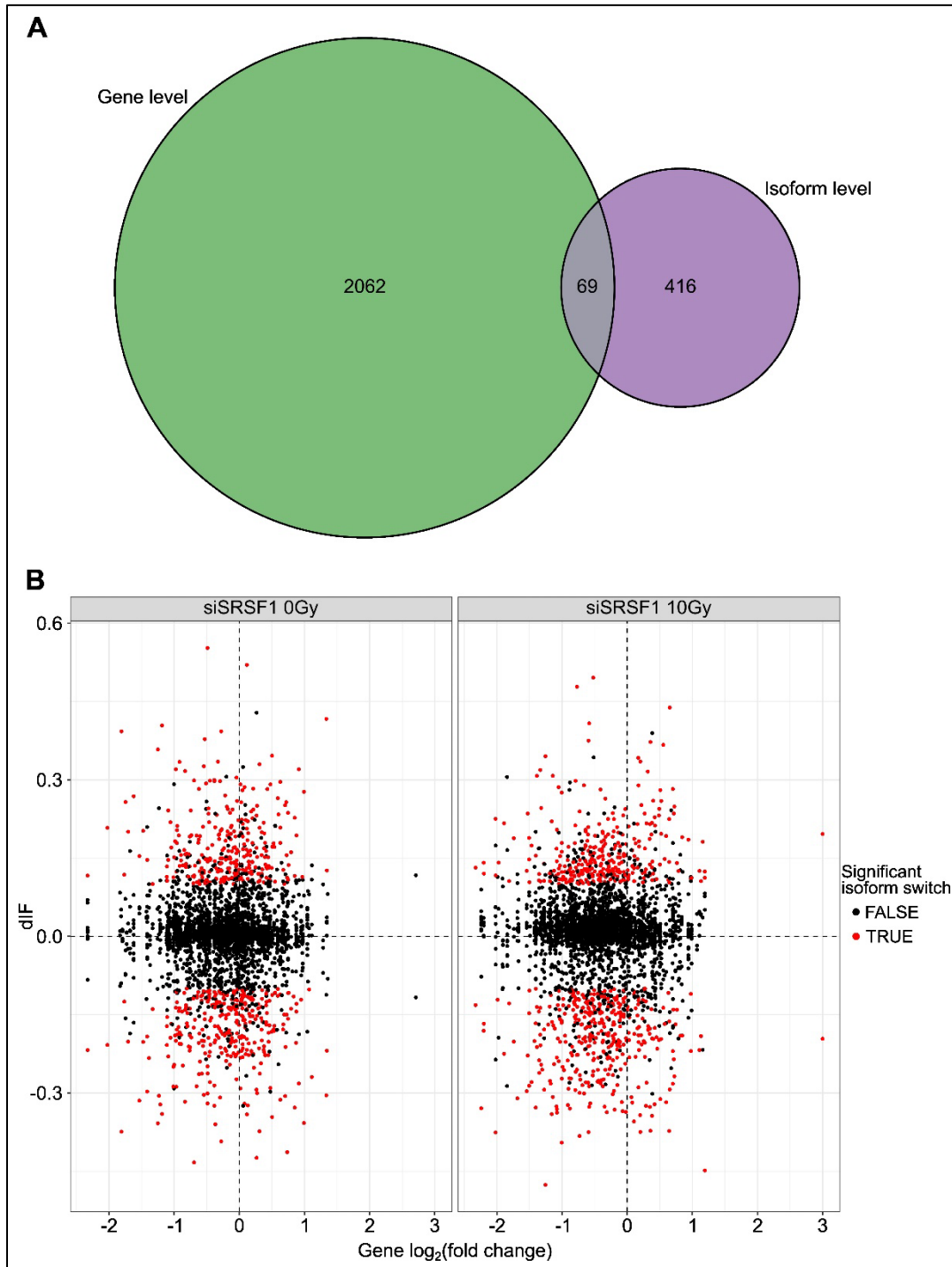
Figure 4.1 Analysis of isoform- and gene-level expression changes in response to SRSF1 knockdown and IR. **A)** Scatter plot of isoform level differential expression (y-axis)

against gene-level differential expression (x-axis) in response to knockdown alone (left panel) or knockdown and 10 Gy IR (right panel), compared to the control (siControl 0Gy). The timepoint used was 48 hours after IR or mock IR. Each dot represents one gene. dIF, differential isoform fraction; significance was defined as dIF ≥ ±0.1 and False Discovery Rate < 0.05. **B)** Venn diagram of genes that are differentially expressed at the gene level (green) and genes with differentially expressed isoforms (purple) in response to IR and or SRSF1 knockdown.

4.3.2 SRSF1 Knockdown Combined with Ionising Radiation Affects the Expression of Nearly 2,000 Genes

I investigated the gene-level response to SRSF1 knockdown and IR, and identified 2,232 differentially expressed genes. Although radiation alone significantly ($\log_2$(fold change) ≥ ±1 and FDR < 0.05) affected the expression of 117 genes, knockdown of SRSF1 had a much larger effect, significantly altering the expression of 800 genes (Table 4.1). The combination of the two, however, had the biggest impact on gene expression, affecting 1,993 genes (Table 4.1). This illustrates how combining the two treatments is larger than the sum of their individual effects (by over 1,000 genes), indicating the effect of combining SRSF1 depletion with IR is not merely additive. KD alone and KD + IR both upregulated genes related to similar biological processes, such as response to stress, immune response (e.g. response to cytokine and response to bacterium), and apoptosis (Figure 4.2). There were no enriched Gene Ontology terms in the genes upregulated after IR alone compared to the control (no IR).

Table 4.1 The number of differentially expressed genes across three comparisons of four conditions. Differentially expressed genes are defined as having a log2(fold change) of ±1 and adjusted p-value < 0.05. IR, ionising radiation; KD, knockdown; FC, fold change.

Effect being examined	IR alone	KD alone	IR + KD	Total
Comparison	siControl 0 Gy vs. siControl 10 Gy	siControl 0 Gy vs. siSRSF1 0 Gy	siControl 0 Gy vs. siSRSF1 10 Gy	
Overall	117	800	1933	2131
Upregulated (range $\log_2$(FC))	109 (1.00 – 5.47)	273 (1.00 – 5.73)	1138 (1.00 – 6.35)	1229
Downregulated (range log2(FC))	8 (-1.73 – -1.00)	527 (-6.55 – -1.00)	795 (-6.23 – -1.00)	905
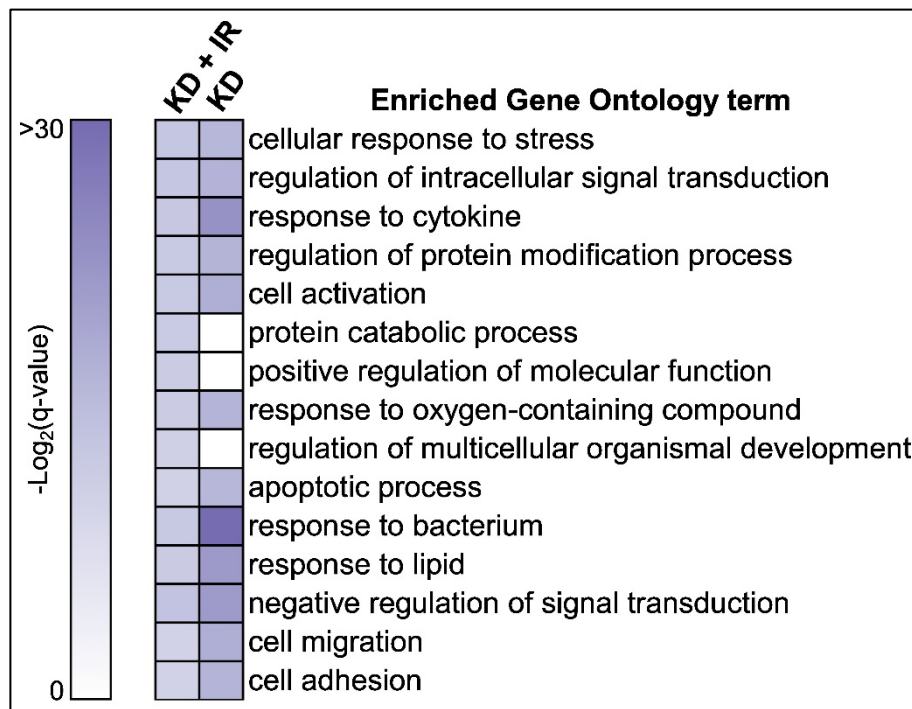
Figure 4.2 Heatmap of the topmost significant Gene Ontology Biological Process terms associated with genes upregulated at siSRSF1 10 Gy (IR + KD) and at siSRSF1 0 Gy (KD) relative to siControl 0 Gy. Color gradient represents the significance based on Benjamini–Hochberg-corrected p-value. KD, knockdown; IR, ionising radiation.

4.3.3 Radiation Alone Does Not Affect Isoform Switching in Triple Negative Breast Cancer Cells

I next sought to understand the unique effects of IR alone, knockdown alone, and the combination of the two on isoform switching in TNBC cells. To my surprise, only one isoform was differentially expressed in response to IR alone, which was the long non-coding RNA *LOC101927060* (Table 4.2). Notably, *SRSF1* levels in these cells are not reduced 48 hours after IR. Artificially reducing them with siRNA knockdown, on the other hand, had a significant impact on the number of switching isoforms, as 619 isoforms switched expression between siControl and siSRSF1 at 0 Gy (Table 4.2). When knockdown and IR were combined, this led, as anticipated, to the largest number of isoform switches (670 isoforms) (Table 4.2). Although it is tempting to assume that all 619 isoforms switching under knockdown alone are also switching under the combination of knockdown and IR, 30% (188/619) of them are in fact unique (Figure 4.3A).

Table 4.2 The number of differentially expressed isoforms across three treatments. Differentially expressed isoforms must have differential isoform fraction (dIF) $\geq \pm 1.0$ and adjusted p-value < 0.05. IR, ionising radiation; KD, knockdown.

Effect being examined	IR alone	KD alone	IR + KD	Total
Comparison	siControl 0 Gy vs. siControl 10 Gy	siControl 0 Gy vs. siSRSF1 0 Gy	siControl 0 Gy vs. siSRSF1 10 Gy	
Overall	1	619	670	1290
Upregulated (range dIF)	0	293 (0.10 – 0.55)	304 (0.10 – 0.50)	397
Downregulated (range dIF)	1 (-0.20)	326 (-0.43 – -0.10)	366 (-0.48 – -0.10)	419
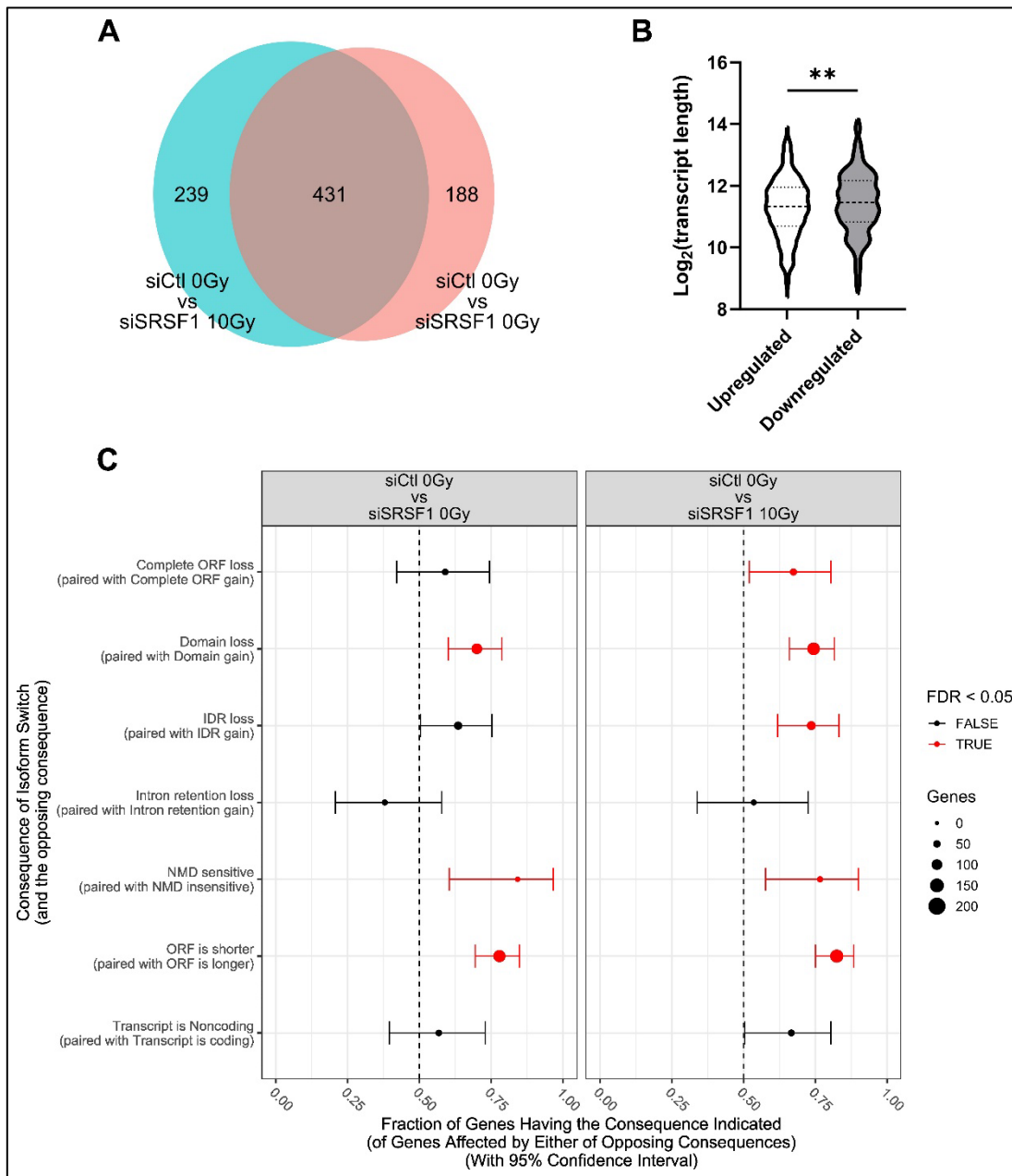
Figure 4.3 Overview of the effect of SRSF1 depletion on isoform switching. **A)** Overlap of differentially expressed isoforms after knockdown alone (siCtl 0Gy vs. siSRSF1 0Gy) and after knockdown plus 10 Gy ionising radiation (siCtl 0Gy vs. siSRSF1 10Gy). **B)** Length (in $\log_2$(number of nucleotides)) distribution of transcripts upregulated (n = 597 isoforms) or downregulated after siSRSF1 knockdown with or without IR. Dashed line is the median, dotted lines are upper and lower quartiles. ** indicates $p < 0.01$. **C)** Enrichment or depletion of consequences due to isoform switches between siCtl and siSRSF1 without (left) or with (right) 10 Gy of ionising radiation. The x-axis represents the proportion of switches leading

to the consequence listed on the y-axis. Dashed line denotes the threshold of no enrichment or depletion. Color coding signifies whether the false discovery rate (FDR) is < 0.05. Error bars represent 95% confidence interval. siCtl, non-targeting control-transfected cells; ORF, open reading frame; IDR, intrinsically disordered region; NMD, nonsense-mediated decay.

Similarly to what I observed in the previous chapter (Figure 2.2 A and B and Figure 2.3C), I found that SRSF1 depletion leads to the generation of significantly shorter transcripts (Figure 4.3B). While the average length of upregulated isoforms is 3,102 nucleotides, downregulated isoforms are about 12% longer, at 3,510 nucleotides on average (p < 0.01; rank-sum test). The reduction in transcript length after SRSF1 knockdown is associated with 1) the loss of domain-encoding regions in transcripts, 2) the production of NMD-sensitive transcripts, and 3) shorter open reading frames (Figure 4.3C; left), which suggests these are the mechanisms leading to shorter transcripts. However, additional experiments to confirm this will be required. The addition of IR further led to the complete loss of open reading frames, in addition to loss of intrinsically disordered regions (Figure 4.3C; right).

I questioned whether the resultant proteins might harbor unique sequences that form neoantigens. To check this, I aligned the sequences of the upregulated isoforms in each condition (KD alone + KD + IR + both) to the sequences of other expressed isoforms with the Clustal-Omega sequence alignment tool and extracted the sequences unique to the upregulated isoforms using Python (see Materials and Methods 4.2.10 for detailed steps). The total number of unique putative amino acid sequences was 37, all from different genes. Seven of them were upregulated at KD alone, 14 at KD + IR, and 16 were upregulated in both, compared to the control (siControl 0Gy). This was a promising start, but to check if any of these transcripts produce neoantigens, I had to perform immunopeptidomics profiling.

4.3.4 Differentially Presented Peptides After IR or SRSF1 Knockdown

I profiled the immunopeptidome of the same cells on which the RNA-seq was performed. This was done by pulling down MHC-I complexes, purifying the peptides bound to them, and analyzing their sequences using mass liquid chromatography coupled to tandem mass spectrometry. From this experiment, 20,664 peptides were identified. The third and final replicate of the knockdown (KD) alone condition was excluded from the analyses because it was an outlier, with the lowest intensity values (Figure 4.4A), the largest number of missing values (Figure 4.4B), and clustering very far from its other replicates, as well as other samples (Figure 4.4C). I then filtered the data to peptides that were quantified at least twice in at least one condition, which left 10,816 peptides. As expected, peptide lengths were consistent with the length distribution of MHC-I antigens, with the majority of peptides 9 amino acids long (Figure 4.4D).
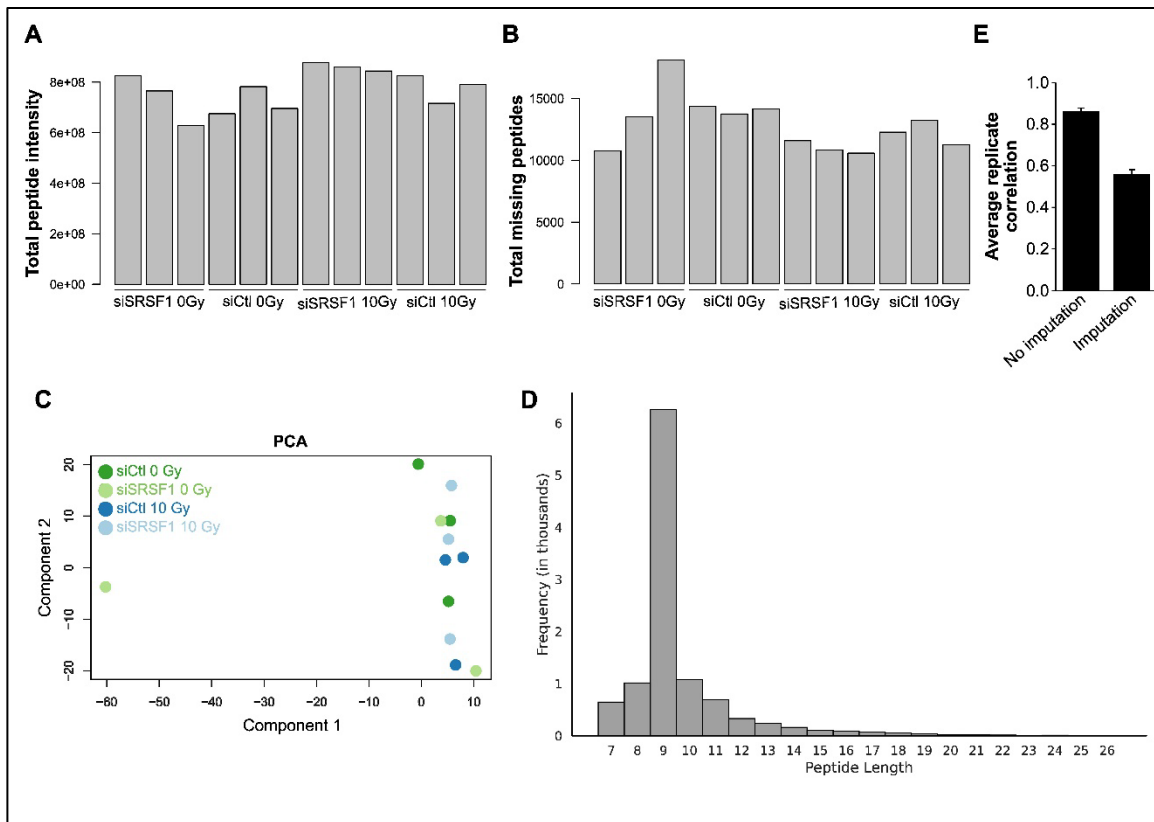
Figure 4.4 Quality control of immunopeptidomics samples. **A)** Summed peptide intensity values for each sample. **B)** The total number of missing values in each sample. **C)** Principal component analysis (PCA) plot showing the clustering of samples in two dimensions. **D)** Histogram of quantified (intensity > 0 in at least 2 replicates of at least 1 condition) peptide lengths. **E)** Average correlation across biological replicates of the same condition with or without imputation.

Due to the large number of missing values, I was not able to impute them without significantly reducing the correlation between biological replicates of the same sample from > 0.8 to < 0.6 (Figure 4.4E). Therefore, I performed ANOVA on the data as is, and identified 4 differentially presented peptides from the following proteins: DAPLE, CAN2, MYEF2, and DNLI1.

The peptides from DAPLE, MYEF2, and DNLI1 have previously been found in both normal and cancerous tissues. However, the peptide from CAN2, AEISAFEL, has not been

profiled as an antigen before in either type of tissue, according to the Cancer Antigen Atlas (caAtlas) and the HLA Ligand Atlas, even though the sequence is part of the canonical protein. This peptide is significantly increased on the MHC complex after SRSF1 knockdown, but only at the peptide level, as opposed to the RNA level (*CAPN2* gene), where it does not change expression or exhibit isoform switching. This indicates that SRSF1 knockdown indirectly increases the production and presentation of this antigen.

However, these four peptides are not the whole picture. What about peptides that were quantified in all replicates of specific conditions, but are completely absent in all other conditions/replicates? To identify those, I performed a qualitative analysis and the results are shown in Figure 4.5B and Appendix

Table 7.3.
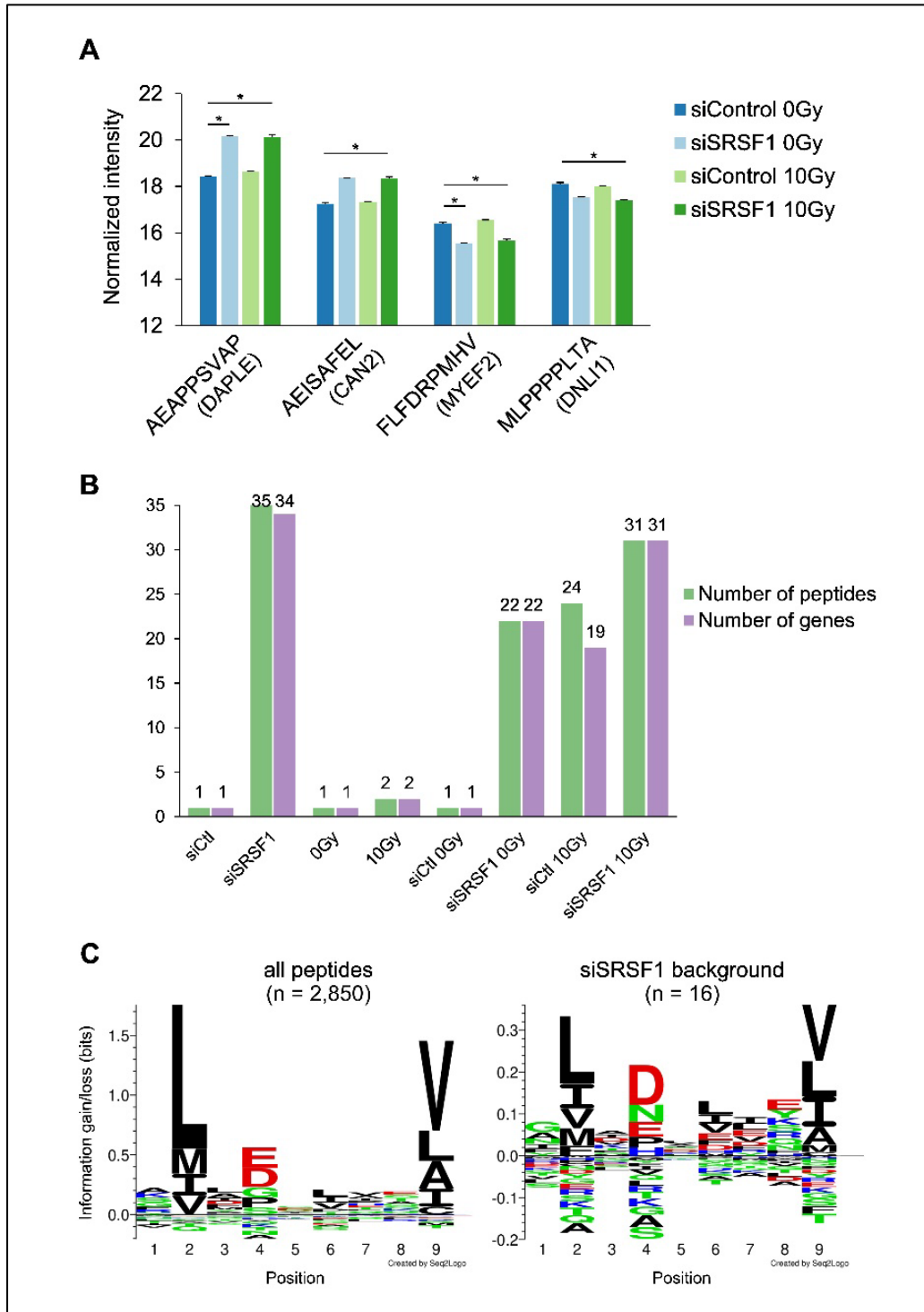
Figure 4.5 Overview of differentially presented peptides in immunopeptidomics data.
A) Normalized intensity values of differentially presented peptides, with the protein name in parenthesis beneath the sequence. * indicates p < 0.05 from Šídák's multiple comparisons test after mixed effects analysis. Error bars depict the standard error of the mean from 2-3

biological replicates. **B)** Histogram of the number of peptides, and the genes they are derived from, that are presented exclusively in the conditions specified. siCtl represents peptides presented in the siControl condition with and without radiation; siSRSF1 represents peptides presented in the knockdown condition with and without radiation; 0Gy represents peptides presented without radiation with and without knockdown; 10 Gy represents peptides presented after radiation with and without knockdown. **C)** HLA-A*02:01 allele binding motifs from all expressed peptides (left) or after SRSF1 knockdown (right). A peptide is defined as expressed if its intensity is > 0 in ≥ 2 replicates of ≥ 1 condition.

The qualitative analysis reveals that SRSF1 knockdown produces 88 unique antigens in total, almost all from different genes. Many (35/88) of these peptides are presented regardless of radiation. However, a close number (31/88) are presented only after 10 Gy IR in the knockdown background. The effect of radiation alone is not to be underestimated either, as it generates 24 antigens from 19 genes that are not presented in any other condition. However, only 2 peptides are presented in both the siControl and the siSRSF1 backgrounds, indicating that IR produces starkly different antigens, depending on the genetic background.

In total, there are 117 antigens exclusively presented in some conditions compared to others (Figure 4.5B and Appendix

Table 7.3). I looked each of those up in caAtlas and the HLA Ligand Atlas, and found that 74% (86/117) of them are potential neoantigens, not documented in any normal cells. Of those, 61 seem to be novel, not documented in cancerous cells either (Appendix

Table 7.3).

Interestingly, antigens presented after SRSF1 knockdown seem to have unique sequence specificities when binding to HLA-A*02:01. The peptides that bind HLA-A*02:01 typically display specific sequence characteristics. For example, they are usually 8-25 amino acids

long, but mostly 9 (Figure 4.4D). In addition, the HLA-A*02:01 binding pocket requires two positional anchors, a leucine or methionine in position 2 and an isoleucine, leucine, or valine in position 9 (Vita et al. 2019). I found that position 4 in peptides presented after SRSF1 knockdown favours glutamic acid (E), aspartic acid (D), and asparagine (N). Although the presence of acidic residues (E and D) is normal in this position (as shown in the left panel of Figure 4.5C), the presence of an asparagine residue is uncommon (Vita et al. 2019). In fact, asparagine is underrepresented when all expressed peptides are considered (Figure 4.5C, left panel). However, this result must be interpreted with caution, given that only 16 peptides contributed to the motif.

4.3.5 Integration of the Transcriptome and the Immunopeptidome

Of the 2,559 genes that are differentially expressed at either the isoform or gene level, only 605 (~24%) generated antigens (Figure 4.6A), which indicates that antigen presentation is not necessarily correlated with gene expression. To further test that, I examined the fold change of 605 presented antigens that are derived from differentially expressed genes against the fold change of these genes. Indeed, the fold changes were not correlated at all (Figure 4.6B). The 605 overlapping genes were clearly important to the response to radiation and stress in general, though, as they were significantly enriched for Gene Ontology terms like response to stress, cell cycle, and apoptotic process.
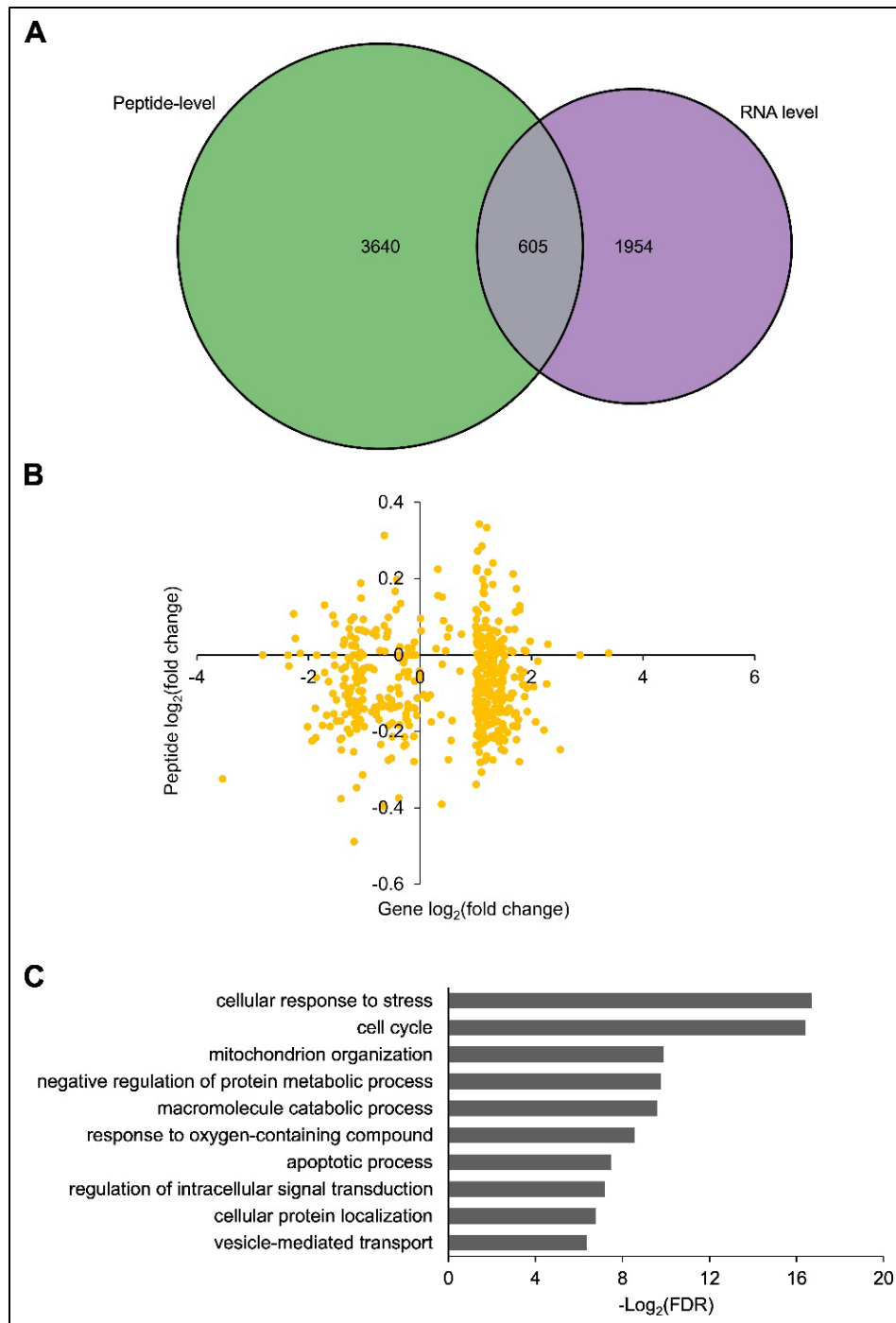
Figure 4.6 Comparison of gene- and peptide-level analyses. **A)** Venn diagram of peptides presented on the MHC-I complex and genes that are differentially expressed at the gene or isoform level. **B)** Scatter plot of the fold change in peptide presentation (y-axis) against the fold change in gene expression (y-axis) among the overlapping set from A. The fold change is in siSRSF1 10Gy relative to siControl 0Gy. **C)** Gene Ontology Biological Process top terms for the 605 overlapping genes from A. FDR, False Discovery Rate.

I then returned to my original question of which peptides are derived from unique sequences in upregulated isoforms (see last paragraph of Section 4.3.3). Of the 37 putative amino acid sequences unique to upregulated transcripts after KD alone or KD + IR, or both, 4 were discovered at the peptide level presented on the MHC-I complex. However, two of these were only quantified in 2 of the 11 samples, so I excluded them. The remaining two peptides are FVYENPISL and REKDDDVVSL (Figure 4.7A), derived from the RHBDD2 and ARHG8 proteins, respectively (corresponding to *RHBDD2* and *NET1* genes; Figure 4.7C). Interestingly, although the transcripts these peptides are derived from significantly change expression after SRSF1 knockdown and/or IR, the peptides are presented at similar levels (Figure 5.8B). According to caAtlas, FVYENPISL is presented in many cancers, including glioblastoma, melanoma, lung cancer, and leukemia. However, according to the HLA Ligand Atlas, this antigen is also present in many normal tissues, namely the adrenal gland, lung, small intestine, spleen, and thymus. REKDDDVVSL, however, has only been identified in cancerous tissues, such as leukemia and breast cancer, and EBV-transformed lymphoblastoid cell lines.
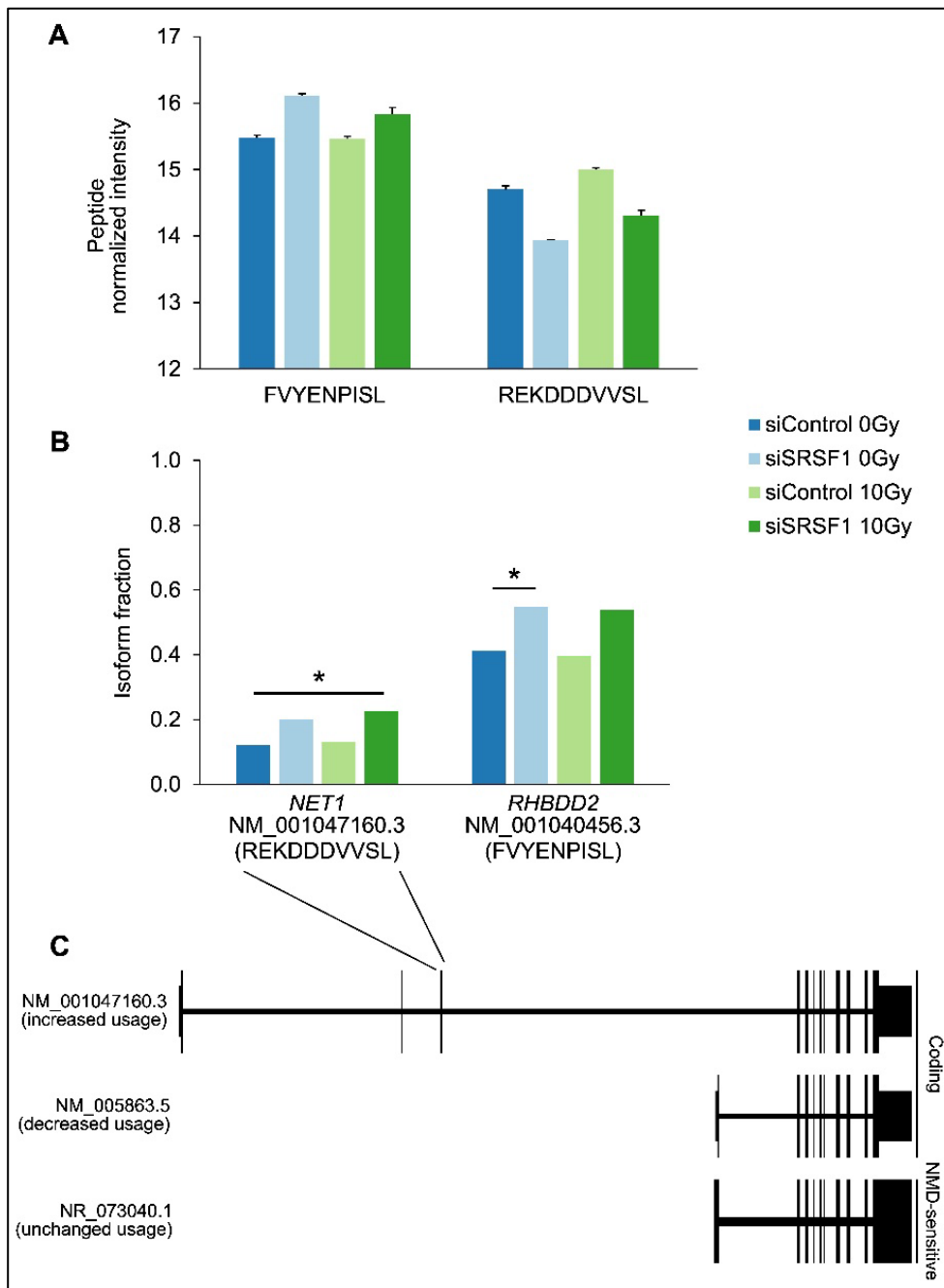
Figure 4.7 Expression of two peptides **(A)** derived from unique sequences in upregulated isoforms, and the isoforms they are derived from **(B)**. **C)** Schematic of the isoforms of NET1, with inset highlighting the unique region from which the potential neoantigen is derived. NMD, nonsense-mediated decay. * indicates $p < 0.05$. Error bars depict the standard error of the mean from 2-3 biological replicates.

4.4 Discussion

Isoform-level analysis provides a nuanced view of gene expression, revealing variations in isoforms that might be missed in traditional gene-level analysis. This approach is crucial in understanding the complex molecular mechanisms underpinning diseases like cancer. I observed a significant discrepancy between gene-level and isoform-level differential expression in TNBC cells (Figure 4.1A and B), indicating the critical role of isoform-level analysis in capturing the complete picture of gene expression regulation. Yet, only 11% of randomly selected RNA-seq studies perform isoform-level analysis (Vitting-Seerup and Sandelin 2017).

A study published in Cancer Cell found that, compared to normal samples, tumours can have 30% more alternative splicing events (Kahles et al. 2018), and that there are more neojunctions (i.e. exon-exon junctions unique to tumours)-derived peptides per tumour than there are single nucleotide variant-derived peptides, emphasizing the importance of alternative splicing in developing immunotherapeutics for cancer. I identified at least one neoantigen derived specifically from a unique sequence in an upregulated isoform of *NET1* after SRSF1 knockdown that was not present in any other condition (Figure 4.7B) and has not been reported in any other normal tissue. Even though it is only one antigen, this number is consistent with the study above, which found an average of 1.7 neojunctions per tumour that are predicted to bind MHC-I.

NET1 is overexpressed in several tumours, including breast cancer (Lahiff et al. 2014). Incidentally, *RHBDD2*, from which I identified an antigen derived from a unique sequence upregulated after SRSF1 KD + IR, is also overexpressed in breast cancer (Abba et al. 2009; Canzoneri et al. 2014). This makes it tempting to assume that overexpressed genes generate

antigens, but this was not the case when I investigated further. Firstly, most differentially expressed genes did not generate detectable antigens, only 24% did (Figure 4.6A). Secondly, even the expression of those 24% was not correlated at all with their presentation (Figure 4.6B). This comes as no surprise, as other studies in the field demonstrated similar results, finding no or modest correlation between the immunopeptidome and the transcriptome (Stopfer et al. 2020; Shraibman et al. 2016; Weinzierl et al. 2007; Kubiniok et al. 2022) and even between the immunopeptidome and proteome (Bourdetsky et al. 2014; Shraibman et al. 2016; Milner et al. 2006), although the expression of a gene or a protein is modestly correlated with the number of peptides presented from it (Chong et al. 2020; Kubiniok et al. 2022; Bassani-Sternberg et al. 2015).

In fact, for the peptides identified from *NET1* and *RHBDD2*, although the RNA sequences they were derived from were only detectable after SRSF1 KD or KD + IR, the peptides were presented at similar levels in all conditions (Figure 5.8A). This indicates posttranslational stability: their transcripts may have been very transiently expressed (so they were not detected), while the peptides were stable and persisted.

A 10 Gy dose of IR did not have a big impact on alternative splicing, compared to SRSF1 KD. This could reflect the radioresistance of triple-negative breast cancer, and MDA-MB-231 cells in particular, which needed a dose of 23 Gy to inhibit colony forming, while 9 Gy of IR actually promoted the formation of radioresistant colonies (Bravatà et al. 2019). On the other hand, combining SRSF1 KD and IR had the largest effect on gene and isoform differential expression (Table 4.1 and Table 4.2), and likewise generated the largest number of antigens not detected in any other condition (Figure 4.5B). This could be because of breast cancer cells' reliance on SRSF1; inhibition of SRSF1 methylation suppresses breast

tumourigenesis (Wj et al. 2023). In addition, its expression is correlated with tumour grade and a poor prognosis (Du et al. 2021). Therefore, decreasing SRSF1 might partially restore the normal splicing landscape, sensitizing the cells to radiation and to immunotherapy, as I observed in Chapter 3.2. It would be worthwhile in future work to test for synergism between SRSF1 depletion and IR based on *in vivo* data (e.g. xenograft models) and rigorous statistical analyses (Duarte and Vale 2022).

Strikingly, although only two RNA sequences uniquely upregulated in SRSF1 knockdown conditions were found to generate antigens, SRSF1 KD was associated with the presentation of 88 peptides from 87 genes. These antigens were not presented in any other condition (not detectable in any replicate of any other condition). SRSF1 is known to have functions beyond alternative splicing, but they have always centred around RNA processing and stability. However, this indicates SRSF1 may function in posttranslational stability as well.

The most important finding in this project is the identification of 86 putative neoantigens exclusively upregulated after SRSF1 KD and/or IR, most (61/86) of which are novel, and most (70/86) of which are derived after SRSF1 KD (Appendix

Table 7.3 and Figure 4.5B). These putative neoantigens must be validated and their immunogenicity tested of course, but if confirmed, they highlight the potential of utilising these antigens in the development of bispecific antibody, vaccine, or CAR-T therapy, especially the ones upregulated after 10 Gy IR (with or without SRSF1 KD).

Important limitations to this work are the fact that I performed it in only one cell line, with one timepoint and one dose only. Further investigations across time, dose, and cell type

are warranted, followed by deeper interrogations in xenograft mouse models, before the clinical potential of this work is confirmed.

In summary, this chapter contributes to our understanding of gene and isoform expression regulation in TNBC. The discovery of unique neoantigens following SRSF1 KD and/or IR open new avenues for immunotherapeutic approaches in this aggressive cancer.

Chapter 5   General Discussion

5.1 Summary and Key Findings

I have investigated the genome-wide RNA isoform switching response to ionising radiation (IR) and its implications for cancer therapy. This entailed investigating in detail the complexities of alternatively spliced-isoform switching and its role in the cellular response to DNA damage (Chapter 2), the significance of the splicing factor SRSF1 in radio- and immunotherapy resistance (Chapter 3), and the generation of neoantigens in triple-negative breast cancer through SRSF1 depletion and/or IR (Chapter 4). My investigations have given insights into the molecular mechanisms underpinning oncogenesis and therapeutic resistance, and underscore the importance of isoform-level analysis in cancer research. This General Discussion gives an overview of my findings, highlighting their collective contribution to the field of precision oncology, while also critically evaluating the broader implications and potential future directions that arise from this work.

5.2 Mechanisms of the Response to IR

5.2.1 Genome-wide Isoform Switching

The work described in Chapter 3 indicates that IR triggers shifts in the isoform usage (isoform switching) of nearly 2,000 isoforms in immortalized B cells from 10 unrelated, seemingly healthy individuals. This isoform switching response occurs mostly by two hours after IR, but about 200 isoforms did not switch expression until 6 hours after IR.

In contrast, the work outlined in Chapter 4 indicates that the same IR dose (10 Gy) alone has practically no impact on isoform switching in 3 biological replicates of the triple-

negative breast cancer cell line, MDA-MB-231. This could at least partly be due to technical reasons, such as use of different tools to quantify isoform switching (MISO vs. DEXSeq within IsoformSwitchAnalyzeR; see 2.2.5 and 4.2.5, respectively), which is worth testing in the future. However, a number of biological reasons could underlie, including differences in the cell types and the timepoints used. In the B cells, relatively early timepoints were used: 2 and 6 hours after IR, while for the MDA-MB-231 cells, samples were analysed 48 hours post-IR. Given that between 2 and 6 hours post-IR in B cells, there was already a huge reduction in the number of isoform-switching events (Figure 2.4), it is perhaps not surprising that by 48 hours, isoform switching would be minimal.

In terms of cell type, the response to IR is highly cell type-specific; some cells are more resistant than others. For example, different lymphocytes have different radiosensitivities, with B cells being most sensitive (Heylmann et al. 2021; Paganetti 2023). Even within B cells, naïve B cells are more sensitive than activated B cells (Franiak-Pietryga et al. 2022). In contrast, triple-negative breast cancer in general and MDA-MB-231 cells in particular, are known to be radioresistant (de Faria Bessa and Marta 2022; To et al. 2022).

I propose that this radioresistance is partly due to the failure of MDA-MB-231 cells to downregulate SRSF1 in response to IR (as shown in Figure 3.7D), which means they do not activate the necessary isoform-switching response that facilitates IR-induced cell death. This is consistent with the survival data of breast cancer patients in Figure 3.7C. In addition, in Figure 3.7D, SRSF1 knockdown more than doubled cleaved PARP levels 48 hours after IR, compared to not knocking it down, although the difference was not statistically significant. Thus, even with SRSF1 knockdown, a higher IR dose might be necessary to kill those cells effectively. This is consistent with the findings of Bravatà et al. that MDA-MB-231 colony

formation was only inhibited after 23 Gy (Bravatà et al. 2019). It would be useful to test whether this dose might be significantly reduced if IR is combined with SRSF1 knockdown, especially given that SRSF1 depletion combined with 10 Gy increased isoform switching more than SRSF1 depletion alone.

5.2.2 Generation of Shorter Transcripts

In Chapter 3, I found that, in response to IR, introns and exons retained are shorter than those spliced out, upstream last exons are favoured over downstream ones, and that more exons were spliced out than included. Taken together, these findings indicate that shorter transcripts are produced in response to IR. This is consistent with findings from other groups that other type of stress generate shorter transcripts (Boutz et al. 2015; Chang et al. 2015; Sadek et al. 2019; Fontana et al. 2017; McKay et al. 2004; Williamson et al. 2017).

In Chapter 4, I was able to determine transcript length accurately thanks to the use of long-read sequencing. Consistent with my findings in the previous chapter, SRSF1 depletion was associated with an increase of shorter transcripts (Figure 4.3B), with shorter open reading frames and loss of exons that encode protein domains (Figure 4.3C).

Therefore, although SRSF1 functions both as a splicing activator and repressor (Anczuków et al. 2015; Du et al. 2021; Mayeda et al. 1993; Erkelenz et al. 2013; Pandit et al. 2013), depending on the transcript context, my findings so far indicate that it generally promotes exon inclusion, which is consistent with its canonically-ascribed role (Das and Krainer 2014).

5.2.3 Consequences for the Putative Proteome

My findings emphasize the importance of investigating isoform expression in gene expression studies. Despite more and more studies calling for this, it is not the default way to process RNA sequencing data. We can no longer reduce the function of a gene to what one of its isoforms does, when the 20,596 protein-coding genes produce 82,685 distinct proteins (The UniProt Consortium 2023), a ratio of 1:4.

I found that isoform switching can upregulate isoforms of the same gene that facilitate a response to IR where a different isoform would have inhibited this. For example, in response to IR, pro-apoptosis genes are fully spliced, while cell cycle promoting genes are spliced to disable their function, contributing to cell cycle repression. I found a role for SRSF1 in facilitating this, and in MDA-MB-231 cells, its depletion affects the proteome by causing domain loss and intrinsically-disordered region loss, likely affecting many protein-protein interactions for proteins coded for by the isoforms.

5.3 Therapeutic Implications

SRSF1 is an essential gene. Therefore, despite how promising it looks as an anti-cancer target at this stage, specific targeting is a long way off. Until technical advances allow tumour cells to be targeted exclusively, depletion of this factor is unlikely to be useful therapeutically.

However, at least two early-splicing inhibitors, E7107 and H3B-8800, entered clinical trials (Bashari et al. 2023). Unfortunately, neither drug progressed beyond Phase I. Severe toxicity related to blurred vision halted the E7107 trial (Eskens et al. 2013), but the H3B-8800 phase I was completed (Steensma et al. 2021). Although both drugs modulated splicing

as expected, there was no clinical benefit. However, there is evidence that targeting specific tumour molecular subtypes, such as those with SF3B1 mutations or MYC amplifications, may be beneficial (Araki et al. 2023).

Still, these disappointing results highlight that targeting splicing on a large scale (e.g. by targeting an important splicing factor or a key component of the spliceosome) will probably not work without exclusive targeting of the tumour or malignant cells, at which point higher doses can be used, while limiting toxicity to normal cells.

Having said that, this is a promising area of research, with antibody-drug conjugates already in clinical use, including as first-line cancer treatments (Dumontet et al. 2023). After further validation of the promise of SRSF1 depletion in cancer, development of an antibody-drug conjugate that depletes it in combination with radiotherapy and/or immune checkpoint inhibitors may become a feasible and effective treatment option.

On another note, I identified nearly 90 antigens increasingly presented after irradiation and/or SRSF1 depletion that are not presented in normal tissues. Beyond simply being tumour-associated antigens, these could be bona fide neoantigens (exclusive to tumour cells), produced after radiotherapy or in SRSF1-depleted tumours. Once confirmed as neoantigens and their immunogenicity validated, these putative neoantigens open up new avenues for the treatment of triple-negative breast cancer, which has historically been the most aggressive breast cancer and the most challenging to treat. For example, these putative neoantigens can be targeted by immunotherapies such as bispecific T-cell engagers, vaccines, and chimeric antigen receptor (CAR) T-cell therapy. This approach can help in directly attacking cancer cells without harming normal cells, thereby reducing side effects and improving treatment efficacy.

Furthermore, the putative neoantigens can be used to develop anti-cancer vaccines, which would be particularly beneficial for preventing recurrence, especially given the high risk of recurrence in the first 3 years after surgery (Kumar and Aggarwal 2016).

5.4 Limitations and Future Directions

Although I have uncovered some of the mechanisms underlying the transcriptional IR response, more mechanistic questions were prompted by my results. For example, how is SRSF1 depleted after IR in normal cells? How does its depletion increase apoptosis and cell cycle arrest? How does SRSF1 depletion induce potentially novel antigen presentation, given that most of the potentially novel antigens identified were not differentially expressed at the RNA level? Had the Covid-19 pandemic not halted my experiments for about 5 months, I could have tackled some of these questions to complement my bioinformatics analyses.

In addition, because the B-cell project (in Chapter 2 and Chapter 3) was started by others in the Cheung Lab at the University of Michigan, I had to work with what they produced during my DPhil. Some of the things I would have changed are the use of long-read RNA sequencing instead of just short-read (although at the time the project started in the mid-2010s, the technology was not mature enough), because short-read sequencing cannot capture full-length isoforms and relies instead on reads across splice junctions. In addition, I would have used time-matched samples in the study design, as opposed to baseline (right before IR) and 2 and 6 hours after IR, to exclude any confounding with time. Furthermore, the Cheung Lab had chosen B cells because the genetics of these particular cell lines are extensively studied, in the context of IR exposure and generally (see 2.2.1 for details), but that created a major limitation after I picked up the project and proceeded with functional

experiments, such as knockdown and overexpression, which are challenging to perform in suspension cells, and I therefore had to change to solid tumour/tissue cell lines.

As mentioned in the Introduction, SRSF1 depletion promotes R-loop formation. Therefore, R-loops may be behind the SRSF1 depletion-induced cell cycle arrest and apoptosis. RNase H1 overexpression experiments would determine whether R-loop abrogation diminishes SRSF1 depletion-associated cell death. Alternatively, promising SRSF1 targets, such as BFAR and NCK2 could be investigated by isoform-specific knockdown/overexpression experiments. In addition, since SRSF1's roles in alternative splicing and NMD are tightly linked (Aznarez et al. 2018), untangling the impact of the two on SRSF1-depletion-associated isoform switching would be an interesting future direction. This can be done by, for example, using drugs that inhibit NMD, such an SMG1 inhibitor (Li et al. 2022; Cheruiyot et al. 2021), in an SRSF1-depletion background, and checking which isoform switches are rescued as a result. However, even if this was done, SRSF1, like many other RNA-binding proteins, has many functions outside of just NMD and splicing (see Section 1.4.4), such as mRNA transport and translation. Although transport and translation are unlikely to influence transcript levels directly, this broad range of activities should still be considered.

To confirm SRSF1 as a biomarker for radiotherapy and immune-checkpoint treatment, the following will be needed: randomized and blinded experiments on more specimens, randomized clinical trials with rigorous statistical controls, followed by demonstrations of clinical validity and utility, as well as the development of a test for SRSF1 levels to be used in the clinic (Ou et al. 2021).

To confirm the neoantigenicity and immunogenicity of the potential neoantigens I found, paired normal and tumour samples from triple-negative breast cancer should be tested for the specificity of the antigens. Next, the most promising neoantigens should be tested for immunogenicity with antigen-specific *in vitro* T-cell stimulation experiments, followed by vaccinating xenograft mouse models with the neoantigens that produce a specific T-cell response.

5.5 Conclusion

The results described in this thesis make several important contributions towards understanding the role of isoform switching in the DNA damage response and its implications for cancer therapy.

Firstly, I demonstrated that ionising radiation triggers genome-wide isoform switching events that generate shorter transcripts, likely facilitating appropriate cellular responses like apoptosis and cell cycle arrest. My findings reveal isoform switching as an underappreciated hallmark of the DNA damage response.

Second, I uncovered a critical role for the splicing factor SRSF1 in conferring radio- and immunotherapy resistance. Tumours that fail to repress SRSF1 after irradiation have poorer prognosis. Targeting SRSF1 presents a promising therapeutic opportunity in the future to sensitize resistant tumours to radiotherapy.

Finally, this work led to the discovery of putative neoantigens generated specifically after SRSF1 knockdown and/or ionising radiation in triple negative breast cancer cells. The immunogenic potential of these tumour-specific antigens merits further exploration for immunotherapeutic development.

In summary, by highlighting the isoform-specific nature of gene regulation, this thesis advances our understanding of the DNA damage response and reveals new precision oncology targets. The findings presented here pave the way for novel prognostic and therapeutic approaches that incorporate alternatively-spliced isoform profiles. Further mechanistic dissection of isoform switching events promises additional insights into how tumours evolve resistance and how we may prevent it.

Chapter 6   References

Abba MC, Lacunza E, Nunez MI, Colussi A, Isla-Larrain M, Segal-Eiras A, Croce MV, Aldaz CM. 2009. Rhomboid domain containing 2 (RHBDD2): A novel cancer-related gene over-expressed in breast cancer. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1792**: 988–997.

Adinolfi M, Pietrosanto M, Parca L, Ausiello G, Ferrè F, Helmer-Citterich M. 2019. Discovering sequence and structure landscapes in RNA interaction motifs. *Nucleic Acids Research* **47**: 4958–4969.

Agrawal AA, Yu L, Smith PG, Buonamici S. 2018. Targeting splicing abnormalities in cancer. *Curr Opin Genet Dev* **48**: 67–74.

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale A-L, et al. 2013. Signatures of mutational processes in human cancer. *Nature* **500**: 415–421.

Alizadeh E, Orlando TM, Sanche L. 2015. Biomolecular Damage Induced by Ionizing Radiation: The Direct and Indirect Effects of Low-Energy Electrons on DNA. *Annu Rev Phys Chem* **66**: 379–398.

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**: 420–423.

Amundson SA, Myers TG, Scudiero D, Kitada S, Reed JC, Fornace AJ. 2000. An informatics approach identifying markers of chemosensitivity in human cancer cell lines. *Cancer Res* **60**: 6101–6110.

Anczuków O, Akerman M, Cléry A, Wu J, Shen C, Shirole NH, Raimer A, Sun S, Jensen MA, Hua Y, et al. 2015. SRSF1-Regulated Alternative Splicing in Breast Cancer. *Molecular Cell* **60**: 105–117.

Anczuków O, Rosenberg AZ, Akerman M, Das S, Zhan L, Karni R, Muthuswamy SK, Krainer AR. 2012. The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nature Structural & Molecular Biology* **19**: 220–228.

Anders S, Reyes A, Huber W. 2012. Detecting differential usage of exons from RNA-seq data. *Genome Res* **22**: 2008–2017.

Apcher S, Daskalogianni C, Lejeune F, Manoury B, Imhoos G, Heslop L, Fåhraeus R. 2011. Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation. *Proceedings of the National Academy of Sciences* **108**: 11572–11577.

Araki S, Ohori M, Yugami M. 2023. Targeting pre-mRNA splicing in cancers: roles, inhibitors, and therapeutic opportunities. *Front Oncol* **13**. https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2023.1152087/full (Accessed March 31, 2024).

Aznarez I, Nomakuchi TT, Tetenbaum-Novatt J, Rahman MA, Fregoso O, Rees H, Krainer AR. 2018. Mechanism of Nonsense-Mediated mRNA Decay Stimulation by Splicing Factor SRSF1. *Cell Reports* **23**: 2186–2198.

Azzam EI, Jay-Gerin J-P, Pain D. 2012. Ionizing radiation-induced metabolic oxidative stress and prolonged cell injury. *Cancer Lett* **327**: 48–60.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202-208.

Bairoch A. 2018. The Cellosaurus, a Cell-Line Knowledge Resource. *J Biomol Tech* **29**: 25–38.

Barbagallo D, Caponnetto A, Barbagallo C, Battaglia R, Mirabella F, Brex D, Stella M, Broggi G, Altieri R, Certo F, et al. 2021. The GAUGAA Motif Is Responsible for the Binding between circSMARCA5 and SRSF1 and Related Downstream Effects on Glioblastoma Multiforme Cell Migration and Angiogenic Potential. *International Journal of Molecular Sciences* **22**: 1678.

Barker DJ, Maccari G, Georgiou X, Cooper MA, Flicek P, Robinson J, Marsh SGE. 2023. The IPD-IMGT/HLA Database. *Nucleic Acids Research* **51**: D1053–D1060.

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, et al. 2013. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research* **41**: D991–D995.

Bashari A, Siegfried Z, Karni R. 2023. Targeting Splicing Factors for Cancer Therapy. *RNA* rna.079585.123.

Bassani-Sternberg M, Pletscher-Frankild S, Jensen LJ, Mann M. 2015. Mass Spectrometry of Human Leukocyte Antigen Class I Peptidomes Reveals Strong Effects of Protein Abundance and Turnover on Antigen Presentation. *Mol Cell Proteomics* **14**: 658–673.

Becker BV, Majewski M, Abend M, Palnek A, Nestler K, Port M, Ullmann R. 2018. Gene expression changes in human iPSC-derived cardiomyocytes after X-ray irradiation. *Int J Radiat Biol* **94**: 1095–1103.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* **57**: 289–300.

Berget SM, Moore C, Sharp PA. 1977. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *PNAS* **74**: 3171–3175.

Bonnal SC, López-Oreja I, Valcárcel J. 2020. Roles and mechanisms of alternative splicing in cancer — implications for care. *Nature Reviews Clinical Oncology* 1–18.

Bourdetsky D, Schmelzer CEH, Admon A. 2014. The nature and extent of contributions by defective ribosome products to the HLA peptidome. *Proceedings of the National Academy of Sciences* **111**: E1591–E1599.

Boustani J, Grapin M, Laurent P-A, Apetoh L, Mirjolet C. 2019. The 6th R of Radiobiology: Reactivation of Anti-Tumor Immune Response. *Cancers* **11**: 860.

Boutz PL, Bhutkar A, Sharp PA. 2015. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev* **29**: 63–80.

Braunschweig U, Barbosa-Morais NL, Pan Q, Nachman EN, Alipanahi B, Gonatopoulos-Pournatzis T, Frey B, Irimia M, Blencowe BJ. 2014. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res* **24**: 1774–1786.

Bravatà V, Cammarata FP, Minafra L, Musso R, Pucci G, Spada M, Fazio I, Russo G, Forte GI. 2019. Gene Expression Profiles Induced by High-dose Ionizing Radiation in MDA-MB-231 Triple-negative Breast Cancer Cell Line. *Cancer Genomics & Proteomics* **16**: 257–266.

Brinster RL, Allen JM, Behringer RR, Gelinas RE, Palmiter RD. 1988. Introns increase transcriptional efficiency in transgenic mice. *Proc Natl Acad Sci U S A* **85**: 836–840.

Buonanno M, de Toledo SM, Pain D, Azzam EI. 2011. Long-Term Consequences of Radiation-Induced Bystander Effects Depend on Radiation Quality and Dose and Correlate with Oxidative Stress. *Radiat Res* **175**: 405–415.

Burnette BC, Liang H, Lee Y, Chlewicki L, Khodarev NN, Weichselbaum RR, Fu Y-X, Auh SL. 2011. The efficacy of radiotherapy relies upon induction of type i interferon-dependent innate and adaptive immunity. *Cancer Res* **71**: 2488–2496.

Canzoneri R, Lacunza E, Isla Larrain M, Croce MV, Abba MC. 2014. Rhomboid family gene expression profiling in breast normal tissue and tumor samples. *Tumor Biol* **35**: 1451–1458.

Chang J-W, Zhang W, Yeh H-S, de Jong EP, Jun S, Kim K-H, Bae SS, Beckman K, Hwang TH, Kim K-S, et al. 2015. mRNA 3′-UTR shortening is a molecular signature of mTORC1 activation. *Nat Commun* **6**: 7218.

Chen J, Crutchley J, Zhang D, Owzar K, Kastan MB. 2017a. Identification of a DNA Damage-Induced Alternative Splicing Pathway That Regulates p53 and Cellular Senescence Markers. *Cancer Discov* **7**: 766–781.

Chen J, Zhang D, Qin X, Owzar K, McCann JJ, Kastan MB. 2021. DNA-Damage-Induced Alternative Splicing of p53. *Cancers* **13**: 251.

Chen L, Luo C, Shen L, Liu Y, Wang Q, Zhang C, Guo R, Zhang Y, Xie Z, Wei N, et al. 2017b. SRSF1 Prevents DNA Damage and Promotes Tumorigenesis through Regulation of DBF4B Pre-mRNA Splicing. *Cell Reports* **21**: 3406–3413.

Chen M-S, Higashikubo R, Laszlo A, Roti Roti J. 2001. Multiple alternative splicing forms of human RAD17 and their differential response to ionizing radiation. *Gene* **277**: 145–152.

Chen W, Moore MJ. 2015. Spliceosomes. *Current Biology* **25**: R181–R183.

Chen Y, Lun ATL, Smyth GK. 2016. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. https://f1000research.com/articles/5-1438 (Accessed November 27, 2023).

Cheruiyot A, Li S, Nonavinkere Srivatsan S, Ahmed T, Chen Y, Lemacon DS, Li Y, Yang Z, Wadugu BA, Warner WA, et al. 2021. Nonsense-Mediated RNA Decay Is a Unique

Vulnerability of Cancer Cells Harboring SF3B1 or U2AF1 Mutations. *Cancer Research* **81**: 4499–4513.

Choi S, Cho N, Kim KK. 2023. The implications of alternative pre-mRNA splicing in cell signal transduction. *Exp Mol Med* **55**: 755–766.

Chong C, Müller M, Pak H, Harnett D, Huber F, Grun D, Leleu M, Auger A, Arnaud M, Stevenson BJ, et al. 2020. Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nat Commun* **11**: 1293.

Chopra S, Shankavaram U, Bylicky M, Dalo J, Scott K, Aryankalayil MJ, Coleman CN. 2022. Profiling mRNA, miRNA and lncRNA expression changes in endothelial cells in response to increasing doses of ionizing radiation. *Sci Rep* **12**: 19941.

Chow LT, Gelinas RE, Broker TR, Roberts RJ. 1977. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12**: 1–8.

Cohen–Jonathan E, Bernhard EJ, McKenna WG. 1999. How does radiation kill cells? *Current Opinion in Chemical Biology* **3**: 77–83.

Cohen-Jonathan E, Toulas C, Monteil S, Couderc B, Maret A, Bard JJ, Prats H, Daly-Schveitzer N, Favre G. 1997. Radioresistance induced by the high molecular forms of the basic fibroblast growth factor is associated with an increased G2 delay and a hyperphosphorylation of p34CDC2 in HeLa cells. *Cancer Res* **57**: 1364–1370.

Contreras-Ochoa CO, Bahena-Román M, López-Díaz LY, Lagunas-Martínez A, Mojica-Cardoso C, Manzo-Merino J, Torres-Poveda K, Madrid-Marina V. 2022. Fas and Fas ligand

are highly expressed in lymphocytes from cervical intraepithelial neoplasia and cervical cancer patients: A possible role for immune escaping. *Iran J Basic Med Sci* **25**: 383–389.

Corley M, Burns MC, Yeo GW. 2020. How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Molecular Cell* **78**: 9–29.

Criswell T, Klokov D, Beman M, Lavik J, Boothman DA. 2003. Repression of IR-Inducible Clusterin Expression by the p53 Tumor Suppressor Protein. *Cancer Biology & Therapy* **2**: 372–380.

Crossley MP, Bocek M, Cimprich KA. 2019. R-Loops as Cellular Regulators and Genomic Threats. *Molecular Cell* **73**: 398–411.

da Costa PJ, Menezes J, Romão L. 2017. The role of alternative splicing coupled to nonsense-mediated mRNA decay in human disease. *The International Journal of Biochemistry & Cell Biology* **91**: 168–175.

Dalenc F, Drouet J, Ader I, Delmas C, Rochaix P, Favre G, Cohen-Jonathan E, Toulas C. 2002. Increased expression of a COOH-truncated nucleophosmin resulting from alternative splicing is associated with cellular resistance to ionizing radiation in HeLa cells. *International Journal of Cancer* **100**: 662–668.

Das S, Anczuków O, Akerman M, Krainer AR. 2012. ONCOGENIC SPLICING FACTOR SRSF1 IS A CRITICAL TRANSCRIPTIONAL TARGET OF MYC. *Cell Rep* **1**: 110–117.

Das S, Krainer AR. 2014. Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer. *Mol Cancer Res* **12**: 1195–1204.

Davarinejad H. 2015. Quantifications of western blots with ImageJ. *University of York*.

de Faria Bessa J, Marta GN. 2022. Triple-negative breast cancer and radiation therapy. *Rep Pract Oncol Radiother* **27**: 545–551.

de Jong Y, Monderer D, Brandinelli E, Monchanin M, van den Akker BE, van Oosterwijk JG, Blay JY, Dutour A, Bovée JVMG. 2018. Bcl-xl as the most promising Bcl-2 family member in targeted treatment of chondrosarcoma. *Oncogenesis* **7**: 1–9.

Deng L, Liang H, Xu M, Yang X, Burnette B, Arina A, Li X-D, Mauceri H, Beckett M, Darga T, et al. 2014. STING-Dependent Cytosolic DNA Sensing Promotes Radiation-Induced Type I Interferon-Dependent Antitumor Immunity in Immunogenic Tumors. *Immunity* **41**: 843–852.

Ding F, Elowitz MB. 2019. Constitutive splicing and economies of scale in gene expression. *Nat Struct Mol Biol* **26**: 424–432.

Dodson H, Wheatley SP, Morrison CG. 2007. Involvement of centrosome amplification in radiation-induced mitotic catastrophe. *Cell Cycle* **6**: 364–370.

Dong C, Cesarano A, Bombaci G, Reiter JL, Yu CY, Wang Y, Jiang Z, Zaid MA, Huang K, Lu X, et al. 2021. Intron retention-induced neoantigen load correlates with unfavorable prognosis in multiple myeloma. *Oncogene* 1–9.

Du J-X, Luo Y-H, Zhang S-J, Wang B, Chen C, Zhu G-Q, Zhu P, Cai C-Z, Wan J-L, Cai J-L, et al. 2021. Splicing factor SRSF1 promotes breast cancer progression via oncogenic splice switching of PTPMT1. *J Exp Clin Cancer Res* **40**: 171.

Duarte D, Vale N. 2022. Evaluation of synergism in drug combinations and reference models for future orientations in oncology. *Curr Res Pharmacol Drug Discov* **3**: 100110.

Dujardin G, Lafaille C, de la Mata M, Marasco LE, Muñoz MJ, Le Jossic-Corcos C, Corcos L, Kornblihtt AR. 2014. How slow RNA polymerase II elongation favors alternative exon skipping. *Mol Cell* **54**: 683–690.

Dumontet C, Reichert JM, Senter PD, Lambert JM, Beck A. 2023. Antibody–drug conjugates come of age in oncology. *Nat Rev Drug Discov* **22**: 641–661.

Dvinge H, Bradley RK. 2015. Widespread intron retention diversifies most cancer transcriptomes. *Genome Med* **7**: 45.

Dvinge H, Kim E, Abdel-Wahab O, Bradley RK. 2016. RNA splicing factors as oncoproteins and tumour suppressors. *Nat Rev Cancer* **16**: 413–430.

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**: 207–210.

El Marabti E, Younis I. 2018. The Cancer Spliceome: Reprograming of Alternative Splicing in Cancer. *Front Mol Biosci* **5**. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6137424/ (Accessed May 8, 2019).

Erdős G, Pajkos M, Dosztányi Z. 2021. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Research* **49**: W297–W303.

Erkelenz S, Mueller WF, Evans MS, Busch A, Schöneweis K, Hertel KJ, Schaal H. 2013. Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *RNA* **19**: 96–102.

Eskens FALM, Ramos FJ, Burger H, O'Brien JP, Piera A, de Jonge MJA, Mizui Y, Wiemer EAC, Carreras MJ, Baselga J, et al. 2013. Phase I Pharmacokinetic and Pharmacodynamic Study of the First-in-Class Spliceosome Inhibitor E7107 in Patients with Advanced Solid Tumors. *Clinical Cancer Research* **19**: 6296–6304.

Feng H, Bao S, Rahman MA, Weyn-Vanhentenryck SM, Khan A, Wong J, Shah A, Flynn ED, Krainer AR, Zhang C. 2019. Modeling RNA-Binding Protein Specificity In Vivo by Precisely Registering Protein-RNA Crosslink Sites. *Molecular Cell* **74**: 1189-1204.e6.

Flynn RL, Centore RC, O'Sullivan RJ, Rai R, Tse A, Songyang Z, Chang S, Karlseder J, Zou L. 2011. TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature* **471**: 532–536.

Fontana GA, Rigamonti A, Lenzken SC, Filosa G, Alvarez R, Calogero R, Bianchi ME, Barabino SML. 2017. Oxidative stress controls the choice of alternative last exons via a Brahma-BRCA1-CstF pathway. *Nucleic Acids Res* **45**: 902–914.

Formenti SC, Demaria S. 2009. Systemic effects of local radiotherapy. *The Lancet Oncology* **10**: 718–726.

Formenti SC, Rudqvist N-P, Golden E, Cooper B, Wennerberg E, Lhuillier C, Vanpouille-Box C, Friedman K, Ferrari de Andrade L, Wucherpfennig KW, et al. 2018. Radiotherapy induces responses of lung cancer to CTLA-4 blockade. *Nat Med* **24**: 1845–1851.

Forrester HB, Li J, Hovan D, Ivashkevich AN, Sprung CN. 2012. DNA repair genes: alternative transcription and gene expression at the exon level in response to the DNA damaging agent, ionizing radiation. *PLoS One* **7**: e53358.

Franiak-Pietryga I, Miyauchi S, Kim SS, Sanders PD, Sumner W, Zhang L, Mundt A, Califano JA, Sharabi AB. 2022. Activated B-cells and Plasma cells are Resistant to Radiation Therapy. *Int J Radiat Oncol Biol Phys* **112**: 514–528.

Fu X-D, Ares M. 2014. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat Rev Genet* **15**: 689–701.

Gameiro SR, Jammeh ML, Wattenberg MM, Tsang KY, Ferrone S, Hodge JW. 2014. Radiation-induced immunogenic modulation of tumor enhances antigen processing and calreticulin exposure, resulting in enhanced T-cell killing. *Oncotarget* **5**: 403–416.

García-Moreno JF, Romão L. 2020. Perspective in Alternative Splicing Coupled to Nonsense-Mediated mRNA Decay. *International Journal of Molecular Sciences* **21**: 9424.

Gerstberger S, Hafner M, Tuschl T. 2014. A census of human RNA-binding proteins. *Nat Rev Genet* **15**: 829–845.

Gleave ME, Miyake H, Zellweger T, Chi K, July L, Nelson C, Rennie P. 2001. Use of antisense oligonucleotides targeting the antiapoptotic gene, clusterin/testosterone-repressed prostate message 2, to enhance androgen sensitivity and chemosensitivity in prostate cancer. *Urology* **58**: 39–48.

Golden EB, Apetoh L. 2015. Radiotherapy and Immunogenic Cell Death. *Seminars in Radiation Oncology* **25**: 11–17.

Goldman M, Craft B, Hastie M, Repečka K, McDade F, Kamath A, Banerjee A, Luo Y, Rogers D, Brooks AN, et al. 2019. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. *bioRxiv* 326470.

Gonçalves V, Jordan P. 2015. Posttranscriptional Regulation of Splicing Factor SRSF1 and Its Role in Cancer Cell Biology. *Biomed Res Int* **2015**: 287048.

Grabski DF, Broseus L, Kumari B, Rekosh D, Hammarskjold M-L, Ritchie W. 2021. Intron retention and its impact on gene expression and protein diversity: A review and a practical guide. *WIREs RNA* **12**: e1631.

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017–1018.

Grzybowska EA. 2013. Human Intronless Genes and their Associated Diseases. In *Encyclopedia of Life Sciences*, John Wiley & Sons, Ltd https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470015902.a0025005 (Accessed December 5, 2023).

Győrffy B. 2023. Discovery and ranking of the most robust prognostic biomarkers in serous ovarian cancer. *Geroscience* **45**: 1889–1898.

Hanahan D. 2022. Hallmarks of Cancer: New Dimensions. *Cancer Discov* **12**: 31–46.

Hanahan D, Weinberg RA. 2011. Hallmarks of Cancer: The Next Generation. *Cell* **144**: 646–674.

Hanahan D, Weinberg RA. 2000. The Hallmarks of Cancer. *Cell* **100**: 57–70.

Heylmann D, Ponath V, Kindler T, Kaina B. 2021. Comparison of DNA repair and radiosensitivity of different blood cell populations. *Sci Rep* **11**: 2478.

Hill M, Ullrich R. 2019. Ionizing Radiation. In *Tumour Site Concordance and Mechanisms of Carcinogenesis* (eds. R. Baan, B. Stewart, and K. Straif), International Agency for Research on Cancer.

Ho S-Y, Chang C-M, Liao H-N, Chou W-H, Guo C-L, Yen Y, Nakamura Y, Chang W-C. 2023. Current Trends in Neoantigen-Based Cancer Vaccines. *Pharmaceuticals* **16**: 392.

Howard FM, Olopade OI. 2021. Epidemiology of Triple-Negative Breast Cancer: A Review. *The Cancer Journal* **27**: 8.

Huber SM, Butz L, Stegen B, Klumpp L, Klumpp D, Eckert F. 2015. Role of ion channels in ionizing radiation-induced cell death. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1848**: 2657–2664.

Jacob AG, Smith CWJ. 2017. Intron retention as a component of regulated gene expression programs. *Hum Genet* **136**: 1043–1057.

Jacquet K, Banerjee SL, Chartier FJM, Elowe S, Bisson N. 2018. Proteomic Analysis of NCK1/2 Adaptors Uncovers Paralog-specific Interactions That Reveal a New Role for NCK2 in Cell Abscission During Cytokinesis. *Mol Cell Proteomics* **17**: 1979–1990.

Jarosz-Biej M, Smolarczyk R, Cichoń T, Kułach N. 2019. Tumor Microenvironment as A "Game Changer" in Cancer Radiotherapy. *Int J Mol Sci* **20**: 3212.

Jen K-Y, Cheung VG. 2005. Identification of Novel p53 Target Genes in Ionizing Radiation Response. *Cancer Research* **65**: 7666–7673.

Jen K-Y, Cheung VG. 2003. Transcriptional Response of Lymphoblastoid Cells to Ionizing Radiation. *Genome Res* **13**: 2092–2100.

Ji X, Zhou Y, Pandit S, Huang J, Li H, Lin CY, Xiao R, Burge CB, Fu X-D. 2013. SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell* **153**: 855–868.

Jiang L, Huang J, Higgs BW, Hu Z, Xiao Z, Yao X, Conley S, Zhong H, Liu Z, Brohawn P, et al. 2016. Genomic Landscape Survey Identifies SRSF1 as a Key Oncodriver in Small Cell Lung Cancer. *PLOS Genetics* **12**: e1005895.

Jiang W, Chen L. 2021. Alternative splicing: Human disease and quantitative analysis from high-throughput sequencing. *Computational and Structural Biotechnology Journal* **19**: 183–195.

Jolma A, Zhang J, Mondragón E, Morgunova E, Kivioja T, Laverty KU, Yin Y, Zhu F, Bourenkov G, Morris Q, et al. 2020. Binding specificities of human RNA-binding proteins toward structured and linear RNA sequences. *Genome Res* **30**: 962–973.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240.

Joruiz SM, Bourdon J-C. 2016. p53 Isoforms: Key Regulators of the Cell Fate Decision. *Cold Spring Harb Perspect Med* **6**: a026039.

Juan CA, Pérez de la Lastra JM, Plou FJ, Pérez-Lebeña E. 2021. The Chemistry of Reactive Oxygen Species (ROS) Revisited: Outlining Their Role in Biological Macromolecules (DNA, Lipids and Proteins) and Induced Pathologies. *International Journal of Molecular Sciences* **22**: 4642.

Jung H, Lee D, Lee J, Park D, Kim YJ, Park W-Y, Hong D, Park PJ, Lee E. 2015. Intron retention is a widespread mechanism of tumor-suppressor inactivation. *Nat Genet* **47**: 1242–1248.

Kaabinejadian S, Barra C, Alvarez B, Yari H, Hildebrand WH, Nielsen M. 2022. Accurate MHC Motif Deconvolution of Immunopeptidomics Data Reveals a Significant Contribution of DRB3, 4 and 5 to the Total DR Immunopeptidome. *Frontiers in Immunology* **13**. https://www.frontiersin.org/articles/10.3389/fimmu.2022.835454 (Accessed November 27, 2023).

Kahles A, Lehmann K-V, Toussaint NC, Hüser M, Stark SG, Sachsenberg T, Stegle O, Kohlbacher O, Sander C, Rätsch G, et al. 2018. Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer Cell* **34**: 211-224.e6.

Kang Y-J, Yang D-C, Kong L, Hou M, Meng Y-Q, Wei L, Gao G. 2017. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Research* **45**: W12–W16.

Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR. 2007. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol* **14**: 185–193.

Karousis ED, Mühlemann O. 2019. Nonsense-Mediated mRNA Decay Begins Where Translation Ends. *Cold Spring Harb Perspect Biol* **11**: a032862.

Katz Y, Wang ET, Airoldi EM, Burge CB. 2010. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* **7**: 1009–1015.

Kędzierska H, Piekiełko-Witkowska A. 2017. Splicing factors of SR and hnRNP families as regulators of apoptosis in cancer. *Cancer Lett* **396**: 53–65.

Kędzierska H, Popławski P, Hoser G, Rybicka B, Rodzik K, Sokół E, Bogusławska J, Tański Z, Fogtman A, Koblowska M, et al. 2016. Decreased Expression of SRSF2 Splicing Factor Inhibits Apoptotic Pathways in Renal Cancer. *International Journal of Molecular Sciences* **17**: 1598.

Keitel U, Scheel A, Thomale J, Halpape R, Kaulfuß S, Scheel C, Dobbelstein M. 2014. Bcl-xL mediates therapeutic resistance of a mesenchymal breast cancer cell subpopulation. *Oncotarget* **5**: 11778–11791.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler and D. 2002. The Human Genome Browser at UCSC. *Genome Res* **12**: 996–1006.

Khodarev NN, Park JO, Yu J, Gupta N, Nodzenski E, Roizman B, Weichselbaum RR. 2001. Dose-dependent and independent temporal patterns of gene responses to ionizing radiation in normal and tumor cells and tumor xenografts. *Proc Natl Acad Sci USA* **98**: 12665–12670.

Kis E, Szatmári T, Keszei M, Farkas R, Esik O, Lumniczky K, Falus A, Sáfrány G. 2006. Microarray analysis of radiation response genes in primary human fibroblasts. *Int J Radiat Oncol Biol Phys* **66**: 1506–1514.

Klokov D, Criswell T, Leskov KS, Araki S, Mayo L, Boothman DA. 2004. IR-inducible clusterin gene expression: a protein with potential roles in ionizing radiation-induced adaptive responses, genomic instability, and bystander effects. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **568**: 97–110.

Kornblihtt AR. 2007. Coupling transcription and alternative splicing. *Adv Exp Med Biol* **623**: 175–189.

Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ. 2013. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nature Reviews Molecular Cell Biology* **14**: 153–165.

Kovács SA, Fekete JT, Győrffy B. 2023. Predictive biomarkers of immunotherapy response with pharmacological applications in solid tumors. *Acta Pharmacol Sin* **44**: 1879–1889.

Krainer AR, Maniatis T. 1985. Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. *Cell* **42**: 725–736.

Kretova M, Selicky T, Cipakova I, Cipak L. 2023. Regulation of Pre-mRNA Splicing: Indispensable Role of Post-Translational Modifications of Splicing Factors. *Life* **13**: 604.

Kubiniok P, Marcu A, Bichmann L, Kuchenbecker L, Schuster H, Hamelin DJ, Duquette JD, Kovalchik KA, Wessling L, Kohlbacher O, et al. 2022. Understanding the constitutive presentation of MHC class I immunopeptidomes in primary tissues. *iScience* **25**. https://www.cell.com/iscience/abstract/S2589-0042(22)00038-4 (Accessed December 21, 2023).

Kumar P, Aggarwal R. 2016. An overview of triple-negative breast cancer. *Arch Gynecol Obstet* **293**: 247–269.

Ladomery M. 2013. Aberrant Alternative Splicing Is Another Hallmark of Cancer. *International Journal of Cell Biology* **2013**: e463786.

Lahiff C, Schilling C, Cathcart MC, Mulligan N, Doran P, Muldoon C, Murray D, Pidgeon GP, Reynolds JV, Macmathuna P. 2014. Prognostic significance of neuroepithelial transforming gene 1 in adenocarcinoma of the oesophagogastric junction. *Br J Surg* **101**: 55–62.

Lavin MF, Khanna KK. 1999. ATM: the protein encoded by the gene mutated in the radiosensitive syndrome ataxia-telangiectasia. *Int J Radiat Biol* **75**: 1201–1214.

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for Computing and Annotating Genomic Ranges. *PLOS Computational Biology* **9**: e1003118.

Leader Y, Lev Maor G, Sorek M, Shayevitch R, Hussein M, Hameiri O, Tammer L, Zonszain J, Keydar I, Hollander D, et al. 2021. The upstream 5′ splice site remains associated to the transcription machinery during intron synthesis. *Nat Commun* **12**: 4545.

Leclair NK, Brugiolo M, Urbanski L, Lawson SC, Thakar K, Yurieva M, George J, Hinson JT, Cheng A, Graveley BR, et al. 2020. Poison Exon Splicing Regulates a Coordinated Network of SR Protein Expression during Differentiation and Tumorigenesis. *Molecular Cell* **80**: 648-665.e9.

Lei G, Zhang Y, Koppula P, Liu X, Zhang J, Lin SH, Ajani JA, Xiao Q, Liao Z, Wang H, et al. 2020. The role of ferroptosis in ionizing radiation-induced cell death and tumor suppression. *Cell Res* **30**: 146–162.

Leskov KS, Klokov DY, Li J, Kinsella TJ, Boothman DA. 2003. Synthesis and Functional Analyses of Nuclear Clusterin, a Cell Death Protein*. *Journal of Biological Chemistry* **278**: 11590–11600.

Leva V, Giuliano S, Bardoni A, Camerini S, Crescenzi M, Lisa A, Biamonti G, Montecucco A. 2012. Phosphorylation of SRSF1 is modulated by replicational stress. *Nucleic Acids Res* **40**: 1106–1117.

Lewis BP, Green RE, Brenner SE. 2003. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proceedings of the National Academy of Sciences* **100**: 189–192.

Lhuillier C, Rudqvist N-P, Yamazaki T, Zhang T, Charpentier M, Galluzzi L, Dephoure N, Clement CC, Santambrogio L, Zhou XK, et al. 2021. Radiotherapy-exposed CD8[+] and CD4[+] neoantigens enhance tumor control. *J Clin Invest*. https://www.jci.org/articles/view/138740 (Accessed January 26, 2021).

Li X, Manley JL. 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**: 365–378.

Li Y, Wu M, Zhang L, Wan L, Li H, Zhang L, Sun G, Huang W, Zhang J, Su F, et al. 2022. Nonsense-mediated mRNA decay inhibition synergizes with MDM2 inhibition to suppress TP53 wild-type cancer cells in p53 isoform-dependent manner. *Cell Death Discov* **8**: 1–12.

Liao Y, Liu S, Fu S, Wu J. 2020. HMGB1 in Radiotherapy: A Two Headed Signal Regulating Tumor Radiosensitivity and Immunity. *Onco Targets Ther* **13**: 6859–6871.

Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. 2019. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* **47**: W199–W205.

Liu J, Li M, Ran X, Fan J, Song J. 2006. Structural Insight into the Binding Diversity between the Human Nck2 SH3 Domains and Proline-Rich Proteins,. *Biochemistry* **45**: 7171–7184.

Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV, et al. 2018. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **173**: 400-416.e11.

Liu Y, Gonzàlez-Porta M, Santos S, Brazma A, Marioni JC, Aebersold R, Venkitaraman AR, Wickramasinghe VO. 2017. Impact of Alternative Splicing on the Human Proteome. *Cell Reports* **20**: 1229–1241.

Lu SX, De Neef E, Thomas JD, Sabio E, Rousseau B, Gigoux M, Knorr DA, Greenbaum B, Elhanati Y, Hogg SJ, et al. 2021. Pharmacologic modulation of RNA splicing enhances anti-tumor immunity. *Cell*. https://www.sciencedirect.com/science/article/pii/S0092867421006905 (Accessed July 19, 2021).

Macaeva E, Saeys Y, Tabury K, Janssen A, Michaux A, Benotmane MA, De Vos WH, Baatout S, Quintens R. 2016. Radiation-induced alternative transcription and splicing events and their applicability to practical biodosimetry. *Sci Rep* **6**. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4725928/ (Accessed December 16, 2018).

Maier P, Hartmann L, Wenz F, Herskind C. 2016. Cellular Pathways in Response to Ionizing Radiation and Their Targetability for Tumor Radiosensitization. *Int J Mol Sci* **17**: 102.

Maity A, McKenna WG, Muschel RJ. 1994. The molecular basis for cell cycle delays following ionizing radiation: a review. *Radiotherapy and Oncology* **31**: 1–13.

Mak QXC, Wick RR, Holt JM, Wang JR. 2023. Polishing De Novo Nanopore Assemblies of Bacteria and Eukaryotes With FMLRC2. *Molecular Biology and Evolution* **40**: msad048.

Marasco LE, Kornblihtt AR. 2022. The physiology of alternative splicing. *Nat Rev Mol Cell Biol* 1–13.

Marcu A, Bichmann L, Kuchenbecker L, Kowalewski DJ, Freudenmann LK, Backert L, Mühlenbruch L, Szolek A, Lübke M, Wagner P, et al. 2021. HLA Ligand Atlas: a benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy. *J Immunother Cancer* **9**: e002071.

Martín-Villa JM, Vaquero-Yuste C, Molina-Alejandre M, Juarez I, Suárez-Trujillo F, López-Nares A, Palacio-Gruber J, Barrera-Gutiérrez L, Fernández-Cruz E, Rodríguez-Sainz C, et al. 2022. HLA-G: Too Much or Too Little? Role in Cancer and Autoimmune Disease. *Frontiers in Immunology* **13**. https://www.frontiersin.org/articles/10.3389/fimmu.2022.796054 (Accessed December 5, 2023).

Maslon MM, Heras SR, Bellora N, Eyras E, Cáceres JF. 2014. The translational landscape of the splicing factor SRSF1 and its role in mitosis. *eLife* **3**. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4027812/ (Accessed October 17, 2020).

Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER, Hurov KE, Luo J, Bakalarski CE, Zhao Z, Solimini N, Lerenthal Y, et al. 2007. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316**: 1160–1166.

Mayeda A, Helfman DM, Krainer AR. 1993. Modulation of Exon Skipping and Inclusion by Heterogeneous Nuclear Ribonucleoprotein Al and Pre-mRNA Splicing Factor SF2/ASF. *MOL CELL BIOL* **13**: 9.

McCann JJ, Fleenor DE, Chen J, Lai C-H, Bass TE, Kastan MB. 2023. Participation of ATM, SMG1, and DDX5 in a DNA Damage-Induced Alternative Splicing Pathway. *rare* **199**: 406–421.

McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* **40**: 4288–4297.

McKay BC, Stubbert LJ, Fowler CC, Smith JM, Cardamore RA, Spronck JC. 2004. Regulation of ultraviolet light-induced gene expression by gene size. *Proc Natl Acad Sci U S A* **101**: 6582–6586.

Middleton R, Gao D, Thomas A, Singh B, Au A, Wong JJ-L, Bomane A, Cosson B, Eyras E, Rasko JEJ, et al. 2017. IRFinder: assessing the impact of intron retention on mammalian gene expression. *Genome Biol* **18**: 51.

Milner E, Barnea E, Beer I, Admon A. 2006. The Turnover Kinetics of Major Histocompatibility Complex Peptides of Human Cancer Cells *. *Molecular & Cellular Proteomics* **5**: 357–365.

Miyake H, Gleave ME, Miyake H, Hara I, Kamidono S. 2001. Synergistic chemsensitization and inhibition of tumor growth and metastasis by the antisense oligodeoxynucleotide targeting clusterin gene in a human bladder cancer model. *Clinical Cancer Research* **7**: 4245–4252.

Montes M, Sanford BL, Comiskey DF, Chandler DS. 2019. RNA Splicing and Disease: Animal Models to Therapies. *Trends Genet* **35**: 68–87.

Monteuuis G, Wong JJL, Bailey CG, Schmitz U, Rasko JEJ. 2019. The changing paradigm of intron retention: regulation, ramifications and recipes. *Nucleic Acids Research* **47**: 11497–11513.

Muñoz MJ, Santangelo MSP, Paronetto MP, Mata M de la, Pelisch F, Boireau S, Glover-Cutter K, Ben-Dov C, Blaustein M, Lozano JJ, et al. 2009. DNA Damage Regulates Alternative Splicing through Inhibition of RNA Polymerase II Elongation. *Cell* **137**: 708–720.

Nagar P, Islam MR, Rahman MA. 2023. Nonsense-Mediated mRNA Decay as a Mediator of Tumorigenesis. *Genes* **14**: 357.

Naro C, Bielli P, Pagliarini V, Sette C. 2015. The interplay between DNA damage response and RNA processing: the unexpected role of splicing factors as gatekeepers of genome stability. *Front Genet* **6**. https://www.frontiersin.org/articles/10.3389/fgene.2015.00142/full (Accessed December 23, 2018).

Nayak RR, Bernal WE, Lee JW, Kearns MJ, Cheung VG. 2014. Stress-induced changes in gene interactions in human cells. *Nucleic Acids Research* **42**: 1757–1771.

Neefjes J, Jongsma MLM, Paul P, Bakke O. 2011. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol* **11**: 823–836.

Nguyen AT, Szeto C, Gras S. 2021. The pockets guide to HLA class I molecules. *Biochem Soc Trans* **49**: 2319–2331.

Oltean S, Bates DO. 2014. Hallmarks of alternative splicing in cancer. *Oncogene* **33**: 5311–5318.

Ou F-S, Michiels S, Shyr Y, Adjei AA, Oberg AL. 2021. Biomarker Discovery and Validation: Statistical Considerations. *J Thorac Oncol* **16**: 537–545.

Owen-Schaub L. 2001. Soluble Fas and cancer. *Clinical Cancer Research* **7**: 1108–1109.

Paganetti H. 2023. A review on lymphocyte radiosensitivity and its impact on radiotherapy. *Frontiers in Oncology* **13**. https://www.frontiersin.org/articles/10.3389/fonc.2023.1201500 (Accessed December 18, 2023).

Palumbo S, Comincini S. 2013. Autophagy and ionizing radiation in tumors: The "survive or not survive" dilemma. *Journal of Cellular Physiology* **228**: 1–8.

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics* **40**: 1413–1415.

Pandit S, Zhou Y, Shiue L, Coutinho-Mansfield G, Li H, Qiu J, Huang J, Yeo GW, Ares M, Fu X-D. 2013. Genome-wide Analysis Reveals SR Protein Cooperation and Competition in Regulated Splicing. *Molecular Cell* **50**: 223–235.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**: 417–419.

Paulsen RD, Soni DV, Wollman R, Hahn AT, Yee M-C, Guan A, Hesley JA, Miller SC, Cromwell EF, Solow-Cordero DE, et al. 2009. A Genome-wide siRNA Screen Reveals

Diverse Cellular Processes and Pathways that Mediate Genome Stability. *Molecular Cell* **35**: 228–239.

Paz S, Ritchie A, Mauer C, Caputi M. 2020. The RNA binding protein SRSF1 is a master switch of gene expression and regulation in the immune system. *Cytokine & Growth Factor Reviews*. http://www.sciencedirect.com/science/article/pii/S1359610120302240 (Accessed November 7, 2020).

Pishesha N, Harmand TJ, Ploegh HL. 2022. A guide to antigen processing and presentation. *Nat Rev Immunol* **22**: 751–764.

Quintens R, Verreet T, Janssen A, Neefs M, Leysen L, Michaux A, Verslegers M, Samari N, Pani G, Verheyde J, et al. 2015. Identification of novel radiation-induced p53-dependent transcripts extensively regulated during mouse brain development. *Biol Open* **4**: 331–344.

Rainer J, Gatto L, Weichenberger CX. 2019. ensembldb: an R package to create and use Ensembl-based annotation resources. *Bioinformatics* **35**: 3151–3153.

Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A, et al. 2013. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**: 172–177.

Rearick D, Prakash A, McSweeny A, Shepard SS, Fedorova L, Fedorov A. 2011. Critical association of ncRNA with introns. *Nucleic Acids Research* **39**: 2357–2366.

Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. 2006. GenePattern 2.0. *Nat Genet* **38**: 500–501.

Reits E, Griekspoor A, Neijssen J, Groothuis T, Jalink K, van Veelen P, Janssen H, Calafat J, Drijfhout JW, Neefjes J. 2003. Peptide Diffusion, Protection, and Degradation in Nuclear and Cytoplasmic Compartments before Antigen Presentation by MHC Class I. *Immunity* **18**: 97–108.

Reits EA, Hodge JW, Herberts CA, Groothuis TA, Chakraborty M, K.Wansley E, Camphausen K, Luiten RM, de Ru AH, Neijssen J, et al. 2006. Radiation modulates the peptide repertoire, enhances MHC class I expression, and induces successful antitumor immunotherapy. *Journal of Experimental Medicine* **203**: 1259–1271.

Rieger KE, Chu G. 2004. Portrait of transcriptional responses to ultraviolet and ionizing radiation in human cells. *Nucleic Acids Res* **32**: 4786–4803.

Risso D, Ngai J, Speed TP, Dudoit S. 2014. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology* **32**: 896–902.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**: e47.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**: R25.

Robinson TJ, Freedman JA, Abo MA, Deveaux AE, LaCroix B, Patierno BM, George DJ, Patierno SR. 2019. Alternative RNA Splicing as a Potential Major Source of Untapped

Molecular Targets in Precision Oncology and Cancer Disparities. *Clin Cancer Res* clincanres.2445.2018.

Rock KL, Reits E, Neefjes J. 2016. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends Immunol* **37**: 724–737.

Rødningen OK, Overgaard J, Alsner J, Hastie T, Børresen-Dale A-L. 2005. Microarray analysis of the transcriptional response to single or multiple doses of ionizing radiation in human subcutaneous fibroblasts. *Radiother Oncol* **77**: 231–240.

Rogalska ME, Vivori C, Valcárcel J. 2022. Regulation of pre-mRNA splicing: roles in physiology and disease, and therapeutic prospects. *Nat Rev Genet* 1–19.

Roninson IB, Broude EV, Chang BD. 2001. If not apoptosis, then what? Treatment-induced senescence and mitotic catastrophe in tumor cells. *Drug Resist Updat* **4**: 303–313.

Rückert M, Flohr A-S, Hecht M, Gaipl US. 2021. Radiotherapy and the immune system: More than just immune suppression. *Stem Cells* **39**: 1155–1165.

Sadek J, Omer A, Hall D, Ashour K, Gallouzi IE. 2019. Alternative polyadenylation and the stress response. *Wiley Interdiscip Rev RNA* **10**: e1540.

Sandell LL, Zakian VA. 1993. Loss of a yeast telomere: arrest, recovery, and chromosome loss. *Cell* **75**: 729–739.

Sanford JR, Wang X, Mort M, VanDuyn N, Cooper DN, Mooney SD, Edenberg HJ, Liu Y. 2009. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res* **19**: 381–394.

Santivasi WL, Xia F. 2014. Ionizing Radiation-Induced DNA Damage, Response, and Repair. *Antioxidants & Redox Signaling* **21**: 251–259.

Savage KI, Gorski JJ, Barros EM, Irwin GW, Manti L, Powell AJ, Pellagatti A, Lukashchuk N, McCance DJ, McCluggage WG, et al. 2014. Identification of a BRCA1-mRNA Splicing Complex Required for Efficient DNA Repair and Maintenance of Genomic Stability. *Mol Cell* **54**: 445–459.

Schaue D, McBride WH. 2015. Opportunities and challenges of radiotherapy for treating cancer. *Nat Rev Clin Oncol* **12**: 527–540.

Schwenzer H, Abdel Mouti M, Neubert P, Morris J, Stockton J, Bonham S, Fellermeyer M, Chettle J, Fischer R, Beggs AD, et al. 2021. LARP1 isoform expression in human cancer cell lines. *RNA Biology* **18**: 237–247.

Schwerk C, Schulze-Osthoff K. 2005. Regulation of apoptosis by alternative pre-mRNA splicing. *Mol Cell* **19**: 1–13.

Sciarrillo R, Wojtuszkiewicz A, Assaraf YG, Jansen G, Kaspers GJL, Giovannetti E, Cloos J. 2020. The role of alternative splicing in cancer: from oncogenesis to drug resistance. *Drug Resistance Updates* 100728.

Seiler M, Peng S, Agrawal AA, Palacino J, Teng T, Zhu P, Smith PG, Buonamici S, Yu L. 2018. Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. *Cell Rep* **23**: 282-296.e4.

Sha D, Jin Z, Budzcies J, Kluck K, Stenzinger A, Sinicrope FA. 2020. Tumor Mutational Burden (TMB) as a Predictive Biomarker in Solid Tumors. *Cancer Discov* **10**: 1808–1825.

Shanbhag NM, Rafalska-Metcalf IU, Balane-Bolivar C, Janicki SM, Greenberg RA. 2010. ATM-dependent chromatin changes silence transcription in cis to DNA double-strand breaks. *Cell* **141**: 970–981.

Shaul O. 2017. How introns enhance gene expression. *The International Journal of Biochemistry & Cell Biology* **91**: 145–155.

Shenasa H, Bentley DL. 2023. Pre-mRNA splicing and its cotranscriptional connections. *Trends in Genetics* **0**. https://www.cell.com/trends/genetics/abstract/S0168-9525(23)00122-1 (Accessed June 5, 2023).

Sheng J, Zhao Q, Zhao J, Zhang W, Sun Y, Qin P, Lv Y, Bai L, Yang Q, Chen L, et al. 2018. SRSF1 modulates PTPMT1 alternative splicing to regulate lung cancer cell radioresistance. *EBioMedicine* **38**: 113–126.

Shkreta L, Michelle L, Toutant J, Tremblay ML, Chabot B. 2011. The DNA damage response pathway regulates the alternative splicing of the apoptotic mediator Bcl-x. *J Biol Chem* **286**: 331–340.

Shraibman B, Kadosh DM, Barnea E, Admon A. 2016. Human Leukocyte Antigen (HLA) Peptides Derived from Tumor Antigens Induced by Inhibition of DNA Methylation for Development of Drug-facilitated Immunotherapy *. *Molecular & Cellular Proteomics* **15**: 3058–3070.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**: 539.

Singh G, Kucukural A, Cenik C, Leszyk JD, Shaffer SA, Weng Z, Moore MJ. 2012. The Cellular EJC Interactome Reveals Higher-Order mRNP Structure and an EJC-SR Protein Nexus. *Cell* **151**: 750–764.

Smart AC, Margolis CA, Pimentel H, He MX, Miao D, Adeegbe D, Fugmann T, Wong K-K, Van Allen EM. 2018. Intron retention is a source of neoepitopes in cancer. *Nat Biotechnol* **36**: 1056–1058.

Smirnov DA, Brady L, Halasa K, Morley M, Solomon S, Cheung VG. 2012. Genetic variation in radiation-induced cell death. *Genome Res* **22**: 332–339.

Smirnov DA, Morley M, Shin E, Spielman RS, Cheung VG. 2009. Genetic analysis of radiation-induced changes in human gene expression. *Nature* **459**: 587–591.

Smith AO, Ju W, Adzraku SY, wenyi L, Yuting C, Qiao J, Xu K, Zeng L. 2021. Gamma Radiation Induce Inflammasome Signaling and Pyroptosis in Microvascular Endothelial Cells. *J Inflamm Res* **14**: 3277–3288.

Sohrabi-Jahromi S, Söding J. 2021. Thermodynamic modeling reveals widespread multivalent binding by RNA-binding proteins. *Bioinformatics* **37**: i308–i316.

Somma MP, Andreyeva EN, Pavlova GA, Pellacani C, Bucciarelli E, Popova JV, Bonaccorsi S, Pindyurin AV, Gatti M. 2020. Moonlighting in Mitosis: Analysis of the Mitotic Functions of Transcription and Splicing Factors. *Cells* **9**: 1554.

Soneson C, Love MI, Robinson MD. 2016. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. https://f1000research.com/articles/4-1521 (Accessed December 12, 2023).

Sprung CN, Li J, Hovan D, McKay MJ, Forrester HB. 2011. Alternative Transcript Initiation and Splicing as a Response to DNA Damage. *PLoS One* **6**. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3198437/ (Accessed January 2, 2020).

Stanley RF, Abdel-Wahab O. 2022. Dysregulation and therapeutic targeting of RNA splicing in cancer. *Nat Cancer* **3**: 536–546.

Steel GG, McMillan TJ, Peacock JH. 1989. The 5Rs of Radiobiology. *International Journal of Radiation Biology* **56**: 1045–1048.

Steensma DP, Wermke M, Klimek VM, Greenberg PL, Font P, Komrokji RS, Yang J, Brunner AM, Carraway HE, Ades L, et al. 2021. Phase I First-in-Human Dose Escalation Study of the oral SF3B1 modulator H3B-8800 in myeloid neoplasms. *Leukemia* **35**: 3542–3550.

Stopfer LE, Mesfin JM, Joughin BA, Lauffenburger DA, White FM. 2020. Multiplexed relative and absolute quantitative immunopeptidomics reveals MHC I repertoire alterations induced by CDK4/6 inhibition. *Nat Commun* **11**: 2760.

Sundaramoorthy S, Vázquez-Novelle MD, Lekomtsev S, Howell M, Petronczki M. 2014. Functional genomics identifies a requirement of pre-mRNA splicing factors for sister chromatid cohesion. *The EMBO Journal* **33**: 2623–2642.

Syljuåsen RG, Jensen S, Bartek J, Lukas J. 2006. Adaptation to the ionizing radiation-induced G2 checkpoint occurs in human cells and depends on checkpoint kinase 1 and Polo-like kinase 1 kinases. *Cancer Res* **66**: 10253–10257.

Tacke R, Manley JL. 1995. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *The EMBO Journal* **14**: 3540–3551.

Tam AS, Stirling PC. 2019. Splicing, genome stability and disease: splice like your genome depends on it! *Curr Genet* **65**: 905–912.

Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR. 1999. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* **96**: 2907–2912.

Tan K, Stupack DG, Wilkinson MF. 2022. Nonsense-mediated RNA decay: an emerging modulator of malignancy. *Nat Rev Cancer* **22**: 437–451.

Thariat J, Hannoun-Levi J-M, Sun Myint A, Vuong T, Gérard J-P. 2013. Past, present, and future of radiotherapy for the benefit of patients. *Nature Reviews Clinical Oncology* **10**: 52–60.

The ENCODE Project Consortium. 2012. An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* **489**: 57–74.

The UniProt Consortium. 2023. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* **51**: D523–D531.

Thomas JD, Polaski JT, Feng Q, De Neef EJ, Hoppe ER, McSharry MV, Pangallo J, Gabel AM, Belleville AE, Watson J, et al. 2020. RNA isoform screens uncover the essentiality and tumor-suppressor activity of ultraconserved poison exons. *Nat Genet* **52**: 84–94.

To NH, Nguyen HQ, Thiolat A, Liu B, Cohen J, Radosevic-Robin N, Belkacemi Y, On behalf of the TransAtlantic Radiation Oncology Network (TRONE) & Association of

Radiotherapy and O of the MA (AROME). 2022. Radiation therapy for triple-negative breast cancer: emerging role of microRNAs as biomarkers and radiosensitivity modifiers. A systematic review. *Breast Cancer Res Treat* **193**: 265–279.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511–515.

Tsang LWK, Hu N, Underhill DA. 2010. Comparative Analyses of SUV420H1 Isoforms and SUV420H2 Reveal Differences in Their Cellular Localization and Effects on Myogenic Differentiation. *PLoS One* **5**. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3012056/ (Accessed June 10, 2019).

Tu Y, Li F, Wu C. 1998. Nck-2, a novel Src homology2/3-containing adaptor protein that interacts with the LIM-only protein PINCH and components of growth factor receptor kinase-signaling pathways. *Mol Biol Cell* **9**: 3367–3382.

Tusher VG, Tibshirani R, Chu G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *PNAS* **98**: 5116–5121.

Tzelepis K, De Braekeleer E, Aspris D, Barbieri I, Vijayabaskar MS, Liu W-H, Gozdecka M, Metzakopian E, Toop HD, Dudek M, et al. 2018. SRPK1 maintains acute myeloid leukemia through effects on isoform usage of epigenetic regulators including BRD4. *Nat Commun* **9**: 5378.

Van Nostrand EL, Freese P, Pratt GA, Wang X, Wei X, Xiao R, Blue SM, Chen J-Y, Cody NAL, Dominguez D, et al. 2020. A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**: 711–719.

Vignard J, Mirey G, Salles B. 2013. Ionizing-radiation induced DNA double-strand breaks: A direct and indirect lighting up. *Radiotherapy and Oncology* **108**: 362–369.

Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. 2019. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res* **47**: D339–D343.

Vitting-Seerup K, Porse BT, Sandelin A, Waage J. 2014. spliceR: an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data. *BMC Bioinformatics* **15**: 81.

Vitting-Seerup K, Sandelin A. 2019. IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics* **35**: 4469–4471.

Vitting-Seerup K, Sandelin A. 2017. The Landscape of Isoform Switches in Human Cancers. *Mol Cancer Res* **15**: 1206–1220.

Wagner K-U. 2022. Know thy cells: commonly used triple-negative human breast cancer cell lines carry mutations in RAS and effectors. *Breast Cancer Research* **24**: 44.

Wagner RE, Frye M. 2021. Noncanonical functions of the serine-arginine-rich splicing factor (SR) family of proteins in development and disease. *BioEssays* **43**: 2000242.

Wahba A, Ryan MC, Shankavaram UT, Camphausen K, Tofilon PJ. 2017. Radiation-induced alternative transcripts as detected in total and polysome-bound mRNA. *Oncotarget* **9**: 691–705.

Wahid M, Pratoomthai B, Egbuniwe IU, Evans HR, Babaei-Jadidi R, Amartey JO, Erdelyi V, Yacqub-Usman K, Jackson AM, Morris JC, et al. 2023. Targeting alternative splicing as a new cancer immunotherapy-phosphorylation of serine arginine-rich splicing factor (SRSF1) by SR protein kinase 1 (SRPK1) regulates alternative splicing of PD1 to generate a soluble antagonistic isoform that prevents T cell exhaustion. *Cancer Immunol Immunother*. https://doi.org/10.1007/s00262-023-03534-z (Accessed November 21, 2023).

Wan Y, Zheng X, Chen H, Guo Y, Jiang H, He X, Zhu X, Zheng Y. 2015. Splicing function of mitotic regulators links R-loop–mediated DNA damage to tumor cell killing. *Journal of Cell Biology* **209**: 235–246.

Wang B. 2014. Analyzing cell cycle checkpoints in response to ionizing radiation in mammalian cells. *Methods Mol Biol* **1170**: 313–320.

Wang B-D, Lee NH. 2018. Aberrant RNA Splicing in Cancer and Drug Resistance. *Cancers (Basel)* **10**: E458.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**: 470–476.

Wang H-H, Wu Z-Q, Qian D, Zaorsky NG, Qiu M-H, Cheng J-J, Jiang C, Wang J, Zeng X-L, Liu C-L, et al. 2018a. Ablative Hypofractionated Radiation Therapy Enhances Non-Small

Cell Lung Cancer Cell Killing via Preferential Stimulation of Necroptosis In Vitro and In Vivo. *Int J Radiat Oncol Biol Phys* **101**: 49–62.

Wang JR, Holt J, McMillan L, Jones CD. 2018b. FMLRC: Hybrid long read error correction using an FM-index. *BMC Bioinformatics* **19**: 50.

Wang Y, Chen D, Qian H, Tsai YS, Shao S, Liu Q, Dominguez D, Wang Z. 2014. The splicing factor RBM4 controls apoptosis, proliferation, and migration to suppress tumor progression. *Cancer Cell* **26**: 374–389.

Weinzierl AO, Lemmel C, Schoor O, Müller M, Krüger T, Wernet D, Hennenlotter J, Stenzl A, Klingel K, Rammensee H-G, et al. 2007. Distorted relation between mRNA copy number and corresponding major histocompatibility complex ligand density on the cell surface. *Mol Cell Proteomics* **6**: 102–113.

Wen C, Tian Z, Li L, Chen T, Chen H, Dai J, Liang Z, Ma S, Liu X. 2022. SRSF3 and HNRNPH1 Regulate Radiation-Induced Alternative Splicing of Protein Arginine Methyltransferase 5 in Hepatocellular Carcinoma. *International Journal of Molecular Sciences* **23**: 14832.

Westholm JO, Lai EC. 2011. Mirtrons: microRNA biogenesis via splicing. *Biochimie* **93**: 1897–1904.

Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York https://ggplot2.tidyverse.org.

Willforss J, Chawade A, Levander F. 2019. NormalyzerDE: Online Tool for Improved Normalization of Omics Expression Data and High-Sensitivity Differential Expression Analysis. *J Proteome Res* **18**: 732–740.

Williams AJ, Blacklow SC, Collins T. 1999. The zinc finger-associated SCAN box is a conserved oligomerization domain. *Mol Cell Biol* **19**: 8526–8535.

Williamson L, Saponaro M, Boeing S, East P, Mitter R, Kantidakis T, Kelly GP, Lobley A, Walker J, Spencer-Dene B, et al. 2017. UV Irradiation Induces a Non-coding RNA that Functionally Opposes the Protein Encoded by the Same Gene. *Cell* **168**: 843-855.e13.

Withers HR. 1975. The Four R's of Radiotherapy. In *Advances in Radiation Biology* (eds. J.T. Lett and H. Adler), Vol. 5 of *Advances in Radiation Biology*, pp. 241–271, Elsevier https://www.sciencedirect.com/science/article/pii/B9780120354054500128 (Accessed December 4, 2023).

Wj L, Y H, Ya L, Bd Z, My L, Yq Z, Gs H, Yh H, Jj Y, Bl X, et al. 2023. Targeting PRMT1-mediated SRSF1 methylation to suppress oncogenic exon inclusion events and breast tumorigenesis. *Cell reports* **42**. https://pubmed.ncbi.nlm.nih.gov/37938975/ (Accessed November 13, 2023).

Wu JY, Tang H, Havlioglu N. 2003. Alternative Pre-mRNA Splicing and Regulation of Programmed Cell Death. In *Regulation of Alternative Splicing* (ed. P. Jeanteur), *Progress in Molecular and Subcellular Biology*, pp. 153–185, Springer, Berlin, Heidelberg https://doi.org/10.1007/978-3-662-09728-1_6 (Accessed March 29, 2020).

Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**: 873–881.

Xiao H, Huang B, Yuan Y, Li D, Han L-F, Liu Y, Gong W, Wu F-H, Zhang G-M, Feng Z-H. 2007. Soluble PD-1 Facilitates 4-1BBL–Triggered Antitumor Immunity against Murine H22 Hepatocarcinoma In vivo. *Clinical Cancer Research* **13**: 1823–1830.

Xiao SH, Manley JL. 1998. Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. *EMBO J* **17**: 6359–6367.

Xu X, Zhou Y, Wei H. 2020. Roles of HLA-G in the Maternal-Fetal Immune Microenvironment. *Frontiers in Immunology* **11**. https://www.frontiersin.org/articles/10.3389/fimmu.2020.592010 (Accessed December 5, 2023).

Yang C-R, Leskov K, Hosley-Eberlein K, Criswell T, Pink JJ, Kinsella TJ, Boothman DA. 2000. Nuclear clusterin/XIP8, an x-ray-induced Ku70-binding protein that signals cell death. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 5907–5912.

Yang X, Coulombe-Huntington J, Kang S, Sheynkman GM, Hao T, Richardson A, Sun S, Yang F, Shen YA, Murray RR, et al. 2016. Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell* **164**: 805–817.

Yi X, Liao Y, Wen B, Li K, Dou Y, Savage SR, Zhang B. 2021. Abstract 1895: caAtlas: An immunopeptidome atlas of human cancer. *Cancer Research* **81**: 1895–1895.

Yin L, Duan J-J, Bian X-W, Yu S. 2020. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Res* **22**: 61.

Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M, et al. 2011. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**: 64–69.

Yoshimi A, Abdel-Wahab O. 2017. Molecular Pathways: Understanding and Targeting Mutant Spliceosomal Proteins. *Clin Cancer Res* **23**: 336–341.

Zhang H, Xu Q, Krajewski S, Krajewska M, Xie Z, Fuess S, Kitada S, Pawłowski K, Godzik A, Reed JC. 2000. BAR: An apoptosis regulator at the intersection of caspases and Bcl-2 family proteins. *PNAS* **97**: 2597–2602.

Zhang L, Cheng F, Wei Y, Zhang L, Guo D, Wang B, Li W. 2019a. Inhibition of TAZ contributes radiation-induced senescence and growth arrest in glioma cells. *Oncogene* **38**: 2788–2799.

Zhang Q-S, Manche L, Xu R-M, Krainer AR. 2006. hnRNP A1 associates with telomere ends and stimulates telomerase activity. *RNA* **12**: 1116–1128.

Zhang Y, Liu L, Qiu Q, Zhou Q, Ding J, Lu Y, Liu P. 2021a. Alternative polyadenylation: methods, mechanism, function, and role in cancer. *Journal of Experimental & Clinical Cancer Research* **40**: 51.

Zhang Y, Qian J, Gu C, Yang Y. 2021b. Alternative splicing and cancer: a systematic review. *Signal Transduction and Targeted Therapy* **6**: 1–14.

Zhang Y, Yan L, Zeng J, Zhou H, Liu H, Yu G, Yao W, Chen K, Ye Z, Xu H. 2019b. Pan-cancer analysis of clinical relevance of alternative splicing events in 31 human cancers. *Oncogene* **38**: 6678–6695.

Zhang Z, Krainer AR. 2004. Involvement of SR proteins in mRNA surveillance. *Mol Cell* **16**: 597–607.

Zhao Y, Mir C, Garcia-Mayea Y, Paciucci R, Kondoh H, LLeonart ME. 2022. RNA-binding proteins: Underestimated contributors in tumorigenesis. *Seminars in Cancer Biology* **86**: 431–444.

Zheng CL, Fu X-D, Gribskov M. 2005. Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. *RNA* **11**: 1777.

Zheng J, Li B, Wu Y, Wu X, Wang Y. 2023. Targeting Arginine Methyltransferase PRMT5 for Cancer Therapy: Updated Progress and Novel Strategies. *J Med Chem* **66**: 8407–8427.

Zheng X, Peng Q, Wang L, Zhang X, Huang L, Wang J, Qin Z. 2020. Serine/arginine-rich splicing factors: the bridge linking alternative splicing and cancer. *Int J Biol Sci* **16**: 12.

Zhou BB, Elledge SJ. 2000. The DNA damage response: putting checkpoints in perspective. *Nature* **408**: 433–439.

Zhou X, Wang R, Li X, Yu L, Hua D, Sun C, Shi C, Luo W, Rao C, Jiang Z, et al. 2018. Splicing factor SRSF1 promotes gliomagenesis via oncogenic splice-switching of *MYO1B*. *J Clin Invest* **129**: 676–693.

Zhu M, Yang M, Zhang J, Yin Y, Fan X, Zhang Y, Qin S, Zhang H, Yu F. 2021. Immunogenic Cell Death Induction by Ionizing Radiation. *Frontiers in Immunology* **12**. https://www.frontiersin.org/article/10.3389/fimmu.2021.705361 (Accessed April 27, 2022).

Zuo P, Manley JL. 1993. Functional domains of the human splicing factor ASF/SF2. *EMBO J* **12**: 4727–4737.

 IEDB.org: Free epitope database and prediction resource. http://www.iedb.org (Accessed December 17, 2023).

Chapter 7   Appendix

Table 7.1 potentially novel IR-responsive genes, exhibiting only isoform-level differential expression in response to IR and have not been linked to IR response previously.

Gene Name	Event Type	Event ID	q-value	Avg PSI No IR	Avg PSI 2hr Post IR	Avg PSI 6hr Post IR	IR Consequence
C14orf37	Cassette Exon	chr14:57834426:57834610:-@chr14:57830237:57832282:-@chr14:57828106:57828178:-	0.0013	0.69	0.80	0.69	Cassette Exon Included
FDX1L	Retained Intron	chr19:10287381:10287435:-@chr19:10287062:10287168:-	0.0024	0.46	0.51	0.31	Intron Retained
LOC339192	Cassette Exon	chr17:40654499:40654915:-@chr17:40654018:40654064:-@chr17:40653418:40653604:-	0.0011	0.21	0.42	0.26	Cassette Exon Included
C14orf80	Cassette Exon	chr14:105029961:105030116:+@chr14:105031217:105031315:+@chr14:105034739:105035359:+	0.0002	0.82	0.69	0.87	Cassette Exon Spliced Out
CMC1	Cassette Exon	chr3:28258088:28258307:+@chr3:28270226:28270260:+@chr3:28279786:28279875:+	0.0009	0.13	0.26	0.18	Cassette Exon Included
AP1G2	Alt Last Exon	8906@uc010aks.1uc012730.uc001wkn.1	0.0005	0.36	0.24	0.32	Downstream last exon used
MRPL32	Alt 5' Splice Site	chr7:42941079:42943365|42941260:+@chr7:42943446:42943981:+	0.0001	0.22	0.35	0.27	N/A
WDR73	Alt 5' Splice Site	chr15:82990583:82990419|82990208:-@chr15:82989706:82990071:-	0.0002	0.40	0.25	0.41	N/A
DNAAF5	Alt First Exon	54919@uc003sja.2@uc010krz.1uc003siz.2	0.0004	0.26	0.41	0.44	Upstream first exon used
DPH7	Alt First Exon	92715@uc010ncl.1@uc004cnk.1	0.0010	0.42	0.49	0.56	Upstream first exon used

PREPL	Alt First Exon	9581@uc002rui.2@uc002ruk.1	0.0006	0.62	0.57	0.45	Downstream first exon used
UBE3C	Alt First Exon	11065@uc002xpq.1@uc002xpl.1uc002xpm.1uc002xpo.1uc002xpp.1	0.0002	0.50	0.34	0.47	Upstream first exon used
SSU72	Alt Last Exon	29101@uc001age.1@uc001agd.1	0.0003	0.49	0.61	0.44	Upstream last exon used
TMEM201	Alt Last Exon	199953@uc001apy.1@uc001apz.1	0.0000	0.42	0.52	0.39	Upstream last exon used
ZFAND4	Alt Last Exon	93550@uc001jcr.2@uc001jco.2	0.0005	0.07	0.24	0.09	Upstream last exon used
ZFAND4	Alt Last Exon	93550@uc001jcr.2@uc001jcl.2uc001jcp.2uc001jcm.2uc009xmu.1uc001jcn.2	0.0004	0.08	0.26	0.11	Upstream last exon used
ABHD11	Cassette Exon	chr7:72789828:72790064:-@chr7:72789195:72789376:-@chr7:72788360:72788957:-	0.0003	0.90	0.76	0.92	Cassette Exon Spliced Out
BAZ2B	Cassette Exon	chr2:160043332:160043478:-@chr2:160018370:160018558:-@chr2:160012999:160013166:-	0.0003	0.94	0.82	0.93	Cassette Exon Spliced Out
COA1	Cassette Exon	chr7:43645573:43645805:-@chr7:43638349:43639532:-@chr7:43637136:43637906:-	0.0002	0.63	0.50	0.65	Cassette Exon Spliced Out
GUSBP1	Cassette Exon	chr5:21497649:21497732:+@chr5:21507328:21507467:+@chr5:21509223:21510322:+	0.0002	0.48	0.20	0.26	Cassette Exon Spliced Out
GUSBP1	Cassette Exon	chr5:21497649:21497732:+@chr5:21507328:21507512:+@chr5:21509223:21510322:+	0.0009	0.51	0.26	0.36	Cassette Exon Spliced Out
PACRGL	Cassette Exon	chr4:20323797:20324260:+@chr4:20335529:20335609:+@chr4:20338006:20339761:+	0.0012	0.59	0.74	0.65	Cassette Exon Included

PLEKHJ1	Cassette Exon	chr19:2187154:2187951:-@chr19:2186761:2186989:-@chr19:2184152:2185060:-	0.0013	0.17	0.17	0.28	Cassette Exon Included
RWDD2A	Cassette Exon	chr6:83960750:83961090:+@chr6:83961633:83961709:+@chr6:83962033:83965373:+	0.0013	0.17	0.33	0.18	Cassette Exon Included
SNX21	Cassette Exon	chr20:43897005:43897155:+@chr20:43902494:43902504:+@chr20:43902685:43905321:+	0.0004	0.50	0.47	0.31	Cassette Exon Spliced Out
SPDL1	Cassette Exon	chr5:168942935:168943471:+@chr5:168946781:168946954:+@chr5:168947976:168948425:+	0.0008	0.39	0.59	0.39	Cassette Exon Included
TESPA1	Cassette Exon	chr12:53664636:53664797:-@chr12:53664045:53664217:-@chr12:53654451:53654658:-	0.0012	0.50	0.38	0.43	Cassette Exon Spliced Out
WDR73	Cassette Exon	chr15:82990419:82990583:-@chr15:82990208:82990320:-@chr15:82989706:82990071:-	0.0000	0.28	0.12	0.28	Cassette Exon Spliced Out
ZCCHC10	Cassette Exon	chr5:132390088:132390195:-@chr5:132386432:132386497:-@chr5:132370350:132370511:-	0.0009	0.51	0.70	0.49	Cassette Exon Included
ZNF266	Cassette Exon	chr19:9390893:9390975:-@chr19:9390438:9390571:-@chr19:9390181:9390307:-	0.0008	0.55	0.36	0.51	Cassette Exon Spliced Out
ZNF821	Cassette Exon	chr16:70455542:70455646:-@chr16:70453180:70453346:-@chr16:70451084:70452076:-	0.0001	0.87	0.64	0.86	Cassette Exon Spliced Out
ZSCAN32	Cassette Exon	chr16:3387193:3387828:-@chr16:3383650:3383815:-@chr16:3380424:3380518:-	0.0013	0.82	0.67	0.86	Cassette Exon Spliced Out
CELF1	Mutually Exclusive Exons	chr11:47531120:47531367:-@chr11:47484986:47485025:-@chr11:47477581:47477732:-@chr11:47466965:47467152:-	0.0008	0.85	0.68	0.88	Upstream exon used

Gene	Event Type	Event ID					IR Consequence
COQ8A	Mutually Exclusive Exons	chr1:225194561:225194723:+@ chr1:225194833:225194904:+@ chr1:225203807:225203959:+@ chr1:225215701:225215886:+	0.0010	0.44	0.42	0.29	Upstream exon used
MRPL32	Retained Intron	chr7:42941079:42941260:+@chr7:42943288:42943365:+	0.0026	0.54	0.71	0.66	Intron Retained
TIMM21	Retained Intron	chr18:69973360:69973422:+@chr18:69973521:69973618:+	0.0014	0.20	0.36	0.22	Intron Retained
TMEM208	Retained Intron	chr16:65819899:65820061:+@chr16:65820201:65820285:+	0.0048	0.30	0.43	0.32	Intron Retained
TMEM208	Retained Intron	chr16:65819899:65820039:+@chr16:65820201:65820285:+	0.0027	0.40	0.59	0.44	Intron Retained
WDR73	Retained Intron	chr15:82990419:82990583:-@chr15:82989706:82990071:-	0.0024	0.50	0.36	0.47	Intron Spliced Out
WDR73	Retained Intron	chr15:82992772:82992860:-@chr15:82992125:82992189:-	0.0016	0.82	0.65	0.78	Intron Spliced Out
CC2D1B	Tandem 3'UTR	chr1:52590794:52591777:-@chr1:52588855:52590793:-	0.0004	0.42	0.56	0.47	N/A
SENP3-EIF4A1	Cassette Exon	chr17:7415415:7415521:+@chr17:7416326:7416906:+@chr17:7418302:7418350:+	0.0011	0.36	0.57	0.36	Cassette Exon Included
LINC00667	Alt 3' Splice Site	chr18:5228083:5229048:+@chr18:5230186|5231698:5231811:+	0.0002	0.51	0.38	0.51	N/A
LINC00847	Mutually Exclusive Exons	chr5:180190563:180190835:+@ chr5:180191342:180191412:+@ chr5:180192279:180192466:+@ chr5:180193918:180195796:+	0.0008	0.24	0.48	0.23	Downstream exon used

Event ID - This column has the MISO notation for identifying a particular event. All coordinates and isoforms are based on genome assembly hg18

For Cassette Exon: upstream exon coordinates:strand @cassette exon coordinates:strand @downstream exon coordinates:strand
For Retained Intron: upstream exon coordinates:strand @downstream exon coordinates:strand
For Alt First/Last Exon: UCSC isoform(s) using one first/last exon @isoform(s) using a different first/last exon. Refer to 'IR Consequence' column to know which isoform the PSI corresponds to.

For Mutually Exclusive Exons: coordinates of upstream exon @coordinates of upstream mutually exclusive exon @coordinates of downstream mutually exclusive exon @coordinates of downstream exon. Refer to 'IR Consequence' column to know which exon the PSI corresponds to.

For Alt 3' Splice Site: coordinates of upstream exon @cooridnate of upstream splice site|coordinate of downstream splice site @downstream coordinate of the exon with the 3' splice site

For Alt 5' Splice Site: coordinates of downstream exon @cooridnate of downstream splice site|coordinate of upstream splice site @upstream coordinate of the exon with the 5' splice site

For Tandem 3' UTR: coordinates of the first 3' UTR @coordinates of the alternative 3' UTR

q-value - Benjamini-Hochberg adjusted p-value with FDR of 5%

Avg PSI Baseline - Average PSI of an event before irradiation (baseline timepoint)

Avg PSI 2hr Post IR - Average PSI of an event 2 hours after irradiation

Avg PSI 6hr Post IR - Average PSI of an event 6 hours after irradiation

Table 7.2 List of (Ray et al. 2013) 80 RNA-binding proteins considered in this thesis

RNA-binding proteins				
A1CF	HNRNPA1	MATR3	RBM28	SNRPA
ANKHD1	HNRNPA1L2	MBNL1	RBM3	SRSF1
BRUNOL4	HNRNPA2B1	MSI1	RBM4	SRSF10
BRUNOL5	HNRNPC	PABPC1	RBM41	SRSF2
BRUNOL6	HNRNPCL1	PABPC3	RBM42	SRSF7
CNOT4	HNRNPH2	PABPC4	RBM45	SRSF9
CPEB2	HNRNPK	PABPC5	RBM46	TARDBP
CPEB4	HNRNPL	PABPN1	RBM5	TIA1
DAZAP1	HNRPLL	PCBP1	RBM6	TUT1
ENOX1	HuR	PCBP2	RBM8A	U2AF2

ESRP2	IGF2BP2	PPRC1	RBMS1	YBX1
FMR1	IGF2BP3	PTBP1	RBMS3	YBX2
FUS	KHDRBS1	QKI	SAMD4A	ZC3H10
FXR1	KHDRBS2	RALY	SART3	ZC3H14
FXR2	KHDRBS3	RBFOX1	SFPQ	ZCRB1
G3BP2	LIN28A	RBM24	SNRNP70	ZNF638

Table 7.3 Average (unnormalized) intensity values for 117 peptides that are presented in all replicates of one or two conditions, and absent from all replicates of all other conditions. The last two columns specify whether the peptide is a potential neoantigen (i.e. not previously identified in normal cells) and whether it is novel (i.e. not previously identified in cancer cells either).

Peptide Sequence	siSRSF1 0Gy	siCtl 0Gy	siSRSF1 10Gy	siCtl 10Gy	Gene	exclusive to	potential neoantigen	novel
ALADLSVAV	67696	57084	0	0	HTR7	0Gy	yes	yes
ILISSVASV	0	0	37356	40670	CUL4B	10Gy	yes	yes
SVLPSLPAI	0	0	179338	160943	MYPN	10Gy	yes	yes
QNSYKPY	0	25871	0	0	HNRNPD	siCtl 0Gy	yes	yes
RLPLQDVYKI	0	0	0	38555	EEF1A1	siCtl 10Gy	yes	yes
SGSGGSTYYADSVKGRFTI	0	0	0	7876	IGHV3-23	siCtl 10Gy	yes	yes
TPLTSMVVTKPE	0	0	0	29307	ITIH3	siCtl 10Gy	yes	yes
TFQQRFDARQKI	0	0	0	15895	POLDIP3	siCtl 10Gy	yes	yes
VAYPEALTL	0	0	0	57850	SLC6A9	siCtl 10Gy	yes	yes
DSKTLAA	0	0	0	22755	SMC2	siCtl 10Gy	yes	yes
ALIESYQNL	11890	0	16844	0	ABI1	siSRSF1	yes	yes
TEMVNGYEA	10090	0	14203	0	ANKRD13B	siSRSF1	yes	yes
GEVIVDGDVV	23513	0	28813	0	ATP2C1	siSRSF1	yes	yes
IEYPTLHVV	12708	0	16906	0	ZNHIT6	siSRSF1	yes	yes
GESGCTFLV	28591	0	27010	0	CAPN2	siSRSF1	yes	yes
GEWVEVVV	87039	0	94616	0	CAPN2	siSRSF1	yes	yes
ADITVVC	108512	0	100367	0	EEF1G	siSRSF1	yes	yes

SELTGFITT	30781	0	33204	0	ERGIC1	siSRSF1	yes	yes
SDVILEVL	55129	0	69312	0	GNL3L	siSRSF1	yes	yes
GEVIGINTL	34928	0	44148	0	HTRA3	siSRSF1	yes	yes
NEAAILSSL	51585	0	56153	0	IPO11	siSRSF1	yes	yes
HENIISYAA	17033	0	25126	0	LMLN	siSRSF1	yes	yes
YEDQYGVAL	24472	0	32905	0	NIBAN2	siSRSF1	yes	yes
VLGVIWGV	35717	0	29942	0	RAB5IF	siSRSF1	yes	yes
GLADVEANYV	24411	0	20753	0	SLC35F2	siSRSF1	yes	yes
HLPAEFPSL	23952	0	40997	0	SH2D5	siSRSF1	yes	yes
SELPVVISL	28821	0	28033	0	SUSD1	siSRSF1	yes	yes
KEFGAVSKVDF	26045	0	27160	0	UTP15	siSRSF1	yes	yes
LETNEIPSL	61899	0	77846	0	WRNIP1	siSRSF1	yes	yes
VEGTVKA	34978	0	0	0	ACACA	siSRSF1 0Gy	yes	yes
ILEELQKV	29748	0	0	0	ANKRD17	siSRSF1 0Gy	yes	yes
QIIRETFHL	24621	0	0	0	AP3S1	siSRSF1 0Gy	yes	yes
AMGLFCL	30703	0	0	0	ATP5MC1	siSRSF1 0Gy	yes	yes
MEMGAVAA	14175	0	0	0	BAG3	siSRSF1 0Gy	yes	yes
YGKTEVV	34762	0	0	0	CASKIN2	siSRSF1 0Gy	yes	yes
KKAAAKKK	16223	0	0	0	H1-4	siSRSF1 0Gy	yes	yes
GEADPSIQL	494192	0	0	0	HEATR6	siSRSF1 0Gy	yes	yes
SEFQISVV	55865	0	0	0	NXT1	siSRSF1 0Gy	yes	yes
EAEFLQKL	18331	0	0	0	PIP5K1A	siSRSF1 0Gy	yes	yes
VDDYTVRV	34201	0	0	0	PSMD14	siSRSF1 0Gy	yes	yes
NDFIQKI	55086	0	0	0	SGK1	siSRSF1 0Gy	yes	yes
SDKTAVL	26202	0	0	0	TACC1	siSRSF1 0Gy	yes	yes
SMIDPDIYL	18612	0	0	0	UBR3	siSRSF1 0Gy	yes	yes
QESRHSYPP	0	0	44544	0	AJUBA	siSRSF1 10Gy	yes	yes
KETERPSI	0	0	38976	0	ANLN	siSRSF1 10Gy	yes	yes
GDYEITLL	0	0	90568	0	ANXA3	siSRSF1 10Gy	yes	yes
SHVSPEV	0	0	33550	0	ARFGAP3	siSRSF1 10Gy	yes	yes
YEHKNTWSA	0	0	16173	0	ARMCX6	siSRSF1 10Gy	yes	yes
RELENSLHEA	0	0	21555	0	CENPE	siSRSF1 10Gy	yes	yes
VEFTSSGSVL	0	0	28090	0	DDX42	siSRSF1 10Gy	yes	yes
QEKAIPLAL	0	0	27901	0	DDX56	siSRSF1 10Gy	yes	yes
SEVQIAQL	0	0	98202	0	ESYT2	siSRSF1 10Gy	yes	yes
RELAGHTGYL	0	0	28288	0	GNB1	siSRSF1 10Gy	yes	yes

SGRGLGVGFGSGGGSSS	0	0	20819	0	KRT5	siSRSF1 10Gy	yes	yes
WEQQQGAVA	0	0	10333	0	NOP9	siSRSF1 10Gy	yes	yes
PLPAPSSPP	0	0	48727	0	NPDC1	siSRSF1 10Gy	yes	yes
LEDKPPAPP	0	0	36145	0	PAK2	siSRSF1 10Gy	yes	yes
SLPSKSFNI	0	0	29381	0	POLK	siSRSF1 10Gy	yes	yes
REKLPSSEVV	0	0	22300	0	TIFA	siSRSF1 10Gy	yes	yes
GEQIAQLIA	0	0	69370	0	TLN1	siSRSF1 10Gy	yes	yes
GENPEVPFP	0	0	69099	0	ATP6V0A1	siSRSF1 10Gy	yes	yes
NIRASMQQQQQLA	0	0	0	15185	CHTOP	siCtl 10Gy	yes	no
GPKFLKSGDAAIV	0	0	0	19978	EEF1A1	siCtl 10Gy	yes	no
VAAGVGEFEA	0	0	0	40310	EEF1A1	siCtl 10Gy	yes	no
YAYPQASAV	0	0	0	24864	FKBP15	siCtl 10Gy	yes	no
ILNPVNTNL	0	0	0	24348	KIF13B	siCtl 10Gy	yes	no
MPVNSEV	0	0	0	66230	NUF2	siCtl 10Gy	yes	no
REKIEASRNEL	13026	0	17015	0	ARL8B	siSRSF1	yes	no
AQTEVIATL	45131	0	62152	0	CXCL1	siSRSF1	yes	no
ILDESHERV	31900	0	32120	0	LSM8	siSRSF1	yes	no
ALAPVTIEV	67328	0	73377	0	DCHS1	siSRSF1	yes	no
YLNETFSEL	14150	0	15859	0	CAD	siSRSF1	yes	no
VEVDTFMEA	17882	0	17282	0	RRP7A	siSRSF1	yes	no
GMPDFLEKL	52769	0	31113	0	STARD7	siSRSF1	yes	no
SEIQNNISL	35573	0	45489	0	TANC2	siSRSF1	yes	no
YGKFFVT	60088	0	0	0	ARHGAP17	siSRSF1 0Gy	yes	no
HGFREGTTPKPK	21510	0	0	0	RPL37	siSRSF1 0Gy	yes	no
FILKKLDSI	20949	0	0	0	TOMM5	siSRSF1 0Gy	yes	no
YEYRHVML	0	0	16331	0	CKS2	siSRSF1 10Gy	yes	no
AELLMTSGV	0	0	26774	0	CNOT8	siSRSF1 10Gy	yes	no
AEVDLERTFTF	0	0	19797	0	GMPR	siSRSF1 10Gy	yes	no
GEIKRDFIA	0	0	32454	0	MAP4	siSRSF1 10Gy	yes	no
GMTHIVREV	0	0	31716	0	RPL3	siSRSF1 10Gy	yes	no
AEYKRYLEM	0	0	22888	0	SF3A3	siSRSF1 10Gy	yes	no
RDFQHVEK	0	0	23797	0	STAP2	siSRSF1 10Gy	yes	no
KEIDYQRELL	0	0	19602	0	UTP15	siSRSF1 10Gy	yes	no
TESAFHYEA	0	9093	0	11951	ITGB4	siCtl	no	no
ALDFEQEMAT	0	0	0	27974	ACTB	siCtl 10Gy	no	no
LLPEYLPYA	0	0	0	24295	AMBRA1	siCtl 10Gy	no	no

GVIKAVDKKAA	0	0	0	16064	EEF1A1	siCtl 10Gy	no	no
TYIKKIGYNPDTV	0	0	0	46132	EEF1A1	siCtl 10Gy	no	no
SRPGRGEPRFI	0	0	0	53081	HLA-A	siCtl 10Gy	no	no
QLKRQPAPPREA	0	0	0	12480	RPL21	siCtl 10Gy	no	no
VKVVKNKAYFKRY	0	0	0	18144	RPL5	siCtl 10Gy	no	no
LLLGKERFAGV	0	0	0	21930	RPS16	siCtl 10Gy	no	no
NAIINSGPREDST	0	0	0	25919	RPS5	siCtl 10Gy	no	no
RIVELISRV	0	0	0	23602	TOM1L2	siCtl 10Gy	no	no
YASSPGGVYATRSS	0	0	0	15961	VIM	siCtl 10Gy	no	no
MFETFNTPA	0	0	0	57167	ACTB	siCtl 10Gy	no	no
LEFEGGEVSL	28951	0	20284	0	AHNAK	siSRSF1	no	no
IIMDDEFQL	14610	0	15324	0	ARL2BP	siSRSF1	no	no
YLKDLIEEV	16423	0	30831	0	ATF4	siSRSF1	no	no
HEFDFIHDV	35818	0	37549	0	MFGE8	siSRSF1	no	no
GEVTNDFVM	18043	0	20430	0	RPL3	siSRSF1	no	no
NLDAAVYQV	34285	0	34015	0	SLC35A1	siSRSF1	no	no
TEIDPICAL	17802	0	24102	0	AHCYL1	siSRSF1	no	no
NQFNSFISV	28511	0	32843	0	SCYL2	siSRSF1	no	no
ALQDFLLSV	40254	0	0	0	FANCG	siSRSF1 0Gy	no	no
GLFEDTNL	41245	0	0	0	H3C1	siSRSF1 0Gy	no	no
QMISRIEYV	48352	0	0	0	CSNK1A1	siSRSF1 0Gy	no	no
GLVDQLVKA	25419	0	0	0	SARS1	siSRSF1 0Gy	no	no
SEILQANQL	61606	0	0	0	TBC1D23	siSRSF1 0Gy	no	no
CEGPKAVAA	0	0	36105	0	CSDE1	siSRSF1 10Gy	no	no
WEISQKTVL	0	0	13121	0	RRM1	siSRSF1 10Gy	no	no
GQGIHHAAGQVGKEAE	0	0	520789	0	SBSN	siSRSF1 10Gy	no	no
ILQEREYRL	0	0	50539	0	SMARCA4	siSRSF1 10Gy	no	no
FLKNELDNV	0	0	16990	0	TRAIP	siSRSF1 10Gy	no	no