MANCHESTER 1824
The University of Manchester

In silico modelling of radiobiological mechanisms in proton beam therapy

A thesis submitted to the University of Manchester for the degree of
Doctor of Philosophy
in the Faculty of Biology, Medicine and Health

2021

Samuel P Ingram
Division of Cancer Science
School of Medical Sciences

Contents

Word Count: 45,077

List of Figures

List of Tables

Abstract

Radiotherapy is premised on being able to deliver lethal doses to the tumour whilst sparing the healthy tissue surrounding it. Proton therapy is an increasingly utilised modality of radiotherapy which has a favourable dose distribution potentially lowering doses to normal tissues. A key question of radiobiology surrounds how the DNA responds to damage following irradiation of varying radiation qualities. A driver of this is to fully understand the relative biological effectiveness between photon and proton irradiation. If achieved, this would allow the leverage of the many more years of clinical experience we have with photons to be applied to proton therapy. However, the pursuit of quantifying relative biological effectiveness has been met with large uncertainties when explored experimentally, limiting clinical confidence. In turn, this has prompted a more mechanistic understanding of how the characteristics of different incident radiation damaging the DNA alters the cellular response. This mechanistic exploration challenges many areas of fundamental radiobiology as it requires 'solid foundations' to build upon.

In this thesis, several models were developed to tackle the issues in fundamental radiobiology modelling. Firstly, I develop a DNA Mechanistic Repair Simulator (DaMaRiS) model to describe how the two dominant DNA double-strand break repair pathways (Homologous Recombination and Non-Homologous End Joining) can function together. This model demonstrated how different repair pathways should not be thought of as purely antagonistic, as some of the steps in each pathway can be complementary, resulting in an entwined pathway. Secondly, a new methodology of modelling the DNA arrangements for radiation simulations was introduced in the form of solving Hi-C experimental data. This work, which culminated in the G-NOME solver software, demonstrated how cell types have significant differences in how their chromatin is arranged, leading to changes in how DNA damage would be distributed which is known to impact cellular fate. Finally, we purpose a computational methodology for characterising the miss-identification (miscounting) of DNA damage foci in immunofluorescence work. The developed tool, called PyFoci, identifies how the miscounting error is not constant and varies across experimental conditions, which when not accounted for, can result in perceived changes even when non-exist. These models in culmination allow for more clarity in the continued development to describe the DNA damage response. They present tools that can be adopted by others in the field for both scientific exploration and validation of their own approaches. This work continues the development of models which can offer clinical utility and promote the use of biological optimisation for patient treatments.

The University of Manchester
Samuel Peter Ingram

Doctor of Philosophy
"*In silico* modelling of radiobiological mechanisms in proton beam therapy"
21^{st} September 2021

Declaration of originality

I hereby confirm that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

i The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made *only* in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see `http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420`), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see `http://www.library.manchester.ac.uk/about/regulations/`) and in The University's policy on Presentation of Theses.

Acknowledgements

I would like to thank my supervisors, Michael Merchant, Karen Kirkby, Norman Kirkby and Ranald Mackay. Their continued support has enabled me to become a better scientist and gave me the confidence to pursue the less obvious routes. I look forward to continuing my relationship with each of them. I would also like to acknowledge the support of my academic advisor Adam Aitkenhead who has always been supportive of my well being. I would like to also acknowledge the financial support of the Science and Technology Facilities Council and the Christie charity.

Thank you to all my friends and colleagues in the PRECISE group. I have been very fortunate to be surrounded by such a friendly group which were always available for critical procrastination sessions. A special thanks to Ed. K. Smith who was always ready to discuss and compare ideas and opinions, the latter being the most entertaining. Thanks must also be given to John-William Warmenhoven and Nick Henthorn, both of which laid the foundation for much of the work I have carried out. We've lost countless afternoons discussing and arguing over modelling ideas, again here with the latter being the most entertaining.

I must thank Stephen McMahon, Jan Schuemann, Aimee McNamara and Harald Paganetti who have been incredible people to collaborate with. Their willingness to help, listen and encourage one another are perfect role models for how collaboration should be done.

Thank you to my friends and family for all of the unwavering support, especially in the times that prioritised work a little too much. Thanks to my Dad and Mum for all the check-ups and phone calls and the constant praise. A special thanks to my late grandparents for getting me here, I will forever live by your words of wisdom - "just do your best, that's all anyone can ask".

My final thanks, although words simply do not suffice, must be given to Victoria. Who has willingly sacrificed so much to support me in my pursuits. I would not be where I am now without you.

Chapter 1

Aims

Cancer is a multi-faceted field of study which draws in researchers from a whole range of fields looking to find a way of contributing something to the community. Therefore, it is not unheard of and is typically encouraged that interested researchers consider "discipline hoping" within their study of cancer. This can often bring new ways of thinking into a discipline and potentially bring with it new insights. This thesis is one such example of this, where I have attempted to leverage my background as a physicist to contribute to the field of radiobiology. Radiobiology is a topic of interest due to both its role as an inducer, but also as a treatment methodology through the use of radiotherapy. It is the latter that drew me to the area and is thus the topic of this thesis.

The delivery of radiotherapy has dramatically improved over the recent years, with technologies like intensity-modulated radiotherapy (IMRT), image-guided radiotherapy (IGRT) and proton beam scanning (PBS) enabling treatment planners to design increasingly optimised plans for the patient. The role of a clinical physicist in the UK is that of optimisation of treatment which can be in the form of dosimetric, delivery technique and logistics. Currently, the role of biological optimisation is typically carried out by the clinician, prior to treatment planning, in the form of patient stratification, dose tolerances and fractionation regimens. However, as there is an increasing desire to biologically optimise treatments for specific clinical endpoints, it would not be unreasonable to foresee an increased emphasis on biological optimisation at the treatment planning stage. To allow for this there is a requirement of ways to generalise plan metrics and patient metrics (e.g. age) to a desired clinical endpoint so they can be applied and evaluated within treatment plans. This requires the creation of both detailed and robust models which can be readily deployed and evaluated in the setting of busy radiotherapy departments.

Computational models, sometimes referred to as *in silico* models, are incredibly useful tools for modelling complex systems and as such are prime candidates for radiobiological modelling. To capture the required detail to perform endpoint specific prediction whilst remaining robust to the variety of patient metrics this thesis leverages mechanistic modelling approaches which attempt to detail biological steps rather than broad relationships. Therefore, to achieve this in the context of fundamental radiobiology the models developed aim to capture details about how radiation interacts with DNA. The DNA is a key target for damage when delivering radiotherapy to achieve the desired cell kill effect. Typically this is then reviewed clinically at a tissue level, but there is no doubt that mechanisms at the cell level are major drivers of the observed outcome. Therefore, we arrive at a starting point for our desired modelling - to bridge the gap between radiation in and cellular response. However, whilst the input and output of the model have been defined there are major outstanding research questions that hinder its success. Firstly, the cellular response will be dictated by the success or failure of the DNA damage response. A key part of this response is how DNA damage is repaired as this will heavily influence the cellular fate. This leads to questions of how various DNA repair pathways are activated and how they interface with one another. Through better mechanistic models of multiple DNA repair pathways, allows for a more sophisticated emulation of the DNA damage response and should inform subsequent

effects at both a cellular and tissue level. We must also appreciate the variation in the cellular response to radiation from different cell types, this is a topic that is well observed empirically, but has major mechanistic drivers which remain unclear. Furthermore, it is important to assess how the behaviours of these models are fitted, "good" experimental data is the linchpin to any "good" model, but there is often a disconnect between the robustness required for modelling an effect and that which is collected by the experimentalist. In culmination these make up the aims of which this thesis attempts to address:

- **DNA Repair -** develop a mechanistic model of how different DNA repair pathways interact with one another.
- **Cell-type Specificity -** within mechanistic models develop a way of accounting for differences in cellular radio-sensitivity.
- **Experimental Data Uncertainty -** propose a methodology for assessing the accuracy of key experimental data used in mechanistic modelling.

The following chapters will give a broader overview of these topics, explain how these aims were addressed and discuss the further developments needed. This thesis will be presented in journal format: chapter 2 is an introduction to the topic, chapter 3 is a published paper on DNA repair, chapter 4 is a published paper on cell-type specificity and chapter 5 is a draft paper on experimental data uncertainty. The final chapters summarise the work with final discussions and then the thesis is concluded.

Chapter 2

Introduction

2.1 Cancer Therapy

2.1.1 Biology of Cancer

Hallmarks of Cancer

Cancer is a disease that results in the uncontrolled proliferation of abnormal cells. It evades the arsenal of the biological defence mechanisms which have arrived from millions of years of evolution and progresses with selfish intent at the cost of the integrity of its host. This single-minded goal of cancer has allowed researchers to identify a series of characteristics that cells must possess to successfully continue their aberrant behaviour. These characteristics were outlined in the seminal work by Hanahan and Weinberg [1] in a review article called "The Hallmarks of Cancer". This review outlined six hallmarks that characterise the traits for cells to become cancerous and provides a strong fundamental description of the biology of cancer. Hanahan and Weinberg suggest that the six hallmarks of cancer are: self-sufficiency in growth signals, insensitivity to anti-growth signals, evasion of programmed cell death, limitless replicate potential, sustained angiogenesis, and tissue invasion and metastasis. However, it should be noted that the first five hallmarks are essential alterations to create tumours, which is defined as the abnormal growth of cells, with only the last hallmark being the difference between benign and cancerous tumours.

The first hallmark of cancer is the self-sufficiency in growth signals. Mitogenic growth signals are required to cause normal cells to leave a quiescent state and progress through the cell cycle and ultimately leading to cell division during mitosis. In normal cells, these growth signals are typically stimulated by exogenous stimuli, whereas in tumour cells this is thought to be reversed, with a larger dependence on endogenous growth signals. This enables the tumour cells to gain growth signal autonomy allowing self-driven cell proliferation to be achieved. The ability of the tumour cell to synthesise its own growth factors creates a positive feedback loop. However, it is also thought that tumour cells do have some exogenous stimuli which are caused by modification to the cell surface receptors. These surface receptors, which transduce growth signals into the cell interior, can alter during tumour development making them hyper-responsive to even ambient levels of growth factors. It has also been shown that some structural alterations of receptors found on some cancer cells can cause ligand-independent signalling [2]. Cancer cells can also modify their extracellular matrix receptor (integrin) expression to prioritise growth signalling [3]. These changes in the cell surface receptors and integrins activate the important SOS-Ras-Raf-MAPK pathway [3, 4], which is a highly characterised signal transduction pathway. The SOS-Ras-Raf-MAPK pathway is the signalling pathways that direct the internal machinery stimulating cell proliferation, along with many other important cell functions such as wound healing, tissue repair, cell motility and stimulation of angiogenesis [5]. It is ultimately the activation of the SOS-Ras-Raf-MAPK pathway, which has likely been deregulated in a variety of manners [1, 6], which in-acts the continued pressing for cell growth, division and differentiation. Understanding these growth factors and their role in tumour re-

population is not just important in the formation of cancer, but also in how tumours respond during the course of treatment.

The second hallmark of cancer is the insensitivity to anti-growth signals. As the first hallmark of cancer is often analogised to the accelerator pedal being stuck on, this second hallmark is best analogised as the breaks not working. Normal tissues can resist additional growth signals from causing unwanted cell proliferation by an ensemble of counter-acting anti-proliferative signals. This balance in signalling allows for cells to maintain their quiescent state and overall tissue homeostasis. The anti-growth signals can be found as either soluble growth inhibitors or inhibitors embedded on the extracellular matrix or surfaces of other proximal cells. The inhibitors signal through cell surface receptors and activate corresponding intra-cellular anti-proliferation pathways. To block proliferation the cell can be forced into a quiescent or postmitotic state, both of which would be detrimental to tumour growth. Proliferation management is often carried out as a cell cycle blockade at G1, stopping the replication of the DNA during S-phase. It is believed that this is carried out by the retinoblastoma protein (pRb) undergoing hyper-phosphorylation altering the function of E2F transcription factors, which hinders the progression from G1 to S phase [7]. It is thought that extrinsic factors which govern the successful activation of the pRb signalling pathway are disrupted in human tumours. An example of an extrinsic factor is TGFβ which is a soluble signalling molecule thought to prevent the inactivation of pRb. However, in some tumours, TGFβ responsiveness has been down-regulated due to reductions in viable receptors [8] or reduced signal transducer efficiency [9]. In either case, it has been shown that TGFβ is acting as a regulator for some DNA repair pathways and its alterations in cancers cells is a likely driver in the sensitivity of the cells to treatments [10].

The third hallmark of cancer is the evasion of programmed cell death. Whilst effective proliferation is important to help form a tumour mass, it is just as important for tumour cells to evade the programmed cell death pathway called apoptosis. Apoptosis is an important cellular mechanism available to most cells throughout the human body, able to trigger at will to eliminate defective, redundant or virus-infected cells. Therefore, it is thought that apoptosis provides a barrier for cancer induction, which in turn highlights why cancer cells evasion of apoptosis is necessary for the formation of tumours. The activation of apoptosis is carried out by a series of sensors that go on to trigger effector components that carry out the self-destructive pathway. The sensors monitor both the intra- and extra-cellular conditions along with pro-survival signals from cell-matrix or cell-cell adherence, disruption in either of these areas can elicit apoptosis. A key effector pathway for apoptosis is that mediated by the tumour suppressor protein p53, which is defective in many cancer types [11]. The apoptotic network is complex and it has been shown that p53 uses simultaneous singling to multiple points within the network to ensure regulation of apoptosis triggering [12].

The fourth hallmark of cancer is the limitless replicative potential. Whilst, the three previous hallmarks all focus on decoupling the cell fate from intra- or extra-cellular signalling, researchers identified a more intrinsic limitation to endless proliferation [13]. It was seen that cells could only undergo a finite amount of doubling before losing the ability to con-

tinue, arriving at a state called cell senescence. Cells entering senescence can be circumvented by deactivation of pRb and p53 tumour suppressor proteins, but after a duration, this results in a crisis state, in which the chromosomes become aberrant. This aberrant result is thought to be due to the loss of telomeres, which provide end of chromosome protection and get naturally eroded over successive cell divisions, which once devoid results in karyotypic disarray [14]. However, in cancer cells, it has been shown that continued cell doubling is enabled by telomere maintenance, where the telomeres are continually lengthened which aids in avoiding both senescence and chromosome aberration [15].

The fifth hallmark of cancer is sustained angiogenesis. The provision of oxygen and nutrients to cells are necessary for their survival. With most normal tissue cells being within 100 μm of a capillary blood vessel, it is clear that controlled spatial distribution of blood flow is important. Once the tissue has formed a process called angiogenesis controls additional blood vessel growth, regulated through environmental signalling. Typically this regulation is managed through a series of angiogenic promoters and inhibitors with a careful balance being reached in normal tissues. However, in tumours, the increasing size of cell mass requires an in-balance between angiogenic factors allowing the promotion of angiogenesis. For tumours to be successful in their aim to continually grow they require continued promotion of angiogenesis signalling. The inhibition of angiogenesis to keep up with tumour growth is one of the limiting factors to the size of tumours and has been poised as a possible mechanism to exploit in cancer treatments [16]. The chaotic distribution of blood vessels results in a large range of cells at varying oxic conditions which subsequently affects the cellular response to various treatment modalities.

The sixth hallmark of cancer is tissue invasion and metastasis. A key developmental point of cancer is when cells, called pioneers, are spawned with the express intent to spread to other areas of the body causing metastasis. Some sources report that this metastatic spread of the cancer is believed to account for up to 90% of cancer deaths [1, 17]. However, in a more recent analysis in Norway, it was found that the regional cancer deaths which had metastatic disease were closer to 66.7% [18]. In either case, the majority of cancer deaths are in patients who have metastatic spread, which is possible due to cancer cells innate ability to invade and metastases. Whilst invasion and metastatic spread are distinct processes they utilise the same mechanisms to enact their task. Broadly, this breaks down to either changing the cellular adherence properties or activation of extracellular proteases. The former involves modification of cell-to-cell adhesion molecules in an attempt to nullify or even reverse their operation causing cellular displacement from the tumour mass. The latter is the increased production of matrix-degrading proteases, which modify the surrounding extracellular matrix allowing for cancer cell invasion. It is in combination that these processes allow for the observed tissue invasion and metastasis seen in cancerous tumours.

Emerging Characteristics of Cancer

The initial six hallmarks of cancer were able to encapsulate the majority of cancer characteristics known at the time it was published [1]. However, over the last 20 years, additional characteristics have been identified as playing a pivotal role in cancer growth and induction. Some of which were identified in the follow-up review - "Hallmarks of Cancer: The Next Generation" published in 2011 by Hanahan and Weinberg [19]. In the updated review there was the addition of two emerging hallmarks and two enabling characteristics of cancer. The two additional emerging hallmarks are the reprogramming of cellular metabolism and the evasion of immune destruction, which are thought to be distinctly different to the core hallmarks, but irrefutably important steps in the pathogenesis of cancer. The two enabling characteristics identified are genome instability and tumour-promoting inflammation, which help facilitate the acquisition of the hallmarks of cancer.

The reprogramming of cellular metabolism can be thought of as ensuring the constant provision of fuel to keep up with the unregulated cellular proliferation in cancer cells. This is achieved by cancer cells reprogramming their glucose metabolism to rely on the glycolytic pathways to provide the required energy for tumour growth and maintenance [20]. At first, this metabolic switch appears to be somewhat counter-intuitive, as the glycolytic pathways have a reduced efficiency for energy production compared to the normal respiration process, thus requiring tumour cells to up-regulate glucose transporters to keep up with the energy consumption needs of the tumour [21]. However, whilst the production of energy is less efficient using the glycolytic pathways the process can take place at a much faster rate, outweighing the efficiency at the cost of overall glucose consumption. Furthermore, it has also been shown that these glycolytic pathways produce intermediate biosynthetic precursors, which aid in the production of macromolecules and organelles required for the production of new cells [22]. These two modifications to glucose metabolism are believed to outweigh the efficiency reduction to facilitate tumour growth.

Evading immune destruction is a pertinent characteristic for any solid tumour, as it is the immune system that monitors cells and tissues for aberrant behaviour. Therefore it becomes obvious that the immune system should have a role in resisting and/or eradicating the formation of tumours, halting the parthenogenesis to late-stage tumours and metastatic spread. The mechanisms by which tumours avoid immune detection are still being discovered and the reversal of some of the known mechanisms encapsulate an entire treatment modality called immunotherapy. The blanket description of cancer cells avoiding the immune system is too general, as the immune system does effectively eliminate the large majority of cells which produce significant immunogenic signalling [23]. However, due to the aggressive rate of proliferation, this creates a biasing that only allows the weakly immunogenic cells to colonise forming tumours predominately formed by immuno-eluding cells. This has resulted in a growing interest in using treatment modalities to damage cells, releasing DNA fragments following DNA repair into the micro-environment in the hope of re-triggering the immune response.

Genomic instability and mutations are the beneficiary groundwork that cancer can thrive in, with the disturbance of one or multiple genes involved in ensuring genome integrity, the likelihood of tumour formation can increase substantially. Defects in these "caretaker" genes include functional roles in the detection of DNA damage, DNA repair activation, and machinery that inactivates mutagens [24]. With the extraordinary ability of the entwined mechanisms to ensure the genome is maintained and free of defects, when one gene becomes mutated it can have a cascading effect leading to an increased chance of other genes mutating. It is this increasing cumulative probability with each caretaker gene mutation that can result in a widespread mutation of the genome. The mutation rate within a population of cells can be increased both by defects in the mechanisms which deal with mutagens and mechanisms which are involved with the cellular surveillance systems, ensuring cells programmed death is enacting for aberrant cells. Through the increased availability of genome sequencing, there have been many insights made of shared cancer genome landscapes that persist between cancer types. There are over 140 genes that promote tumour formation, with the primary driver genes being involved in core cellular functions, such as cell fate, cell survival and genome maintenance [25]. With the continued improvements in the techniques, resolution and availability of genome sequencing we should expect the continued discovery of the cancer genome landscape and the exploitation of this knowledge to trickle down into optimisation of cancer treatments. This will be supported by models which can incorporate details on both the genomic landscape along how alterations would alter the response to treatment.

Tumour-promoting inflammation is the utilisation of the cellular inflammatory response at the benefit of tumour pathogenesis. It has been known for some time that tumours contain immune cells at varying levels, it was originally thought that this was an attempted immune response, but was later discovered that these immune cells promote cancer pathogenesis. The infiltrated immune cells cause additional inflammation which triggers a response within the tumour microenvironment. These responses include the supply of molecular factors which aid in proliferation, survival, angiogenesis, invasion and metastasis [26]. Furthermore, in some cases it has been shown that inflammation promotes the initial stages of cancer pathogenesis accelerating the necessary production of malignant cell clones and induce pro-survival pathways, resulting in why inflammation is seen as an enabling characteristic.

Tumour Microenvironment

Initially, the majority of research was focused on understanding the intra-cellular mechanisms of cancer cells. However, when examining most tumours it becomes apparent that some of the mass is comprised of a whole range of normal non-malignant cells. In a proportion of cancers over 50% of the tumour mass is formed from these non-malignant cells [27]. It is the interactions between these non-malignant cells and malignant cells which forms the tumour microenvironment (TME), which has been demonstrated to have a large influence on the tumour pathogenesis [28]. This developed understanding of the multi-faceted

role of the TME has reinforced the necessity of considering cancers as tissue or an organ rather than just a mass of malignant cells. The non-malignant cell types often found within the TME are related to the immune, vasculature, lymphatics systems, as well as fibroblasts, pericytes and adipocytes. The heterogeneous selection of cells makes it difficult to uncover their roles, especially as some of their functions become aberrant due to the conditions of the TME. However, there have been some mechanisms elucidated within the literature and these will be briefly discussed here.

The T lymphocytes (T cells) are a key component in the immune system, with a direct role in killing infected cells and regulating the immune response. These cells are often located at the invasive tumour margin and in draining lymphoid organs. They are normally considered tumour antigens, but with most tumours claiming the hallmark of being able to evade the immune system, these cells may not be sensitive to the native malignant cells within the TME. However, large numbers of certain T cells (e.g. CD8+ memory T cells) are strongly associated with a good prognosis from cancer treatment [29]. Whilst T cells are most renowned for their immune activation roles, the T regulatory cells (Tregs) have immunosuppressive roles, this results in the observed correlation of high numbers of Tregs resulting in a worse prognosis in many cancers [27].

The B lymphocytes (B cells) are a type of white blood cell that produce antibodies to help the immune system kill infected cells. They are often found at the invasive tumour margin, draining lymph nodes or lymphoid structures. B cells are often described as having a "love-hate" relationship with cancer, playing a wide array of roles in tumour immunity which results in either tumour growth or an anti-tumour immune response [30]. Some of the difficulties in the consensus of the role for B cells is that it becomes difficult to distinguish functional and bystander B cells within the TME leading to contradictory correlations seen between B cell population and prognosis. However, recent studies are starting to elucidate the differences between pro- and anti-tumour roles for the B cells found in tertiary lymphoid structures which depend on if they are located in immature or mature compartments [31]. In immature structures, it is believed that B cells have fewer interactions with T cells, leading to a reduces T cell activity and this is reversed in mature structures. Whilst requiring further validation, there were promising results for this hypothesis in a recent study looking at sarcoma patients, in which B cells within tertiary lymphoid structures were the strongest prognostic factor, even in the context of high or low CD8 T memory cells [32].

The natural killer (NK) and natural killer T (NKT) cells are natural cytotoxic lymphocytes that play a major role in limiting tumour spread and surrounding tissue damage. They can be found infiltrated in the tumour stroma, but are usually not in direct contact with tumour cells. For many cancers, their presence has been shown to predict a good prognosis [33]. However, their presence in the TME does not express their own innate tumour-killing function and it is believed that this disablement is induced by malignant cell-derived transforming growth factor-beta (TGF-β) [29].

The myeloid cells are one of the most abundant cell lineages in the TME, with four main

subtypes: tumour-associated macrophages (TAMs), myeloid-derived suppressor cells (MDSCs), tumour-associated neutrophils (TANs) and dendritic cells. The TAMs are abundant in most human cancers and are often linked to pro-tumour functions, they like to accumulate in hypoxic and/or necrotic areas of the tumour and promote angiogenesis [34]. As expected this leads to the association between TAM abundance and poor prognosis [35] and there is evidence that TAM transcriptional signatures could be a suitable predictor of patient survival [36, 37]. The MDSCs are potent suppressors of various immune cells with a broad phenotype, they can increase the development of Tregs and can differentiate into TAMs. Therefore, MDSCs are commonly thought of as being pro-tumour, even if the mechanisms of action are not very direct. There is also some evidence for non-immunological functions of MDSC which include promotion of tumour angiogenesis, tumour-cell invasion and metastasis [38]. The TANs role in tumour pathogenesis is somewhat more controversial, there is evidence for both pro- and anti-tumour functions in both a direct and indirect manner. Whilst it is understood that TANs play a role in the initiation and development of cancer, most attempts of using them as predictors have resulted poorly. However, the importance of their location within the tumour is starting to be realised and they are still outlined as an important area of research for new immunotherapeutic targets [39]. Dendritic cells are important for antigen processing and presentation for the immune system, but it is thought that the ones within the TME are defective and cannot perform this role. It is thought that the hypoxic and inflammatory conditions of the TME impair the dendritic cell functions and can even suppress T cell responses [27].

The cancer-associated fibroblasts (CAFs) are fibroblasts that have been differentiated into myofibroblasts, a common process seen when tissues become injured. These CAFs are abundant in many TMEs, they secrete multiple types of growth factors, including the immunosuppressant TGF-β. The CAFs also secrete extracellular matrix components and remodelling enzymes which directly impact the available physical scaffolding for the TME. The distribution of CAFs can vary between cancers, in some cases, CAFs have been observed to branch throughout the tumour mass, whereas in others, they localise into dense regions around the malignant cells, this latter arrangement impedes the ability of anti-cancer drugs to reach the malignant cells [27]. The overall immune suppression and growth factor secretion mean that CAFs are typically thought of as pro-tumour cells within the TME.

Adipocytes are fat cells that are commonly found in adipose tissue and well are suited for the storing of energy. In breast cancer, adipocytes can make up a large proportion of the TME. Due to the energy-storing nature of fat cells, dysfunction adipocyte metabolism, which is closely linked to inflammation can form a more aggressive TME and stimulate the progression of tumour pathogenesis. Whereas, in some abdominal tumours it has been seen that adipocytes can actively aid in the recruitment and growth of malignant cells [40].

The pericyte cell wraps around endothelial cells providing structural support to blood vessels. Whilst pericytes provide key structural support of the vasculature required for advancing tumour growth, the low amounts of the cells create leaky blood vessels which correlates with poor prognosis and metastases. Hence, the pro-tumour effects of pericyte availabil-

ity, in terms of tumour growth, is thought to be better than the negative impact on prognosis which occurs when pericytes coverage is lacking.

Vascular endothelial cells are positioned as a monolayer on the inner lumen of blood vessels, in healthy tissues these cells help maintain the flow of blood. However, in tumours hypoxia and growth factors can cause these endothelial cells to become dysfunctional, which coupled with the chaotic network of blood vessels become leaky with abnormal blood flow. These abnormalities culminate in pro-tumour growth signalling and metastatic spread through the bloodstream [41].

Lymphatic endothelial cells form the structure of lymphatic vessels which become important as many tumours drive the production of new lymphatic vessels through lymphangiogenesis. Whilst, tumour cells can invade pre-existing nearby lymphatic systems, malignant cells and macrophages can trigger further lymphangiogenesis in the TME, this may make it easier for malignant cells to disseminate through the body to form distant metastases. This has resulted in both periphery lymphatics and lymphatics within TME being important when trying to develop anti-metastatic treatments [42]. Furthermore, it has also been acknowledged that the increased lymph flow in the TME can cause mechanical stress-induced changes in the surrounding cells and stiffens the extracellular matrix, which in turn alters/reduces the immune response [43].

2.1.2 Cancer Prevalence

Global

Cancer is a disease not bound to a geographic location. Globally it is believed that cancer is the second leading cause of death, responsible for about 1 in 6 deaths [44]. The Global Cancer Observatory (GCO) estimated that cancer caused 9.6 million deaths in 2018 and is expected to rise to 29.5 million deaths by 2040 [45]. These figures put into perspective both the humanitarian and financial burden of cancer across the world. The global cancer incidence across the world (Fig. 2.1) is not homogeneous with elevated levels predominately found in the western world. This remains true even when the incidence rates are age-standardised, controlling for the increase in average life expectancy for western counties [45], eluding to this not being just a problem of an ageing population. Cancer incidence increase can come from multiple aetiologies, both environmental and lifestyle factors have been shown to impact the rate of cancer induction [46]. Developing countries have lower levels of cancer incidence, but recent work has shown this may be changing rapidly [47]. This also outlines a major concern, that whilst the cancer rates increase in rapidly developing countries there may be disproportionate resources for the treatment of cancer when compared to the western world [48, 49].

Incidence rates are only one side of cancer, due to a concerted effort in the developments in the treatment of cancer, survival has increased in the vast majority of cancer sites [50]. This is evidenced by the increased incidence rates of the western world not corresponding

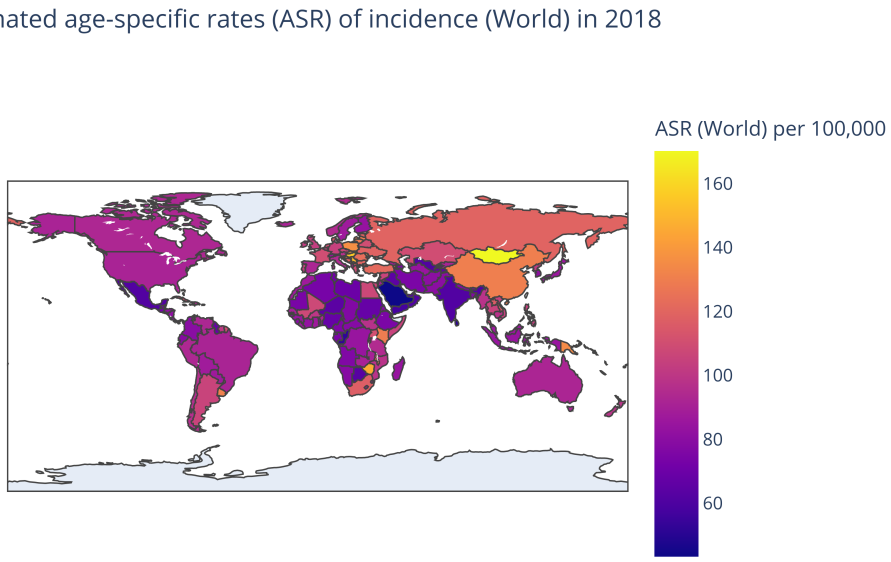

Figure 2.1: Global cancer age-standardised incidence rates, where the age groups weighting is based on the WHO standard population figures. Source data is from the global cancer observatory.

to elevated mortality rates (Fig. 2.2) when compared to the rest of the world. Whilst, there is some difference in the incident cancer types in these populations, the reduced mortality rate is likely due to larger financing for combating cancer within these countries [49]. This increased financing can allow for early-detection schemes, the widespread gold standard treatment being available and shorter wait times on treatment [51].

To compare a countries mortality rate and incidence rate you can evaluate the mortality-to-incidence rate (MIR) (Fig. 2.3). The MIR is the number of cancer deaths divided by the number of new cases for a given time period and can be a useful metric for comparison, but it is a poor analogue for survival [52]. The western countries have comparatively lower MIR than other countries with higher levels of MIR found in Africa and some regions of Asia. Whilst, there are some cancer type variations, it has been shown that areas with lower MIR tend to have increased financial provisions put towards cancer care [53].

The prognosis of cancer varies significantly across the various sites it can occur in. Whilst, number of deaths from a site is strongly influenced by the corresponding number of cases, some treatment sites have a significantly worse prognosis than others. For example, in a study looking at trends in cancer survival within the UK the 10-year survival of breast and lung cancer was 78% and 5% respectively in 2011 [50]. However, whilst this difference in prognosis is stark, it is only relevant if the treatment for better prognosis cancers is widely available. Therefore, it can be of interest to look at both the incidence and mortality broken down for the various cancer types (Fig. 2.4) geographically to see if this can be observed. This breakdown identifies at a global level, several cases have a strong correlation to the number of deaths. However, when the UK based prognosis example is compared to

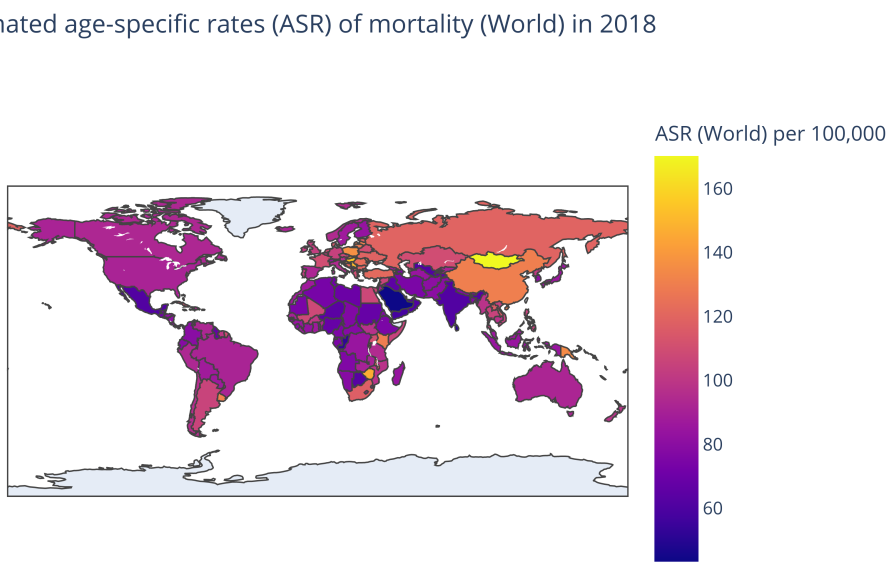

Figure 2.2: Global cancer age-standardised mortality rates, where the age groups weighting is based on the WHO standard population figures. Source data is from the global cancer observatory.

the global figures, whilst breast does have a reduced proportion of the number of deaths when compared to lung, these differences are not as large as may be expected. This highlights how the availability of early detection and treatment options have a larger impact on the prognosis of cancer which is not uniform across countries.

There is also an age component to the prevalence of certain cancer types. As an example, the global incidence and mortality rates were examined for a paediatric and teenage population (Fig. 2.5). In this population, there is an increased level of blood and central nervous system (CNS) cancers, with lung and breast cancer being extremely rare. This outlines another factor when talking about the prevalence of cancer, in that the landscape of cancers seen changes with age.

National - United Kingdom

In the United Kingdom (UK) the lifetime risk, which is the likelihood that a human will develop cancer over the entire span of their life, was estimated to be 54% and 48% for males and females respectively for those born in 1960 [54]. With over half of the adult population expected to be diagnosed with cancer at some point in their lifetime it is becoming increasingly unlikely that any person is not directly or indirectly be impacted by cancer. Furthermore, with the incidence rates of cancer increasing year on year [45] we should not expect this lifetime burden to lessen organically. The breakdown of cancer sites within the UK (Fig. 2.6) remains similar to that seen at a global level (Fig. 2.4), with a prominent incidence of breast, lung and colorectum (combination of colon, rectum and anus) cancers.

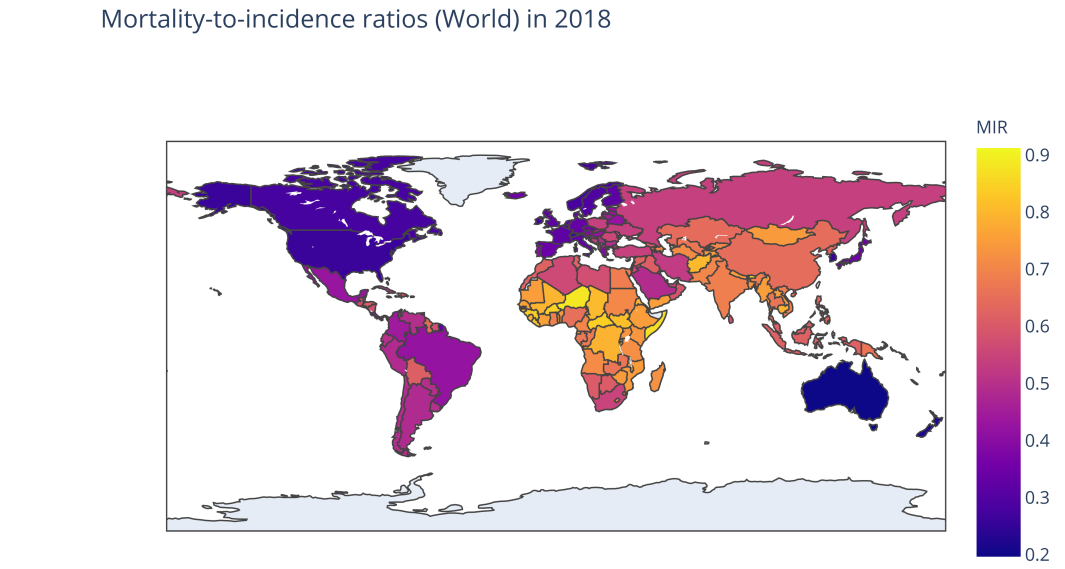

Figure 2.3: Global cancer mortality-to-incidence ratios, where the ratio is the age-standardised rates of mortality and incidence with age groups weighting being based on the WHO standard population figures. Source data is from the global cancer observatory.

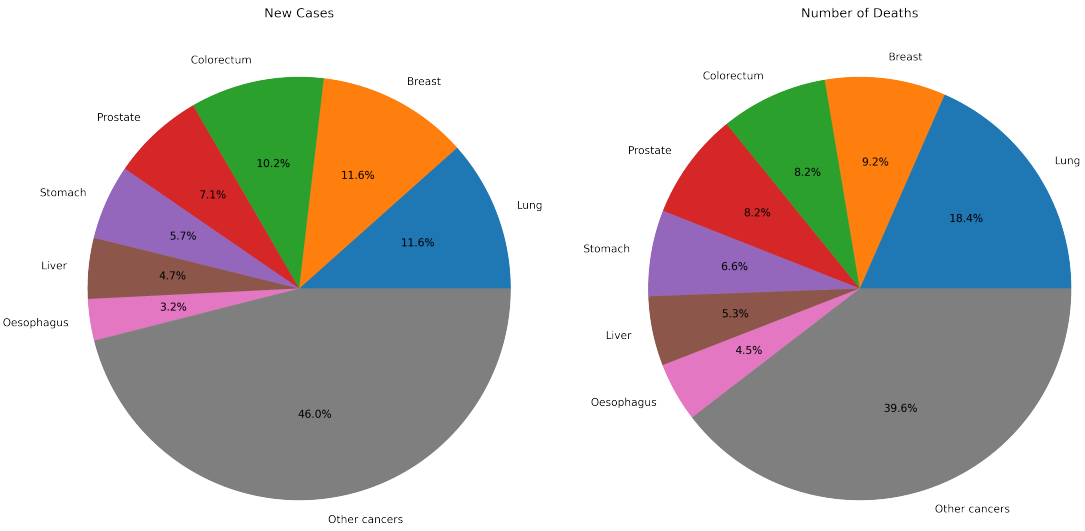

Figure 2.4: Global estimated number of new cases in 2018 for all ages. Source data is from the global cancer observatory. For the new cases N=18,078,957 and for the number of deaths N=9,555,027.

The MIR for the breakdown of sites is visualised in Fig. 2.7 for the UK. Some treatment sites show the UK to have very similar MIR to the levels seen in a global analysis. An example of this is lung cancer where the UK has a MIR of 0.81 and globally the MIR is 0.84. Whereas, sites such as Breast has a noticeably reduced MIR than the global level, with the UK MIR being 0.26 and the global MIR being 0.42. These differences identify aspects of variations in site prognosis which can be masked by different countries, specifically developing countries, availability of early detection and treatment options [48]. This difference comes with a financial cost tied to it, the European Union (EU) spent 51 billion Euros

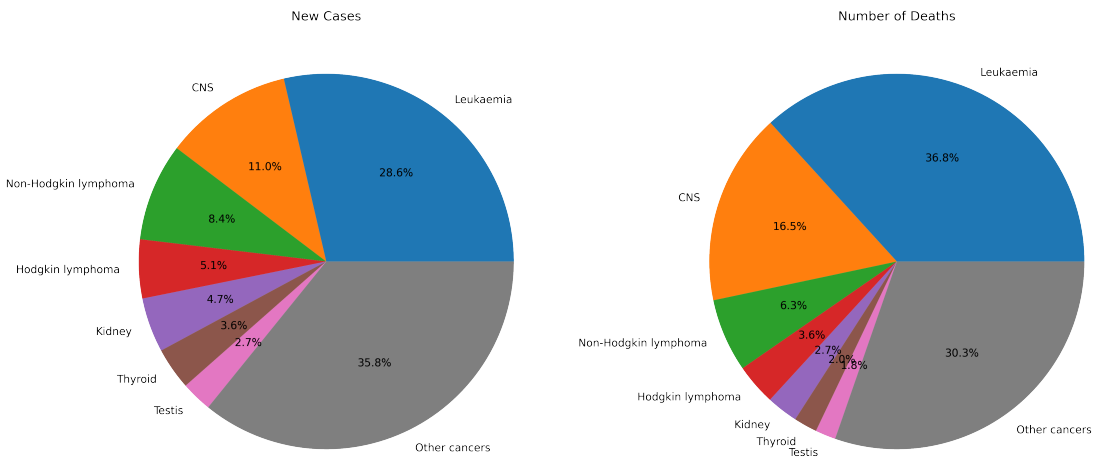

Figure 2.5: Global estimated number of new cases in 2018 for ages 0-19 years old. Source data is from the global cancer observatory. For the new cases N=272,603 and for the number of deaths N=101,724.

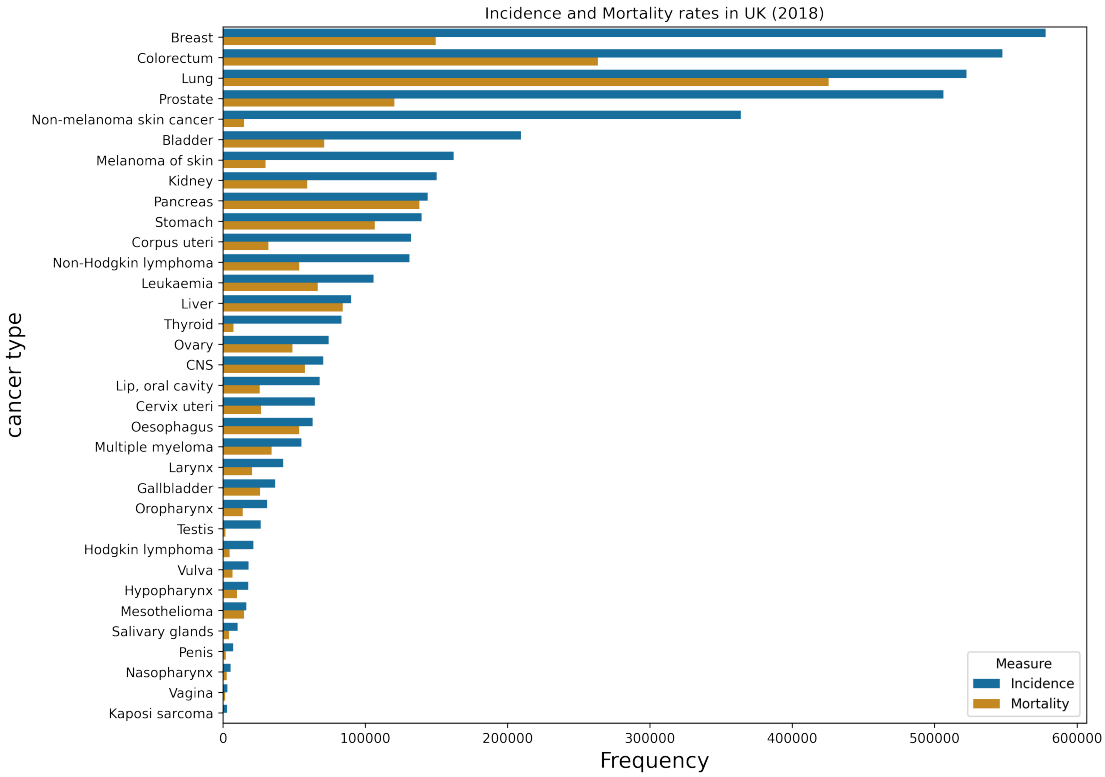

Figure 2.6: Number of incidence and mortality in the UK for 2018 for 34 cancer types. Top to bottom order is by incidence. Source data is from the global cancer observatory.

on cancer-related health care in 2009 making up 4% of total healthcare expenditure. The UK spent 5.2 billion Euros on cancer-related health care which is 3% of its total health-care expenditure. This difference in available spending between developed and developing countries is becoming increasingly apparent with only 5% of global cancer resources being spent by the developing countries which also make up the 80% of disability-adjusted life-years lost to cancer.

In contrast to the bleak image for developing countries, the UK has shown that cancer net survival increased substantially during the 40 years (1971-2011) in England and Wales [50]. Patients diagnosed in 1971-72 had a net 1-year survival chance of 50%, which over

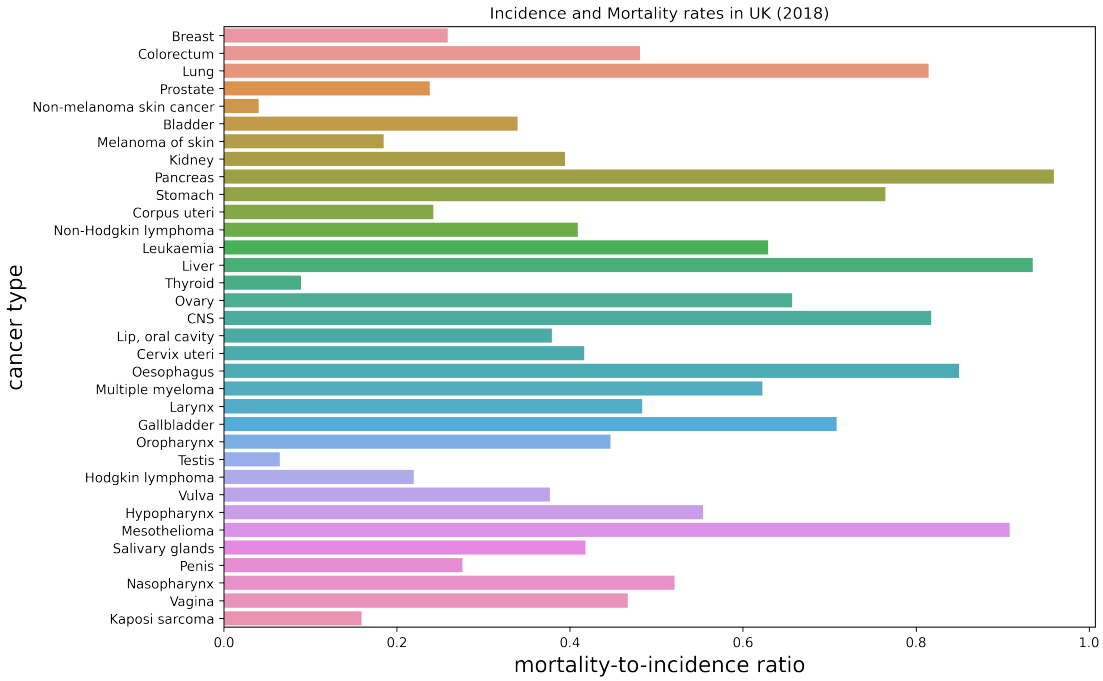

Figure 2.7: The mortality-to-incidence ratio of the same 34 cancer types in Fig. 2.6. Top to bottom order is consistent with Fig. 2.6. Source data is from the global cancer observatory.

40-years was increased to a net 10-year survival chance of 50%. This was achieved by an increased survival seen in the majority of cancer sites, with testis, malignant melanoma, prostate and breast having some of the largest increases in 10-year survival. However, this was not the case across all treatment sites, brain and lung both only saw minimal levels of improved survival, and the pancreas saw no change at all during the 40 years. Whilst, there is still a tremendously difficult road ahead in bettering the survival of cancer, for many treatment sites, the trends of reduced mortality are a major encouragement and promote the incredible work carried out by those working and researching cancer.

2.1.3 Treatment of Cancer

The treatment of cancer varies depending on location, grade, biology and availability. Deciding the best treatment is a complicated topic that has been continually refined and adapted to improve its efficacy. This has been a lengthy pursuit with some of the first records of cancer treatment coming from ancient Egypt 1500 BC, in those times cancer was seen as a terminal disease, but they were attempting to manage the symptoms with palliative treatment [55]. Whereas, more “modern” oncological treatment started to be defined in the 18th century with a surgeon called John Hunter documenting their attempts of surgically treating cancer [56]. There is now a large array of possible treatment techniques which all have their benefits and disadvantages and each of which are substantial clinical/research fields in their own right. These treatments can be prescribed as singular or in combination to target the specific nature of the cancer being targeted. In this section, there will be a brief overview of the most used types of cancer treatment.

Surgery

With surgery being the oldest discipline of oncological practices it has drastically changed since its early years. Some of the most significant changes were concerning the advent of general anaesthesia and antisepsis which drastically decreased the level of mortality related to the surgery itself [57]. This has been coupled with improved biological understanding and technological advances so that surgery is no longer just based on survival, but can often be conducted with the aim of preservation of function and quality of life without impeding on survival. Modern surgery consists of the precise excision of the tumour mass guided by medical imaging and histology and is a viable treatment technique for the majority of cancers. As of the current statistics held by Cancer Research United Kingdom (CRUK), surgery was used to remove their tumour in 45% of cases [58] and is the predominant treatment used for early-stage cancers (Stage I - Stage II).

Surgery remains prevalent, but the benefit from surgery alone has plateaued for several cancer types (e.g. radical mastectomy for breast cancer) which is where multi-modality treatment is being increasingly utilised (e.g. post-surgery radiotherapy) [57] to optimise patient treatment further. This is due to two innate problems with surgery, which are its invasive nature and potent treatment efficiency. The invasive nature of surgery can put a large burden on the patient's well-being as you need to gain physical access to the treatment area, this also makes it limited in how to deal with the metastatic disease which can be spread throughout the body. Performing multiple surgeries at several metastatic sites would place a large strain on the patient's well-being and in most cases may not be tolerable. The potent treatment efficiency can be an issue as during surgery it is a binary decision to either remove or keep tissue, whilst this may not be an issue for bulk tumours, the surrounding microscopic spread usually results in surgeons relying on resection margins. Resection margins by their nature will often remove healthy tissue which is "at-risk" of becoming cancerous but places a further burden on the patient's recovery after surgery. Attempting to understand suitable resection margins is extremely difficult and requires striking a balance between the risk of cancerous cells remaining and risk to the patient due to healthy tissue being removed. Increasingly, this may be tackled with the use of adjuvant therapy, in which a subsequent treatment (e.g. chemotherapy or radiotherapy) is used to kill any remaining cancerous cells. Sometimes this is decided post-surgery if there is the identification of cancer cells at the edge of the resection specimen, this is termed positive surgical margin (PSM) and can necessitate further adjuvant treatment [59].

Chemotherapy

Chemotherapy is the process of using chemical agents to kill cancerous cells and was first documented by the German chemist Paul Ehrlich in the early 1900s. One of the earliest diseases that Ehrlich treated with chemotherapy was syphilis which resulted in the development and use of arsenicals. Whilst there was initial interest in using chemotherapy for cancer treatment the community, including Ehrlich, was sceptical of its usefulness. In fact, it

was reported that the laboratory where chemotherapy for cancer was being developed had a sign over the door that read, “Give up all hope oh ye who enter” [60]. However, the development never stopped on cancer chemotherapy, but it wasn’t until after World War II that major clinical application was found in the form of nitrogen mustard. This was discovered by Alfred Gilman and Louis Goodman, two pharmacologists at Yale University who experimented on mice with transplanted lymphoid tumours. With a bit of convincing, Gilman and Goodman got a thoracic surgeon (Gustaf Lindskog) to administer nitrogen mustard to a patient with non-Hodgkin’s lymphoma. After seeing a marked response to the treatment it was quickly followed up by an initial pilot study in 1943 and then spread rapidly throughout the United States. However, the optimistic rapid uptake of nitrogen mustard as a chemotherapeutic was followed by wider community pessimism due to the inability of achieving cancer remission for any significant amount of time. The field of chemotherapeutics was stunted by this failure, with the majority of cancer treatment in the 1950s onward being surgery and radiotherapy. This was somewhat overcome by the success of using chemotherapeutics in children suffering from acute leukaemia, but it wasn’t until the 1960s that the use of combinational chemo-therapeutics showed how invaluable chemotherapy is for treating this disease. Before this complete remission in acute leukaemia was essentially zero compared to the 80% complete remission rate seen in the first combinational chemotherapeutics trial [60].

The concept of using chemotherapy as the primary treatment remains in leukaemia based diseases, but for other cancer sites, it is more commonly administered as an adjuvant treatment. Early tests of this were carried out in breast cancer, where the majority of patients present with only locoregional disease, but following regional treatment alone would often develop recurrences. Therefore, investigators began to use combinational chemotherapy as an adjuvant treatment in advanced breast cancer in the late 1960s, which had encouraging results. Over time this concept of using chemotherapy to aid as an adjuvant therapy has increased to other cancer sites and at the time of writing CRUK reports, chemotherapy is used in 28% of cancer cases [58] with increased use in late-stage cancers (stage III - stage Iv). With many variations of chemotherapeutic combinations given at all stages of the overall treatment (pre-, alongside- and post-treatment). Chemotherapy has multiple treatment roles in modern cancer therapy from aiding in tumour shrinkage before surgery to attempting to kill metastatic disease post-treatment. Due to the cytotoxic nature of chemotherapeutic agents and its delivery is often intravenously it is usually thought of as a systemic treatment. However, there is increasing research in targeting the delivery of chemotherapeutic agents which may reduce the systemic burden of the treatment and open up new avenues of treatment combinations in the future.

Radiotherapy

The discovery of the X-ray was announced by Roentgen in 1895 who was one of the first to truly demonstrate the possible use case of his discovery. However, whilst the announcement of the X-ray was carried out by Roentgen, the accidental discovery of the X-ray was

carried out by Plucker in 1859. After Roentgen seminal demonstration on the use of X-rays by photographing through his wife's hand, the applicability of using radiation in the medical field became almost instantly recognised for diagnostic purposes. The first record of using X-rays clinically for therapeutic reasons are claimed by Grubbé who in a retrospective study reports it occurred merely 60 days after Roentgen's announcement. Grubbé, who self-inflicted himself with dermatitis through experimentation with X-ray tubes, decided that this biological reaction to X-rays may have beneficial effects in lupus or cancerous lumps. As a second-year medical student Grubbé carried out this hunch on two patients one with carcinoma of the breast and the other with lupus vulgaris. Whilst this development of X-ray therapy continued there was an alternative type of radiotherapy developing in parallel called Curietherapy. Becquerel discovered radioactivity in 1895 at about the same time as Roentgen discovered the X-ray. The Curies discovery of radium and some incidental events of skin reactions to the material ultimately resulted in the radium being supplied to St. Louis Hospital in Paris for experimentation of radium-based therapies. Similarly to the treatments, Grubbé provided, the first clinical record of treatment was by a dermatologist for patients suffering from cutaneous lupus. Radium therapy was delivered in a variety of manners, but due to its expense was often used within multi-use applicators which could be positioned in close proximity, either superficially or via a cavity, to the tumour. The design of the radium applicators evolved to include filtration of the unwanted alpha and beta radiation allowing for clinicians to use the gamma radiation for treatment [61]. This type of treatment led to the development of brachytherapy techniques that are still used clinically today.

Modern radiotherapy most commonly comprises of external beam radiotherapy (EBRT) or brachytherapy for the treatment of a variety of cancer types. EBRT is the modern realisation of the tools used by Grubbé, where particle accelerators have increased the available energies and particles which can be used for treatment. Whereas, brachytherapy has evolved to use sealed-source control devices which can precisely position high-dose-rate sources to exact positions within an applicator for a predefined amount of treatment time. Radiotherapy is used as either a primary and adjuvant treatment in a wide range of cancer types. In the UK radiotherapy is used in approximately 27% of cancer cases [58] and is a suitable treatment for all stages of cancer.

Radiotherapy highly localised nature requires advanced anatomical imaging, such as computed tomography (CT), magnetic resonance imaging (MRI) and ultrasound to direct the radiation to the desired location. Radiotherapy has benefited from the technological advancements in these imaging modalities which are now highly integrated into the treatment process. As radiotherapy is delivered in fractions, which is due to the preferential radiation tolerance of healthy tissues compared to cancerous tissues when allowing DNA damage to repair, ensuring highly localised treatment is being delivered as planned is a difficult task. In the past radiotherapy has relied on rigorous patient immobilisation to ensure patient set-up between when the treatment is designed and when the patient is receiving treatment. However, with the increased availability of on-board (i.e. at treatment) imag-

ing, image-guided radiotherapy (IGRT) has become regular practice for most modern treatment facilities which offer improved quality assurance of patient set-up. The uncertainties related to patient set-up and anatomical changes remains one of the largest areas of potential for improvement in radiotherapy. These issues are aided by the field intrinsically being a technology-focused sector and is continuously involved upon in rapid development. The other area of potential improvement of treatment efficacy is improved understand and leveraging of the biological mechanisms from irradiation, aptly termed radiobiology. This strive to improve our radiobiological understanding is the aim and topic of this thesis and will be discussed in the subsequent sections.

2.2 Radiobiology

2.2.1 Radiation-Induced DNA Damage

Damage to DNA can occur from both endogenous and exogenous sources. Endogenous sources are primarily due to the metabolic function of the cell, whereas exogenous sources are often environmentally based. Exogenous sources include ingestion of chemicals within food, exposure to background radiation (e.g. radon and cosmic rays), exposure to UV radiation and air pollution. Ionising radiation by definition has sufficient ability to cause ionisation within a cell, these events result in the ejection of electrons which are the main cause of biological damage. The biological effects of radiation at the cell scale were captured in critical experiments carried out by Puck and Marcus in 1956, using x-rays they demonstrated an efficient methodology to conduct colony experiments using mammalian cells [62]. As electrons interact in matter they slow down, towards the ends of their track, interactions become more frequent and damage becomes more clustered spatially. This clustering of damage can occur within a few base pairs (bp) of the DNA [63]. Damage clustering like this is a unique characteristic of the ionising radiation paths, often referred to as tracks, and is thought to be a major factor in the modalities ability to cause cell death. Ionising radiation deposits its energy stochastically, but the principal target for radiation-induced cell killing is the DNA. Important experiments carried out by Warters and Hofer in 1977 using labelled iodine-125 showed that the cells environment outside the nucleus could withstand high doses without causing cell death. However, when the labelled iodine-125 was associated with DNA, entering within the nucleus, radiation-induced cell death significantly increased [64].

Though the basis of energy deposition causing ionisations to damage DNA is well established, the translation of damage to cell fate is complex with many steps in the pathway. A major influence is that various types of DNA damage can occur, most broadly damage can be sectioned to either a strand (the sugar-phosphate backbone) or a base (cytosine, guanine, adenine and thymine). Arriving at a coarse standard categorisation of DNA damage as single-strand breaks (SSBs), double-strand breaks (DSBs) and base damages. The type of damage changes the biological response, with DSBs being the most lethal when deal-

ing with ionising radiation, but also the rarest. A DSB is formed by a pair of SSBs within close proximity (< 10bp) on opposing strands. When dealing with DSBs the complexity influences the cells reaction to the damage, early studies in yeast identified that the number of DSBs have the largest impact on the cells fate post-irradiation [65]. This has resulted in numerous experiments focusing on the number of created DSBs to quantify the effects of ionising radiation [66–70]. To appreciate the biological resilience to the damage induced from ionising radiation it can be beneficial to quantify the approximate amount of damage that occurs from an exposure. A general approximation for 1 Gy of irradiation will cause 10^5 ionisations/cell with the following proportions: > 10000 damages to DNA bases, 1000 SSBs and 20-40 DSBs. This amount of DNA damage would kill about 30% of cells for a typical mammalian cell line [63] and highlights the efficiencies of cellular repair to this form of damage. Though this generalised approximation can be useful for demonstration purposes, it eludes any responsiveness to track-structure, radiation type and cellular variation - parameters which are known to influence DNA damage and response [71, 72] and should not be relied upon.

Post DNA damage there is a biological response aptly termed the DNA damage response (DDR) and involves several well-orchestrated pathways to identify and respond to the induced damage. There are two main components of the DDR which are categorised as either sensors or effectors of DNA damage. The sensors can identify damage by consistent surveillance of the genome, which once identified they signal for effectors to determine the fate of the cell. The effector pathways include: programmed cell death, attempted DNA repair and cell-cycle stalling. The DDR is a protein-based reaction which on the identification of damage proteins are recruited to the site and sub-nuclear regions forming radiation-induced foci (RIF). These RIF are instrumental to the experimental observation of DNA damage as some of the key proteins involved can be fluorescently tagged and observed microscopically. One of the key steps a cell must take after DNA damage has been detected is how it might repair the damage which is the topic of the next section.

2.2.2 DNA Repair

The type of damage will largely dictate the repair mechanism; base damage primarily repairs through base excision repair (BER); SSBs primarily repair through single-strand break repair (SSBR) pathways, and DSBs primarily repair through either non-homologous end joining (NHEJ) or homologous recombination (HR). These primary repair pathways also depend on genes available, protein presence, metabolic activity and cell cycle. The next sections will give a brief overview of some of the main DNA repair pathways found in normal human cells.

Base Excision Repair

One of the most prominent types of DNA damage following irradiation are damages to the bases. Base Excision Repair (BER) corrects small base lesions caused by oxidation,

deamination and alkylation, which have not significantly distorted the DNA helix structure [73]. This repair happens frequently and generally with success, as unrepaired base lesions can cause errors during the DNA replication process which can result in mutation or cell death. BER is split into two separate sub-pathways, called short-patching and long-patching. Whilst it is thought the decision depends on the initiating glycosylase and cell type the choice between these pathways is still not fully understood [73]. Short-patching involves the replacement of a single damaged base, whereas long-patching can replace a larger section of bases (2-10 bases on either side) surrounding the damaged base.

The BER process is similar regardless of the sub-pathways being used. Firstly, the base lesion is identified by DNA glycosylases and cleaved from the sugar-phosphate backbone. An endonuclease cuts the sugar-phosphate backbone creating a single-strand DNA which is then cleaned ready for ligation. The crated gap is then filled by a polymerase. Finally, the clean and filled single-strand DNA is ligated by DNA ligase I and III to complete the repair.

Nucleotide Excision Repair

When lesions are bulky and distort the DNA helix structure they require Nucleotide Excision Repair (NER). This type of repair can only act on a single strand and cannot be used to repair damage which includes both strands of the DNA. These types of damages are common following exposure to platinum-based chemotherapeutics or UV light. There are two sub-types of NER: global genome NER (GG-NER) and transcription-coupled NER (TC-NER). TC-NER is a rapid pathway specific to regions being actively transcribed, whereas other regions are typically repaired by GG-NER.

The difference between TC-NER and GG-NER is predominately in the process of how each pathway recognises the damage. As TC-NER takes place in actively transcribed regions its activation is managed by the stalling of RNA polymerases and results in the prompt induction of repair. This is in contrast to GG-NER, which relies upon the XPC-RAD23-CETN2 heterotrimer for damage recognition [74]. Once identified both TC-NER and GG-NER excise the damage contain oligonucleotide (approximately 23-32 nucleotides in length) and general replication factors fill in the remaining gap. The final step in NER is ligation of the newly synthesised patch to the original sequence by DNA ligase I [75].

Mismatch Repair

Mismatch Repair (MMR) is commonly thought of as a post-replication repair pathway used in the conjunction with the replication machinery and can correct for DNA replication errors. MMR explicitly targets replication errors that have resulted from a mismatch of bases on newly synthesised strands of DNA allowing for DNA polymerase to address any found errors. Loss of MMR results in inherited cancer susceptibility, as well as an increased incidence of sporadic cancers [76]. In a similar manner to BER, there are both short- and long-patching sub-pathway versions of MMR.

A newly synthesised strand containing mismatch is targeted for excision, the recognition is thought to be carried out by MutSα and MutSβ. Depending on the size of the mismatch MutSα repairs base-base mispairs of 1-2 nucleotides, whereas MutSβ handles indels of 1-15 bases [76]. After initial recognition MutSα and MutSβ get converted into sliding clamps that can diffuse along with the DNA, carrying out mismatch search. Once the extent of the error is identified the patch can be excised as either a short or long patch. After excision, the site uses high-fidelity polymerase and ligation by DNA ligase I to restore and correct the intact DNA duplex.

Non-Homologous End Joining

Non-Homologous End Joining (NHEJ) is a very fast repair pathway specifically for DSBs. It requires no homology between the break ends nor a template to synthesise from. Therefore, it can be utilised throughout the cell cycle and is one of the most utilised DSB repair pathways in human cells [77]. Whilst the NHEJ pathway is often thought of as the "error-prone" DSB repair pathway when compared to its counterpart homologous recombination (HR), the small deletions which are often caused are quite often innocuous to the cell integrity. NHEJ consists of two sub-pathways split by their requirement of resection: resection-independent NHEJ does not require significant resection of the break ends apart from that which is carried out by DNA-PKcs, whereas resection-dependent NHEJ requires resection of the DNA ends by CtIP/EXO1 which is initiated by Artemis [78].

All NHEJ pathways start similarly, with damage recognition by the Ku70/80 heterodimer which positions itself on the break ends. This is the scaffolding that attracts and enables DNA-PKcs to attach and start its nuclease activity stipping back the broken end to a ligatable state. In resection-dependent NHEJ it is at this point Artemis may be recruited and more significant end resection can be mediated by CtIP/EXO1. In either sub-pathway, this results in two ligatable DNA ends which utilise the XLF-XRCC4-Lig4 to ligate the ends and complete the repair.

Alternative Non-Homologous End Joining

An alternative pathway for NHEJ exists which uses short sections of homology between double-strand breaks to complete its repair. The pathways have multiple names as either alternative NHEJ (alt-NHEJ), alternative end-joining (alt-EJ) or micro-homology mediated NHEJ (MMEJ). The pathway has two sub-pathways available, one which is referred to as "simple" and the other as "synthesis-dependent". The synthesis-dependent pathways are thought to be the most prominent form of alt-NHEJ even though it is more error-prone than the simple pathway [79]. There is evidence that how alt-NHEJ utilises various substrates from other DSB repair pathways, whilst not preserving as much fidelity, may point to the pathway being a backup when the other pathways have failed [80].

The alt-NHEJ pathways all begin with end resection by MRN/CtIP which mirrors the ini-

tial start of HR. If the resection exposes micro-homology between the two break ends it can undergo the "simple" alt-NHEJ and be ligates by either DNA ligase I or III. However, if no micro-homology is exposed or detected then the break ends can undergo synthesis-dependent alt-NHEJ. This utilises polymerase to add nucleotides to the break ends until a micro-homology is formed and it can be ligated in the same manner as the "simple" alt-NHEJ pathway.

Homologous Recombination

The HR repair pathway is considered to be the "error-free" repair mechanism. The process requires the use of homologous undamaged DNA, usually obtained from a sister chromatid, to use as a template to repair the broken portion of DNA. The HR process can be broken down into three sequential phases: presynapsis, synapsis and postsynapsis (Fig. 2.8); where "synapsis" denotes the time point at which two homologous chromosomes are paired for recombination [81].

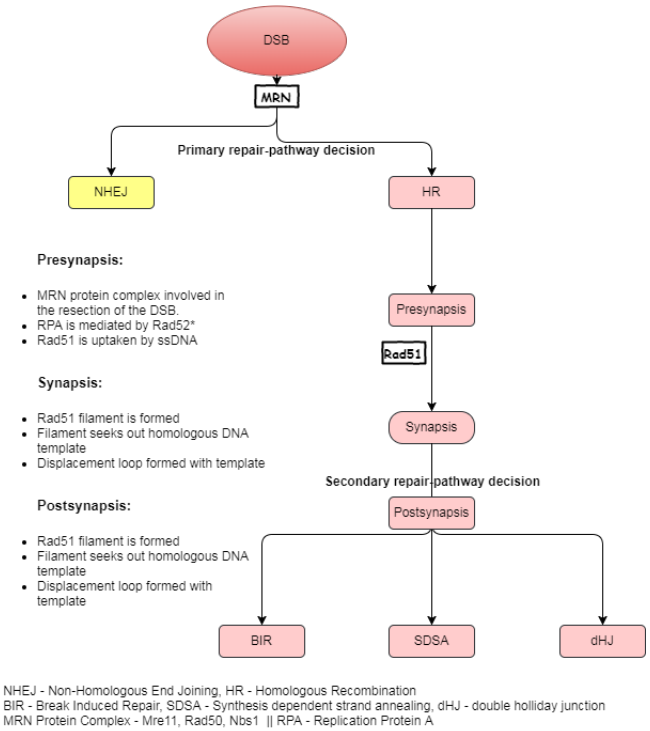

Figure 2.8: Simplified flow diagram of the Homologous Repair Pathway

The presynapsis phase includes the necessary processing of the DSB region to enable Rad51 uptake. The MRN protein complex is thought to play a role in the resection of the DSB to form single-strand regions around each side of the break. The resection makes a 3' overhang on either side of the break and replication protein A (RPA) immediately attempts to coat the single strands to avoid secondary structure formation. Regulation of the RPA formation on single-strand DNA (ssDNA) is carried out by several recombination mediators. If unregulated RPA would saturate the ssDNA and block the required uptake of Rad51 [82]. Recombination mediation is thought to possibly be the role of Rad52s augmentation of BRCA2 or Rad51 paralogues, though this has not been established in humans [82]. At the

point of Rad51 uptake to the ssDNA, the damaged DNA is ready to seek out a template to repair from within the synapsis phase.

The attachment of Rad51 to the ssDNA leads to the formation of a nucleoprotein filament. The filament is thought to 'hold' the damaged DNA strand and search for a homologous piece of DNA to use. The process of the filament finding and connecting the two homologous pieces of DNA is thought to be highly complex with the mechanisms not well established [83]. The strand invasion by the Rad51-ssDNA filament generates a displacement loop (D-loop) with the homologous undamaged DNA. The D-loop formation denotes the end of the synapsis phase and introduces several possible sub-pathways which can be executed in the postsynapsis phase.

The physical repair process takes place in the postsynapsis phase, there are several sub-pathways available when repairing using a homologous undamaged DNA: synthesis-dependent strand annealing (SDSA), break-induced replication (BIR) and double holiday junction (dHJ). In SDSA there are multiple invades, syntheses and disengages with the homologous DNA, one strand at a time until the damaged regions are repaired and then annealed with their second ends. As each strand is dealt with one at a time this method results in non-crossover repair, thus reducing the potential of genomic rearrangement. BIR occurs when damage is on a replication fork, as the strand is unwound for DNA replication, damage to a single strand will result in a DSB when the corresponding backbone is attempted to be synthesised. Without a second end for the synthesising DNA to anneal to, the D-loop forms a new replication fork with the invaded strand and copies the entire distal part of the chromosome resulting in the loss of heterozygosity. The aforementioned sub-pathways have dealt with single strands to resolve a DSB, the final dHJ pathway involves double-strand involvement between the damaged and homologous undamaged DNA. This forms a dHJ which can be resolved using an enzyme called resolvases to form crossover or non-crossover repair. Alternatively, the dHJ can be dissolved by BLM, topoisomerase 3 and cofactors which always results in non-crossover repair [84].

The possible outcomes from HR each have favourable action conditions and different post-repair consequences. Repair with crossover products includes the exchange of genetic material through the physical crossing of a DNA strand with a homologous partner. This crossover recombination event can be detrimental in the event of non-allelic recombination which involves two lengths of DNA that have a high sequence similarity but are not alleles. This confusion can cause deletions, inversions, or translocations which may result in detrimental mutations. An alternative consequence is related to BIR is the loss of heterozygosity which refers to the loss of one of the two versions of alleles. This is an issue when the lost gene is a tumour suppressor or the homologous replication copies a deleterious mutation. It is not entirely clear how each HR sub-pathway is enacted, but this remains an active research topic where some observations have been made. It has been demonstrated in S. cerevisiae that SDSA pathway out-competes BIR in mitotic DSB repair due to its speed of activation [85]. It is also thought that close proximity of the second end suppresses BIR, though this mechanism is unclear about how the information is communicated to the D-loop [86]. In S.

cerevisiae it has also been shown the dHJs only account for low levels of mitotic DSB repair [84]. In contrast, it is reported that dHJ is the most common sub-pathway within mammalian cells [87]. However, this does not mean the other sub-pathways are not critical as BIR deficient cells show significant radiosensitisation [88]. There are contrasting opinions and results seen which underpin the decision-making process within HR, indicating further work is required for a conclusive understanding.

2.2.3 Cell Fate Following Irradiation

Cells which are unable to successfully repair are generally killed or deactivated to ensure possible genomic instability is not proliferated. The killing of cells, similarly to repair, are controlled by a series of pathways each with their own specific role in preserving biological integrity. These pathways of cell death can be split into those which are self-initiated and those which are caused from injury-induced morphology alterations. The mechanism in which radiotherapy completes its therapeutic action utilises these different cell death pathways to varying amounts and is a critical step in the developing understanding of radiobiology.

Programmed Cell Death

The ability of cells to initialise a self-destructive process is important for maintaining tissue homeostasis by being the direct counterpart to cell proliferation. There are several pathways established, but the two most dominant pathways are apoptosis and Autophagy. The two pathways are distinct and have differing roles in radiotherapy.

Apoptosis was discovered in the 1970s and occurs in healthy adult tissues and focal elimination of cells during embryonic development [89]. The pathway can be triggered either intrinsically, through stress-induced mitochondria signalling, or extrinsically where cells can loose adhesion becoming detached from the extracellular matrix. The process of apoptosis takes place in two discrete structural stages. The first is where the cell condenses and becomes fragmented and the second the disposal of these fragments by other cells or shedding from epithelial-lined surfaces [89]. Both of these triggers have been shown to increase following irradiation [90].

Autophagy is generally activated to cope with nutrient deprivation and excessive amounts can be seen in people with neuro-degenerative diseases, infections and cancer. The pathway is subdivided into three major types: macroautophagy, microautophagy, and chaperone-mediated autophagy. Autophagy has a complex role in cancer development and acts as both a tumour promoter and suppressor. There are reported up-regulations of autophagy following radiotherapy, but similar to its role in cancer development, it has mixed reports positively and negatively affecting the therapeutic effect [91]. Therefore, it remains an open research question on how best to modulate the pathway in attempts to improve radiotherapy effectiveness.

Senescence

Cells which go into senescence are no longer able to proliferate, but maintain their metabolic activity. This mechanism was first described in 1961, were it was shown there is a finite number of cell divisions that can be undertaken before the cell halts it's proliferation [92]. This was an observation of replication senescence, whereas "premature" senescence, can be activated by cellular stress, such as non-lethal DNA damage following irradiation. In fact, this is one of the main objectives of this type of senescence and helps ensure damaged cells are not proliferated. Senescent cells are distinguished from normal or quiescent cells by their different morphology (large and flat) and chromatin conformation (additional heterochromatin regions). The latter gives the cell the ability to maintain it's metabolic activity, secreting senescence associated secretory phenotypes (SASP). The SASP include secretions of pro-inflammatory cytokines, chemokines, growth factorsand proteases and allows for the senescent cells to influence the surrounding microenviroment [93].

Senescence as a pathway is crucial when trying to avoid cancer induction as bypassing senescence is typically a required step by cancerous cells. In the context of cancer therapy, there is emerging evidence that the prominence of senescence depends on the cancer type. Cancer types such as lung and glioblastoma are thought to preferentially use senescence over apoptosis as the pathway of cell fate following chemotherapy and/or radiotherapy [93]. Furthermore, the long term implications of inducing senescence from treatment is not fully understood with evidence that these cells, if not cleared by the immune system, may result in a poorer outcome following treatment [94]. Therefore, combinational drug treatments which aid in the removal of the unwanted senescence cells following radiotherapy is an area of interest when trying to optimise cancer therapy.

Mitotic Catastrophe

Mitotic catastrophe is not formally a pathway, but is a situation a cell can find itself which goes onto triggering the apoptotic, senescent or necrotic pathways. As the name suggests, mitotic catastrophe is triggered from an event during mitosis, specifically structurally erroneous or incomplete repair of the DNA when attempting to divide. Typically this is avoided by the various cell-cycle checkpoints which attempt to halt the cell from progressing to mitosis whilst structurally altering DNA damage is present. However, if the checkpoints are defective or unable to regulate these damages when the cell attempts to undergo division it does not meet the standards to successfully do so. Due to the heterogeneous spectrum of DNA damage caused by radiation, it is believed that mitotic catastrophe is an important instigator of cell death following radiotherapy. Furthermore, cancer cells are more likely to have defective checkpoints and may be more susceptible to triggering this type of event [90].

Necrosis

Necrosis is typically thought of as a externally driven pathway of cell death. Though some studies do suggest some internal signalling may allow for the pathway to be orchestrated by a molecular system [95]. Necrosis presents itself distinctly to apoptosis as it causes cellular swelling, rather than shrinkage and result in the deformation of the membrane. It is also thought that it is heavily utilised by mitotic catastrophe and is regularly paired with it when talking about mechanisms of cell death for radiotherapy. Necrotic cells are of particular interest in radiotherapy as cluster of necrotic cells within a tissue are sometimes detectable using post-treatment imaging [96]. This allows for real-world detection of the biological response following treatment and has become quite a focus point for clinical evaluation [97].

2.2.4 Radiobiology in the Context of Radiotherapy

General Principles

Radiotherapy is premised on the ability to leverage the differentials in the radiobiological response of healthy and cancerous tissues. As discussed in the Biology of Cancer Section (2.1.1), for cancer to form it requires key alterations to its biological processes which are not found in healthy tissues. These differences can result in a reduced capacity for cancerous tissues to survive radiation damage. Over the history of radiotherapy, it has been the increasing understanding of radiobiology which has enabled biological optimisation of the way radiation is delivered. Most prominently this is demonstrated in the use of fractionated radiotherapy, whereby the prescription of radiation is delivered over an extended period of time in smaller amounts rather than a single dose. This attempts to leverage the differences in radio-response of healthy and cancerous tissues to increase the lethality of the radiation to the tumour whilst minimising radiation side-effects in the healthy tissue. This differential in response has been characterised as the four [98], five [99] and six [100] Rs of radiobiology.

The original four Rs of radiobiology proposed by Withers in 1975 [98] were redistribution, re-population, repair and re-oxygenation. Redistribution refers to the unsynchronised cell cycle population that is being irradiated. As the cell transitions through the cell cycle, there are intrinsic changes to its radiosensitivity, with S-phase being most resistant, late G2/Mitosis being most sensitive and G1 being less sensitive [101], resulting in preferential cell death for cells in sensitive portions of the cell cycle. If time is not allowed between irradiation's, the cells which are least radiosensitive can become synchronised with one another which may result in subsequent efficacies per irradiation. Through fractionation, the time between irradiation's allows for the cell populations to continue through their cell cycles at differing rates and become asynchronous so that the efficacy of each irradiation is less variable.

Re-population refers to the stimulation of new cells being generated within tissues as a re-

sponse to proximal cell killing. The rate at which this regeneration occurs varies between tissue types, but in terms of concern during a typical treatment length is only accounted for in tumours. This is due to many normal tissues having no appreciable proliferation activity during a standard treatment period (6 weeks) [98]. Whilst some normal tissues have steady-state proliferation's, such as skin, intestinal and oropharyngeal mucosa the benefits of explicitly accounting for their regeneration is not usually done in clinical practice. However, concerns over tumour re-population are accounted for as this can directly impact the efficacy of the treatment when ignored [102, 103]. Therefore, in well-proliferating tumour groups, any elongation of treatment should be accounted for with an additional total dose to obtain the same biological effectiveness as the originally planned treatment [104]. This is always at the cost of additional dose to the normal tissues and is thus avoided where possible through the use of bi-daily fractions or increased dose per fraction.

Repair in the context of radiotherapy refers to the cellular capabilities to survive and resolve sub-lethal damage following irradiation between fractions. The surviving fraction of a population of cells can often be characterised by a linear-quadratic relationship with dose, it is thought that compounded sub-lethal events result in the chance of becoming lethal events when combined. This results in dose-survival curves usually having an initial shoulder before achieving a steeper final slope when cell survival is plotted logarithmically. This relates to fractionation as if sub-lethal events are allowed to repair between fractionated radiation you can exploit the initial shoulder in response to preferentially spare cells with intact repair mechanisms. As these repair mechanisms can often be hindered within cancer cell lines this offers a possible therapeutic gain, wherein normal tissues can be irradiated to much higher total fractionated doses allowing for increased dose delivered to the tumour which can often be less capable of dealing with sub-lethal damage. However, this therapeutic gain cannot be infinitely leveraged due to the re-population effect discussed above. Finally, some cancer cell lines may have increased repair capability [105] or are simply not impeded by sub-lethal damage in the same way as normal cell lines. It is in these cases, that the re-oxygenation effect still offers a potential benefit for fractionation and will be described next.

Tumour hypoxia which is caused through chaotic angiogenesis is known to lessen the radioresponse. These regions of hypoxia can lead to difficulties in achieving tumour local control following treatment. However, through fractionation, it was shown that the repetitive damage and pause methodology allows for the reduction of hypoxia over the treatment duration. An early review by Kallman in 1972 [106] suggested that this was achieved by four possible methods: reduced O_2 metabolism, improved circulation, shrinkage and migration. These mechanisms are triggered by the radiation damage and fractionating radiation dose allows for cells that were previously hypoxic to re-oxygenate to become more radio-sensitive. The benefits of fractionating are variable both within and between different tumour groups and this is still an active area of improvement in clinical radiotherapy with large scale clinical trials used to help steer the community in optimising fractionation patterns.

In 1989 there was the addition of a 5th R of radiobiology in the form of radiosensitivity

[99]. This differs to the radiosensitivity from the factors above and relates more specifically to inherent radiosensitivity. This inherent radiosensitivity should be thought of as the sensitivity to the initial induction of damage, thus separating it from the other Rs of radiobiology. The 6th R of radiobiology is still up for debate, but in some cases, it has been proposed that this should relate to the immuno-modulatory properties of radiotherapy [100]. The way radiation is administered (fraction, dose, timing) all have an impact on the immune response and as it is well known that the immune system is evaded by tumours, there is a potentially large clinical benefit to being able to reactivate the anti-tumour immune response.

Whilst, fractionation is a prominent way of evidencing radiobiology within the context of radiotherapy, one of the most obvious mechanisms of radiobiology is simply the dose-response as a whole. At a high level for most cancer types if you can deliver more dose to the tumour whilst avoiding serious normal tissue complications you can increase your probability of tumour local control. In practice, this fundamental use of radiobiology has been improved significantly, through the adoption of technological advancements. Delivery of radiotherapy using multi-leaf collimators, image-guidance, motion management and improved treatment delivery techniques (e.g. intensity-modulated radiotherapy) to better target the radiation dose and allowing for better dose conformity to the tumour. These advances should not be overlooked in their role of radiobiological improvements to radiotherapy.

Proton Therapy

Proton therapy is a growing subset of radiotherapy which uses high-energy protons as incident radiation rather than x-rays. The main benefit of protons over x-rays comes from the ability to use its characteristic Bragg Peak to minimise dose to normal tissues distal to the target volume. This benefit is ultimately limited by the complexity of the treatment site as some tumour sites (e.g. head and neck) can be surrounded by sensitive healthy tissues and the difficultly of providing high doses to the target volume requires several beams administered from different angles which can reduce the benefit of proton therapy. However, in the context of radiobiology, the sites in which the dose distribution of protons do allow the reduced dose to sensitive normal tissues can open up the possibility of dose-escalation to the tumour [107, 108].

The other aspect of radiobiological differences between protons and x-rays is how DNA damage is distributed. Charged particles can characterise how they release energy to the surrounding area through the physics metric linear energy transfer (LET), this describes the energy per unit length a particle is depositing. With the characteristic Bragg Peak depositing most of the proton energy at the end of the range this also results in an increased LET at the same point. This metric has been shown to be influential at the scales of the cell nucleus [109] and it is thought that this clustering of DNA damage has a direct impact on the cell's ability to repair and in turn cell fate [110]. Whilst, there is gathering evidence on this effect at a clinical level [111, 112], it is predominately not quantified during treatment planning,

but rather qualitatively assessed. Therefore, a large amount of research focus, this thesis included, is trying to better capture the variable biological effect caused by this increased LET at the end of the proton range.

Heavy Ions

Heavy ions, denoted as "heavy" due to the mass of the particle, are also becoming increasingly popular alternative modalities of radiotherapy. Clinically this is mostly in the form of carbon ions, but there is also increasing interest in using helium ions. As these are also charged particles they also exhibit the same beneficial dose distributions as proton therapy which can utilise the Bragg Peak. However, with heavier ions, there is a resultant fragmentation tail of lighter charged ions which does reduce the sharpness of the delivered dose distal to Bragg Peak. In turn, as the particles are heavier they are scattered less laterally out of the beam, resulting in enhanced sharpness at the beam penumbra. A key difference with heavier ions is that the effect of LET becomes indisputably important for the clinical application [113]. Therefore, the effects of increased clustering of DNA damage can be utilised for radiotherapy treatment planning which can offer various benefits for radio-resistant tumours [114, 115]. Furthermore, heavier ions have benefits when treating hypoxic tumours as the damage can become so clustered that there are increased mechanisms of cell death that arise from structural disruptions (e.g. apoptosis) in the DNA [116]. The more apparent correlations between DNA damage distribution and radiosensitivity makes heavy ions a prime candidate to develop our mechanistic understanding from which can potentially be translated to proton therapy where these effects are more subtle.

2.3 Radiobiological Modelling

2.3.1 A Brief History

Modelling the radiation response at a cellular level, which are often referred to as cell fate models, can provide insight into the effects observed on larger biological systems (i.e. tissues and organs toxicities). Target theory was one of the earliest radiation response models used for clinical guidance of radiotherapy. At its foundation it consists of two major principles: 1) radiation is modelled as a series of random projectiles, and 2) cells are made up of targets to be bombarded by these radiation projectiles. From the early application of target theory, work carried out by Marie Skłodowska Curie, gave rise to the so-called quantum radiobiology which purposed that cell death only occurs if a sensitive region absorbs a threshold minimum of projectiles. The conceptual addition of a sensitive region of the cell and a required minimum number of projectiles gave rise to two ballistic models of target theory: the single-target single-hit model and the n-target single-hit model. Both models consider a single hit to the sensitive target is enough to damage that target, the differences arise from the required number of damaged sensitive targets to cause cell death. The single-

target single hit requires just a single impact to a sensitive target to kill the cell. Whereas the n-target single-hit model is based on the cell having n identical sensitive targets, impact to a single target causes a sub-lethal event, after n sub-lethal events the accumulated damage will cause cell death. Both target theory models have been criticised for not resembling experiential survival curves obtained from mammalian cell lines [117].

With target theories inability to model experimental data, Kellerer and Rossi (1972) introduced the empirically-based linear-quadratic (LQ) model based on their theory of dual radiation action. The theory proposes that cell lesions are the result of the interaction of pairs of sub lessons caused by ionising radiation, where the interaction probability of sub-lesions is a function of the distance between the pair [118]. As the name implies the LQ model consists of a linear component (αD) and a quadratic component (βD^2). There have been several purposed mechanisms to explain the linear and quadratic component of the LQ model [118–120], but to date, there is no agreed consensus on the underlying mechanisms to the LQ model components. Even with these criticisms the LQ model is considered one of the best fitting models for survival data and is the most widespread model used clinically. The model is simple to fit with surprising robustness; which further promotes its use as the standard to use in clinical studies (QUANTEC reports) [121]. Though the empirical derivation of the LQ model may give well-fitting survival curves, these fits become less justified as parameters are moved away from the empirical setting the model has been based on. The parameters of influence include high dose per fraction, patient-to-patient variability, inhomogeneous dose, radiation quality and radiation type. To provide an example, the LQ model is generally matched for around 1.5-2 Gy per fraction treatments, there has been debate as to if these same models can be applied for high dose per fraction treatments [122]. In lung stereotactic ablative body radiotherapy (SABR) 12-18 Gy per fraction is the current UK recommendation [123]; this is a substantial increase of 6-9 fold when compared to the empirical dose per fraction and is the foundation of the debated use of the LQ model for high dose per fraction regimes [122, 124, 125]. Furthermore, as the underlying mechanisms are not understood it's not pragmatic to use the model for personalised radiotherapy, fits are often based on cohorts of patients which does not account for patient-to-patient variability.

In the attempt to interpret the mechanisms of the LQ model the idea of sub-lesions prompted the foundation of both the repair-misrepair model (Tobias 1985) [126] and the lethal-potentially lethal model (Curtis 1986) [127]. The repair-misrepair model evaluates the change in number of lesions U over time t, the number of lesions at a given $U(t)$ is reduced by the number of lesions repaired self-repair (non-lethal lesions) and quadratic misrepair (lethal lesions).

The lethal-potentially-lethal model purposes that lesions are categorised as either: lesions that are unrepairable and therefore lethal, and potentially lethal lesions that may be repaired. Both sublesion models are clear attempts at interpreting the LQ model through the use of a repair mechanism. However, in both models misrepaired lesions are considered lethal, something which is now known to be a factor but not a determinant in cell death. A third model, the saturable repair model (Goodhead 1985) [128], was also purposed alongside the sublesion models. The saturable repair model was based on the hypothesis that repair ki-

netics become saturated with the increase of dose, leading to a compromised repair rate for the induced lesions. As the repair kinetics could saturate the model steered away from sub-lesions to explain the quadratic portion of survival curves. The model was seen as an alternative to the repair-misrepair model and lethal-potentially lethal model. The primary criticism of the model is that the availability of repair proteins vastly outnumber the produced lesions; the mechanism of saturation has never been verified and from a theoretical point seemed unlikely. A more recent attempt at mechanising the LQ model has been attempted through the use of nucelo-shuttling of the ATM protein which is believed to be required for recognition of DNA damage. The rate of the propagation of ATM into the nucleus is thought to describe radiosensitivity variation seen in patients [120]. The components α and β are thought to describe these two responses to lethal DSBs: 1) the damage is recognised but non-repaired, and 2) damage is not recognised and therefore non-repaired.

The effort to define mechanisms to the LQ model has been paralleled by an alternative modelling focus - attempted mechanistic modelling at the DNA level. These models tend to be based on using Monte Carlo (MC) simulation frameworks which are capable of modelling particle interactions to very low energies applicable to the scale of DNA. Such simulations are possible due to the improvement in computational power, improved radiobiological understanding and more experimental data. Primarily, the goal of most research groups involved in this area is to be able to build a model that can describe the process between energy deposition and cell fate, which includes the interplay of physics, chemistry and biology.

2.3.2 Modelling approaches

Over its lengthy history, radiobiological modelling has been attempted in a variety of approaches. With clinical applicability often being a primary objective in the development of these models it becomes important to consider the scale for which you want to develop a model. There is radiation that is interacting at the atomic scale ($\sim 1e^{-15}$ m), we are predominately interested in the damage it does to the DNA ($\sim 1e^{-12}$ m), which in turn dictates a cellular response ($\sim 1e^{-6}$ m) which in combination results in a tissue response ($\sim 1e^{-2}$ m) and ultimately impacts the treatment efficacy of the patient (~ 1 m). This 15-orders of magnitude jump between cause and effect can make capturing the finer details of radiobiology incredibly difficult and is compounded when you consider time scales ($\sim 1e^{-9}$ to $1e^{8}$ s) and radio-sensitivity variation within the patient population. Therefore, it is clear to see that within the long history of radiobiological modelling there is plenty to achieve and this is very much still an active research topic. In this next section, we will discuss some of the approaches which can be used when trying to make radiobiological models.

Empirical Modelling

Due to the enormous complexity involved in radiobiological modelling, one of the most utilised approaches is observation based. Empirical modelling has allowed for some of the

highest quality conclusions of cause and effect in clinical radiotherapy. Most clinical trials are empirical experiments which validate or disprove a hypothesis and in turn updates radiotherapy guidance. As clinical trials build from an input/output point of view it can be very approachable and easy to understand. However, clinical trials contain little to no detail on the processes which take you from delivering radiation to the clinical endpoint that is being evaluated. This can make it difficult to extrapolate to other patient populations which are not represented by the cohort in the clinical trial. This is generally accounted for in the trial design and will often require large multi-institutional (preferably multi-country) studies to become the *de facto* evidence for global guidance. Furthermore, clinical trials have evolved beyond their original input/output design to often include aspects like biopsy sampling, improved imaging techniques, dose levels and patient stratification to help shed light on the underlying factors which impact any correlations observed in the study. Finally, due to how impactful and clear clinical trials are for the development of radiotherapy, they are often used as validation steps to solidify the findings of other modelling approaches.

Phenomenological Modelling

Phenomenological modelling are supposed to have better capabilities of extrapolation than empirical models. They tend to align with the known fundamentals of radiobiology, but do not need to be directly derived from theory. Their purpose first and foremost is to best capture the relationship between particular variables, often using statistical techniques to do this. The LQ model is one of the most well-known radiobiological models and it fits within this phenomenological category, it's the ability to capture the relationship between dose and cell survival, along with being able to extrapolate to nearby conditions has made it well suited for clinical use. The main criticism of phenomenological models, the LQ model included, is that you must be cautious extrapolating too far from the founding conditions of the model, without any underlying theory it becomes difficult to understand the realistic limits of the model variables which can result in the relationship breakdown and possibly incorrect conclusions to be made.

Mechanistic Modelling

Mechanistic models tackle the complexity head-on by attempting to capture the relationships of specific biological processes. Each parameter corresponds to a real process and thus can be measured, fitted and validated independently of the other parameters. Whilst this is advantageous in building a robust model which is can generalise more than phenomenological models, it does then require reliable data for every step in the model. Given the complexity of radiotherapy with all the scales at which radiation interacts with biological processes this is a massive undertaking. This complexity is further exhausted by the fact that the theoretical understanding of key biological processes are still being discovered and therefore building a complete mechanistic model is a continuous process. To date, no "complete" mechanistic model of radiobiology exists and it will likely be a very long time till

one will. This is because the mechanistic concept could in theory be extended all the way to subatomic forces. However, in practise, every mechanistic model must include phenomenological elements, to truncate it's smallest scale. This results in a continuum between mechanistic and phenomenological models rather than a discreet boundary. With that distinction in mind there are emerging 'mechanistic' models which contain sufficient description that they can provide radiobiological predictions of some aspects of clinical radiotherapy. Furthermore, some models utilise mechanistic descriptions of some but not all of the biological processes and then use correlations to bridge the gap to a clinical level.

2.3.3 Current State of mechanistic cell fate modelling

This section looks to draw a detailed comparison between several influential cell fate models which have varying levels of mechanistic modelling. The comparison is spread over several modelling subjects and is summarised in Fig. 2.9. The models included are: Monte Carlo Damage Simulation (MCDS) developed by Semenenko and Stewart at Purdue University [129, 130]; Monte Carlo N-Particle (MCNP) which is a developed extension to MCDS headed by Stewart at University of Washington [131]; BIophysical ANalysis of Cell death and chromosome Aberrations (BIANCA) developed by Ballarini at University of Pavia [132]; Mechanistic DNA Repair and Survival (MEDRAS) developed by McMahon from Queen's University Belfast [133, 134]; PARTRAC developed by a group headed by Friedland at the German Research Center for Environmental Health [72, 135–137]; and the Manchester Mechanistic Models which has been developed by the Proton Research at the Christie and Institute of Cancer SciencEs (PRECISE) group at University of Manchester [138–140]. The included models are all subject to continued work and therefore may change in the future. This comparison will use a snapshot of the current reported capabilities of each of the models.

Model Methods

Mechanistic modelling refers to the breaking down of a complex system into its constitute parts. By its nature, it promotes a tangible explanation of how each part works together to form the overall system, something that can often be forgone in broader phenomenological based modelling (e.g. LQ-model). Monte Carlo (MC) simulations are an ideal framework for such mechanistic modelling as each process can often be effectively modelled using a probabilistic approach. An alternative approach is to use an analytical framework that fits formulas to experimental data for each mechanism in the model. To preserve the mechanistic element the formulas and its parameters involved should be based on well-established features of the process rather than using non-descriptive fitting coefficients. All the models use MC to perform some aspect of their modelling. The MEDRAS model has a purely analytical version, with 11 mechanistic fitting parameters, of which 9 characterise DNA repair and 2 describe the cell death rates, and all are cell line independent. Analytical models have the advantage of being less computationally intensive, able to be calculated in a mat-

Model Name	MCDS	MCNP	BIANCA	MEDRAS	PARTRAC	Manchester
Mechanisms Modelled	DNA damage	DNA damage – MCDS extension	DNA damage DNA repair (NHEJ)	DNA repair	DNA damage DNA repair (NHEJ)	DNA damage DNA repair (NHEJ)
Model Methods	Monte Carlo	Monte Carlo	Monte Carlo	Analytical / Monte Carlo	Monte Carlo	Monte Carlo
Track Structure	N	N	N	N	Y	Y
Primary particles	e^- - ^{56}Fe	e^- - ^{56}Fe	Photons, He-, C- and Ne-ions	Photons, H-, He-, C= and N-ion	H-, He-, C-, N-, O- and Ne-ions	Photons, e^-, alpha-particles and H-ions
Target Geometry	Small segment DNA - randomly distribute number of lesion	Small segment DNA - randomly distribute number of lesion	AG1522 nucleus	Scaled to cell, not modelled.	Multi-scale DNA target model	Segment of DNA, Chromosome Territories
Cell scale	Target geometry is scaled for the cell nucleus	Target geometry is scaled for the cell nucleus	Cell nucleus	Scaled to cell, not modelled	Models whole cell nucleus	Target geometry is scaled for the whole cell nucleus
Largest scale Chromosome Territories	Cell nucleus	Tissues	Cell nucleus	Cell nucleus	Cell nucleus	Cell nucleus
Damage Scoring	N	N	Y	Y	Y	Y
	SSB^+, 2 SSB, DSB^+, DSB^{++}, SSB_c, SSB_{cb}, DSB_c, DSB_{cb}	SSB^+, 2 SSB, DSB^+, DSB^{++}, SSB_c, SSB_{cb}, DSB_c, DSB_{cb}	Cluster lesions – a subclass of DSB	Simple DSB and complex DSB	SSB, DSB, DSB sites, DSB clusters, DNA fragments	SSB, DSB, DSB sites, DSB clusters, DSB complexity, DNA fragments
Included Indirect Damage	Y	Y	N	N	Y	Y
Variable oxic conditions	Y	Y	N	N	N	N
Repair Details	N/A	N/A	Threshold: free-ends with initial distance below 3µm undergo NHEJ	Probability of DSB being repaired by NHEJ, HR or MMEJ is calculated	Stochastic kinetics using diffusive motion of nuclear attachment sites	Modelled sub-diffusive movement and repair kinetics which results in probability of repair

SSB – Single Strand Break, SSB^+ - two SSB or more on the same strand, 2 SSB – two or more SSBs on opposite strands but not constituting DSB, DSB – Double Strand Break, DSB^+ - complex with one DSB and one or more additional strand breaks on one strand only, DSB^{++} - at least two DSBs within close proximity, SSB_c – SSB^+ + 2 SSB, DSB_c – DSB^+ + DSB^{++}, SSB_{cb} – the sum of all SSB containing at least one base damage, DSB_{cb} – complex DSB including base damage (DSB_{cb} = DSB_c + DSB_{bd} where DSB_{bd} are those DSB with at least one base damage not included in DSB_c). – Nikjoo et al. (1999)

Figure 2.9: Model overview comparison. References for the model details shown are: MCDS[129, 130], MCNP [131], BIANCA [132], MEDRAS[133], PARTRAC [72, 135, 137, 141] and Manchester Model [110]

ter of seconds-minutes without a negative impact on the validity of the results. MC systems are more computationally intensive often requiring significant time frames which increase with the complexity of the simulation and the required statistical uncertainty. The benefit of MC is that as simulations become more detailed they have a better chance of accurately modelling the mechanisms at play. This accuracy of the real world can aid in proving theo-

retical hypotheses and can potentially identify correlations that have been eluded within the theory.

Though most of the evaluated models use MC there is variation in the mechanisms simulated. A computationally onerous but very accurate mechanism to model is the track structure of energy deposition within the medium. Models like MEDRAS, MCDS, MCNP and BIANCA forgo modelling the energy deposition and opt for random assignment of damaging events. Whilst this is advantageous in computational speed there will be a loss in the detail over what a track structure can give you. This loss in detail may become problematic when trying to relate the complex pattern of damage inherent to the incident particle with more clinically relevant mechanisms (e.g. cell fate). This arises from not having a fully understood definition of how different damage characteristics scale into cellular response and fate.

Geometry

In Monte Carlo simulations the deduction of what energy depositions correspond to DNA damage can be carried out by specifying some form of geometry to be modelled in the simulation space. DNA within the cell has a complex structure, which would be computationally expensive to model its entirety in high detail. Furthermore, the *in vivo* geometry of DNA is not fully understood. Typically smaller portions of DNA are modelled to achieve a sample distribution of damage, which is then scaled to the whole nucleus. The samples are usually produced on a segment of chromatin fibre using one of many purposed geometry models. It has been shown that the nano-dosimetry of physical depositions is not significantly altered between the different chromatin fibre models [142], but it is still necessary to define some level of geometry. Just as the damage patterns are important to achieve a high detail simulation, so is the geometry the damage pattern is overlayed onto. Most of the discussed MC models have calculated damage on a segment of DNA that has been scaled for the cell nucleus. As the BIANCA model does not try to classify the damage into more than clustered, the random population is directly applied on a cell nucleus scale. The MEDRAS model using a random and uniform spread of DSBs which are distributed throughout the nucleus and no sub-nuclear geometry is modelled. The MCNP model is included in the MC models that use a segment of DNA that is scaled but also has a secondary macro geometry for when the model moves beyond its MCDS stage. The macro scale is that of a cell culture being irradiated, represented as a water phantom, though it is purposed that the model would be applicable for other geometries (e.g patients) [131]. Chromosome territories are modelled in both track structure models (PARTRAC and Manchester) along with the BIANCA model. By subdividing the nucleus into chromosome territories it allows for the possibility to categorise damage to a chromosome territory domain, important for deducing the likelihood of chromosome aberrations. This is a favourable biological endpoint for the reasons discussed earlier, the territories are usually modelled with some form of overlap and can be computationally intensive to generate in a random fashion. An issue with the chromosome territories currently modelled is that once generated they are typi-

cally static, in actuality the chromosome territories are rather mobile and vary somewhat between cell phases [143].

Radiation Quality

Damage characteristics are specific to particle species and energy of both the incident primary particles and modelled secondary particles. Most MC frameworks have been validated for a large number of different particles over wide energy regions. However, a snapshot of the reported available primary particles for each model can be seen in Fig. 2.9. It should be noted that restrictions in available particles are likely to be limited by the research groups interests over the capabilities of the models. However, some limitations will arise due to inconsistencies in the basic physics lists being used to model energy depositions. Recent work has shown discrepancy when kinetic energy transfer equation switch between relativistic and classical models [138]. It becomes important to assess if the physics model being used remains viable over the required energy region.

Event Scoring

The conversion of energy deposition in DNA geometry is carried out by the model's event scoring methodology. With the compact geometry of DNA and highly localised energy deposition, there is a multitude of damage configurations possible. As previously mentioned the exact correlation between damage types, cellular response and cellular fate are not fully understood, so picking only a few damage configurations to focus on can be difficult. Damage is typically either single strand, double-strand, or some form of cluster damage. DSBs which have long been considered a key indicator of cell fate [65] can themselves be difficult to define. Definition typically arises from two SSBs within N number of base pairs of one another on opposing strands. A commonly accepted definition of when two SSBs become a DSB is if they are within a threshold proximity of 10 bp of each other [137]. However, this value would be difficult to deduce experimentally and is typically based on within a single helical turn of the DNA. Furthermore, this categorisation might be overlooking the inequality of possible damage; SSBs directly opposite may have a different biological consequence than SSBs 9bp away from one another. Clustered damage poses similar issues as the spatial proximity required for damage to be included in a clustering event is difficult to quantify. Some models use a similar approach to classifying DSBs damage where a proximity threshold defines if a break is included within a cluster, for example, 25bp as used in PARTRAC [137]. It becomes highly likely, using this definition, that cluster damage would form on the same chromosome. However, depending on if and how the chromosome territories are modelled will alter what proximity is reasonable. Another consideration would be regarding how events are considered to be proximal - this can be measured in base pairs or using euclidean distances. The use of euclidean distances means that proximity neglect the known compact helical nature of DNA, which results in a bias towards damage similarly positioned along the helical being treated more proximal than their actual separation along

the chromosome.

A further damage type of interest are events that involve some form of damage complexity. It is believed that this complex damage may be more difficult for DNA repair mechanisms to rectify, which results in lengthened repair duration [144], misrepair [145] and increased cell death [146, 147]. Complexity is a widely applicable term with variable levels, from DSBs with a single base involvement to clustered DSBs with a large number of base damages. With many possible configurations it becomes difficult to identify what forms of complexity is significant and at what point increased complexity might have a reduced impact (i.e. the overkill effect). Complex damage is a topic of interest among modellers, with the most detailed investigations carried out by track structure models, which can simulate different complexity configurations on a given DNA geometry. In the PARTRAC model complexity is examined at a local scale; the mean number of DSB in a DSB cluster is referred to as a DSB multiplicity and characterises the local DSB complexity for a given site. In a paper from PARTRAC [137] their models showed that for protons, complexity in the form of multiplicity does not increase among clinically relevant LET values (< 20 $keV/\mu m$). However, the number of DSB clusters does increase with LET values greater than 20 $keV/\mu m$. In the Manchester model, a comparison can be drawn by examining cluster density, which is defined as the average neighbouring DSBs within a given radius from each DSB, this DSB density is a similar premise to PARTRAC's multiplicity. In the Manchester model, a second-order polynomial increase is seen with an increase of LET for a radius of 50 nm, which was chosen for its best linear agreement between the fraction of misrepair and cluster density [138]. The discrepancy between PARTRAC and the Manchester model may arise from the proximity inclusion criteria, the 50 nm proximity used in the Manchester model is much larger than the 25 bp (8.5 nm) proximity requirement set by PARTRAC. Manchester's 50 nm proximity is backed by the agreement to the biological endpoint of the fraction of misrepair, this may suggest that it would be a more biologically relevant definition for DSB clustering. In non-track structure models, complexity remains a key part of the models. In the BIANCA model only complex damage, termed cluster lesions, is considered for the arise of chromosome aberrations - the biological endpoint of interest [132]. In both the MCDS and MCNP models complex DSBs are specifically classified as at least more than one strand break on either one side and at least two DSBs within proximity; there is also a separate classification for base involvement. The role of complexity in MCNP is postulated as an interesting parameter to compare with biological endpoints [131].

This comparison highlights the variability seen in how modelling groups are reporting types of damage. With the many configurations of possible damage and no definitive explanation to which are biologically significant, it is up to the groups to identify which damage types to report. This space in the literature has encouraged many of these *in silico* models to seek out well-correlating damage types with interesting biological endpoints. However, the variation has drawn attention to the need for further and clearer experimental examination of the effects of damage complexity [144].

Repair Modelling

The PARTRAC, BIANCA, MEDRAS and Manchester groups have included repair mechanisms in their models. The ability to model DNA repair allows for concepts such as misrepair and residual damage to be predicted. Both features are thought to have important biological consequences which influence chromosome aberration yields and cell fate. Modelling DNA repair is often categorised as a biology-based modelling exercise, removed from the physical modelling of energy deposition and particle interactions. However, with increased acknowledgement of the required multi-disciplinary approach to cancer science, groups are looking to model DNA repair using the underlining physical mechanisms at play. The primary aim for the groups that have modelled DNA repair has been towards the DSB repair pathway Non-Homologous End Joining (NHEJ). It has been shown that in human cell lines NHEJ is the dominant repair pathway and is present in all stages of the cell cycle [148]. The alternative primary pathway in DSB break repair, homologous recombination (HR), is thought to be only present in S and G2 phase and plays a smaller role in the total repair carried out post-irradiation. The main feature of why NHEJ is so interesting is that its error-prone pathway can result in misrepair which can lead to chromosome aberrations. However, with many models missing an HR pathway and therefore the error-free pathway, it can be difficult fitting to wild-type cell lines, possibly hindering the predictive power of misrepair. Furthermore, it has been shown that damage of a replication fork can only be repaired by HR [149], which is completely missed in many of the current models. It has been suggested in the literature that complex damage may be favoured by HR due to a more comprehensive attempt of repair using a template [63, 150]. Another interesting avenue is the possibility of modelling multi-modality treatments (e.g. chemotherapy and radiotherapy), this would present with a different resultant damage pattern which would impact repair pathways. Platinum-based chemotherapy, the predominant type of chemotherapy, is known to cause oversized cross-link DNA damage and can also only be repaired by HR [151].

As with the damage scoring the most mechanistic and detailed approach to model, DNA repair is shown in the two-track structure models (PARTRAC and Manchester). Both models look to model the motion of the broken ends as proximity is thought to be the key factor to if the DNA break will be involved in NHEJ. Furthermore, the motion is likely to be an important predictor end-joining occurring between incorrect end partners, which would result in a chromosome aberration. PARTRAC models motion using step-by-step stochastic Brownian motion, whereas the Manchester model utilises sub-diffusive motion and both are modelled in homogeneous liquid water. As the DSB ends are tethered by the rest of the DNA it has been shown that sub-diffusion may be a better model of DSB end mobility [152]. Furthermore, it is known that the MRN protein complex tethers the broken ends together - further restricting motion [153]. The Manchester model has been used to evaluate how residuals and misrepair vary for dose and LET at clinically relevant values [138]. In contrast, the published literature from PARTRAC looked to evaluate misrepair over experimental dose levels which are typically much higher doses than seen in clinical fractionated

treatment [137]. Both models suggest as the dose increases so do the proportion of misrepair. The BIANCA model, which is interested in predicting chromosome aberrations, also models NHEJ. The BIANCA model uses Brownian diffusion to model motion and a threshold approach to repair. The threshold approach is that free-ends with an initial distance below $3\mu m$ undergo NHEJ, and if the damage is complex then a chromosome aberration is formed. The biological endpoint in the BIANCA model is lethal aberrations rather than misrepair, which are related, but one step removed as not all misrepair may end in a lethal event. The advantage of modelling lethal events means that it is possible to produce cell survival curves. The final model for discussion is the MEDRAS analytical model which includes the most complete description of available repair pathways. The MEDRAS model includes NHEJ, HR and MMEJ. MEDRAS produced analytical solutions to model DSB repair fidelity which feeds into an aberrations calculation for both G1 and G2 cell phases, which further assesses the lethality of the lesions. The MEDRAS model also includes a calculation for mutation rates which differ from lethal aberration but are an important biological indicator for induced genetic disorders and cancer formation. The final biological endpoint of the MEDRAS model is to estimate cell survival curves for a given exposure.

The models discussed highlight the benefit of including DNA repair pathways as a step towards being able to establish higher-level biological endpoints (e.g. misrepair and cell survival curves). Misrepair is a direct step from DNA repair failure and can help predict chromosome aberrations. Cell survival curves are an advantageous endpoint to reach due to their familiarity in the field of radiobiology with ample experimental data available. Furthermore, cell survival curves have a direct impact on clinical radiobiology through the LQ-model, so it would be a reasonable assumption that relating more comprehensive mechanistic models to the same endpoint could accelerate any adoption into clinical practice.

Role of Experimental Results

Experimental results play a key role in all forms of modelling, from fitting to validating the data impacts both the quality and robustness of the produced models. Critical examination of the role of experimental results within the different models is key to identify their strengths and weaknesses. A comprehensive, but not exhaustive, examination of the various data sources used within the models in question is given in Fig. 2.10 and Fig. 2.11. The examination shows a clear overlap in data used between models, with some of the newer models basing fits and validation against the older models. This is a pragmatic approach as older pre-validated models are valuable data sources themselves. Experimental results lay the foundations for all of the discussed models; the distinct overlap of the literature used between models that vary by over a decade since their inception, possibly highlights that high-quality data specifically useful for this type of modelling is sparse. It is unclear if this is due to a discontinuity between the topic of interest of the groups carrying out the experiments and the groups hoping to use experimental data to build models, but it is clear that from a modelling perspective high-quality experimental results which examine specific mechanisms of interest would be of great benefit.

Model Name	Experimental Data used for Fitting	Experimental Data used for Validation
MCDS	Method includes a dimensionless scale factor that can be used to adjust absolute yield of DNA lesions to better mimic experimental observations. The linearity of DSB induction with dose derives from: Erixon and Cedervall (1995) who used pulsed-field gel electrophoresis to show a linear response between DNA fragment size and dose using a 4 MeV linear accelerator; and Frankenberg et al. (1981) who used irradiated cell colonies to deduce linear DSB relation to dose using a cobolt-60 source. The observation of DSB formation being approximately proportional to the DNA content of a cell derives qualitatively from Prise et al. (1998).	The model damage output it directly compared with the computations results of Nikjoo et al. (1999). The computational results generated by Nikjoo et al showed good agreement with several experimental results (Table 1 in source paper). The experimental results include various mediums (yeast, V79 and Solid DNA), radiations (protons, α-particles, γ-rays, electrons and ^{125}I) and where reported over various LET.
MCNP	MCDS portion of model as above. The MCDS algorithm was adapted to simulate the effects of radical scavengers by work carried out by Semenenko and Stewart (2006). Furthermore the interplay between particle LET and the concentration of oxygen was included by Stewart et al. (2011). For macro scale modelling, simulations are corrected for spatial variation in dose, particle type and particle fluence. Charged particle transport used the Vavilov energy straggling and neutron interactions were modelled using MCNP6 (CEM) physics models.	MCDS portion of model as above. Though the model is not validated against any other data directly and relies on the validations of MCDS to carry through to the results given, there are some comparisons drawn regarding other propagating biophysical models that have been purposed. Mairani et al. (2010) combined FLUKA with the local effect model (LEM) for carbon ion therapy. Carlson et al. (2008) combined MCDS with the repair-misrepair-fixation (RMF) model. These combination models could be used as points of validation/comparison in further work.
BIANCA	The threshold value for end joining (d) was deduced through trial and error by good agreement with a photon curve. The cluster lesion yields were also derived using trial and error by comparing the simulated surviving fraction at 2 Gy with the corresponding experimental survival. The one-to-one relationship between the average number of "lethal aberrations" per cell and surviving fraction derives from experimental work carried out on AG1522 cells using photons by Cornforth and Bedford (1987).	Simulated survival curves were compared with AG1522 experimental data within the literature. Survival curves for γ-rays, He, C and Ne ions were taken from Hamada et al. (2006) and survival curves for γ-rays and alpha particles were taken from Neti et al. (2004).
MEDRAS	The initial DSB yield approximation of 35 DSBs/Gy in human cells was taken from Rothkamm et al. (2003). There is an assumption of the human diploid cell containing 6.1 Gbp which comes from the Genome Reference Consortium. Fitting parameters arose from several experimental sources for DNA repair, DNA misrepair rates, chromosome aberration yields and mutation rates. DNA repair is fitted using γH2AX fluorescence from Kühne et al. (2004) and Beucher et al. (2009). DNA misrepair rates are fitted using pulsed field gel electrophoresis from Löbrich et al. (2000) and Rydberg et al. (2005). Chromosome aberrations were fitted using Giemsa straining and *In Situ* Hybridisation (FISH) experiments from a large pool of literature. Mutation rates were fitted using induction rates in the *Hprt* gene of Chinese Hamster cells from Morgan et al. (1990), Belli et al. (1998) and Schwartz et al. (2000).	Inter-laboratory variability scaling factors were fitted to each dataset used for validation. Each derived fitting parameter for DNA repair, DNA misrepair rates, chromosome aberration yields and mutation rates are compared with the original data. The fittings and simulated predictions were then compared with a series of hamster and human cell lines. The experimental data selected was done so to have little sensitivity to cycling effects. Data was validated against for both G1 and G2 phases.

Figure 2.10: Experimental data comparison. References for the tabulated authors are: ‖ MCDS - Erixon and Cedervall (1995) [154], Prise et al. (2009) [67] and Nikjoo et al. (1999) [71]. ‖ MCNP - Semenenko and Stewart (2006) [130], Stewart et al. (2011) [155], Mairani et al. (2010) [156] and Carlson et al. (2008) [157]. ‖ BIANCA - Cornforth and Bedford (1987) [158], Hamada et al. (2006) [159] and Neti et al. (2004) [160]. ‖ MEDRAS - Rothkamm et al. (2003) [161], K hne et al. (2004) [162], Beucher et al. (2009) [163], Löbrich et al (2000) [164], Rydberg et al (2005) [165], Morgan et al (1990) [166], Belli et al (1998) [167] and Schwartz et al (2000) [168].

PARTRAC	Repair time constants for enzyme attachment during the pre-synaptic phase of NHEJ were fitted to experimental data from Taghva et al. (2002) and Uematsu et al. (2007), for Ku and DNA-PK_{cs} attachment kinetics respectively. Data from Uematsu et al. (2007) also gave information on enzyme detachment which was inferred from florescence recovery after photobleaching. The background DNA fragmentation is fitted to the measured pulsed-field gel electrophoresis data of Belli et al. (2001). The remainder of the damage simulation is based on charge state-dependent cross sections which has some extrapolation from experimental data, described by Dingfelder et al. (2000).	Simulation is compared with experimental data on DSB induction in human fibroblast cells for γ-rays, He, B, N and Ne ions from work conducted by Höglund et al. (2000). DNA fragment size interval also agrees with experimental work by Stenerlow et al. (2000) and Belli et al. (2001), using pulsed-field gel electrophoresis. Due to an absence of experimental results for DNA damage induced by low-energy ions, a comparison against the MCDS model was carried out.
Manchester	The induced DNA damage methodology is based on work carried out by Francis et al (2011), where various assumptions are made to convert energy deposition to DNA damage. The sensitive region of the nucleus was determined by fitting SSBs to DSB ratios found in the PARTRAC and Nikjoo computational models. Base damage inclusion within clusters derives from computational model work by Watanabe et al. (2015). Overall repair kinetics of the c-NHEJ repair pathways is fitted to experimental laser generated DSB work by Li et al. (2014). Recruitment kinetics of Ku and DNA-PKcs are fitted to work by Uematsu et al. (2007), who used laser irradiation of Xrs6 and V3 hamster cells. The repair kinetics was further adapted to fit to proton irradiated AG01522 human fibroblast cells from work by Chaudhary et al. (2016). Time constants for complexity cleaning before ligation are taken from the Friedland et al. (2010) who made a stochastic NHEJ model.	The induced DNA damage methodology is based on work carried out by Francis et al. (2011); the predicted yields in this work have been validated against the PARTRAC model and experimental data by Leloup et al. (2005), Fulford et al. (2001) and Botchway et al. (1997). The repair model has been used to reproduce experimental results of Uematsu et al. (2007), who carried out fluorescent recovery after photo-bleaching (FRAP) experiments. The modelled relationship between initial DSB yields on LET was compared to work carried out by Chaudhary et al. (2016) which used 53BP1 foci analysis; the model predicted a systematic offset with the compared data. Fraction of misrepaired DSBs produced by the model is thought to explain the differences in cell kill and resultant mutations seen within experiments by Beli et al. (1993).

Figure 2.11: Experimental data comparison (continued). References for the tabulated authors are: ‖ PARTRAC - Taghva et al. (2002) [169], Uematsu et al. (2007) [170], Belli et al. (2001) [171], Dingfelder et al. (2000) [172], Höglund et al. (2000) [173]] and Stenerlow et al. (2000) [174]. ‖ Manchester - Francis et al. (2011) [[175], Watanabe et al. (2015) [176], Li et al. (2014) [146], Uematsu et al. (2007) [170], Chaudhary et al. (2016) [177], Friedland et al. (2010) [141], Leloup et al. (2005) [68], Fulford et al. (2001) [69], Botchway et al (1997) [70], Beli et al. (1993) [167]

Ability to model the 6 Rs

In attempting to develop cell fate models which may provide clinical utility it is important to consider if the radiobiological principles established by the 6 Rs are incorporated.

Repair has been discussed in the previous section, but re-distribution, re-oxygenation, re-population and radio-sensitivity have been demonstrated to be important when talking about radiobiology in the context of radiotherapy [99]. To account for re-distribution the inclusion of cell cycle is required within the model which accounts for the alteration of radiosensitivity as a cell moves through the different phases. At present, it is only the MEDRAS model which has published this capability [178]. To model re-oxygenation effects models are required to be able to adapt to variable oxic conditions, at present it is only the MCDS [129, 130] and MCNP [131] models which have this capabilities. However, with the re-emerged interest around FLASH (ultra-fast dose rate) radiotherapy, this is a capability that most of the other models are actively working on. The interest around re-population is primarily the focus if attempting to model cancer cells, at the time of writing I am unaware of any of the discussed models having this capability. Furthermore, it is important to note typically for treatment planning optimisation, the normal tissues will tend to be the limiting factor for the dose which can be delivered to the target tumour, which may be why this has not been a targeted feature of the models. Radio-sensitivity differences encapsulate a whole range of mechanistic drivers. MEDRAS has published results on fitting several different cell lines split by repair deficiencies or species [178] and Manchester has published results fitting to repair deficiencies as part of this thesis [139]. Modelling cellular radio-sensitivity differences is one of the broadest and challenging capabilities to include as it is driven by both the genetics (e.g. repair deficiencies) and the physical geometry of the DNA, both of which are discussed in the subsequent chapters.

References

1. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100,** 57–70. ISSN: 0092-8674 (2000).
2. Guo, G. *et al.* Ligand-Independent EGFR Signaling. *Cancer Research* **75,** 3436–3441. ISSN: 0008-5472 (2015).
3. Giancotti, F. G. & Ruoslahti, E. Integrin Signaling. *Science* **285,** 1028–1033. ISSN: 0036-8075 (1999).
4. Aplin, A. E., Howe, A., Alahari, S. K. & Juliano, R. L. Signal transduction and signal modulation by cell adhesion receptors: the role of integrins, cadherins, immunoglobulin-cell adhesion molecules, and selectins. *Pharmacological reviews* **50,** 197–263. ISSN: 0031-6997 (1998).
5. Molina, J. R. & Adjei, A. A. The Ras/Raf/MAPK Pathway. *Journal of Thoracic Oncology* **1,** 7–9. ISSN: 1556-0864 (2006).
6. Hunter, T. Oncoprotein Networks. *Cell* **88,** 333–346. ISSN: 0092-8674 (1997).
7. Weinberg, R. A. The retinoblastoma protein and cell cycle control. *Cell* **81,** 323–330. ISSN: 0092-8674 (1995).
8. Markowitz, S. *et al.* Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science* **268,** 1336–1338. ISSN: 0036-8075 (1995).
9. Chin, L., Pomerantz, J. & DePinho, R. A. The INK4a/ARF tumor suppressor: one gene—two products—two pathways. *Trends in Biochemical Sciences* **23,** 291–296. ISSN: 0968-0004 (1998).
10. Liu, Q., Lopez, K., Murnane, J., Humphrey, T. & Barcellos-Hoff, M. H. Misrepair in Context: TGFβ Regulation of DNA Repair. *Frontiers in Oncology* **9,** 799. ISSN: 2234-943X (2019).
11. Olivier, M., Hollstein, M. & Hainaut, P. TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use. *Cold Spring Harbor Perspectives in Biology* **2,** a001008 (2010).
12. Fridman, J. S. & Lowe, S. W. Control of apoptosis by p53. *Oncogene* **22,** 9030–9040. ISSN: 0950-9232 (2003).
13. Hayflick, L. Mortality and immortality at the cellular level. A review. *Biochemistry-New York-English Translation of Biokhimiya* **62,** 1180–1190 (1997).

14. Counter, C. *et al.* Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *The EMBO Journal* **11,** 1921–1929. ISSN: 1460-2075 (1992).

15. Shay, J. & Bacchetti, S. A survey of telomerase activity in human cancer. *European Journal of Cancer* **33,** 787–791. ISSN: 0959-8049 (1997).

16. O'Reilly, M. S. *et al.* Endostatin: An Endogenous Inhibitor of Angiogenesis and Tumor Growth. *Cell* **88,** 277–285. ISSN: 0092-8674 (1997).

17. Gupta, G. P. & Massagué, J. Cancer Metastasis: Building a Framework. *Cell* **127,** 679–695. ISSN: 0092-8674 (2006).

18. Dillekås, H., Rogers, M. S. & Straume, O. Are 90% of deaths from cancer caused by metastases? *Cancer Medicine* **8,** 5574–5576. ISSN: 2045-7634 (2019).

19. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144,** 646–674. ISSN: 0092-8674 (2011).

20. WEINHOUSE, S., WARBURG, O., BURK, D. & SCHADE, A. L. On Respiratory Impairment in Cancer Cells. *Science* **124,** 267–272. ISSN: 0036-8075 (1956).

21. Jones, R. G. & Thompson, C. B. Tumor suppressors and cell metabolism: a recipe for cancer growth. *Genes & Development* **23,** 537–548. ISSN: 0890-9369 (2009).

22. Heiden, M. G. V., Cantley, L. C. & Thompson, C. B. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science* **324,** 1029–1033. ISSN: 0036-8075 (2009).

23. Teng, M. W. L., Swann, J. B., Koebel, C. M., Schreiber, R. D. & Smyth, M. J. Immune-mediated dormancy: an equilibrium with cancer. *Journal of Leukocyte Biology* **84,** 988–993. ISSN: 0741-5400 (2008).

24. Fanale, D., Maragliano, R., Bazan, V. & Russo, A. eLS, 1–10 (2017).

25. Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science* **339,** 1546–1558. ISSN: 0036-8075 (2013).

26. Greten, F. R. & Grivennikov, S. I. Inflammation and Cancer: Triggers, Mechanisms, and Consequences. *Immunity* **51,** 27–41. ISSN: 1074-7613 (2019).

27. Balkwill, F. R., Capasso, M. & Hagemann, T. The tumor microenvironment at a glance. *Journal of Cell Science* **125,** 5591–5596. ISSN: 0021-9533 (2012).

28. Wang, M. *et al.* Role of tumor microenvironment in tumorigenesis. *Journal of Cancer* **8,** 761–773. ISSN: 1837-9664 (2017).

29. Fridman, W. H., Pagès, F., Sautès-Fridman, C. & Galon, J. The immune contexture in human tumours: impact on clinical outcome. *Nature Reviews Cancer* **12,** 298–306. ISSN: 1474-175X (2012).

30. Yuen, G. J., Demissie, E. & Pillai, S. B Lymphocytes and Cancer: A Love–Hate Relationship. *Trends in Cancer* **2,** 747–757. ISSN: 2405-8033 (2016).

31. Bruno, T. C. New predictors for immunotherapy responses sharpen our view of the tumour microenvironment. *Nature* **577,** 474–476. ISSN: 0028-0836 (2020).

32. Petitprez, F. *et al.* B cells are associated with survival and immunotherapy response in sarcoma. *Nature* **577,** 1–5. ISSN: 0028-0836 (2020).

33. Larsen, S. K., Gao, Y. & Basse, P. H. NK Cells in the Tumor Microenvironment. *Critical Reviews in Oncogenesis* **19,** 91–105. ISSN: 0893-9675 (2014).

34. Lin, E. Y. *et al.* Macrophages Regulate the Angiogenic Switch in a Mouse Model of Breast Cancer. *Cancer Research* **66,** 11238–11246. ISSN: 0008-5472 (2006).

35. Bingle, L., Brown, N. J. & Lewis, C. E. The role of tumour-associated macrophages in tumour progression: implications for new anticancer therapies. *The Journal of Pathology* **196,** 254–265. ISSN: 1096-9896 (2002).

36. Ojalvo, L. S., Whittaker, C. A., Condeelis, J. S. & Pollard, J. W. Gene Expression Analysis of Macrophages That Facilitate Tumor Invasion Supports a Role for Wnt-Signaling in Mediating Their Activity in Primary Mammary Tumors. *The Journal of Immunology* **184,** 702–712. ISSN: 0022-1767 (2010).

37. Zabuawala, T. *et al.* An Ets2-Driven Transcriptional Program in Tumor-Associated Macrophages Promotes Tumor Metastasis. *Cancer Research* **70,** 1323–1333. ISSN: 0008-5472 (2010).

38. Gabrilovich, D. I. & Nagaraj, S. Myeloid-derived suppressor cells as regulators of the immune system. *Nature Reviews Immunology* **9,** 162–174. ISSN: 1474-1733 (2009).

39. Shaul, M. E. & Fridlender, Z. G. Tumour-associated neutrophils in patients with cancer. *Nature Reviews Clinical Oncology* **16,** 601–620. ISSN: 1759-4774 (2019).

40. Nieman, K. M. *et al.* Adipocytes promote ovarian cancer metastasis and provide energy for rapid tumor growth. *Nature Medicine* **17,** 1498–1503. ISSN: 1078-8956 (2011).

41. Dudley, A. C. Tumor Endothelial Cells. *Cold Spring Harbor Perspectives in Medicine* **2,** a006536. ISSN: 2157-1422 (2012).

42. Ji, R.-C. Lymphatic endothelial cells, tumor lymphangiogenesis and metastasis: New insights into intratumoral and peritumoral lymphatics. *Cancer and Metastasis Reviews* **25,** 677–694. ISSN: 0167-7659 (2006).

43. Swartz, M. A. & Lund, A. W. Lymphatic and interstitial flow in the tumour microenvironment: linking mechanobiology with immunity. *Nature Reviews Cancer* **12,** 210–219. ISSN: 1474-175X (2012).

44. *World Health Organisation (WHO) - Cancer* `https://www.who.int/health-topics/cancer%5C#tab=tab%5C_1` (2021).

45. *Global Cancer Observatory* `https://gco.iarc.fr/databases.php` (2021).

46. Jemal, A., Center, M. M., DeSantis, C. & Ward, E. M. Global Patterns of Cancer Incidence and Mortality Rates and Trends. *Cancer Epidemiology Biomarkers & Prevention* **19,** 1893–1907. ISSN: 1055-9965 (2010).

47. Malvezzi, M. *et al.* Cancer in developing countries: can the revolution begin? *The Lancet Oncology* **12,** 201. ISSN: 1470-2045 (2011).

48. Vieira, R. A. d. C., Biller, G., Uemura, G., Ruiz, C. A. & Curado, M. P. Breast cancer screening in developing countries. *Clinics* **72,** 244–253. ISSN: 1807-5932 (2017).

49. Luengo-Fernandez, R., Leal, J., Gray, A. & Sullivan, R. Economic burden of cancer across the European Union: a population-based cost analysis. *The Lancet Oncology* **14,** 1165–1174. ISSN: 1470-2045 (2013).

50. Quaresma, M., Coleman, M. P. & Rachet, B. 40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971–2011: a population-based study. *The Lancet* **385,** 1206–1218. ISSN: 0140-6736 (2015).

51. Schlueter, M., Chan, K., Lasry, R. & Price, M. The cost of cancer – A comparative analysis of the direct medical costs of cancer and other major chronic diseases in Europe. *PLOS ONE* **15,** e0241354 (2020).

52. Ellis, L., Belot, A., Rachet, B. & Coleman, M. P. The Mortality-to-Incidence Ratio Is Not a Valid Proxy for Cancer Survival. *Journal of Global Oncology* **5,** 1–9 (2019).

53. Choi, E. *et al.* Cancer mortality-to-incidence ratio as an indicator of cancer management outcomes in Organization for Economic Cooperation and Development countries. *Epidemiology and Health* **39,** e2017006. ISSN: 2092-7193 (2017).

54. Ahmad, A. S., Ormiston-Smith, N. & Sasieni, P. D. Trends in the lifetime risk of developing cancer in Great Britain: comparison of risk for those born from 1930 to 1960. *British Journal of Cancer* **112,** 943–947. ISSN: 0007-0920 (2015).

55. Sudhakar, A. History of Cancer, Ancient and Modern Treatment Methods. *Journal of Cancer Science & Therapy* **01,** i–iv (2009).

56. Dobson, J. JOHN HUNTER S VIEWS ON CANCER. *Annals of the Royal College of Surgeons of England* **25,** 175–181. https://pubmed.ncbi.nlm.nih.gov/13817098 (1959).

57. Wyld, L., Audisio, R. A. & Poston, G. J. The evolution of cancer surgery and future perspectives. *Nature Reviews Clinical Oncology* **12,** 115–124. ISSN: 1759-4774 (2015).

58. *Cancer treatment statistics | Cancer Research UK* https://www.cancerresearchuk.org/health-professional/cancer-statistics/treatment (2021).

59. Orosco, R. K. *et al.* Positive Surgical Margins in the 10 Most Common Solid Cancers. *Scientific Reports* **8,** 5686 (2018).

60. DeVita, V. T. & Chu, E. A History of Cancer Chemotherapy. *Cancer Research* **68,** 8643–8653. ISSN: 0008-5472 (2008).

61. Lederman, M. The early history of radiotherapy: 1895–1939. *International Journal of Radiation Oncology*Biology*Physics* **7,** 639–648. ISSN: 0360-3016 (1981).

62. Puck, T. T. & Marcus, P. I. ACTION OF X-RAYS ON MAMMALIAN CELLS. *The Journal of Experimental Medicine* **103,** 653–666. ISSN: 0022-1007 (1956).

63. Kogel, A. v. d., Joiner, M., Kogel, A. V. D. & Steel, G. *Basic Clinical Radiobiology Fourth Edition* ISBN: 9780340929667 (2009).

64. Warters, R. L. & Hofer, K. G. Radionuclide Toxicity in Cultured Mammalian Cells: Elucidation of the Primary Site for Radiation-Induced Division Delay. *Radiation Research* **69,** 348. ISSN: 0033-7587 (1977).

65. Frankenberg, D., Frankenberg-Schwager, M., Blöcher, D. & Harbich, R. Evidence for DNA double-strand breaks as the critical lesions in yeast cells irradiated with sparsely or densely ionizing radiation under oxic or anoxic conditions. *Radiation research* **88,** 524–32. ISSN: 0033-7587 (1981).

66. Frankenberg-Schwager, M., Frankenberg, D., Blöcher, D. & Adamczyk, C. Effect of dose rate on the induction of DNA double-strand breaks in eucaryotic cells. *Radiation research* **87,** 710–7. ISSN: 0033-7587 (1981).

67. PRISE, K. M. *et al.* A review of dsb induction data for varying quality radiations. *International Journal of Radiation Biology* **74,** 173–184. ISSN: 0955-3002 (2009).

68. Leloup, C. *et al.* Evaluation of lesion clustering in irradiated plasmid DNA. *International Journal of Radiation Biology* **81,** 41–54. ISSN: 0955-3002 (2009).

69. Fulford, J., Nikjoo, H., Goodhead, D. & P, O. Yields of SSB and DSB induced in DNA by Al K ultrasoft X-rays and alpha-particles: comparison of experimental and simulated yields. *Int J Radiat Biol* **77,** 1053–1066. ISSN: 0955-3002 (2001).

70. Botchway, S., Stevens, D., Hill, M. & Radiation …, J. T. Induction and rejoining of DNA double-strand breaks in Chinese hamster V79-4 cells irradiated with characteristic aluminum K and copper L ultrasoft X rays. *Radiation …* (1997).

71. Nikjoo, H., P, O., Terrissol, M. & Goodhead, D. Quantitative modelling of DNA damage using Monte Carlo track structure method. *Radiat Environ Bioph* **38,** 31–38. ISSN: 0301-634X (1999).

72. Friedland, W., Bernhardt, P., Jacob, P., Paretzke, H. & Dingfelder, M. Simulation of DNA damage after proton and low LET irradiation. *Radiat Prot Dosimetry* **99,** 99–102. ISSN: 0144-8420 (2002).

73. Krokan, H. E. & Bjørås, M. Base Excision Repair. *Cold Spring Harbor Perspectives in Biology* **5,** a012583 (2013).

74. Kusakabe, M. *et al.* Mechanism and regulation of DNA damage recognition in nucleotide excision repair. *Genes and Environment* **41,** 2. ISSN: 1880-7046 (2019).

75. Laat, W. L. d., Jaspers, N. G. & Hoeijmakers, J. H. Molecular mechanism of nucleotide excision repair. *Genes & Development* **13,** 768–785. ISSN: 0890-9369 (1999).

76. Hsieh, P. & Zhang, Y. The Devil is in the details for DNA mismatch repair. *Proceedings of the National Academy of Sciences* **114,** 3552–3554. ISSN: 0027-8424 (2017).

77. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. Comparison of nonhomologous end joining and homologous recombination in human cells. *Dna Repair* **7,** 1765–1771. ISSN: 1568-7864 (2008).

78. Biehs, R. *et al.* DNA Double-Strand Break Resection Occurs during Non-homologous End Joining in G1 but Is Distinct from Resection during Homologous Recombination. *Mol Cell* **65,** 671–684.e5. ISSN: 1097-2765 (2017).

79. Kostyrko, K. & Mermod, N. Assays for DNA double-strand break repair by microhomology-based end-joining repair mechanisms. *Nucleic Acids Research* **44,** e56–e56. ISSN: 0305-1048 (2016).

80. Iliakis, G., Murmann, T. & Soni, A. Alternative end-joining repair pathways are the ultimate backup for abrogated classical non-homologous end-joining and homologous recombination repair: Implications for the formation of chromosome translocations. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **793,** 166–175. ISSN: 1383-5718 (2015).

81. Li, X. & Heyer, W.-D. Homologous recombination in DNA repair and DNA damage tolerance. *Cell Research* **18,** cr20081. ISSN: 1748-7838 (2008).

82. Filippo, J., Sung, P. & Klein, H. Mechanism of Eukaryotic Homologous Recombination. *Annu Rev Biochem* **77,** 229–257. ISSN: 0066-4154 (2008).

83. West, S. C. Molecular views of recombination proteins and their control. *Nat Rev Mol Cell Bio* **4,** 435–445. ISSN: 1471-0072 (2003).

84. Heyer, W., Ehmsen, K. & review of genetics, L. J. Regulation of homologous recombination in eukaryotes. *Annual review of genetics* (2010).

85. Malkova, A., Naylor, M. L., Yamaguchi, M., Ira, G. & Haber, J. E. RAD51-Dependent Break-Induced Replication Differs in Kinetics and Checkpoint Responses from RAD51-Mediated Gene Conversion. *Mol Cell Biol* **25,** 933–944. ISSN: 0270-7306 (2005).

86. Jain, S. *et al.* A recombination execution checkpoint regulates the choice of homologous recombination pathway during DNA double-strand break repair. *Genes Dev* **23,** 291–303. ISSN: 0890-9369 (2009).

87. Johnson, R. & Jasin, M. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *EMBO J.* **19,** 3398–407. ISSN: 0261-4189 (2000).

88. Liu, Q. *et al.* Disruption of SLX4-MUS81 Function Increases the Relative Biological Effectiveness of Proton Radiation. *Int J Radiat Oncol Biology Phys* **95,** 78–85. ISSN: 0360-3016 (2016).

89. Kerr, J. F. R., Wyllie, A. H. & Currie, A. R. Apoptosis: A Basic Biological Phenomenon with Wideranging Implications in Tissue Kinetics. *British Journal of Cancer* **26,** 239–257. ISSN: 0007-0920 (1972).

90. Sia, J., Szmyd, R., Hau, E. & Gee, H. E. Molecular Mechanisms of Radiation-Induced Cancer Cell Death: A Primer. *Frontiers in Cell and Developmental Biology* **8,** 41. ISSN: 2296-634X (2020).

91. Li, L., Liu, W.-l., Su, L., Lu, Z.-c. & He, X.-s. The Role of Autophagy in Cancer Radiotherapy. *Current Molecular Pharmacology* **13,** 31–40. ISSN: 1874-4672 (2019).

92. Hayflick, L. & Moorhead, P. The serial cultivation of human diploid cell strains. *Experimental Cell Research* **25,** 585–621. ISSN: 0014-4827 (1961).

93. Tabasso, A., Jones, D., Jones, G. & Macip, S. Radiotherapy-Induced Senescence and its Effects on Responses to Treatment. *Clinical Oncology* **31,** 283–289. ISSN: 0936-6555 (2019).

94. Campisi, J., Andersen, J. K., Kapahi, P. & Melov, S. Cellular senescence: A link between cancer and age-related degenerative disease? *Seminars in Cancer Biology* **21,** 354–359. ISSN: 1044-579X (2011).

95. Belizário, J., Vieira-Cordeiro, L. & Enns, S. Necroptotic Cell Death Signaling and Execution Pathway: Lessons from Knockout Mice. *Mediators of Inflammation* **2015,** 1–15. ISSN: 0962-9351 (2015).

96. Furuse, M. *et al.* Radiological diagnosis of brain radiation necrosis after cranial irradiation for brain tumor: a systematic review. *Radiation Oncology* **14,** 28 (2019).

97. Song, Y. P. & Colaco, R. J. Radiation Necrosis – A Growing Problem in a Case of Brain Metastases Following Whole Brain Radiotherapy and Stereotactic Radiosurgery. *Cureus* **10,** e2037. ISSN: 2168-8184 (2018).

98. Withers, H. R. The Four R's of Radiotherapy. *Advances in Radiation Biology* **5,** 241–271. ISSN: 0065-3292 (1975).

99. Steel, G. G., McMillan, T. J. & Peacock, J. H. The 5Rs of Radiobiology. *International Journal of Radiation Biology* **56,** 1045–1048. ISSN: 0955-3002 (1989).

100. Boustani, J., Grapin, M., Laurent, P.-A., Apetoh, L. & Mirjolet, C. The 6th R of Radiobiology: Reactivation of Anti-Tumor Immune Response. *Cancers* **11,** 860. ISSN: 2072-6694 (2019).

101. Sinclair, W. & Morton, R. X-Ray and Ultraviolet Sensitivity of Synchronized Chinese Hamster Cells at Various Stages of the Cell Cycle. *Biophysical Journal* **5,** 1–25. ISSN: 0006-3495 (1965).

102. Withers, H. R., Taylor, J. M. G. & Maciejewski, B. The hazard of accelerated tumor clonogen repopulation during radiotherapy. *Acta Oncologica* **27,** 131–146. ISSN: 0284-186X (1988).

103. YANG, J., YUE, J.-B., LIU, J. & YU, J.-M. Repopulation of tumor cells during fractionated radiotherapy and detection methods (Review). *Oncology Letters* **7,** 1755–1760. ISSN: 1792-1074 (2014).

104. Jones, B., Dale, R., Deehan, C., Hopkins, K. & Morgan, D. The Role of Biologically Effective Dose (BED) in Clinical Oncology. *Clinical Oncology* **13,** 71–81. ISSN: 0936-6555 (2001).

105. Domogauer, J. D., Toledo, S. M. d., Howell, R. W. & Azzam, E. I. Acquired radioresistance in cancer associated fibroblasts is concomitant with enhanced antioxidant potential and DNA repair capacity. *Cell Communication and Signaling* **19,** 30 (2021).

106. Kallman, R. F. The Phenomenon of Reoxygenation and Its Implications for Fractionated Radiotherapy. *Radiology* **105,** 135–142. ISSN: 0033-8419 (1972).

107. Gomez, D. R. & Chang, J. Y. Accelerated dose escalation with proton beam therapy for non-small cell lung cancer. *Journal of thoracic disease* **6,** 348–55. ISSN: 2072-1439 (2014).

108. Palm, R. F. *et al.* The role of dose escalation and proton therapy in perioperative or definitive treatment of chondrosarcoma and chordoma: An analysis of the National Cancer Data Base. *Cancer* **125,** 642–651. ISSN: 0008-543X (2019).

109. Paganetti, H. & van Luijk, P. Biological Considerations When Comparing Proton Therapy With Photon Therapy. *Semin Radiat Oncol* **23,** 77–87. ISSN: 1053-4296 (2013).

110. Henthorn, N. *et al.* In Silico Non-Homologous End Joining Following Ion Induced DNA Double Strand Breaks Predicts That Repair Fidelity Depends on Break Density. *Sci Reports* **8,** 2654 (2018).

111. Peeler, C. R. *et al.* Clinical evidence of variable proton biological effectiveness in pediatric patients treated for ependymoma. *Radiotherapy and Oncology* **121,** 395–401. ISSN: 0167-8140 (2016).

112. Wang, C.-C. *et al.* End-of-range Radiobiological Effect on Rib Fractures in Patients Receiving Proton Therapy for Breast Cancer. *International Journal of Radiation Oncology*Biology*Physics* **107,** 449–454. ISSN: 0360-3016 (2020).

113. Malouff, T. D. *et al.* Carbon Ion Therapy: A Modern Review of an Emerging Technology. *Frontiers in Oncology* **10,** 82. ISSN: 2234-943X (2020).

114. Schlaff, C. D., Krauze, A., Belard, A., J, O. J. & Camphausen, K. A. Bringing the heavy: carbon ion therapy in the radiobiological and clinical context. *Radiation Oncology* **9,** 1–19. ISSN: 1748-717X (2014).

115. Lee, Y. & Okayasu, R. Strategies to Enhance Radiosensitivity to Heavy Ion Radiation Therapy. *International Journal of Particle Therapy* **5,** 114–121. ISSN: 2331-5180 (2018).

116. Valable, S. *et al.* Impact of Hypoxia on Carbon Ion Therapy in Glioblastoma Cells: Modulation by LET and Hypoxia-Dependent Genes. *Cancers* **12,** 2019. ISSN: 2072-6694 (2020).

117. Bodgi, L. *et al.* Mathematical models of radiation action on living cells: From the target theory to the modern approaches. A historical and critical review. *J Theor Biol* **394,** 93–101. ISSN: 0022-5193 (2016).

118. Kellerer, A. M. & Rossi, H. H. A generalized formulation of dual radiation action. *Radiat. Res.* **178,** AV204–13. ISSN: 0033-7587 (2012).

119. Chadwick, K., in medicine & biology, L. H. A molecular theory of cell survival. *Physics in medicine and biology* **18,** 78–87. ISSN: 0031-9155 (1973).

120. Bodgi, L. & Foray, N. The nucleo-shuttling of the ATM protein as a basis for a novel theory of radiation response: resolution of the linear-quadratic model*. *Int J Radiat Biol* **92,** 117–131. ISSN: 0955-3002 (2016).

121. Marks, L. B. *et al.* Use of Normal Tissue Complication Probability Models in the Clinic. *International Journal of Radiation Oncology*Biology*Physics* **76,** S10–S19. ISSN: 1879-355X (Electronic){\}r0360-3016 (Linking) (2010).

122. Astrahan, M. Some implications of linear-quadratic-linear radiation dose-response with regard to hypofractionation. *Med Phys* **35,** 4161–4172. ISSN: 2473-4209 (2008).

123. RCR. Stereotactic Ablative Body Radiation Therapy (SABR): A Resource. *UK SABR Consortium* (2019).

124. Guckenberger, M. *et al.* Applicability of the linear-quadratic formalism for modeling local tumor control probability in high dose per fraction stereotactic body radiotherapy for early stage non-small cell lung cancer. *Radiother Oncol* **109,** 13–20. ISSN: 0167-8140 (2013).

125. Brenner, D. J. The Linear-Quadratic Model Is an Appropriate Methodology for Determining Isoeffective Doses at Large Doses Per Fraction. *Semin Radiat Oncol* **18,** 234–239. ISSN: 1053-4296 (2008).

126. Tobias, C. A. The repair-misrepair model in radiobiology: comparison to other models. *Radiation research. Supplement* **8,** S77–95. ISSN: 0485-8611 (1985).

127. Curtis, S. B. Lethal and potentially lethal lesions induced by radiation–a unified repair model. *Radiation research* **106,** 252–70. ISSN: 0033-7587 (1986).

128. Goodhead, D. T. Saturable repair models of radiation action in mammalian cells. *Radiation research. Supplement* **8,** S58–67. ISSN: 0485-8611 (1985).

129. Semenenko, V. & Stewart, R. A Fast Monte Carlo Algorithm to Simulate the Spectrum of DNA Damages Formed by Ionizing Radiation. *Radiat Res* **181,** 451–7. ISSN: 0033-7587 (2004).

130. Semenenko, V. & Stewart, R. Fast Monte Carlo simulation of DNA damage formed by electrons and light ions. *Phys Med Biol* **51,** 1693–706. ISSN: 0031-9155 (2006).

131. Stewart, R. D. *et al.* Rapid MCNP simulation of DNA double strand break (DSB) relative biological effectiveness (RBE) for photons, neutrons, and light ions. *Phys Med Biol* **60,** 8249–74. ISSN: 0031-9155 (2015).

132. Ballarini, F. *et al.* The BIANCA model/code of radiation-induced cell death: application to human cells exposed to different radiation types. *Radiat Environ Bioph* **53,** 525–533. ISSN: 0301-634X (2014).

133. J, M. S., Schuemann, J., Paganetti, H. & Prise, K. M. Mechanistic Modelling of DNA Repair and Cellular Survival Following Radiation-Induced DNA Damage. *Sci Reports* **6,** 33290. ISSN: 2045-2322 (2016).

134. McMahon, S. J., McNamara, A. L., Schuemann, J., Paganetti, H. & Prise, K. M. A general mechanistic model enables predictions of the biological effectiveness of different qualities of radiation. *Scientific Reports* **7,** 10790. ISSN: 2045-2322 (2017).

135. Friedland, W., Jacob, P., Bernhardt, P., Paretzke, H. G. & Dingfelder, M. Simulation of DNA damage after proton irradiation. *Radiation research* **159,** 401–10. ISSN: 0033-7587 (2003).

136. Friedland, W., Jacob, P. & Kundrát, P. Stochastic Simulation of DNA Double-Strand Break Repair by Non-homologous End Joining Based on Track Structure Calculations. *Radiation Research* **181,** 677–688. ISSN: 0033-7587 (2010).

137. Friedland, W. *et al.* Comprehensive track-structure based evaluation of DNA damage by light ions from radiotherapy-relevant energies down to stopping. *Sci Reports* **7,** 45161. ISSN: 2045-2322 (2017).

138. Henthorn, N. *et al.* In Silico Non-Homologous End Joining Following Ion Induced DNA Double Strand Breaks Predicts That Repair Fidelity Depends on Break Density. *Scientific Reports* **8,** 2654 (2018).

139. Ingram, S. P. *et al.* Mechanistic modelling supports entwined rather than exclusively competitive DNA double-strand break repair pathway. *Scientific Reports* **9,** 6359. ISSN: 2045-2322 (2019).

140. Warmenhoven, J. W. *et al.* Insights into the non-homologous end joining pathway and double strand break end mobility provided by mechanistic in silico modelling. *DNA Repair* **85,** 102743. ISSN: 1568-7864 (2020).

141. Friedland, W., Jacob, P. & Kundrát, P. Mechanistic simulation of radiation damage to DNA and its repair: on the track towards systems radiation biology modelling. *Radiat Prot Dosim* **143,** 542–8. ISSN: 0144-8420 (2010).

142. Henthorn, N. *et al.* Nanodosimetric Simulation of Direct Ion-Induced DNA Damage Using Different Chromatin Geometry Models. *Radiation Research* **188.** ISSN: 0033-7587 (2017).

143. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* **2,** 35066075. ISSN: 1471-0064 (2001).

144. Carter, R. J. *et al.* Complex DNA Damage Induced by High Linear Energy Transfer Alpha-Particles and Protons Triggers a Specific Cellular DNA Damage Response. *Int J Radiat Oncol Biology Phys* **100,** 776–784. ISSN: 0360-3016 (2018).

145. Schipler, A. & acids research, I. G. DNA double-strand–break complexity levels and their possible contributions to the probability for error-prone processing and repair pathway choice. *Nucleic acids research* **41,** 7589–7605. ISSN: 0305-1048 (2013).

146. Li, Y., Reynolds, P., Peter, O. & Cucinotta, F. A. Modeling Damage Complexity-Dependent Non-Homologous End-Joining Repair Pathway. *Plos One* **9,** e85816 (2014).

147. Wahl, G. M. & Carr, A. M. The evolution of diverse biological responses to DNA damage: insights from yeast and p53. *Nat Cell Biol* **3,** ncb1201–e277. ISSN: 1476-4679 (2001).

148. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle Georget Tex* **7,** 2902–6. ISSN: 1538-4101 (2008).

149. Li, X. & Heyer, W. Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res* **18,** cr20081. ISSN: 1748-7838 (2008).

150. Nickson, C. M., Moori, P., Carter, R. J., Rubbi, C. P. & Parsons, J. L. Misregulation of DNA damage repair pathways in HPV-positive head and neck squamous cell carcinoma contributes to cellular radiosensitivity. *Oncotarget* **8,** 29963–29975. ISSN: 1949-2553 (2017).

151. Nakano, T. *et al.* Homologous Recombination but Not Nucleotide Excision Repair Plays a Pivotal Role in Tolerance of DNA-Protein Cross-links in Mammalian Cells. *J Biol Chem* **284,** 27065–27076. ISSN: 0021-9258 (2009).

152. Girst, S. *et al.* Subdiffusion Supports Joining Of Correct Ends During Repair Of DNA Double-Strand Breaks. *Sci Reports* **3,** 2511 (2013).

153. Lamarche, B. J., Orazio, N. I. & Weitzman, M. D. The MRN complex in double-strand break repair and telomere maintenance. *FEBS Lett.* **584,** 3682–95. ISSN: 0014-5793 (2010).

154. Erixon, K. & research, C. B. Linear induction of DNA double-strand breakage with X-ray dose, as determined from DNA fragment size distribution. *Radiation research* (1995).

155. Stewart, R. D. *et al.* Effects of Radiation Quality and Oxygen on Clustered DNA Lesions and Cell Death. *Radiat Res* **181,** 587–602. ISSN: 0033-7587 (2011).

156. Mairani, A. *et al.* The FLUKA Monte Carlo code coupled with the local effect model for biological calculations in carbon ion therapy. *Phys Med Biol* **55,** 4273–4289. ISSN: 0031-9155 (2010).

157. Carlson, D. J., Stewart, R. D., Semenenko, V. A. & Sandison, G. A. Combined Use of Monte Carlo DNA Damage Simulations and Deterministic Repair Models to Examine Putative Mechanisms of Cell Killing. *Radiat Res* **181,** 447–59. ISSN: 0033-7587 (2008).

158. Cornforth, M. & research, B. J. A quantitative comparison of potentially lethal damage repair and the rejoining of interphase chromosome breaks in low passage normal human fibroblasts. *Radiation research* (1987).

159. Hamada, N., Funayama, T., Wada, S. & Radiation …, S. T. LET-dependent survival of irradiated normal human fibroblasts and their descendents. *Radiation …* (2006).

160. Neti, P. V., de Toledo, S. M., Perumal, V., Azzam, E. I. & Howell, R. W. A Multiport Low-Fluence Alpha-Particle Irradiator: Fabrication, Testing and Benchmark Radiobiological Studies. *Radiat Res* **181,** 732–8. ISSN: 0033-7587 (2004).

161. Rothkamm, K., Krüger, I., Thompson, L. H. & Löbrich, M. Pathways of DNA Double-Strand Break Repair during the Mammalian Cell Cycle. *Mol Cell Biol* **23,** 5706–5715. ISSN: 0270-7306 (2003).

162. Kühne, M. *et al.* A double-strand break repair defect in ATM-deficient cells contributes to radiosensitivity. *Cancer Res.* **64,** 500–8. ISSN: 0008-5472 (2004).

163. Beucher, A. *et al.* ATM and Artemis promote homologous recombination of radiation induced DNA double strand breaks in G2. *Embo J* **28,** 3413–3427. ISSN: 1460-2075 (2009).

164. Löbrich, M., Kühne, M., Wetzel, J. & Rothkamm, K. Joining of correct and incorrect DNA double-strand break ends in normal human and ataxia telangiectasia fibroblasts. *Genes Chromosomes Cancer* **27,** 59–68. ISSN: 1098-2264 (2000).

165. Rydberg, B., Cooper, B., Cooper, P. & Radiation …, H. W. Dose-dependent misrejoining of radiation-induced DNA double-strand breaks in human fibroblasts: experimental and theoretical study for high-and low-LET radiation. *Radiation …* (2005).

166. Morgan, T. L. *et al.* Molecular characterization of X-ray-induced mutations at the HPRT locus in plateau-phase Chinese hamster ovary cells. *Mutat Res Fundam Mol Mech Mutagen* **232,** 171–182. ISSN: 0027-5107 (1990).

167. Belli, M. *et al.* RBE-LET relationships for cell inactivation and mutation induced by low energy protons in V79 cells: further results at the LNL facility. *Int J Radiat Biol* **74,** 501–509. ISSN: 0955-3002 (1998).

168. Schwartz, J., Jordan, R., Sun, J., Ma, H. & Research, H. A. Dose-dependent changes in the spectrum of mutations induced by ionizing radiation. *Radiation Research* **181,** 312–317. ISSN: 0033-7587 (2000).

169. Alexander, T., Ma, Y. & R, L. M. Analysis of the Kinetic and Equilibrium Binding of Ku Protein to DNA. *J Theor Biol* **214,** 85–97. ISSN: 0022-5193 (2002).

170. Uematsu, N. *et al.* Autophosphorylation of DNA-PKCS regulates its dynamics at DNA double-strand breaks. *J Cell Biology* **177,** 219–229. ISSN: 0021-9525 (2007).

171. Belli, M. *et al.* DNA Fragmentation in mammalian cells exposed to various light ions. *Adv Space Res* **27,** 393–399. ISSN: 0273-1177 (2001).

172. Dingfelder, M., Inokuti, M. & Paretzke, H. G. Inelastic-collision cross sections of liquid water for interactions of energetic protons. *Radiat Phys Chem* **59,** 255–275. ISSN: 0969-806X (2000).

173. Höglund, E., Blomquist, E., Carlsson, J. & Stenerlöw, B. DNA damage induced by radiation of different linear energy transfer: initial fragmentation. *Int J Radiat Biol* **76,** 539–547. ISSN: 0955-3002 (2000).

174. Stenerlöw, B., Höglund, E., Carlsson, J. & Blomquist, E. Rejoining of DNA fragments produced by radiations of different linear energy transfer. *Int J Radiat Biol* **76,** 549–557. ISSN: 0955-3002 (2000).

175. Francis, Z. *et al.* Molecular scale track structure simulations in liquid water using the Geant4-DNA Monte-Carlo processes. *Appl Radiat Isotopes* **69,** 220–226. ISSN: 0969-8043 (2011).

176. Watanabe, R., Rahmanian, S. & Nikjoo, H. Spectrum of Radiation-Induced Clustered Non-DSB Damage – A Monte Carlo Track Structure Modeling and Calculations. *Radiat Res* **183,** 525–540. ISSN: 0033-7587 (2015).

177. Chaudhary, P. *et al.* Variations in the Processing of DNA Double-Strand Breaks Along 60-MeV Therapeutic Proton Beams. *Int J Radiat Oncol Biology Phys* **95,** 86–94. ISSN: 0360-3016 (2016).

178. McMahon, S. J. & Prise, K. M. A Mechanistic DNA Repair and Survival Model (Medras): Applications to Intrinsic Radiosensitivity, Relative Biological Effectiveness and Dose-Rate. *Frontiers in Oncology* **11,** 689112. ISSN: 2234-943X (2021).

Chapter 3

Mechanistic modelling supports entwined rather than exclusively competitive DNA double-strand break repair pathway

Samuel P Ingram[1,2,*], John W Warmenhoven[1,3], Nicholas T Henthorn[1,3], Edward A K Smith[1,2,*], Amy L Chadwick[1,3], Neil G Burnet[1,3], Ranald I Mackay[2,1], Norman F Kirkby[1,3], Karen J Kirkby[1,3], Michael J Merchant[1,3],

1 Division of Cancer Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, UK
2 Christie Medical Physics and Engineering, The Christie NHS Foundation Trust, Manchester, UK
3 The Christie NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

* samuel.ingram@postgrad.manchester.ac.uk

Published as Ingram, S. P. et al. Mechanistic modelling supports entwined rather than exclusively competitive DNA double-strand break repair pathway. Sci Rep-uk 9, 6359 (2019). Slight modifications to formatting for consistency.

3.1 Abstract

Following radiation induced DNA damage, several repair pathways are activated to help preserve genome integrity. Double Strand Breaks (DSBs), which are highly toxic, have specified repair pathways to address them. The main repair pathways used to resolve DSBs are Non-Homologous End Joining (NHEJ) and Homologous Recombination (HR). Cell cycle phase determines the availability of HR, but the repair choice between pathways in the G2 phases where both HR and NHEJ can operate is not clearly understood. This study compares several *in silico* models of repair choice to experimental data published in the literature, each model representing a different possible scenario describing how repair choice takes place. Competitive only scenarios, where initial protein recruitment determines repair choice, are unable to fit the literature data. In contrast, the scenario which uses a more entwined relationship between NHEJ and HR, incorporating protein co-localisation and RNF138-dependent removal of the Ku/DNA-PK complex, is better able to predict levels of repair similar to the experimental data. Furthermore, this study concludes that co-localisation of the Mre11-Rad50-Nbs1 (MRN) complexes, with initial NHEJ proteins must be modelled to accurately depict repair choice.

3.2 Introduction

Radiation is known to damage DNA both directly and indirectly. This DNA damage elicits a range of biological responses to ensure the preservation of genome integrity. If left unrepaired, DNA damage can lead to cell death [1]. Therefore, a key function of the DNA Damage Response (DDR) is the attempt to preserve the cell's function. There are several repair pathways available for DNA repair, each with its own collection of biological conse-

quences. For example, Base Excision Repair (BER) is responsible for the removal of non-helix-distorting base lesions, whereas Nucleotide Excision Repair (NER) is responsible for removal of the bulky helix-distorting lesions. Other pathways are thought to compete for the same types of damage, such as the availability of both Non-Homologous End Joining (NHEJ) and Homologous Recombination (HR) for the repair of DNA Double Strand Breaks (DSBs) during the G2 phase of the cell cycle. There are also "back-up" pathways for situations where primary repair pathways are unavailable: for example Microhomology-Mediated End Joining (MMEJ) is sometimes used instead of NHEJ, but will often cause large deletion mutations. This variety of repair pathways, acting in unison to address DNA damage, accounts for the observed biological resilience in dealing with natural levels of endogenous and exogenous damage resulting from everyday life.

The exogenous generation of DNA damage is not intrinsically bad; it is utilised in the form of cancer treatments where the aim is to achieve cell death within a cancer target while avoiding too much damage to healthy surrounding tissue and critical organs. For instance, radiotherapy is used in the treatment of 40-50% of patients diagnosed with cancer [2, 3], because of its non-invasive ability to damage DNA and can be geographically localised. Radiotherapy works by creating high amounts of DNA damage, much greater and more complex than natural levels, leading to increased cell death. Radical radiotherapy aims to deliver enough radiation to the cancerous target volume to gain local control while minimising radiation dose to healthy tissues and critical organs and thereby avoiding compromises in quality of life. Although it is understood that elevated levels of cell death are caused through radiotherapy, there is a clinical need to understand and quantify how both DNA damage and repair impact patient treatment for different radiation modalities, energies and spectra, and dose rates.

Treatment planning is based on physical dose, with dose prescriptions guided by clinical experience. Biological effects are considered through phenomenological modelling. For example, the Linear-Quadratic model is used clinically to guide both fractionation and organ dose constraints. However, at the planning stage biological optimisation of dose deposition is not considered. While aspects of not accounting for these biological factors are counteracted by more than 100 years of experience in X-ray radiotherapy, in other treatment modalities, such as proton beam therapy, not accounting for these variations in radiobiological effectiveness (RBE) could potentially limit treatment success [4, 5]. There has been a recent resurgence of interest in modelling of DNA damage and repair [6–13], which aims to give quantifiable depictions of the DNA damage response following irradiation, with some models **[7, 9]** able to account for some of the aforementioned radiation properties. Some *in silico* models **[10, 12]** attempt to evaluate the biological response through observations of DNA damage alone, although they neglect the variations known to exist in repair function between cell type, organ type and patient **[14–16]**. By including models of the DNA repair, it is possible to account for some of these differences, and predict their relationship with different DNA-level endpoints (e.g. repaired, un-repaired damage and misrepair) [7], and potentially relate these to cell fate [8]. Although more onerous, including DNA repair in

in silico models is the first step towards achieving patient-specific, biologically optimised, treatment plans.

In radiotherapy, radiobiological models of DNA repair have typically been focused on pathways that interact with DSBs, since DSBs have a significant role in cell fate [1]. It has been established that, for a population of cells, NHEJ is the most frequently used repair pathway as it is present through all phases of the cell cycle [17, 18], whereas HR becomes more pronounced during the S and G2 phases due to the availability of a sister chromatid supplying proximal homology. Therefore, in most work, NHEJ is the focus of the modelling efforts [6–8, 19], with HR being partially or entirely omitted from models. Although HR makes up a smaller proportion of repair, several publications have suggested that HR may be preferential for the repair of proton-induced DSBs, and that in cohorts of HR impaired patients there are significant clinically observable effects [20–23]. Futhermore, HR is of interest for drug radiosensitizers, HR-deficiencies have been shown to be especially lethal when paired with a poly(ADP-ribose) polymerase (PARP) inhibitor [24, 25], which may be exploited for treatment of some cancer types. Finally, incorporation of more than one repair pathway is important as it is believed that many subsets of cancers include perturbations to various DNA repair pathways, altering their response to radiation [23].

It is well established that the availability of prominent HR repair is regulated by the cell cycle and sister chromatid availability, which the pathway uses as a homologous template [26]. However, the processes of repair choice are less well established (i.e. during G2 where both NHEJ and HR are available). To model the mechanism of repair choice in an *in silico* step-by-step model, like the one proposed within this study, requires the explicit inclusion of the stage or stages within the repair pathway where repair choice is possible. It has been generally believed that NHEJ and HR act in a predominately competitive fashion **[27],** with initial protein attachment directing repair fate. However, other repair choice models have been suggested within the literature **[27–30]** and it is under these different scenarios that our *in silico* models will be compared against experimental literature data. Four scenarios are tested: Scenario A) 'NHEJ first' approach, Scenario B) 'no way back' approach, Scenario C) 'continuous competition' approach, and Scenario D) 'entwined pathway' approach. The repair pathways NHEJ and HR have different biological consequences with NHEJ being more error prone than HR [29]. Therefore, understanding how to model the repair choice *in silico* is required in any mechanistic attempt to quantify the amount and fidelity of repair which takes place for a given DNA damage pattern. It has been shown that while NHEJ and HR share some of the DSB caseload, when each pathway is removed in turn, the remaining pathway can compensate for its deficient counterpart to a differing extent [31]. It is the purpose of this paper to evaluate this complex behaviour which suggests a level of overlap between the repair pathways.

3.3 Methods

The *in silico* framework used to produce the models for this study is the DNA Mechanistic Repair Simulator (DaMaRiS) framework [6, 7], developed at University of Manchester which has been built using the Monte Carlo Geant4-DNA simulation toolkit **[32]** (version 10.4). The produced DaMaRiS model requires an input of the distribution of DNA DSB damage following irradiation, this was supplied by an in-house model of DNA damage [7], but in principle could be supplied by other DNA damage models which output in the Standard DNA Damage (SDD) Data Format [33]. The DaMaRiS framework (version 0.3) is designed to be versatile, enabling the user to define, build and test models of DNA repair. The mechanistic nature of the DaMaRiS framework requires every parameter in the model should have a biological definition, thereby allowing for model alteration by the addition and/or removal of biological mechanisms as they are elucidated within the literature. The applied mechanisms, in unison, form entire repair pathways (e.g. NHEJ and HR), making DaMaRiS well suited for the multi-scenario modelling presented in this study. Each mechanistic step of the model depicts how proteins attach and react in order to process broken DNA ends from break recognition through to repair completion. The framework provides temporal and information on how the DNA damage is processed along the repair pathways at any time point. The DaMaRiS models in this study consist of NHEJ and the initial steps of HR, with the intent to add further repair pathways to provide a full depiction of the DDR at the DNA level. Since the DaMaRiS models are mechanistic, where each core step of a pathway is incorporated in the model, it is necessary to understand at which point or points the repair pathway choice is carried out (i.e. areas of repair pathway where the DNA end can transfer from being actively repaired by one available pathway to another). Furthermore, it is important to assess, in the case of repair failure, if alternative pathways become available or even enforced, as depicted in Scenario A below. This study is the first time HR has been modelled in our DNA repair work; all previous work has focused on NHEJ with no repair choice. Furthermore, we have used the model to describe protein deficient systems which are a new potential area of the DaMaRiS framework being explored. A schematic overview of the methodology process has been outlined in Fig. 3.1.

3.3.1 Repair choice scenarios.

The repair choice scenarios (Fig. 3.2) modelled, as well as a brief summary of their justification, include:

A) The 'NHEJ first' approach - NHEJ processes so much faster than HR and is therefore always attempted first with only failure allowing HR repair to be used [28, 29].

The concept of NHEJ being the first attempted repair pathway originates from the experimentally observed faster repair kinetics [34, 35]. Therefore, this scenario is analogous to NHEJ being so much faster that it is always attempted first, rather than a competitive ap-

proach used in the other scenarios. The model was adapted so that DSBs entering the system could only progress down NHEJ, and only upon failure at the synapsis complex stage is HR attempted (i.e. DNA-PKcs auto-phosphorylation and disassociation) [35]. The HR repair following NHEJ failure was enforced (i.e. HR was the only available pathway upon NHEJ failure), removing the possibility of multiple attempts of NHEJ for the same break.

B) The 'no way back' approach competitive directed repair with no cross-over of pathways, the initially recruited protein locks the broken end into its corresponding repair pathway which it can either succeed or fail to repair [27].

NHEJ and HR repair pathways being directed by competition is a commonly used hypothesis within the literature [36, 37]. However, when implementing a competition-based approach the concept of pathway cross-over (i.e. DSB ends being able to transfer from one pathway to another) needs to be addressed. In Scenario B there is no cross-over allowed; this is realised by the first protein attachment to the broken DNA end locking the end into the respective pathway, effectively inhibiting the other pathway from even attempting repair at any time point. Whilst this scenario's description of cross-over is not explicitly stated within the literature, it does represent the commonly held assumption of repair pathway cross-over (i.e. no cross-over) shown in textbook descriptions [38]. Furthermore, the scenario acts as a point of completeness when comparing to the other cross-over description, in which, upon failure of repair, the DNA end becomes available for re-competition.

C) The 'continious competition approach' same as Scenario B: but upon failure of repair the end is available for competition again [30].

This scenario depicts NHEJ and HR competition in combination with allowed cross-over of repair pathways upon repair failure. Upon all cases of dissociation, there is re-competition between initial HR and NHEJ proteins. This re-competition enables HR proteins to have a chance to interact with isolated DNA ends which are not suitable for NHEJ but can be repaired by HR. This scenario, along with Scenario D, is seen as the least directed system and is largely dictated by protein recruitment kinetics.

D) The 'entwined pathway' approach: similar to Scenario C, but includes mechanisms of protein co-localisation and RNF138 mediated Ku70/80 removal.

In this scenario, additional mechanisms from the recent literature were added which resulted in a substantial change for the simulated repair kinetics and partialy addressed the issue of Ku70/80 and CtBP-interacting protein (CtIP) competition detailed above. Scenario D is similar to Scenario C as it utilises continuous competition, but includes the three additional mechanisms:

- MRN co-localisation with Ku70/80, DNA-PK complex (Ku70/80 and DNA-PKcs) and the DNA-PK synapsis complex [36, 39].

- RNF138-dependent Ku removal from the DNA ends with MRN attached to allow for an additional point of resection, thereby promoting HR [40].
- MRN independent resection to allow for naked DSB ends to undergo resection [41].

These additional mechanisms describe a more entwined initial DNA-repair response than the previously perceived hypothesis that exclusive competition directs the repair choice [27, 30, 37].

These scenarios are assumed to span the most likely possibilities for the mechanism of repair choice. However, this field is still being actively developed; therefore the framework remains available for the evaluation of further drivers of repair choice. Aspects such as DNA density (i.e. euchromatin and heterochromatin) and damage complexity which are believed to also impact repair choice [42, 43] will be assessed independently of this study.

3.3.2 Model construction.

The DaMaRiS models are altered versions of one another with the removal and/or addition of available progression points from the initial DSB end, describing each of the proposed repair choice scenarios (Fig. 3.2). The transition from one state to another is controlled through a series of time constants acted upon by the random number generator of Geant4, generating a distribution of time delays for progression. Each time constant and its generated distribution of time delays have been derived by fitting to experimental protein recruitment kinetics, giving an accurate progression rate of step-by-step repair (see Supplementary Data - Figures S1 & S2). The only time-constant not fitted to experimental data is the progression rate between a "resected end" to "repaired by HR" (τ_{RR}). Other steps, i.e. the rad51-filament formation, homology-seeking along the sister chromatid, DNA replication and double Holliday Junction resolution have not been explicitly modelled, so a combined large time constant is used instead. This loss of detail, after resection within the HR pathway, does not affect the repair choice process, since it has been established that canonical NHEJ cannot be utilised following post-resection from CtIP [44]. Although new work regarding NHEJ resection mediated repair in G1 has been proposed [45], this study focuses in G2 where the assumption is that resection-mediated repair is predominately HR. Therefore, the time constant τ_{RR} was allowed to vary in order to achieve the best fit to the experimental data. If a scenario did not fit the experimental data whilst varying τ_{RR} it was concluded that the scenario is not fit for purpose (see Supplementary Data - Figures S3-S6).

3.3.3 Model evaluation.

Repair kinetics of DSBs for each scenario were compared to phosphorylated form of histone H2AX (γ-H2AX) foci, a commonly used experimental marker of DSBs. It is well established that on average HR takes significantly more time than NHEJ to repair a break, and this has given rise to the extensively used dual component description of slow (HR) and fast

(NHEJ) repair to describe the repair rates seen [31, 45]. The repair choice of each scenario will alter the ratio of HR and NHEJ and since each pathway has specific progression rates, the overall shape of the repair rate is changed. This allows the agreement between the generated data and the experimental data to be compared. Furthermore, these quantities can be altered by the scenarios whilst remaining predominately independent of protein recruitment kinetics. However, to ensure a fair comparison the recruitment kinetics have been optimised for each scenario (Supplementary Data - Figures S1 & S2), but fixed when evaluating against the experimental γ-H2AX foci data.

To gain further confidence for each scenario the repair rates in two deficient cell lines were also reviewed. This review was possible because the mechanistic step-by-step model can have progression points removed (inhibited) at any point along the pathway, allowing accurate depictions of different deficiencies within the same repair pathway. This results in four scenario models being produced, each with the three configurations (Fig. 3.2): wild-type cells - all progression points intact, XLF-deficient cells - all progression points intact excluding synapsis complex stabilisation between two DNA-PKcs-loaded ends within NHEJ [46], and Lig4-deficient cells - all progression points intact excluding formed synapsis complex ligation creating repaired DNA through NHEJ [47].

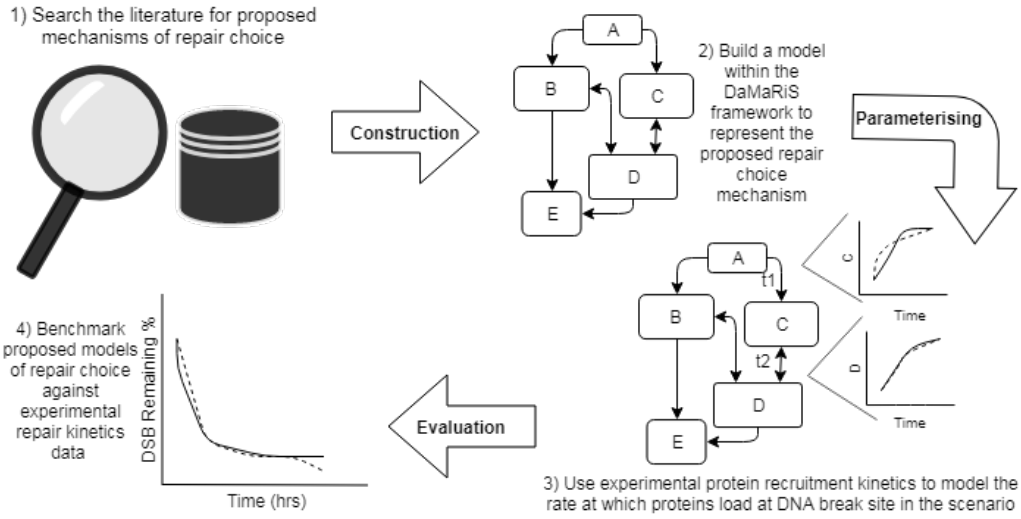

Figure 3.1: Schematic overview of this study's methodology. 1) The literature was reviewed to establish possible mechanisms of repair choice. 2) These mechanisms were pieced together to form cohesive possible scenarios of DNA double-strand break repair choice. The proposed scenarios and the required protein recruitment steps were formed in the DaMaRiS framework. 3) The created scenarios were parameterised to allow for the addition of protein recruitment time constants which emulate protein recruitment kinetics similar to experimental literature data. 4) The produced models were used to simulate individual cells undergoing DNA repair following radiation damage and the overall repair kinetics were measured for each proposed scenario. The measured repair kinetics were then evaluated against experimental γ-H2AX foci data to benchmark each scenario.

3.3.4 Experimental benchmarks.

The overall repair kinetics were compared to the γ-H2AX foci data extracted from Beucher *et al.* [31] using WebPlotDigitizer (https://automeris.io/WebPlotDigitizer [48]) , to evalu-

ate the modelled repair choice scenarios. The goodness of fit was compared between the experimental data points and corresponding simulation data points. The comparison was carried out using Reduced Chi-Square (χ^2_{red}), Root Mean Square Error (RMSE) and Dynamic Time Warping (DTW) [49]. Both χ^2_{red} and RMSE evaluate data points at the same time point, whereas DTW can evaluate possible time discontinuities between the simulated and experimental data sets. In the Beucher *et al.* experiment, the DNA damage was induced by a 2 Gy exposure from a Cs-137 source. The Beucher *et al.* experimental data set used was from G2-phase human fibroblast (HF) and mouse embryonic fibroblasts (MEFs) cells and included variations of wild-type (WT) cells (C2886 and "WT2" for human and mouse respectively), XLF-deficient cells (2BN HF) and Lig4-deficient cells ("Lig4-/-" MEFs). The Beucher data set was used to evaluate the simulations agreement to experiment. This is due to it being the only dataset used which has been G2 synchronised. Further experimental data was sourced and added to evaluate how the repair rate may vary through different experimental parameters. Experiments carried out by Kuhne *et al.* [50] delivered 2 Gy of 90 kVp X-rays, filtered by 1 mm of aluminium. The data extracted from Kuhne *et al.* includes a Wild-Type Lung HF (48BR) and an additional Lig4-deficient cell line (411BR). However, the Lig4-deficient line was derived from a patient with Lig4 syndrome, which results in a prominent but not complete removal of available Lig4 protein. The experimental data extracted from Wu *et al.* 2012 [51] is for 2 Gy of 7.5 MeV protons, which corresponds to a reported linear energy transfer (LET) of 6 keV/μm (calculated by the authors using SRIM), delivered to Normal Human Lung Fibroblast (NHLF) cells. All the experimental repair rates were deduced from γ-H2AX foci experiments and were self-normalised. The *in silico* models were configured to simulate the total amount of repair through NHEJ and HR of DSBs at 8-hours post-irradiation (matching the Beucher *et al.* data set), recording the progression of the simulation every 30 seconds. The DSB pattern used as an input for the simulation was created from the model described by Henthorn *et al.* [7] and simulates 2 Gy to the cell nucleus using a 34 MeV mono-energetic proton beam (equivalent to a track-averaged LET of 1.77 keV/μm). Relatively low LET protons were selected as experimental data has shown that the repair kinetics are similar to that of photon irradiation [52, 53].

3.3.5 Data availability

The data sets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

3.4 Results

Scenarios are analysed for best fit with the experimental data from the literature. The allowed variation in τ_{RR} was implemented to investigate if the Beucher et al. experimental data points at the various time points could be achieved through simulation. The ability to do so is indicative of a suitable repair choice scenario, while the deviation of fit when using

an average value of τ_{RR} (Fig. 3.3) may be an indication of missing mechanisms within the current model (see discussion). The goodness of fit metrics for each Scenario in each cell system are presented in Table 3.1.

Simulated System	Reduced Chi Sq	Root Mean Square Error	Dynamic Time Warping
Scenario A - Lig4	0.42	4.31	16.16
Scenario A - XLF	4.35	13.18	49.73
Scenario A - HF WT	4.18	7.92	30.90
Scenario A - MEF WT	6.89	10.42	37.18
Scenario A Average	**3.96**	**8.96**	**33.49**
Scenario B - Lig4	13.61	23.38	100.59
Scenario B - XLF	13.26	23.44	102.65
Scenario B - HF WT	5.13	10.97	37.38
Scenario B - MEF WT	3.88	6.68	23.84
Scenario B Average	**8.97**	**16.12**	**66.11**
Scenario C - Lig4	2.82	10.63	43.37
Scenario C - XLF	0.40	4.09	16.36
Scenario C - HF WT	6.72	9.86	42.24
Scenario C - MEF WT	4.76	10.48	40.62
Scenario C Average	**3.68**	**8.77**	**35.65**
Scenario D - Lig4	0.61	5.16	16.19
Scenario D - XLF	3.44	11.72	44.20
Scenario D - HF WT	2.85	6.71	26.67
Scenario D - MEF WT	4.76	8.64	28.97
Scenario D Average	**2.92**	**8.06**	**29.01**

Table 3.1: Summary of goodness-of-fit metrics between simulated and the experimental Beucher data set [31] and the purposed repair choice scenarios. Each cell system from each scenario has been evaluated against the experimental data set. For every experimental data point a corresponding simulation point was evaluated against. Reduced Chi Square (χ^2), Root Mean Square Error (RMSE) and Dynamic Time Warping (DTW) were used to quantify the goodness-of-fit. The values presented in bold show the mean goodness-of-fit for each Scenario.

3.4.1 Scenario A - The 'NHEJ first' approach.

Scenario A has a relatively good agreement ($\bar{\chi}^2_{red}$ = 3.96) with the experimental data from each tested cell system (i.e. WT, Lig4-deficient and XLF-deficient). In the WT-system (Fig. 3.3b) the repair rate up to 2-hours agrees with the experimental human fibroblast data and demonstrates a slower but similar shape until 8 hours. In the XLF-deficient system (Fig. 3.3c) the simulation predicts a higher amount of repair beyond 4 hours than is seen in the experimental data. Furthermore, the repair rate appears to be missing any bi-phasic behaviour resulting in a predominately linear result and negates the required curvature to match the experimental data. In the Lig4-deficient system (Fig. 3.3d) Scenario A gives the best agreement over all the tested scenarios (χ^2_{red} = 0.42). However, this is only marginally better than Scenario D. The agreement demonstrates the requirement of Lig4 being present for final ligation and the premise of a stabilised DNA-PKcs synaptic complex [46, 54], which in the Lig4-deficient cell system would result in increased un-repaired DNA ends within the NHEJ pathway.

3.4.2 Scenario B - The 'no way back' approach.

The goodness-of-fit metrics, shown in Table 3.1, highlight that the concept of the 'no way back' approach has the least agreement with the evaluated Beucher *et al.,* data set ($\bar{\chi}^2_{red}$ = 8.97). In the WT-system (Fig. 3.3b) the simulated results of Scenario B demonstrates a situation where the absolute differences from the experimental points are not extensive, but the overall shape shows clear disagreement with all data sets. The simulation shows very fast repair kinetics before the 2-hour time point and a plateau beyond this. This shape is due to NHEJ dominating the system and DNA ends which become isolated, without a partner end to form a synaptic, will remain as such for the rest of the simulation. In both the XLF- and Lig4-deficient systems (Fig. 3.3 c & d) the NHEJ dominance is evident, with almost no repair seen. In principle, whilst not fitting with the experimental results, this is the expected result from the model. This dominance occurs since Ku70/80, the initial responder for NHEJ repair, is known to be highly abundant **[55]** and has been shown to have fast recruitment kinetics [34, 35]. These protein characteristics have been implemented within the *in silico* model, resulting in almost all repair being locked into the NHEJ repair pathway within an NHEJ-deficient cell system, giving rise to a large number of persistent unrepaired ends.

3.4.3 Scenario C - The 'continuous competition' approach.

Scenario C has better goodness of fit metrics than Scenario B, but worse than Scenarios A and D (Table 3.1). In both the WT and Lig4-deficient systems (Fig. 3.3b & d) it is seen that the simulated data for Scenario C have demonstrated a slower repair rate than the experimental data. Though the amount of repair in the deficient systems is in better agreement with the experimental data than Scenario B, there is still not enough total HR repair within the 8-hour time frame. These slower kinetics are due to re-competition between Ku70/80 and CtIP still being predominantly won by Ku70/80's faster recruitment kinetics. This results in several attempts of NHEJ before CtIP can attach and resect the end allowing HR to attempt repair. The repair kinetics simulated for the XLF-deficient system is in strong agreement (χ^2_{red} = 0.40) with the experiential data set.

3.4.4 Scenario D - The 'entwined pathway' approach.

The simulated results for Scenario D are similar, in kinetics and shape, to those from Scenario A with the advantage of not having to be a directed system. Overall, Scenario D has the best goodness-of-fit metrics across all tests ($\bar{\chi}^2_{red}$ = 2.92, RMSE = 8.06, DTW =29.01). In the WT-system (Fig. 3.3b) the scenario closely mimics the shape of the experimental data with improved goodness-of-fit metrics than Scenario A. In the XLF-deficient system (Fig. 3.3c) the results closely mimic that of Scenario A with slightly slower repair rates. In a similar manner of Scenario A the repair rate post 4 hours is faster than the experimental

data which may be due to the HR model missing any form of stalling mechanisms post resection. In the Lig4-deficient system (Fig. 3.3d) there is again a similar shape to Scenario A with larger deviations at later time points, including noticeably slower kinetics at 8 hours. Compared to Scenario A, there is a higher chance of DNA-PKcs synapsis stabilisation due to the lack of enforcement of HR after a failure of NHEJ and thus increased un-repaired stable complexes without Lig4 to ligate the break.

3.4.5 Scenario D - Implications of the additional mechanisms.

The competition, in Scenario C, between Ku70/80 and CtIP recruitment kinetics does not allow for HR repair at the same rate seen in the experimental data. It is known that the Mre11-Rad50-Nbs1 (MRN) complex is also required for CtIP activation and resection [41, 56]. However, explicit modelling of the recruitment kinetics of MRN was not established within Scenario A, B, or C and instead explicit modelling of CtIP was included with the assumption that MRN would be present at this point. Furthermore, the previous scenarios were unable to match with experimental results of Mre11 recruitment. However, upon enabling the recruitment of MRN alongside attempting the NHEJ pathway, the simulation recruitment kinetics have good agreement with the experimental recruitment kinetics (Fig. 3.4c). This addition of MRN co-localisation with NHEJ proteins results in a naturally occurring bias, that if MRN has attached during the NHEJ repair attempt, then the CtIP attachment is more likely to attach at points of re-competition.

The incorporation of RNF138-dependent removal of Ku70/80 and the DNA-PK complex within Scenario D has subtle effects when compared to those of co-localisation (Fig. 3.4e). Dissociation of Ku70/80 and the DNA-PK complex which shields the DNA-end from CtIP-based resection increases the likelihood of HR repair taking place. Furthermore, as RNF138 is recruited in an MRN-dependent manner, the effects are increased with the MRN co-localisation mechanism applied. In Scenario D, the RNF138 protein is recruited at a time constant approximately 80 times larger than that of Ku and DNA-PKcs recruitment and 3 times larger than MRN recruitment. This larger time constant results in RNF138 increasing at later time points of the simulation. This translates to an increased bias towards HR at later time points in the simulation. There is a complex behaviour between the mechanisms and CtIP recruitment (Fig. 3.4d): without co-localisation, there is not the required increasing relationship with time required to match the experimental data, whereas if RNF138 is removed the increasing relationship with time is present, but the initial recruitment is too fast. The best fit to the CtIP recruitment data is achieved through the inclusion of both co-localisation and RNF138 (Fig. 3.5).

3.5 Discussion

The analysis of each scenario in turn, when compared to experimental literature results, has highlighted the poor fit of competitive only pathways (Scenarios B and C) to model the re-

pair choice process (Table 3.1). However, we show that with the inclusion of co-localisation of the MRN complex, a complex typically thought of as being involved in HR, with initial NHEJ proteins (Scenario D), it is possible to better mimic the repair kinetics of the Beucher *et al.,* experimental data. Scenario D is not exclusively competitive and utilises co-localisation of proteins to allow for explicit modelling of MRN recruitment. Although Scenario D is the most favoured repair choice, it should be stated that Scenario A, the 'NHEJ first' approach, also fits the experimental results well. However, Scenario A is heavily directed as we force the model to use NHEJ and then utilise HR only if the repair fails. Furthermore, while Scenario A can mimic repair rates similar to Scenario D, without co-localisation the explicit modelling of MRN recruitment is not possible and CtIP protein recruitment is delayed in the simulation which deviates from the experimental recruitment data (see Supplementary Data - Figure S1). The inability to fit the MRN and CtIP recruitment kinetics could be easily overlooked in less mechanistic approaches where protein level progression is being omitted. In contrast, Scenario D retains the known faster recruitment kinetics of Ku whilst not having to apply system constraints that cannot be substantiated by literature-reported mechanisms.

The introduction of additional mechanisms to Scenario D required careful consideration of all the implications on the model. RNF138-dependent removal of Ku and DNA-PKcs at DNA ends with MRN attached was fitted to the immunofluorescence recruitment of RNF138 (Fig. 3.4 f). Whilst the removal of Ku70/80 from the DNA end was more clearly set-out within the literature [40], the mechanism of removal of DNA-PKcs along with Ku70/80 back to a naked DNA end was established from an amalgamation of literature results. *Firstly,* it is widely understood that DNA-PKcs is recruited in a Ku-dependent manner [35, 57]. *Secondly,* it is believed that Ku70/80, under regular cellular conditions[58], helps retain DNA-PKcs at the end of the DSB *[59]. Finally,* Ismail *et al.,* 2015 [40] demonstrated that the recruitment of RNF138 is not dependent on Ku70/80 foci formation, resulting in our interpretation that if Ku70/80 was somehow protected by DNA-PKcs from RNF138-dependent removal, then it would be adversely affected by the DNA-PKcs recruitment that arises from the Ku70/80 foci formation. The establishment of MRN complex co-localisation was also an amalgamation of literature findings. *Firstly* Britton *et al.,* 2013 [39] showed co-localisation through simultaneous visualisation of Ku and NBS1 (part of the MRN complex) foci, and *secondly* Zhou *et al.,* 2017 [36] found the presence of Ku/DNA-PKcs does not impact the recruitment of MRN/ATM, a premise further supported by earlier work from Hartlerode *et al.,* 2015 [34] who showed MRN and Ku recruitment to be independent of one another. The observed agreement in Mre11 recruitment kinetics post-adjustment of the model for co-localisation gives further confidence in the added mechanism (Fig. 3.4). The addition of these mechanisms has only been applied for Scenario C's 'continuous competition' approach to form Scenario D. However, these additions were evaluated for both Scenario A and B to evaluate if further simulations were required. In Scenario A's 'NHEJ first' approach, the co-localisation would be non-consequential due to the directed nature of enforcing NHEJ before HR. Also, the additional dissociation points would lead to a minimal change in the WT and Lig4-deficient systems with slightly quicker repair rates. Therefore,

it is likely to result in noticeable differences in XLF-deficient cells which would quicken the repair rate further beyond the already to fast kinetics observed (Fig. 3.3c), negatively impacting the fit. In Scenario B's 'no way back' approach, both the co-localisation and additional dissociation points would be non-consequential as there is no cross-talk between post initial protein loading. For these reasons, the additional mechanisms incorporated were limited to Scenario C's 'continuous competition' approach to form Scenario D with the observed effects shown in Fig. 3.3.

This work has demonstrated how subtle changes in the repair choice can lead to substantial changes in the repair kinetics and hence its importance in repair modelling. Furthermore, the study has demonstrated that new mechanisms proposed in the literature can be applied and evaluated in a timely manner using the DaMaRiS framework. The difficulty in modelling using a mechanistic approach is that as each step is modelled independently, missing or incorrectly modelled steps can lead to the system deviating from experimental results. As no single group have published work for every required mechanism included in this *in silico* model, the experimental data used to model each step comes from various sources and thus inter-experimental uncertainty will be present. Furthermore, the final data set used to examine our *in silico* model is again from a different source to those used to build the model, which whilst being beneficial in avoiding systematic error is hindered by further inter-experimental uncertainties. The advantage of a step-by-step mechanistic model is that each step can be independently validated against experimental data.

Though some causes of the differences between the simulated and experimental data are explained above, others can be explained with the examination of the model response to the different scenarios and cell systems tested. Within the XLF-deficient system, the progression to synapsis stabilisation is removed [46, 54]; this means that DNA-PKcs synapsis dissociation will occur given enough simulation time. Upon dissociation of DNA-PKcs synapsis, HR is enforced (within Scenario A) or is allowed to re-compete (within Scenario C & D), and repair, whilst being longer, will be achieved given enough time. This resultant repair produces the observed linearity of the simulation data and the overall deviation from the experimental data beyond 2 hours. To be able to emulate the results seen in the XLF-deficient system either further mechanisms for un-repaired DNA ends to form should be present within HR, or XLF-deficiency is not an absolute determinant of synapsis stabilisation, but does perturb the ability to ligate the break. Whilst XLF is known to be important within NHEJ repair, its purpose is not as clearly defined [60] as that of Lig4; this may be why the simulation reproduces better the Lig4-deficient system than the XLF-deficient system. Therefore, the choice to attempt modelling XLF-deficiency in this study required extra consideration. *Firstly*, it has been shown that XLF availability impacts NHEJ and not HR [61], so the alterations should only be made to the NHEJ pathway. *Secondly*, XLF is not merely a tool to quicken NHEJ, as when depleted it can significantly impact the repair capabilities of the system [31, 60]. *Thirdly*, XLF is thought to provide additional stability to Lig4 and XRCC4 at the final steps of NHEJ repair [61], is likely to be present prior to Lig4 [60] and may even be present early enough to help achieve synapsis formation [54,

62]. Given the ambiguity regarding the explicit use of XLF in NHEJ repair, additional simulations of XLF-deficient systems are presented in the Supplementary Data (Figure S7). The study included the addition of experimental literature data from two WT and one Lig4 deficient cell lines for analysis of the variation one may expect from different laboratory set-ups (Fig. 3.3). The WT cell system data set from Kuhne *et al.* and Wu *et al.* are examples of both faster and slower repair kinetics respectively. Interestingly the Lig4 syndrome cell system from the Kuhne *et al.* data set shows substantially faster kinetics than those observed through simulation. This is most likely due to the cell line not being entirely devoid of the Lig4 protein which would allow for a small amount of NHEJ repair [50]. In order to evaluate this data set accurately, partial progression inhibition would be needed within the model**.**

The analysis of repair choice between each scenario is predominately unaffected by the modelling post resection within the HR repair pathway. However, the simplification of modelling the progression of DNA ends that have been resected through to being repaired using a single time constant, τ_{RR}, may be responsible for some of the discontinuities when compared with experimental literature data. It is expected that a more detailed model of HR repair post-resection, which includes the repair rate implications of Rad51-filament seeking out the sister chromatid, may help to better mimic the shape of the repair kinetics seen in experimental results whilst being able to maintain the mechanistic approach. However, modelling HR repair in such fine detail to uphold the current step-by-step mechanistic modelling approach is a large project and to our knowledge, no such model currently exists. Furthermore, several of the steps in HR are lacking detailed descriptions, such as filament seeking [63], misrepair [64] and sub-pathway choice [44], increasing the challenge of modelling and further encouraging studies such as this one to improve our mechanistic understanding of DNA repair and therefore how best to model it.

The requirement for all DSBs to be repaired via either NHEJ or HR is a simplification. In practice there are several other, less commonly utilised, repair pathways able to address DSBs. Alternatively, DSB repair could be classified as either being resection-dependent or resection-independent: within the scope of the presented work this would be analogous to HR and NHEJ respectively. As the current model does not include Alternative NHEJ (Alt-NHEJ) or Single Strand Annealing (SSA), both of which are resection-dependent DSB repair pathways, any presence within the experimental data would be classified by the model as HR. Whilst Alt-NHEJ and SSA are believed to be less utilised, there is an increase of use for deficient cell lines, specifically NHEJ deficient cell lines [65, 66], which are present in this study. Furthermore, before any analysis of the fidelity of repair from the model can be made, the biological impact of these additional resection-dependent DSB repair mechanisms should be considered. Whilst HR, Alt-NHEJ, and SSA are commonly categorised as resection-dependent repair, their processes are very different. It is thought that HR is a relatively error-free process, while both Alt-NHEJ and SSA are considered to be extremely error-prone, often resulting in large deletions [67].

Though the mechanisms investigated in this study encompass how the repair choice should

be interfaced within an *in silico* model, there are additional factors which are thought to influence repair choice which have not been included. The additional factors are either tied to the DNA geometry or the detail of the damage that is trying to be repaired. The currently simulated model takes place on an entirely homogeneous spread, whereas it is known that the cell nucleus has a heterogeneous spread of DNA geometry formed by euchromatin and heterochromatin regions. It is believed that these regions have biases towards particular repair pathways [42], an aspect omitted from the current study. Furthermore, the DNA density will also have an impact on the proximity between DSBs, which has been previously scored through the 'Cluster Density' [7], a known aspect of repair fidelity, but one which will also impact the quantities of repair as further proximity-based mechanisms are incorporated within the model (e.g. sister chromatid seeking). Additional to the DNA geometry, a known factor in the production of DNA damage density is the incident radiation source, dose and the linear energy transfer (LET). This study utilises a range of experimental setups using various radiation sources, but the simulations are all based on 1.77 keV/μm LET mono-energetic proton irradiation of 2 Gy. There is an established belief within the literature that repair choice is influenced by the damage complexity [43], although the mechanisms which produce this effect are still to be established. At present, there is nothing within our model that would lead us to observe this sensitivity, though the literature is being carefully observed for new mechanisms to be proposed which can be incorporated into our *in silico* models and further exploration of repair choice can be carried out.

3.6 Conclusion

This study has demonstrated the complexities of modelling repair choice within a mechanistic *in silico* model. The ability to match the recruitment kinetics of key proteins whilst piecing together the possible progression and dissociation of said proteins, based on what has been seen experimentally, is challenging. Overall, this study has demonstrated that the 'entwined pathway' approach (Fig. 3.5), for modelling the competition between the two main pathways of DSB repair, results in the most effective fit to the evaluated published data. Furthermore, it has been shown in this study that co-localisation of MRN with initial NHEJ proteins and RNF138-dependent removal of Ku are required in order to achieve a model which can emulate both the protein recruitment and repair kinetics from experiments. Through the explicit modelling of protein recruitment the DaMaRiS model is able to meet more constraints whilst remaining mechanistic. This approach allows the combination of many experimentally proposed mechanisms into a single system to explore and evaluate how they interact with one another.

3.7 Acknowledgements

This research was co-funded by the NIHR Manchester Biomedical Research Centre (grant number: BRC-1215-20007), STFC Global Challenge Network+ in Advanced Radiother-

apy (grant number: ST/N002423/1) and The Christie Charitable funds - The Christie NHS Foundation Trust, UK. K.J.K., N.G.B., M.J.M. and A.L.C. were supported by NIHR Manchester Biomedical Research Centre. The computational element of this research was achieved using The Computational Shared Facility (CSF) at the University of Manchester.

3.8 Author Contributions

S.P.I. developed the DNA repair choice models, gathered the data, and analysed the data. J.W.W. developed the DNA repair model of NHEJ which was modified by S.P.I. for this work. J.W.W. developed the framework DaMaRiS and provided technical support in its use. N.T.H. provided the simulated DNA damage data. M.J.M., K.J.K., N.F.K. and R.I.M. provided day to day supervision. A.L.C. provided information on the biological relevance. N.G.B. and E.A.K.S gave input on the clinical relevance. S.P.I. drafted the manuscript based on discussions with the authors. All authors reviewed and approved the manuscript.

3.9 Competing Interests

The authors declare no competing interests.

References

1. Frankenberg, D., Frankenberg-Schwager, M., Blöcher, D. & Harbich, R. Evidence for DNA double-strand breaks as the critical lesions in yeast cells irradiated with sparsely or densely ionizing radiation under oxic or anoxic conditions. *Radiation research* **88,** 524–32. ISSN: 0033-7587 (1981).

2. Burnet, N. G. *et al.* Applying physical science techniques and CERN technology to an unsolved problem in radiation treatment for cancer: the multidisciplinary 'Vox-Tox' research programme. *CERN Ideasq J Exp Innov* **1,** 3–12. ISSN: 2413-9505 (2017).

3. Jena, R., Mee, T., Kirkby, N. & Williams, M. Quantifying Uncertainty in Radiotherapy Demand at the Local and National Level using the Malthus Model. *Clinical Oncology* **27,** 92–98. ISSN: 0936-6555 (2015).

4. Paganetti, H. Relative biological effectiveness (RBE) values for proton beam therapy. Variations as a function of biological endpoint, dose, and linear energy transfer. *Phys Med Biol* **59,** R419–R472. ISSN: 0031-9155 (2014).

5. Underwood, T. & Paganetti, H. Variable Proton Relative Biological Effectiveness: How Do We Move Forward? *Int J Radiat Oncol Biology Phys* **95,** 56–58. ISSN: 0360-3016 (2016).

6. Warmenhoven, J. *et al.* An In Silico Model of DNA Repair for Investigation of Mechanisms in Non-Homologous End Joining. *Biorxiv,* 318139 (2018).

7. Henthorn, N. *et al.* In Silico Non-Homologous End Joining Following Ion Induced DNA Double Strand Breaks Predicts That Repair Fidelity Depends on Break Density. *Sci Reports* **8,** 2654 (2018).

8. J, M. S., Schuemann, J., Paganetti, H. & Prise, K. M. Mechanistic Modelling of DNA Repair and Cellular Survival Following Radiation-Induced DNA Damage. *Sci Reports* **6,** 33290. ISSN: 2045-2322 (2016).

9. Friedland, W. *et al.* Comprehensive track-structure based evaluation of DNA damage by light ions from radiotherapy-relevant energies down to stopping. *Sci Reports* **7,** 45161. ISSN: 2045-2322 (2017).

10. Stewart, R. D. *et al.* Rapid MCNP simulation of DNA double strand break (DSB) relative biological effectiveness (RBE) for photons, neutrons, and light ions. *Phys Med Biol* **60,** 8249–74. ISSN: 0031-9155 (2015).

11. Belov, O. V., Krasavin, E. A., Lyashko, M. S., Batmunkh, M. & Sweilam, N. H. A quantitative model of the major pathways for radiation-induced DNA double-strand break repair. *J Theor Biol* **366,** 115–130. ISSN: 0022-5193 (2015).

12. Ballarini, F. *et al.* The BIANCA model/code of radiation-induced cell death: application to human cells exposed to different radiation types. *Radiat Environ Bioph* **53,** 525–533. ISSN: 0301-634X (2014).

13. Friedland, W., Jacob, P., Bernhardt, P., Paretzke, H. G. & Dingfelder, M. Simulation of DNA damage after proton irradiation. *Radiat. Res.* **159,** 401–10. ISSN: 0033-7587 (2003).

14. Barnett, G. C. *et al.* Normal tissue reactions to radiotherapy: towards tailoring treatment dose by genotype. *Nature Reviews Cancer* **9,** nrc2587. ISSN: 1474-1768 (2009).

15. Khanna, K. & Jackson, S. DNA double-strand breaks: signaling, repair and the cancer connection. *Nat Genet* **27,** ng0301_247. ISSN: 1546-1718 (2001).

16. Mohrenweiser, H. & Jones, I. Variation in DNA repair is a factor in cancer susceptibility: a paradigm for the promises and perils of individual and population risk estimation? *Mutat Res Fundam Mol Mech Mutagen* **400,** 15–24. ISSN: 0027-5107 (1998).

17. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle Georget Tex* **7,** 2902–6. ISSN: 1538-4101 (2008).

18. Rothkamm, K., Krüger, I., Thompson, L. H. & Löbrich, M. Pathways of DNA Double-Strand Break Repair during the Mammalian Cell Cycle. *Mol Cell Biol* **23,** 5706–5715. ISSN: 0270-7306 (2003).

19. Friedland, W., Jacob, P. & Kundrát, P. Mechanistic simulation of radiation damage to DNA and its repair: on the track towards systems radiation biology modelling. *Radiat Prot Dosim* **143,** 542–8. ISSN: 0144-8420 (2010).

20. Liu, Q. *et al.* Disruption of SLX4-MUS81 Function Increases the Relative Biological Effectiveness of Proton Radiation. *Int J Radiat Oncol Biology Phys* **95,** 78–85. ISSN: 0360-3016 (2016).

21. Liu, Q. *et al.* Lung Cancer Cell Line Screen Links Fanconi Anemia/BRCA Pathway Defects to Increased Relative Biological Effectiveness of Proton Radiation. *International Journal of Radiation Oncology*Biology*Physics* **91,** 1081–1089. ISSN: 0360-3016 (2015).

22. Mao, Z., Jiang, Y., Liu, X., Seluanov, A. & Gorbunova, V. DNA Repair by Homologous Recombination, But Not by Nonhomologous End Joining, Is Elevated in Breast Cancer Cells. *Neoplasia* **11,** 683–IN3. ISSN: 1476-5586 (2009).

23. Willers, H. *et al.* Toward A variable RBE for proton beam therapy. *Radiotherapy Oncol J European Soc Ther Radiology Oncol.* ISSN: 0167-8140 (2018).

24. McCabe, N. *et al.* Deficiency in the Repair of DNA Damage by Homologous Recombination and Sensitivity to Poly(ADP-Ribose) Polymerase Inhibition. *Cancer Research* **66,** 8109–8115. ISSN: 0008-5472 (2006).

25. Patel, A. G., Sarkaria, J. N. & Kaufmann, S. H. Nonhomologous end joining drives poly(ADP-ribose) polymerase (PARP) inhibitor lethality in homologous recombination-deficient cells. *Proceedings of the National Academy of Sciences* **108,** 3406–3411. ISSN: 0027-8424 (2011).

26. Johnson, R. & Jasin, M. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *EMBO J.* **19,** 3398–407. ISSN: 0261-4189 (2000).

27. Shrivastav, M., Haro, L. P. & Nickoloff, J. A. Regulation of DNA double-strand break repair pathway choice. *Cell Res* **18,** cr2007111. ISSN: 1748-7838 (2008).

28. Kakarougkas, A. & Jeggo, P. DNA DSB repair pathway choice: an orchestrated handover mechanism. *Br J Radiology* **87,** 20130685. ISSN: 0007-1285 (2014).

29. Chiruvella, K. K., Liang, Z. & Wilson, T. E. Repair of double-strand breaks by end joining. *Cold Spring Harbor perspectives in biology* **5,** a012757 (2013).

30. Symington, L. S. & Gautier, J. Double-Strand Break End Resection and Repair Pathway Choice. *Annu Rev Genet* **45,** 247–271. ISSN: 0066-4197 (2011).

31. Beucher, A. *et al.* ATM and Artemis promote homologous recombination of radiation induced DNA double strand breaks in G2. *Embo J* **28,** 3413–3427. ISSN: 1460-2075 (2009).

32. Incerti, S. *et al.* The Geant4-DNA project. *International Journal of Modeling, Simulation, and Scientific Computing* **1,** 1–27 (Jan. 2010).

33. Schuemann, J. *et al.* A New Standard DNA Damage (SDD) Data Format. *Radiation Research.* ISSN: 0033-7587 (2018).

34. Hartlerode, A. J., Morgan, M. J., Wu, Y., Buis, J. & Ferguson, D. O. Recruitment and activation of the ATM kinase in the absence of DNA-damage sensors. *Nat. Struct. Mol. Biol.* **22,** 736–43. ISSN: 1545-9985 (2015).

35. Uematsu, N. *et al.* Autophosphorylation of DNA-PKCS regulates its dynamics at DNA double-strand breaks. *J Cell Biology* **177,** 219–229. ISSN: 0021-9525 (2007).

36. Zhou, Y. *et al.* Regulation of the DNA Damage Response by DNA-PKcs Inhibitory Phosphorylation of ATM. *Mol Cell* **65,** 91–104. ISSN: 1097-2765 (2017).

37. Chapman, R. J., Taylor, M. & Boulton, S. J. Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Mol Cell* **47,** 497–510. ISSN: 1097-2765 (2012).

38. Joiner, M. C. & van der Kogel, A. *Basic Clinical Radiobiology Fourth Edition* eng. Mar. 2009. `https://www.123library.org/ebook/isbn/9781444109610/`.

39. Britton, S., Coates, J. & Jackson, S. P. A new method for high-resolution imaging of Ku foci to decipher mechanisms of DNA double-strand break repair. *J Cell Biology* **202,** 579–595. ISSN: 0021-9525 (2013).

40. Ismail, I. *et al.* The RNF138 E3 ligase displaces Ku to promote DNA end resection and regulate DNA repair pathway choice. *Nat Cell Biol* **17,** 1446–1457. ISSN: 1465-7392 (2015).

41. Nimonkar, A. *et al.* BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. English (US). *Genes and Development* **25,** 350–362. ISSN: 0890-9369 (Feb. 2011).

42. Jeggo, P. A., Geuting, V. & Löbrich, M. The role of homologous recombination in radiation-induced double-strand break repair. *Radiother Oncol* **101,** 7–12. ISSN: 0167-8140 (2011).

43. Yajima, H. *et al.* The complexity of DNA double strand breaks is a critical factor enhancing end-resection. *Dna Repair* **12,** 936–946. ISSN: 1568-7864 (2013).

44. Heyer, W., Ehmsen, K. & review of genetics, L. J. Regulation of homologous recombination in eukaryotes. *Annual review of genetics* (2010).

45. Biehs, R. *et al.* DNA Double-Strand Break Resection Occurs during Non-homologous End Joining in G1 but Is Distinct from Resection during Homologous Recombination. *Mol Cell* **65,** 671–684.e5. ISSN: 1097-2765 (2017).

46. Ahnesorg, P., Smith, P. & Jackson, S. P. XLF Interacts with the XRCC4-DNA Ligase IV Complex to Promote DNA Nonhomologous End-Joining. *Cell* **124,** 301–313. ISSN: 0092-8674 (2006).

47. Reid, D. A. *et al.* Organization and dynamics of the nonhomologous end-joining machinery during DNA double-strand break repair. *Proceedings of the National Academy of Sciences* **112,** E2575–E2584 (2015).

48. Rohatgi, A. & ZlatanStanojevic. *WebPlotDigitizer: Version 3.9 of WebPlotDigitizer* Oct. 2015. `https://doi.org/10.5281/zenodo.32375`.

49. Stan Salvador, P. C. FastDTW: Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis 11.5,* 561–580 (2007).

50. Kühne, M. *et al.* A double-strand break repair defect in ATM-deficient cells contributes to radiosensitivity. *Cancer Res.* **64,** 500–8. ISSN: 0008-5472 (2004).

51. Wu, J. *et al.* Repair rates of DNA double-strand breaks under different doses of proton and gamma-ray irradiation. *Nucl Instruments Methods Phys Res Sect B Beam Interactions Mater Atoms* **276,** 1–6. ISSN: 0168-583X (2012).

52. Fontana, A. O. *et al.* Differential DNA repair pathway choice in cancer cells after proton- and photon-irradiation. *Radiother Oncol* **116,** 374–380. ISSN: 0167-8140 (2015).

53. Grosse, N. *et al.* Deficiency in Homologous Recombination Renders Mammalian Cells More Sensitive to Proton Versus Photon Irradiation. *Int J Radiat Oncol Biology Phys* **88,** 175–181. ISSN: 0360-3016 (2014).

54. Graham, T., Walter, J. C. & Loparo, J. J. Two-Stage Synapsis of DNA Ends during Non-homologous End Joining. *Mol Cell* **61,** 850–858. ISSN: 1097-2765 (2016).

55. Blier, P., Griffith, A., Craft, J. & Hardin, J. Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. *J. Biol. Chem.* **268,** 7594–601. ISSN: 0021-9258 (1993).

56. You, Z. *et al.* CtIP Links DNA Double-Strand Break Sensing to Resection. *Mol Cell* **36,** 954–969. ISSN: 1097-2765 (2009).

57. Jackson, S. P. DNA-dependent protein kinase. *The International Journal of Biochemistry and Cell Biology* **29,** 935–938 (1997).

58. Drouet, J. *et al.* Interplay between Ku, Artemis, and the DNA-dependent Protein Kinase Catalytic Subunit at DNA Ends. *Journal of Biological Chemistry* **281,** 27784–27793 (2006).

59. Pawelczak, K. S., Bennett, S. M. & Turchi, J. J. Coordination of DNA-PK activation and nuclease processing of DNA termini in NHEJ. *Antioxidants and redox signaling* **14,** 2531–43 (2010).

60. Menon, V. & Povirk, L. F. XLF/Cernunnos: An important but puzzling participant in the nonhomologous end joining DNA repair pathway. *DNA Repair* **58,** 29–37 (2017).

61. Liu, P. *et al.* Akt-Mediated Phosphorylation of XLF Impairs Non-Homologous End-Joining DNA Repair. *Molecular Cell* **57,** 648–661 (2015).

62. Brouwer, I. *et al.* Sliding sleeves of XRCC4–XLF bridge DNA and connect fragments of broken DNA. *Nature* **535,** 566–569 (2016).

63. Van der Heijden, T. *et al.* Real-time assembly and disassembly of human RAD51 filaments on individual DNA molecules. *Nucleic Acids Res* **35,** 5646–5657. ISSN: 0305-1048 (2007).

64. JosÃe, G., Lambert, S., Bertrand, P. & Lopez, B. S. Is homologous recombination really an error-free process? *Frontiers Genetics* **5,** 175. ISSN: 1664-8021 (2014).

65. Rodgers, K. E. & McVey, M. Error-Prone Repair of DNA Double-Strand Breaks. *Journal of cellular physiology* **231 1,** 15–24 (2016).

66. Sfeir, A. & Symington, L. Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends Biochem Sci* **40,** 701–714. ISSN: 0968-0004 (2015).

67. Bhargava, R., Onyango, D. & Stark, J. Regulation of Single-Strand Annealing and its Role in Genome Maintenance. *Trends Genet* **32,** 566–575. ISSN: 0168-9525 (2016).

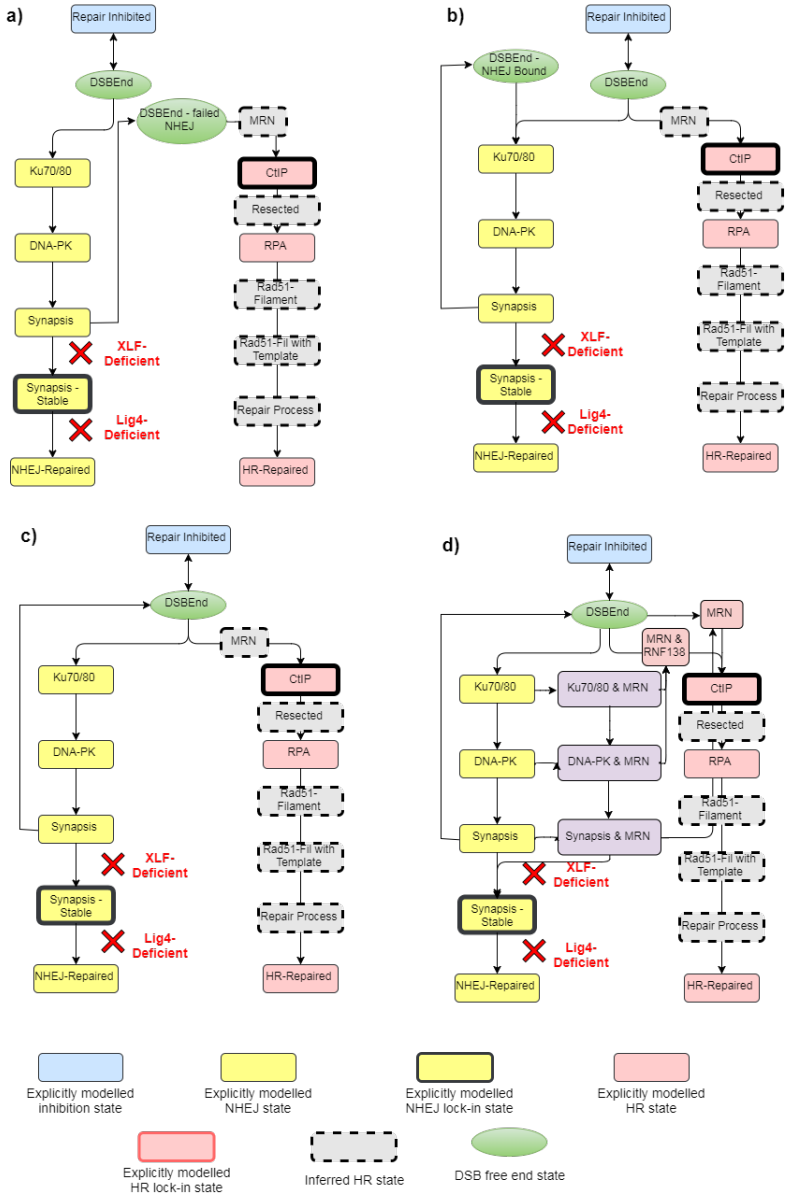

Figure 3.2: Scenario schematics - a) Scenario A - NHEJ is attempted first and only upon failure can HR be attempted. b) Scenario B - initial competition dictates the repair progression as there is no cross-over allowed between pathways; the pathway which has the end will be attempted until the end of simulation time. c) Scenario C - initial competition dictates the attempt of repair, and upon failure, re-competition takes place for the next attempt of repair, this is repeated until the end of simulation time. d) Scenario D 'Entwined pathway approach' - a similar competition to Scenario C with the addition of MRN co-localisation with initial NHEJ proteins, Ku dissociation and PKcs dissociation; it should be noted that dissociation of NHEJ proteins where MRN is co-localised results in the fall back to a DNA end with MRN attached. The steps in the model are either explicitly modelled (yellow NHEJ and red HR) with progression and reaction rates deduced from experimental data (Supplementary Data - Figures S1 and S2), or there are several inferred HR steps (grey with dashed external lines) which are assumed to have taken place by the time the next explicitly modelled step is reached. The emboldened steps within the schematic show points at which the DNA-end gets modified such that repair can only progress in the corresponding pathway (the point at which a DSB end is locked into the pathway). The green circular steps represent DSB free ends with none of the modelled repair proteins loaded, which occur either at the start of the simulation (i.e. as repair is about to start) or through DNA-PKcs synapsis dissociation. The blue step represents the DNA-end being inhibited from protein loading and represents aspects of end cleaning before proteins can successfully load on. Finally, the red crosses represent the progression points that are deactivated for working in a protein deficient cell system; the text of XLF- or Lig4-deficient indicates at what point the DNA end progression is removed for the corresponding system.

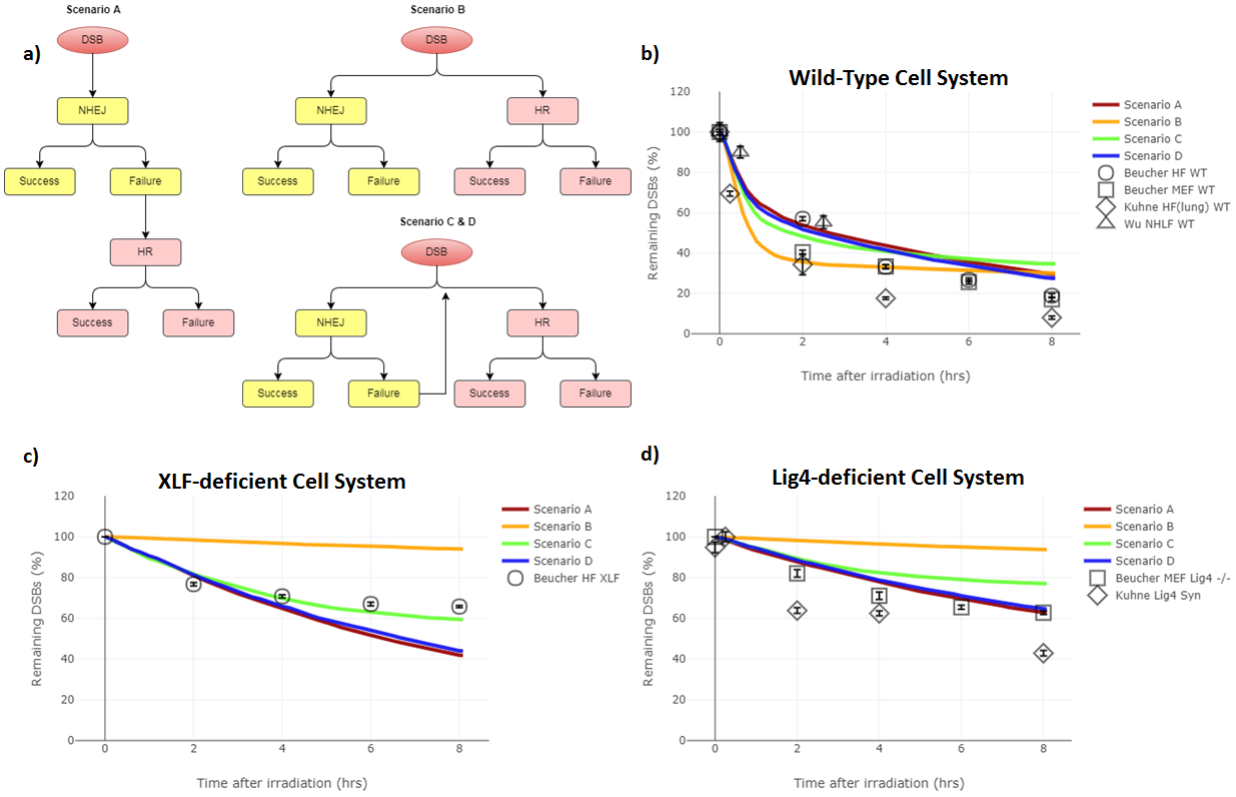

Figure 3.3: Scenario Comparison - **a)** simplistic flow diagrams of the repair choice scenarios tested. Scenario A represents the 'NHEJ first' approach. Scenario B represents the 'no way back' approach between pathways. Scenario C represents the 'continuous competition' approach between pathways. Scenario D represents the 'entwined pathway' approach. The results for the wild-type, XLF-deficient and Lig4-deficient systems are shown in b), c) and d) respectively. The diamond and the triangle symbols are for comparison between experimental data and are not included in the quantitative analysis of fit between simulated and experimental data. The error bars are the reported $\pm$SEM from the experimental data set (black data points). The error in the simulated data is the $\pm$SEM is displayed as the line width for 50 repeated simulations each with their own independent exposures on different cells.

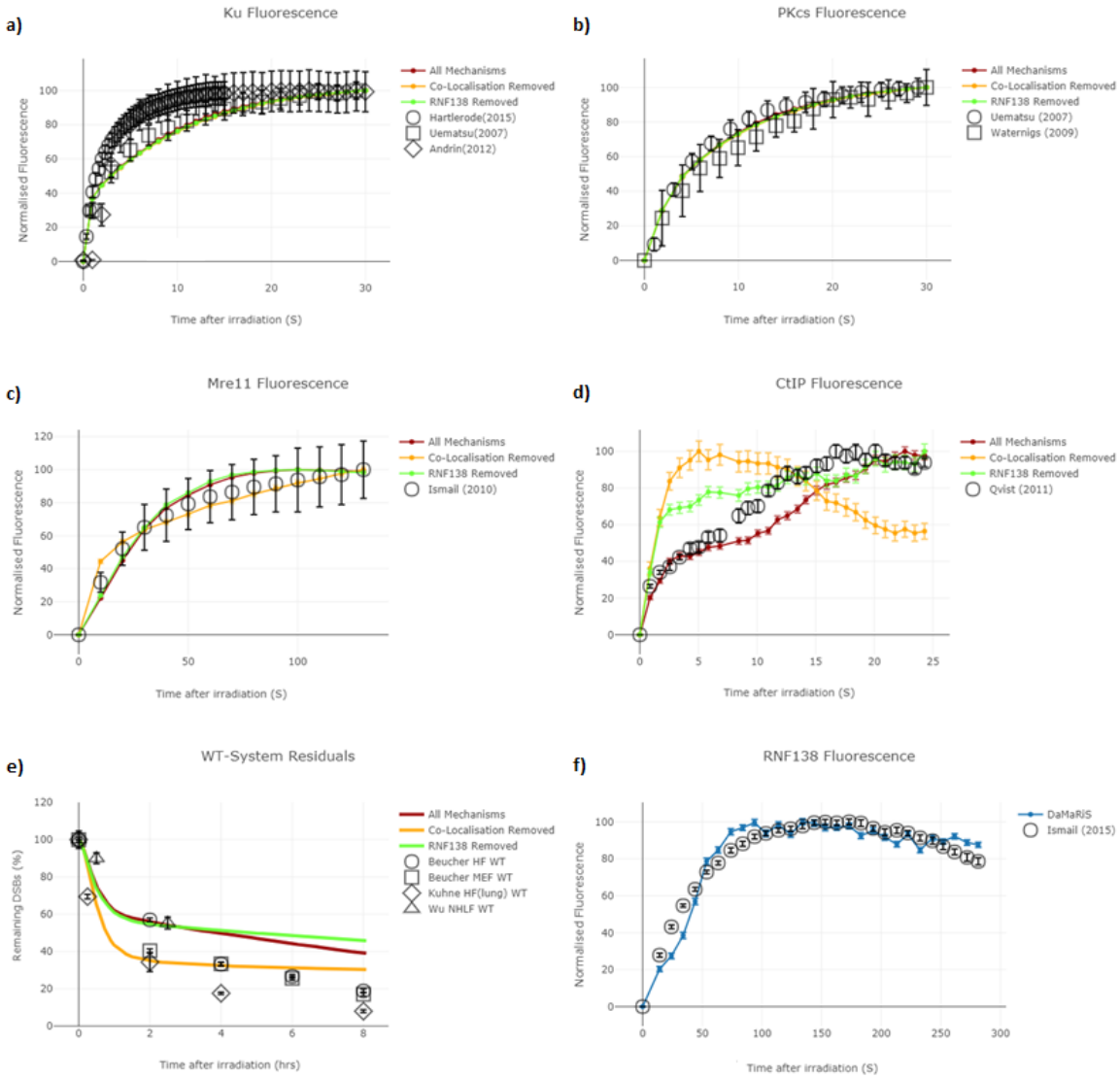

Figure 3.4: Scenario D mechanism evaluation - the differences in model behaviour are shown for "All Mechanisms" (red line) which includes both MRN co-localisation and RNF138 recruitment, "Co-Localisation Removed" (orange line) which has the role of MRN co-localisation removed and "RNF138 Removed" (green line) which has the role of RNF138 removed. a) Ku immunofluorescence recruitment graph, within the simulation Ku is assumed to persist until synapsis stabilisation or is explicitly removed by RNF138. b) DNA-PKcs immunofluorescence recruitment graph, within the simulation DNA-PKcs is assumed to persist until synapsis stabilisation or is removed along with Ku by RNF138. c) Mre11 (which is representative of the MRN complex) immunofluorescence recruitment graph. d) CtIP immunofluorescence recruitment graph. e) Residual DNA damage graph, experimental data is from γ-H2AX foci data and simulation data represents number of breaks which has not yet been repaired. f) RNF138 immunofluorescence recruitment graph. All error bars are the $\pm$SEM for both simulated and experimental data. In Sub-Figure f) the $\pm$SEM is represented as the width of the line. The error displayed is representative of 250 repeated simulations for protein recruitment data and 50 repeated simulations for un-repaired data, each with their own independent exposures on different cells.

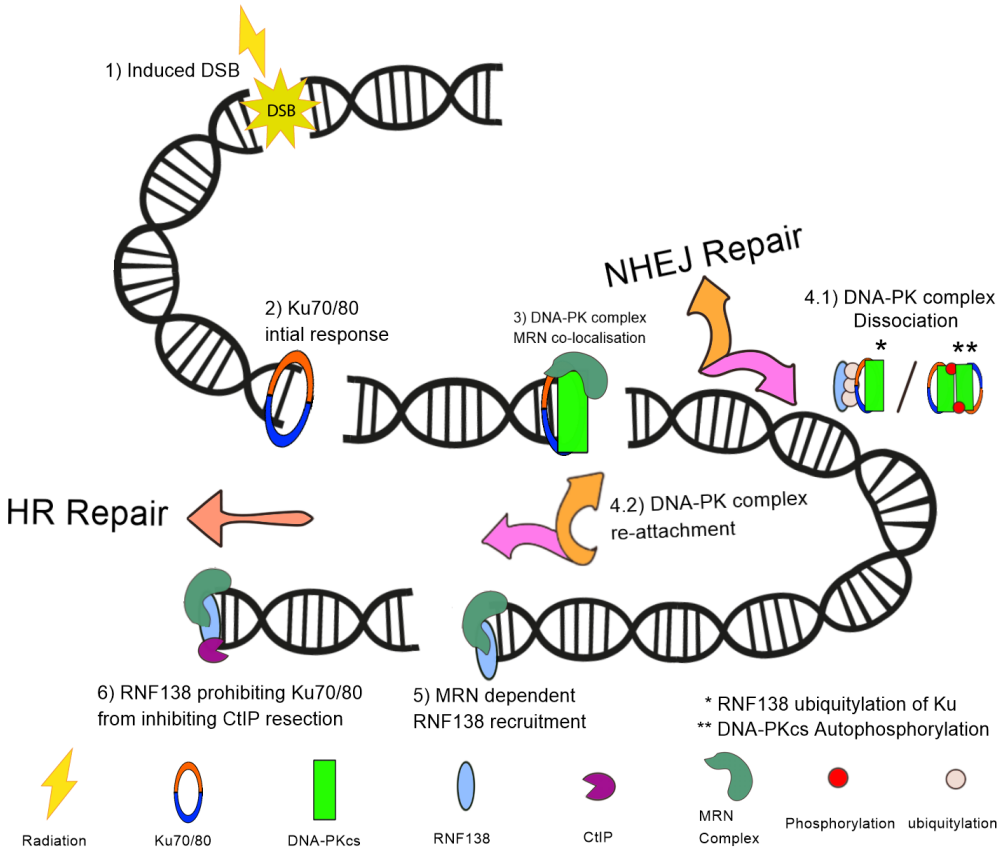

Figure 3.5: Schematic of Scenario D – The 'entwined pathway' approach of repair choice. Positions 1-6 represent steps along the DNA DSB repair pathway. 1) Radiation induces DSB damage. 2) Due to its high abundance and affinity for DNA ends, on average Ku provides the first response to the DSB. 3) DNA-PKcs can attach in a Ku-dependent manner and/or MRN can co-localise with Ku/DNA-PKcs. 4.1) Either the DNA-PK synapsis can be stabilised allowing for NHEJ repair to follow or the DNA-PK complex can dissociate through either autophosphorylation or RNF138 ubiquitylation. 4.2) Either NHEJ can be attempted again by the recruitment of Ku and then DNA-PKcs, or steps can be taken towards resected-mediated repair. 5) RNF138 is recruited in an MRN dependent manner. 6) RNF138 prohibits Ku attachment providing enough time for CtIP resection to take place and HR repair to follow. As the repair model is Monte Carlo-based, with various progression points at each step, the above is one possible repair route available in the entwined pathway model. However, this route has been demonstrated to be important for matching protein recruitment kinetics (Fig. 4) and highlights an NHEJ and HR symbiotic relationship.

Chapter 4

Hi-C implementation of genome structure for *in silico* models of radiation-induced DNA damage

Samuel P Ingram[1,2,*,+], Nicholas T Henthorn[1,3,+], John W Warmenhoven[1,3], Norman F Kirkby[1,3], Ranald I Mackay[2,1], Karen J Kirkby[1,3], Michael J Merchant[1,3],

1 Division of Cancer Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, UK
2 Christie Medical Physics and Engineering, The Christie NHS Foundation Trust, Manchester, UK
3 The Christie NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

+ These authors contributed equally to this work.

* samuel.ingram@postgrad.manchester.ac.uk

Published as Ingram, S. P. et al. Hi-C implementation of genome structure for in silico models of radiation-induced DNA damage. Plos Comput Biol 16, e1008476 (2020). Slight modifications to formatting for consistency.

4.1 Abstract

Developments in the genome organisation field has resulted in the recent methodology to infer spatial conformations of the genome directly from experimentally measured genome contacts (Hi-C data). This provides a detailed description of both intra- and inter-chromosomal arrangements. Chromosomal intermingling is an important driver for radiation-induced DNA mis-repair. Which is a key biological endpoint of relevance to the fields of cancer therapy (radiotherapy), public health (biodosimetry) and space travel. For the first time, we leverage these methods of inferring genome organisation and couple them to nano-dosimetric radiation track structure modelling to predict quantities and distribution of DNA damage within cell-type specific geometries. These nano-dosimetric simulations are highly dependent on geometry and are benefited from the inclusion of experimentally driven chromosome conformations. We show how the changes in Hi-C contract maps impact the inferred geometries resulting in significant differences in chromosomal intermingling. We demonstrate how these differences propagate through to significant changes in the distribution of DNA damage throughout the cell nucleus, suggesting implications for DNA repair fidelity and subsequent cell fate. We suggest that differences in the geometric clustering for the chromosomes between the cell-types are a plausible factor leading to changes in cellular radiosensitivity. Furthermore, we investigate changes in cell shape, such as flattening, and show that this greatly impacts the distribution of DNA damage. This should be considered when comparing *in vitro* results to *in vivo* systems. The effect may be especially important when attempting to translate radiosensitivity measurements at the experimental *in vitro* level to the patient or human level.

4.2 Author summary

We have used a technique which allows us to understand how parts of our DNA are organised within a cell nucleus. This technique has previously shown differences in the organisation between different cell-types. In this study, we show that these differences produce significant change in the way our DNA is damaged when exposed to radiation. This is important to understand as one of the primary ways we treat cancer is using radiotherapy. However, whilst we attempt to target the cancer with radiation, some healthy tissue also receives radiation. It is the radiation delivered to the healthy tissue which limits how much radiation we can safely give to the cancer without causing significant side effects in patients. To know how much radiation we can give, over time, we have learnt generally safe amounts of radiation that can be given to healthy tissue. Even so, sometimes patients will still have worse side effects than what we would have predicted. If we want to further improve our treatments and patient safety, we need to better understand how this safe limit varies between each patient. The first step in to fully understanding this process comes from a better understanding of how different cell-types are affected by radiation, which is partly driven by DNA organisation, shown in this work.

4.3 Introduction

The research fields of radiobiology and DNA structure have shared a symbiotic past. Radiation has been used to infer the presence of chromosome territories [1], structural cytotoxic responses [2, 3] and examinations of the chromatin dynamics [4]. In turn, as we gain a better description of the DNA and chromatin structure we observe an intrinsic relationship with the radiobiological properties of a cell [5, 6]. This is due to the radiobiological response being majorly driven by damage to the DNA structure. The formation of double-strand breaks (DSBs), which is where the sugar-phosphate backbone of the DNA is broken on both sides in close proximity ($<$10 bp) has been shown to correlate strongly with cellular survival [7, 8]. This is thought to be partially caused by DSBs giving rise to the possibility of chromosomal interchanges, whereby chromosomes that misrepair can form a whole variety of chromosome aberrations [9]. It is believed that a major factor for misrepair events is the mobility and spatial distribution of DNA break ends [10–12], increasing the importance of chromosomal arrangement on cell fate. One of the more lethal chromosome aberration types are those that involve two different chromosomes being misrepaired and plays a major role in the radiation-induced cell death [13, 14]. The probability for these types of chromosomal translocation to occur has been shown to relate to the intermingling of the different chromosomes [15]. Therefore, there is an innate interest in being able to accurately predict the spatial distribution of interchromosomal DNA break ends as this will inform us to the probability of inducing interchromosomal aberrations.

The field of chromosome organisation has made major strides, many of which surround the microscopy technique fluorescence *in situ* hybridization (FISH). However, this cytogenetic

approach is ultimately limited by its sensitivity and resolution. The field has culminated in the emergence and development of inference based methods to examine genome organisation called chromosome conformation capture techniques [16]. Chromosome conformation capture uses DNA cross-linking to capture interactions between proximal regions of the genome, the frequencies at which sections of the genome are captured proximal to one another gives a mechanism by which to infer spatial proximity. One such method is Hi-C, which uses high-throughput sequencing to capture proximal regions for the entire genome. The Hi-C technique has been rapidly developing, from initial work of experimental design [17] through to the complex inferences, on both nuclear structure and function [18–22], which can be made from the gathered data [23, 24]. Whilst, the field is still developing it has become clear that an area of interest is the inference of the three-dimensional genome structure from the contact probabilities observed in Hi-C experiments. This has been attempted using multi-dimensional scaling [24], polymer [23, 25–27] and statistical [28, 29] type models with a varying structural focus.

On the other side, radiobiological models of DNA damage and repair are a concept that have been developing for several decades to better understand the effects of radiation in areas such as, healthcare, public health and space travel. Some of these models have proved themselves critical for informing the use of clinical radiotherapy [30, 31]. The clinical efficacy of radiotherapy treatment is limited by the compromise between cell kill in the tumour and in normal tissue, resulting in tumour control and normal tissue complication respectively [32]. Radiobiological models aim to provide insights here, with the intention of better informing the clinician when confronted with this compromise through exploitation of the 5 R's of radiobiology (repair, repopulation, reoxygenation, redistribution and radiosensitivity) [33]. In a subset of radiobiological models, radiation track structure is simulated with an interpretation of the genome structure [34–36], providing a detailed representation of the energy deposition at the DNA level. These structural interpretations have varying levels of complexity but are most commonly relatively uniform and mathematically driven. Furthermore, if cell-type specific geometric models are used, they will generally be the same for all simulations [37] absent of any cellular variation that is found in populations of cells. This variation is prevalent within the radiobiological experimental results and creates ambiguity when interpreting data for clinical decision making [38, 39]. The results of DNA damage models can be used in radio-response models [40] which try to identify effects at the cellular level [12, 41–43]. Ultimately, radiobiological models aim to describe cellular radio-response, with the hope of transference to patient-level response to better predict outcomes of radiation-based treatments.

In this study, we extend the overlap of these research areas by incorporating the chromosome structure, in the form of Hi-C data, in Monte Carlo radiation track structure simulations of DNA damage. This improved geometric representation is crucial to accurately model types of chromosome aberration and the subsequent effects of cell death, senescence or possible radiation-induced mutations. These biological endpoints are key drivers of clinical outcome following radiotherapy that will be benefited by an improved understanding

of their origin. To evaluate the effects of a Hi-C geometry we have developed G-NOME (G-NOME - Nuclear Organisation Modelling Environment) to infer the geometry from the Hi-C data. G-NOME is a highly extensible python library that allows for geometry inference using a Markov-chain Monte Carlo polymer model with the ability to be directly fed into tool-kits of radiation track-structure, such as Geant4-DNA [44] and TOPAS-nBio [45]. Through the combination of inferred genome structure from experimental Hi-C data and radiation track structure, it is possible to better encapsulate and therefore understand some of the biological variation seen within experimental radiobiology. Furthermore, this study has highlighted that Hi-C derived geometries for different cell lines have varying amounts of chromosomal intermingling, which are likely to be fundamental drivers for differences in observed cellular radiosensitivity.

4.4 Results

Definition of Terms

In the purpose of these definitions an object can refer to either polymer beads (when describing geometry distribution) or DSBs (when describing DNA damage distribution). Cluster radius: the spherical radius of inclusion space which counts the objects (e.g. beads or DSBs) within it. clustering: the number of objects that fall within a given radius averaged for all objects in the simulation. Interchromosomal clustering: the number of objects that fall within the cluster radius and *are not* on the same chromosome (including homologous chromosomes) averaged for all objects in the simulation. Intrachromosomal clustering: the number of objects that fall within the cluster radius and *are* on the same chromosome (not including homologous chromosomes) averaged for all objects in the simulation. Inter/Intrachromosomal clustering ratio: the ratio of the averaged interchromosomal clustering and the intrachromosomal clustering for the object being examined.

Effects of Using LADs and Ellipsoid Nuclei in Hi-C Geometries

The IMR90 (human fetal lung fibroblast) variant conformations, inclusion of Lamina-associated domains (LADs) and flattening of the nucleus to form an ellipsoid, were solved using the G-NOME software and viewed for gross abnormalities (Fig 4.1A). For each of the three tested cell types, 200 geometries were created. All ellipsoid and Lamina-associated domain (LAD) geometries were solved for 4 million iterations using the G-NOME software, this was chosen to achieve comparable nucleus-bead outliers compared to the spherical geometry, which were solved for 2 million iterations (Fig 4.1B).

To examine differences in interchromosomal proximity within the modelled cell nucleus the bead clustering of each cell variant is plotted (Fig 4.1C) for varying cluster inclusion radii. The single value of interchromosomal bead clustering is the result of analysing every bead in the model for the number of interchromosomal beads that are within the tested cluster

radius of the bead being analysed, the per bead clustering value is then averaged to obtain a per cell geometry bead clustering. This allows for a coarse examination of differences in the "intermingling" between chromosomes for each of the cells. When incorporating LAD objectives in the geometries there is a slight increase in the amount of intermingling of the chromosomes. Whereas, when the geometry is solved as an ellipsoid the intermingling decreases.

The optimised distribution of the chromosomes are reviewed by according to the positioning of their constituent beads, scored between central or peripheral of the modelled nucleus (Fig 4.1D). Larger chromosomes (chr1-chr9) are predominately situated at the periphery and smaller chromosomes (chr13-chrX) are situated centrally. This is similar to the chromosomal ordering seen in other models. [23] The analysis shows an increased peripheral and decreased peripheral bead positioning for IMR90 LADs and ellipsoid respectively when compared to the standard IMR90 variant (no LADs and spherical). Although all analysis methods halved the volumes to define the central and peripheral regions, in the ellipsoid geometry it becomes harder for the optimiser to place beads within the periphery that doesn't incur a cost due to the constraint placed on fitting beads within the nucleus. This promotes central placement of the beads within the cell nucleus as indicated by the percentage of DNA content placed in the periphery in Fig 4.1D.

To identify which chromosomes have beads which are consistently proximal across multiple inferences of chromatin arrangements, a series of chord plots (Fig 4.1E) were generated from the 200 cell geometries. Each linking line represents chromosomes that share at least one proximal ($<$500nm) interchromosomal bead in at least X% of the examined 200 geometries, where X can be a particular threshold. When analysing consistent chromosomes with proximal interchromosomal beads, there are no such examples which occur within 50% of the ellipsoid sample analysed. Therefore, all variants have been analysed at a 40% threshold to examine the difference between the conformations. There are more chromosomes that share proximal interchromosomal beads in the spherical geometries than the ellipsoid geometry. Interestingly, the chromosomes included in the chord plot for the ellipsoid geometries occur in the spherical geometries also, suggesting these are interaction driven from the experimental Hi-C contact maps rather than compromises due to other bead constraints. There are the same number of chromosomes that share proximal interchromosomal beads for both the geometries with and without LADs. However, there is an increase in the chord density with the inclusion of LADs, suggesting an increased amount of consistent intermingling at the analysed 500 nm radius, which is reasonable given the increase in constraints for these geometries.

Variations in Hi-C Geometries for Different Cell-Type

All spherical cell-type geometries (IMR90 - human fetal lung fibroblast, HMEC - human adult mammary epithelial and GM12878 - human B-lymphocyte), were solved using the GNOME software with an optimisation limit of 2 million iterations of successful movements.

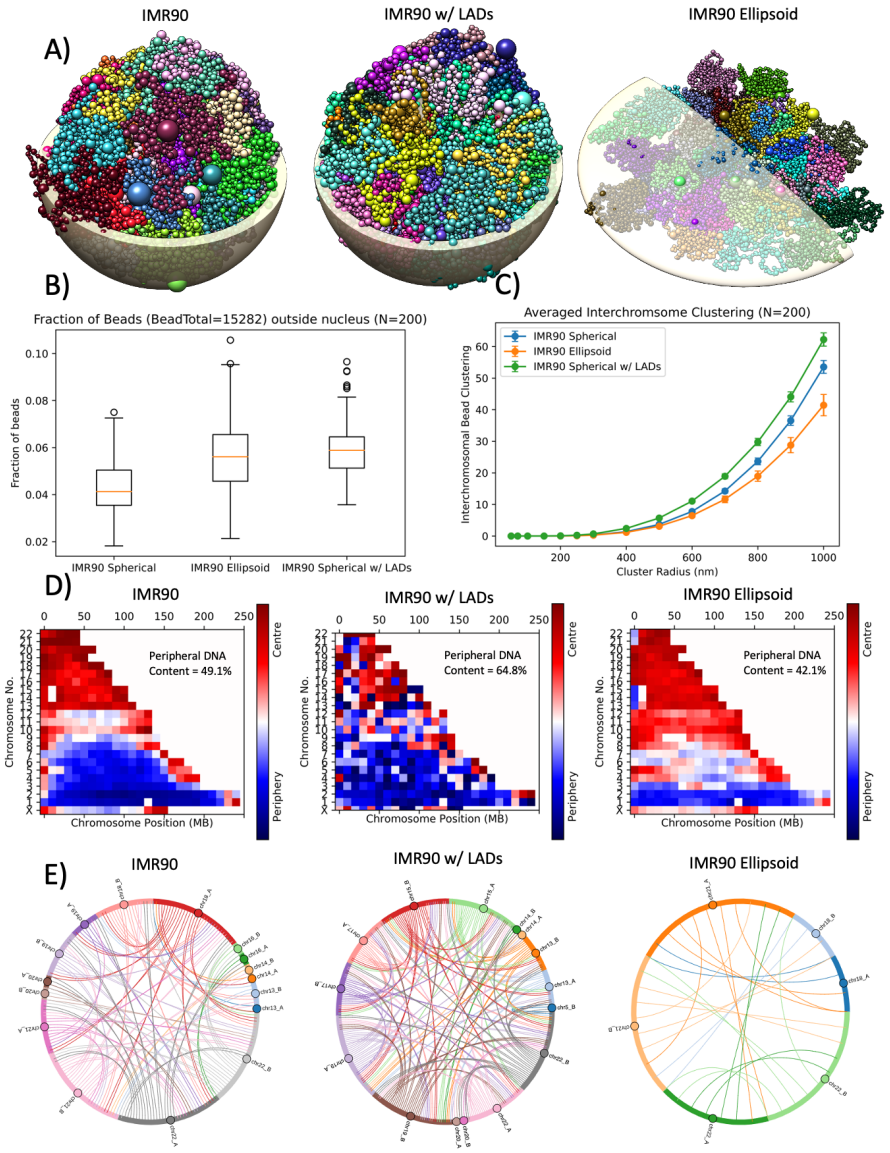

Figure 4.1: **Geometrical properties of an inferred encapsulation of the 3D spatial chromatin arrangement. A)** rendered 3D example of IMR90 Hi-C data solved for variants of adding LADs or optimising for an ellipsoid geometry. Each colour represents different chromosomes and the bead size represents the amount of DNA content. Rendered nuclear shell is to give an idea of scale for a 5 μm radius spherical nucleus for IMR90 and IMR90 w/ LADs, the ellipsoid nucleus is shown for 1.0x11.8x11.8μm. **B)** comparison of the fraction of beads that were not encapsulated by the optimiser objective nucleus. **C)** interchromosomal bead cluster analysis for the IMR90 variants. **D)** chromatin positioning analysis for beads that have been scored based on being within the central or peripheral half of the nuclear volume and is averaged over a 10 Mbp bin. The overall percentage of DNA content placed in the periphery is displayed within each plot. **E)** chromosomes that have interchromosomal beads within a 500 nm radius of one another. Lines are only displayed for the chromosomes that share proximal beads in at least 40% of the examined geometries. The data shown is averaged for 200 inferred geometries for each of the cell variants. Error bars are the standard deviation of 200 geometries.

For each of the three tested cell types, 200 geometries were solved. The resultant conformations were visually examined to check for gross abnormalities in the same manner as the cell variants (Fig 4.2A). To examine the optimisation of the three cell-type Hi-C datasets, which have different number of constraints, the average cost per constraint for the geometries are plotted for comparison (Fig 4.2B). Whilst there is quite a large variation observed within the same cell type, there is an overlap between the interquartile ranges of the differ-

ent cell types, suggesting suitable optimisation of the different datasets and allowing direct comparison.

Cell type changes in the interchromosomal bead clustering are shown in Fig 4.2C. It can be observed that the cell line GM12878 (human B-lymphocyte) had increased levels of intermingling than the other two cell lines.

The spatial positioning analysis (Fig 4.2D) shows the same resultant distributions largely follow the same pattern between the different cell-variants. Larger chromosomes (chr1-chr9) are predominately situated at the periphery and smaller chromosomes (chr13-chrX) are situated centrally.

The cell types were analysed for consistent chromosomes with proximal interchromosomal beads (Fig 4.2E) in the same manner as the cell variants. However, due to the increase of consistent shared proximal beads in comparison to the ellipsoid geometry the threshold was increased to 50% to better compare amongst the three cell types. The inferred GM12878 geometries share a higher amount of proximal chromosomes than both IMR90 and HMEC. The observed number of chromosomes that share a 500 nm proximity matches the observed amount of interchromosomal bead clustering, highlighting that this difference in Fig 4.2C is due to having an increased number of intermingling chromosomes. The chromosomes that, on average, have mutual proximal beads are those of the smaller and centrally located chromosomes observed in Fig 4.2D.

Comparison of Hi-C Genome Organisation and Pseudo-Random Organisation

To better quantify the genomic organisation of solved Hi-C geometries we compared the interchromosomal bead clustering of the different cell-types and variants against a set of pseudo-random geometries (Fig 4.3A). These pseudo geometries have no bead-bead constraints and can be placed anywhere within the specified nuclear volume (5 μm radius). Through the removal of organisation constraints the geometries become devoid of chromosome territories, resulting in the upper achievable level of chromosome intermingling (Fig 4.3B). The corresponding bead clustering values from the pseudo-random geometries were used to obtain a normalised value for all the Hi-C cell-types and variants (Fig 4.3C). The lack of chromosome territories results in a linear log-log relationship for the pseudo-random geometries between interchromosomal bead clustering and cluster radius, which is not seen by the geometries formed from Hi-C data suggesting that they are not mathematically repeating arrangements (Fig 4.3D). This aids in the definition of an upper limit of comparison for the structured Hi-C solved geometries and shows that solved geometries contain order. To spatially analyse the geometries the normalised Ripley-K function was calculated for the 3D geometries (Fig 4.3E). The pseudo random geometry resulted in a straight y=0 line for all cluster radii, suggesting it is completely spatially random. The IMR90 Ellipsoid geometry displayed an increase in clustering with a positive non-linear trend at a magnitude that was distinct to the other geometries. Whereas, the remaining geometries showed a lesser, but still non-linear separation to the y=0 line of complete spatial

randomness, suggesting some amounts of clustering, but less than the IMR90 ellipsoid geometry.

To further evaluate the spatial arrangements of the 3D geometries compared to the pseudo random geometry a proximity score was produced based on the solved 3D geometries ability to appease the Hi-C derived of TAD (bead) contact constraints. The proximity score is an averaged measure of euclidean distance for each bead with all of it's contact constraints. A lower value equivalent to a better optimisation score, for the Hi-C constraints, but ignores the constraints of all the beads being within the nucleus or lamina-based constraints. In theory this can equal zero if the analysed bead is next to all other beads it has a contact for, but given that beads have a physical size and cannot overlap this value is likely to always have some value > 0. The proximity score analysing (Fig 4.3F) shows similar levels of optimisation for all cell type and variants which are much less than the pseudo random geometry which represents a non-optimised solution. This helps justify that the optimiser is in fact promoting the constraints derived from the Hi-C data. The spatial positioning of DNA content was also evaluated (Fig 4.3G) to get an idea of spread in either the central or peripheral half of the nucleus volume. For geometries the three cell type geometries (GM12878, HMEC, IMR90) the distribution is relatively balanced. Wheras, the IMR90 Ellipsoid cell has increased placement of DNA content within the central portion of the cell nucleus. The IMR90 LADs have an increased amount of DNA content places in the periphery which is expected given the lamina-based constraints being applied.

Simulated DNA Damage Yields in Hi-C Geometries

The solved Hi-C geometries for the three cell types (GM12878, HMEC and IMR90) along with the two variants of IMR90 (IMR90-LADs and IMR90-Ellipsoid) were built in Geant4-DNA. The built Geant4 geometries were subjected to irradiation from protons, helium- and carbon-ions and the resultant energy depositions were classified into DNA damage (Fig 4.4A). The yields of both DSB and Single-Strand Break (SSB) per unit dose (Gy) as a function of Linear-Energy Transfer (LET) (keV/μm) shows the expected increase in DSBs and a corresponding decrease in SSBs with increasing LET values (Fig 4.4). It can be seen that across the LET range investigated all yields remain within error bars of one another and the cell-type geometry or geometry alterations (LADs and ellipsoid shaping) do not affect resultant damage yields.

We investigated the markedly large drop in DSBs at $>$1000 keV/μm, which was due to increased damage clustering which results in DSBs consisting of more than two backbones, forming complex DNA damage (S1 Fig). To evaluate if there is substantial change on DNA DSB yield between different geometries of the same cell type we investigated the DSB/Gy yields as a function for each geometry (S2 Fig).

The yields per geometry were sorted (smallest to largest yields) as there was no clear correlation to specific geometries within the same cell types leading to offsets in yields. Whilst, there is some cell to cell DSB yield variation for the same cell-type it does not appear to be

a significant (S3 Fig), with most points falling within the standard error of the mean from different exposures. Furthermore, to check if there were portions of the genome that were geometrically more vulnerable to DSB induction, we examined the DSB/Gy/BasePair for each chromosome, no differences were identified between cell types (S4 Fig).

To analyse how the yields of damage changes relative to the position within the cell nucleus both the DSB normalised frequency and DSB density has been plotted for each cell type (S5 Fig), which show minimal change between the cell types. The effects of adding lamina-based constraints in the IMR90-LADs geometries were evaluated against their counterpart IMR90 geometries, which do not have such constraints, in the same manner (S6 Fig). The increased positioning of beads towards the periphery does cause a notable effect for the damage distribution, with higher levels of DSB density located at the periphery.

Simulated DNA Damage Distributions in Hi-C Geometries

The spatial pattern of the DSBs within the cell geometry were analysed for clustering at various radii, this provided a description of the total DSB clustering (S7 Fig), the inter-chromosomal DSB clustering (S8 Fig) and the intrachromosomal DSB clustering (S9 Fig). These metrics can then be used to discern the intrachromosomal DSB clustering and the ratio of inter-/intra- chromosome clustering (Fig 4.5). Total DSB clustering, comprimised of both inter- and intra- chromosome DSBs, increases with LET, whilst the inter/intra clustering ratio has the opposite relationship with LET. This relationship is caused by the distribution of breaks making up the 1 Gy of absorbed dose within the cell nucleus. At higher LETs the distribution of damage becomes more localised, increasing the corresponding intrachromosomal DSB clustering. Conversely, at lower LETs the damage becomes more distributed, which increases the interchromosomal DSB clustering.

To evaluate the relationship between the chromosome intermingling, in the form of bead clustering (Fig 4.2C), and DSB interchromosomal clustering we plotted these metrics for each of the 200 geometries per cell-type category against one another (Fig 4.6). Whilst, there is a spread in the relationship between these two metrics it can be seen at 1 Gy of Co-60 (Fig 4.6A) there is a subtle increase in the mean value (denoted by the intersection of the black lines) for both geometry and damage metrics. However, as the radiation is only damaging within a small proportion of beads it was expected that the underlying relationship would strengthen as the number of damages increases (essentially sampling the geometry further). To evaluate this we also irradiated the geometries with 100 Gy of Co-60 and to show a decrease in the variation due to the exposure and a stronger observable relationship between the metrics at both a geometric and damage level. When analysing the relationship between geometry and DSB interchromosomal clustering for a range of LET values (Fig 4.6B) the relationship persists within a particular LET band (colour marked). However, with the inclusion of the two variations on the IMR90 geometry (Fig 4.6B-right) there is a separation from the relationship seen between cell-types (Fig 4.6B-left). These results identify that whilst differences between cell-types are able to be detected at the DNA damage

level, the changes at this level from resultant cell shape (i.e. flattened ellipsoid cell) are noticeably different to the pattern seen across spherical shapes. The same can be said for the inclusion of LADs when solving the geometries, but on a smaller scale than that of solving for a flattened ellipsoid cell.

To examine if changes in interchromosomal DSB clustering were significant each group was tested at all LET and cluster radius values shown previously (Fig 4.7). This identifies that changes in the interchromosomal bead clustering (Fig 4.1C and Fig 4.2C), shown to be statistical significance (S11 Fig), translate to significant differences in the induced damage. Variation between the cell-types (Fig 4.7A, Fig 4.7B and Fig 4.7C) are typically statistically significant at higher values of LET and cluster radius (excluding carbon-ions at 1034.69 keV/μm). The largest difference when comparing cell-types is between the solved GM12878 and HMEC geometries and the smallest difference is between the IMR90 and HMEC geometries, matching the differences observed in (Fig. 4.6). Whereas, statistically significant differences can be seen for the majority of tested LETs and cluster radii in the variant comparisons (Fig 4.7D, Fig 4.7E and Fig 4.7F).

4.5 Discussion

We have shown that the changing of cell-type, addition of LADs, or solving for an ellipsoid does not have noticeable influences on the yields of DSB and SSB DNA damage, but there is a difference in the pattern of damage. The resultant DNA DSB interchromosomal clustering (Fig 4.6B) suggests that it is possible to detect statistically significant differences between cell types (Fig 4.7), but the addition of LADs and especially ellipsoid shaping causes a distinct alteration in how the interchromosomal geometric description propagates through to the damage distribution. It is in this case that the normalised Ripley-K may be a better predictor of the DSB interchromosomal clustering when you depart from spherical geometries as this accounts for non-overlapping cluster radii volume and nucleus volume through boundary corrections. These variations between cell-types are detectable in all types of clustering (interchromosomal, intrachromosomal and total), with the magnitude of the values still dictated by radiation parameters (e.g. LET). The observed changes in inter/intra chromosomal DSB clustering ratios is an interesting parameter to analyse as it may be a predictor of inter- and intra-chromosomal misrepair, which has a differing biological response [46]. We believe that interchromosomal DSB clustering may be a good predictor of interchromosomal chromosome aberrations, which are often toxic to the cell [47]. Experimental work has previously shown increased chromosomal intermingling correlates with increased chromosomal translocation [15], we are able to explore these findings at their intermediate step of DNA damage distributions.

To resolve if it is feasible to detect differences at a DNA damage level when using different Hi-C cell-type datasets we ensured that all geometries were solved for the same nuclear volume and the total amount of DNA (6 Gbp). However, from the literature, it is apparent that cell nucleus [48] size can vary substantially both between and within cell-type. This

change in cell size has also been linked to a change in the amount of DNA content [49], these features have not been encapsulated here due to simulation time, but are within the capability of the G-NOME software and our damage model. Due to the differences noted from changes in cell shape, we would expect the differences in both size and DNA content to have a marked response in the damage distribution produced. Furthermore, through changing the amount of DNA content there is an alteration in the number of sensitive targets within the cell, which would be expected to lead to a difference in the yields of damage.

In this study, we have included the effects of solving the geometry of the same volume, but for a flattened ellipsoid shape (which may be predominant *in vitro*) and comparing it to the corresponding spherical geometry solved using the same Hi-C data. This was perfomed to understand any radiation damage differences that may occur in a flattened nucleus, a phenomenon that is observed during cell spreading when cells are plated in a 2D system [50]. As it is common to plate cells in some irradiation experiments the substantial change in the damage distribution within the nucleus may indicate differences in the expected radio-response between 2D and 3D cell system experimental techniques. The possibility to evaluate the geometric effects of nuclear shaping in attached 2D versus 3D cell systems may help discern experimental variability observed across different techniques analysing the same endpoint [51]. There are further implications of this when we move towards modelling patient radio-response, previous work has shown that there is improved predictive power of clinical efficacy when using 3D over 2D *in vitro* systems [52]. This is clearly an issue as a large quantity of previous and continuing experimental radiobiological work are carried out in 2D cell systems, here we are able to understand the differences in radio-response due to nuclear shape change and open the possibility of translation factors to be derived.

The geometries produced are representative of a single cell but are derived from a highly averaged population dataset. By enforcing the polymer model to arrange itself through a series of movements we are able to use the Hi-C data as a guide to achieving realistic single-cell conformations, rather than positioning beads with no heuristic limitation to achieve the smallest cost function. With the increased development in single-cell Hi-C [53] a development of the G-NOME software would be to use data to perform single-cell conformation modelling. This would allow a method of validation for using the current population-based method as a suitable approximation for single-cell conformations. However, the benefits of single-cell Hi-C extend beyond simple validation, it may allow for the ability to describe different cell sub-types, which includes cell-cycle specific states [20, 54]. The ability to discern cell-cycle specific geometries is a pertinent goal in radiobiological modelling, as there is a well establish variation in radiosensitivity at different cell cycle phases [55]. Leveraging differences in chromatin conformation along with alterations in the active repair pathways could be a viable method of modelling these effects.

The current model is limited in terms of its genomic resolution, at present the size of each bead is defined by topologically associated domains (TADs), which are used widely as discerning DNA segments of self-interacting regions. There may be a benefit in trying to build

models at smaller genomic resolutions, but there should be caution in the decreased signal-to-noise ratio and the possibility of mis-classifying significant contact points. Although, as the field of Hi-C moves to increasingly finer resolutions, including the recent inception of Micro-C [56], there may be opportunity to model sub-structures (e.g. nucleosome) of the genome. These sub-structures may be especially important when attempting to model the more granular configuration of the DNA damage, along with the ability to model regions of heterochromatin and euchromatin.

In order to accurately model differences in intrinsic radio-sensitivity between cell-types we must evaluate the the characteristic differences, starting with DNA damage. We have shown that there is differences in DNA damage that are geometry-driven and believe this is key to understanding the effects of LET on relative biological effectiveness through the explicit modelling of chromosome aberration probability. Furthermore, through examination of the effects of LET and cluster radius leading to significant differences in simulated interchromosomal DSB clustering of different cell types (Fig 4.7), it is possible to use this information to design experiments to evaluate DSB end motion, a subject that will drive chromosome aberration probability, but remains elusive with proposed DSB end motion being in the range of 70 [57], 100 [58] and 500 nm [59]. The related topic of chromatin dynamics following irradiation, that has recently been shown to be cell-type specific [60], will also need to be examined to predict the downstream biological effects from the cell-type specific damage patterns. Finally, we acknowledge that to fully encapsulate the radiobiological differences between cell-types we must include descriptions of the characteristic variation that are omics-driven [61].

The combination of Hi-C data to inform geometric structures for *in silico* modelling of radiation track structure and DNA damage has been shown to be feasible in this study. Through leveraging the genomic element of the Hi-C data it is possible to enrich the descriptions of radiation damage and quantify the overall damage distribution. This will be used in subsequent models of DNA repair, where cell-type differences of proliferation rate, protein expression and micro-environment can be encapsulated to further investigate the observed variation in cellular radio-response. Furthermore, through this expansion into Hi-C modelling for normal cells in an attempt to better understand normal tissue response, we can look towards incorporating the alterations in genomic structure found in cancer cells to improve our understanding of target response [62–64]. It is thought that this relationship will work symbiotically, as radiobiological response is strongly dependent on geometry (e.g. translocations) [15] and can be used as an additional experimental technique to validate the Hi-C geometries. Finally, though the use of radiobiological modelling there forms a new translational pathway to how the improved geometric understanding of Hi-C could benefit cancer patients at a clinical level.

4.6 Materials and Methods

Preparation of Hi-C data

This introduced model has been designed to utilise the same input gtrack file format as previous 3D inference model "Chrom3D", enabling previous protocol work [65] that outlines how to analyse and process gtrack files to still be applicable for the G-NOME software. Furthermore, this also enables the ability to incorporate lamina-associated domains (LADs) data into the optimisation objectives to promote areas of the genome which have higher interaction with the nuclear lamina to the periphery of the solved geometry [66, 67].

Whilst the previous protocol work [65] should be referred to for a detailed explanation of how to process a Hi-C dataset to get a gtrack file for subsequent G-NOME solving, we provide a short overview here. The gtrack file is a line by line summary of the TADs identified within the Hi-C data, it includes the chromosome identity, genomic start position, genomic end position, unique ID, contacts and optional periphery status (binary). To generate this information tools such as aidenlab's juicer (https://github.com/aidenlab/juicer/) should be used to pre-process the raw contact data into *.hic files. After which the Arrowhead algorithm should be used to deduce contact domains and will provide the starting files for the protocol outlined by Paulsen et al [65]. These domains can then be examined at a intra-chromosomal and interchromosomal level, to establish TAD interaction counts ready for analysis and resulting in the creation of a BEDPE-format file. This process can take significant time and will depend on the resolution of the Hi-C matrix. It is at this point to account for domains with centromeres and unmappable regions to avoid artifacts in the statistical tests. The P-values and odds ratios are calculated for all the contacts detailed in the BEDPE files using the non-central hypergeometric (NCHG). The data is then filtered using both the calculated false discovery rate (FDR) and the odds ratio. It is at this point that the data can be formatted in the gtrack file format, along with the chance to aggregate corresponding Lamina-associated domain (LAD) data if required.

When using the protocol for generating the gtrack files, we ensured that variation in the analysed intra-chromosomal (down to 5 kbp) and inter-chromosomal (100 kbp - 1 Mbp) interaction matrix gave no discernible difference in the geometric analysis when solving the geometry at a TAD resolution (S12 Fig). All cell-type and variant geometries shown in the results have been solved using the 50 kbp intra-chromosomal and 1 Mbp inter-chromsomal matrix for the production of the gtrack file.

This study uses the Hi-C data published by Rao et al., [68] (GEO Accession GSE63525), which has been used widely within the literature as an exemplar test dataset. However, the proposed model is suitable for inferring 3D geometries from other Hi-C datasets.

Hi-C Solver

The G-NOME software uses a Markov-Chain Monte Carlo (MCMC) polymer model, which can be optimised through both the metropolis-hastings and simulated annealing algorithms. The resultant geometry is based on a series of optimised objectives applied to polymer beads, which represent a defined number of base pairs of DNA along a chromosome (TADs), being arranged so that areas of strong contact probability are spatially proximal. In order to arrange the geometry a series of iterations attempt several bead movement types at random, if the movement results in an improved objective score the iteration can be accepted and the next iteration may take place. Running through many iterations allows for the overall solved geometry to be realised. The movement types are as follows:

1. Crankshaft - rotates a random number of beads around two fixed beads points.
2. Arm Rotation - from a random position along the chromosome rotate all beads to the end or start of the the chromosome. The start or end is chosen by random.
3. Arm Wiggle - from a random position along the chromosome re-position (via a self avoiding walk) beads to the end or start of the chromosome. The start or end is chosen by random.
4. Translation - move the whole chromosome by a random x, y, z between 0 and 1um.
5. Rotation - rotate the whole chromosome by a random amount.

The G-NOME software developed is a re-implementation of the Chrom3D model [23], but includes modifications to better interface with track-structure models of DNA damage [57, 69]. The resultant 3D structures from G-NOME have been compared with Chrom3D for DNA-content spatial placement (S13 Fig) and the optimisation score it delivers (S14 Fig) to ensure no major deviation between the models given for the same set-up. This improved interfacing allows for geometric solving and DNA damage models to be coupled, resulting in generation of unique cell geometries for every run of the DNA damage model. Further differences between the introduced G-NOME software and Chrom3D include: change in the programming language to Python, improved compute efficiency resulting in faster solving times (S15 Fig), the ability to optimise for non-spherical geometries (here we demonstrate ellipsoid) and custom optimisation routines. Custom optimisation routines refers to the ability to be able to dynamically alter optimisation constraints during a run time session. We believe this gives us additional flexibility that may be required in future work with G-NOME and radiobiological modelling.

As Hi-C data is the averaged result of a large cell population it becomes apparent that the solution space for solving these geometries may contain many suitable solutions of chromosome conformation. This has been shown in several studies [23, 24] and is the underpinning for the hypothesis of encapsulating the biological variation by using different solutions outputted from the model. Variation in the output arises from the Monte Carlo approach of

using different random seeding for the initial distribution of the chromosomes, since the initial distribution has a strong influence on the end result.

In total 1,000 geometries (200 geometries per group) have been calculated which include: three cell types (IMR90 - human fetal lung fibroblast, HMEC - human adult mammary epithelial and GM12878 - human B-lymphocyte), two cell shapes (spherical and ellipsoid) and inclusion of LADs for the IMR90 cell line. All spherical cells were solved for the same target nucleus size of 5 μm radius. The ellipsoid cells were solved for the target nucleus size of 1.0x11.8x11.8μm radii. All cells were modelled with the same amount of DNA content (6Gbp) and as normal diploid human cells (46 chromosomes). All geometries have been optimised with 2 million iterations of successful movements unless optimised solutions were found before this or stated otherwise. All renderings of the 3D spatial chromatin arrangement were made using the 3D visualisation tool Chimera [70] by loading the "*.cmm" files created from the G-NOME software. All geometries were solved using additional nuclear boundary constraints, which adds a cost based on if the beads were confined to the user-defined nuclear boundary. The costs applied to these constraints are 0 if the bead is within the cell nucleus and only occur cost on positioning outside of the nucleus based on the euclidean distance from the nuclear boundary. This constraint can be toggled when using the simulation run script provided through the flag "–ConstrainNucleus". These additional constraints were required as we wanted to preserve total volume across all cell geometries for the subsequent Geant4 simulation.

DNA Damage Simulation

Details of the polymer beads, produced by G-NOME, are read into our DNA damage application [57, 69]. Each bead is placed as a spherical geometry object in the Monte Carlo toolkit Geant4 (geant4 10.5.1) [71], using the X, Y, Z and variable bead radius, within a bounding nucleus volume. Simulation of cell nucleus irradiation is performed within Geant4, using the default Geant4-DNA physics list [44]. Within Geant4 the track structure of a radiation source is simulated as a series of interaction limited steps through a specified geometry, with each step updating the energy and trajectory of the primary or secondary particle. For particle-induced DNA damage, energy depositions occurring within beads are recorded. Two conditions are applied to convert energy depositions into strand breaks. Firstly, a spatial sampling of 14.1% is applied to the bead. Secondly, an energy range probability is applied, from 0 at 5 eV to 1 at 37.5 eV. Once passing both conditions an energy deposition is accepted as a strand break and is randomly assigned to strand 1 or 2 of the double helix, with equal probability. The chromosome of damage is directly assigned from the G-NOME bead. The position along the chromosome is informed by the G-NOME bead. Since each bead contains a portion of the chromosome a minimum and maximum base pair position is known, with the actual position of damage taken as a random point between the two. Following the simulation of all particles required to deliver a dose the list of damages is analysed through a clustering algorithm. Double Strand Breaks are formed by two or more strand breaks that are on opposite strands and separated by 3.2 nm or less (equivalent to 10

bp). Strand breaks that don't form a DSB are classified as Single-Strand Breaks.

The spatial sampling of 14.1% corresponds to a sensitive fraction of the bead and was determined to reproduce DSB yields seen in our previous work [57]. This fit is in good agreement with other models [72, 73] which have been validated against experimental data. The energy-based probability of damage induction was adapted from PARTRAC [73] and is based on studies of DNA strand breakage after exposure to very low energy electrons [74].

For photon induced DNA damage, DSB induction is assumed to follow a Poisson distribution with an average of 25 DSBs/Gy. For each DSB the chromosome is chosen at random, with probability weighted according to the chromosome size relative to total genome size. Similarly, a bead within that chromosome is chosen randomly, with probability weighted according to on bead size relative to the sum of all beads forming the chromosome. A random X, Y, Z within the selected is bead is assigned to the DSB. All DNA damages are recorded in the Standard DNA Damage (SDD) format [40].

In this work, the G-NOME cell models are irradiated with 1 Gy of photons, protons (3 MeV – 67 MeV), helium ions (4 MeV – 80 MeV), or carbon ions (10 MeV – 213 MeV). Ensuring a range of particle type and track-averaged Linear Energy Transfer, with an overlap in LET between the particles. A table of all particles, particle energies and calculated LET values are given in the S1 Table. A reader for these geometries will be implemented in "TOPAS-nbio" to further accessibility.

Statistical information

The statistical analysis used in Fig 4.7 was an adjusted P-value using the using the Benjamini-Hochberg correction to control the false discovery rate (type I error). The P-values calculated for adjustment are from a Kolmogorov–Smirnov test between different cell-type or variant groups. The test was two-sided with the null hypothesis indicating that the two samples are drawn from the same continuous distribution. The adjusted P-values was presented for all cell-types and variants groups for a range of track averaged LET values and cluster radii. A threshold of $P < 0.05$ was used to determine statistically significant differences between tested samples. The Kolmogorov–Smirnov test was calculated using Python and the scipy (v1.4.1) package. The Benjamini-Hochberg correction was calculated using Python and the statsmodels (v0.12.0) package. All tests had 200 geometries in each group.

The Ripley-K function was calculated for the 3D distrubtion of polymer beads for each 3D geometry produced by G-NOME. The calculation methodology follows the 3D implementation carried out by Jafari-Mamaghani et al [75]. The equation is detailed below:

$$K(CR) = V_{Nuc}\frac{\sum_{i=1}^{n}\sum_{i\neq j} I[D(i,j) \leq CR]}{V_s n^2} \tag{4.1}$$

where V_{Nuc} is the volume of the nucleus, CR is the cluster radius, n is the number of DSBs in the nucleus, V_s the edge correction term is the fraction of overlapping volume of the CR

volume and the V_{Nuc} and I is the indicator function which will be either 1 if the condition $D(i,j) \leq t$ is true or will be 0, $D(i,j)$ is the euclidean distance from DSB i to DSB j.

4.7 Supporting information

S1 Fig **DSB Damage Complexity**. Average number of DSBs/Gy (filled symbols) and backbones per DSBs (empty symbols) for a range of LET values across different particle types. Error bars are displayed as the standard error of the mean for 100 repeats.

S2 Fig **DSB Yield Variation Across Generated Geometries**. Yields of DSB per Gy of dose for the each of the 200 geometries created. Results have been sorted from smallest to largest yields to allow for easier interpretation. Different cell-types are shown as different line types with each radiation quality presented as a different colour. Errors are the transparent area around the line and are the standard error in the mean for 50 independent exposures per geometry.

S3 Fig **DSB Yield Distribution**. Double-strand break yield histograms for 200 geometries of each cell-type and variant.

S4 Fig **DSB Yield Chromosome Distribution**. Double-strand break per 1Gy of dose per DNA basepair on each of the modelled 46 chromosomes for each cell-type and variant. Error bars are displayed as the standard error of the mean for 200 geometries for each cell-type and variant with each geometry having 50 independent exposures.

S5 Fig **Spatial distribution of DNA DSB yields for different cell types**. Dual axis plot – left y-axis shows the histogram plot of the Normalised DSB frequency and right y-axis is the corresponding average DSB density for the same x-axis bin per geometry. Both are given as a function of distance from the nucleus centre. The cell types are all solved for a spherical nucleus and do not include LADs. The DSB frequency was normalised to the maximum number of DSBs within any bin for a given cell type. DSB density is calculated as the average number of DSBs per geometry (N=200) within a bin divided by the volume (μm^3) of the spherical shell of the bin. Error bars in the DSB density are the standard error in the mean for all 200 geometries for each cell type.

S6 Fig **Spatial distribution of DNA DSB yields for the addition of LADs**. Dual axis plot – left y-axis shows the histogram plot of the Normalised DSB frequency and right y-axis is the corresponding average DSB density for the same x-axis bin per exposure. Both are given as a function of distance from the nucleus centre. Comparison between IMR90 with and without LADs constraints for a spherical nucleus. The DSB frequency was normalised to the maximum number of DSBs within any bin for a given cell variant. DSB density is

calculated as the average number of DSBs per geometry (N=200) within a bin divided by the volume (μm^3) of the spherical shell of the bin. Error bars in the DSB density are the standard error in the mean for all 200 geometries for each cell variant.

S7 Fig **DSB clustering per Radiation Quality Plots**. Double-strand break clustering as a function of the cluster radius for all cell-types and variants. Error bars are displayed as the standard error of the mean for 200 geometries for each cell-type and variant with each geometry having 50 independent exposures.

S8 Fig **DSB Interchromosomal clustering per Radiation Quality Plots**. Double-strand break interchromosomal clustering as a function of the cluster radius for all cell-types and variants. Error bars are displayed as the standard error of the mean for 200 geometries for each cell-type and variant with each geometry having 50 independent exposures.

S9 Fig **DSB Intrachromosomal clustering per Radiation Quality Plots**. Double-strand break intrachromosomal clustering as a function of the cluster radius for all cell-types and variants. Error bars are displayed as the standard error of the mean for 200 geometries for each cell-type and variant with each geometry having 50 independent exposures.

S10 Fig **DSB Inter/Intra Clustering per Radiation Quality Plots**. Double-strand break inter/intra chromosomal clustering as a function of the cluster radius for all cell-types and variants.

S11 Fig **Statistical Differences between Interchromsomal Bead clustering of Different Cell Lines**. False discovery rate adjusted P-values from a 2-sided Kolmogorov-Smirnov test on the interchromosomal bead clustering values for the different cell-types and variants. Colour coding for adjusted P-values at varying thresholds: red ($P > 0.05$), purple ($0.05 > P > 0.01$), yellow ($0.01 > P > 0.001$) and green ($P < 0.001$). In this case, distributions with adjusted P-values < 0.05 will be considered as having significant statistical difference to one another. Each of the tested distributions had 200 geometries per cell-type or variant group.

S12 Fig **Nuclear Positioning of Beads at Varying Hi-C Contact Resolutions**. Bead positioning between periphery and central locations for a range of different interchromosomal contact resolutions at the finest available intrachromosomal contact resolution. Each category consists of 200 geometries created from the corresponding gtrack file created from using different analysis resolutions.

S13 Fig **DNA content position model comparison**. Box plots of the DNA content positioned either in the peripheral half or central half of the cell nucleus volume. These results

are for 50 geometries from G-NOME and 50 geometries from Chrom3D (v1.0.2). In both models the same input IMR90 noLADs gtrack file was optimised for 1 million iterations, 5-micron nuclear radius and 0.15 occupancy volume.

S14 Fig **Proximity score model comparison**. Box plots of the proximity scores which is the average Euclidean distance between TADs which have a constraint to be proximal other TADs (lower value indicates a better optimisation of the contact constraints). To put these differences into perspective for a randomly distributed geometry where the proximity score is approximately 12. These results are for 50 geometries from G-NOME and 50 geometries from Chrom3D (v1.0.2). In both models the same input IMR90 noLADs gtrack file was optimised for 1 million iterations, 5-micron nuclear radius and 0.15 occupancy volume.

S15 Fig **Nominal evaluation of speed performance**. Timing performance for a nominal single IMR90 spherical cell 3D geometry generation using both G-NOME and Chrom3D (v1.0.2). In this case for 2 million iterations (the number used for the evaluation of different cell types) is 23.5 hours in Chrom3D and 7.6 hours in G-NOME.

S1 Table **Incident Particle Range** Examination of particle range travelling through 20μm of water. Simulation was carried out for each energy with the furthest depth of each particle scored and averaged for 100 single-particle transversals. Particles which transverse beyond the 20μm of water are simply signified as having ranges beyond 20μm.

S1 File. **Cell Type Interchromosomal Clustering** - Interactive 3D plot of interchromosomal DSB clustering for a range of cluster radii and LET for the three cell types: IMR90, GM12878 and HMEC.

S2 File. **Cell Type Intrachromosomal Clustering** - Interactive 3D plot of intrachromosomal DSB clustering for a range of cluster radii and LET for the three cell types: IMR90, GM12878 and HMEC.

S3 File. **Cell Type Inter/Intra Ratio Clustering** - Interactive 3D plot of the ratio of interchromosomal/intrachromosomal DSB clustering for a range of cluster radii and LET for the three cell types: IMR90, GM12878 and HMEC.

S4 File. **Cell Type Clustering** - Interactive 3D plot of the total DSB clustering for a range of cluster radii and LET for the three cell types: IMR90, GM12878 and HMEC.

S5 File. **Cell Variation Interchromosomal Clustering** - Interactive 3D plot of interchromosomal DSB clustering for a range of cluster radii and LET for the three IMR90 cell variants: IMR90, IMR90 with LADs and IMR90 ellipsoid.

S6 File. **Cell Variation Intrachromosomal Clustering** - Interactive 3D plot of intrachromosomal DSB clustering for a range of cluster radii and LET for the three IMR90 cell variants: IMR90, IMR90 with LADs and IMR90 ellipsoid.

S7 File. **Cell Variation Inter/Intra Ratio Clustering** - Interactive 3D plot of the ratio of interchromosomal/intrachromosomal DSB clustering for a range of cluster radii and LET for the three IMR90 cell variants: IMR90, IMR90 with LADs and IMR90 ellipsoid.

S8 File. **Cell Variation Clustering** - Interactive 3D plot of total DSB clustering for a range of cluster radii and LET for the three IMR90 cell variants: IMR90, IMR90 with LADs and IMR90 ellipsoid.

4.8 Data Availability

All datasets are available at https://data.mendeley.com/datasets/kzycj3n2mm. The G-NOME source code is available at https://gitlab.com/PRECISE-RT/releases/g-nome.

4.9 Acknowledgments

This project utilised a significant amount of computational resources (approx. 120 CPU/years). The authors would like to acknowledge the significant amount of assistance given by University of Manchester Research IT, with special appreciation to Dr. Daniel Corbett. The computational work was carried out with the use of both the Computational Shared Facility and Condor at The University of Manchester. Molecular graphics and analyses performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco.

4.10 Funding

K.J.K, R.I.M, N.F.K and M.J.M were funded by the NIHR Manchester Biomedical Research Centre [grant number: BRC-1215-20007]. S.P.I, M.J.M, K.J.K and R.I.M were funded by the STFC Global Challenge Network+ in Advanced Radiotherapy [grant number: ST/N002423/1] N.F.K, K.J.K, R.I.M, M.J.M, N.T.H and J.W.W were funded by the European Union Horizon 2020 Research and Innovation [grant number: 730983 – INSPIRE]. K.J.K, R.I.M, N.F.K and M.J.M were funded by the United Kingdom Engineering and Physical Science Research Council [grant No.: EP/S024344/1 – BioProton]. S.P.I was funded by The Christie Charitable funds - The Christie NHS Foundation Trust (https://www.christie.nhs.uk/the-christie-charity). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Cremer, T. & Cremer, M. Chromosome Territories. *Cold Spring Harbor Perspectives in Biology* **2,** a003889 (2010).
2. Zhang, Y. *et al.* Radiation Induced Chromatin Conformation Changes Analysed by Fluorescent Localization Microscopy, Statistical Physics, and Graph Theory. *PLOS ONE* **10,** e0128555 (2015).
3. Roti, J. L. R., Wright, W. D. & Taylor, Y. C. DNA Loop Structure and Radiation Response. *Advances in Radiation Biology* **17,** 227–259. ISSN: 0065-3292 (1993).
4. Shukron, O., Seeber, A., Amitai, A. & Holcman, D. Advances Using Single-Particle Trajectories to Reconstruct Chromatin Organization and Dynamics. *Trends in Genetics* **35,** 685–705. ISSN: 0168-9525 (2019).
5. Takata, H. *et al.* Chromatin Compaction Protects Genomic DNA from Radiation Damage. *PLoS ONE* **8,** e75622 (2013).
6. Schneider, U., Vasi, F. & Besserer, J. The Impact of the Geometrical Structure of the DNA on Parameters of the Track-Event Theory for Radiation Induced Cell Kill. *PLOS ONE* **11,** e0164929 (2016).
7. Dugle, D. L., Gillespie, C. J. & Chapman, J. D. DNA strand breaks, repair, and survival in x-irradiated mammalian cells. *Proceedings of the National Academy of Sciences* **73,** 809–812. ISSN: 0027-8424 (1976).
8. Wang, W. *et al.* Modelling of Cellular Survival Following Radiation-Induced DNA Double-Strand Breaks. *Scientific Reports* **8,** 16202 (2018).
9. Savage, J. R. Classification and relationships of induced chromosomal structual changes. *Journal of Medical Genetics* **13,** 103. ISSN: 0022-2593 (1976).
10. Sachs, R. K., Chen, A. M. & Brenner, D. J. Review: Proximity effects in the production of chromosome aberrations by ionizing radiation. *International Journal of Radiation Biology* **71,** 1–19. ISSN: 0955-3002 (2009).
11. McMahon, S. J., Schuemann, J., Paganetti, H. & Prise, K. M. Mechanistic Modelling of DNA Repair and Cellular Survival Following Radiation-Induced DNA Damage. *Scientific Reports* **6,** 33290. ISSN: 2045-2322 (2016).
12. Warmenhoven, J. W. *et al.* Insights into the non-homologous end joining pathway and double strand break end mobility provided by mechanistic in silico modelling. *DNA Repair* **85,** 102743. ISSN: 1568-7864 (2020).
13. Ballarini, F. From DNA Radiation Damage to Cell Death: Theoretical Approaches. *Journal of Nucleic Acids* **2010,** 350608. ISSN: 2090-0201 (2010).

14. Sia, J., Szmyd, R., Hau, E. & Gee, H. E. Molecular Mechanisms of Radiation-Induced Cancer Cell Death: A Primer. *Frontiers in Cell and Developmental Biology* **8,** 41. ISSN: 2296-634X (2020).

15. Branco, M. R. & Pombo, A. Intermingling of Chromosome Territories in Interphase Suggests Role in Translocations and Transcription-Dependent Associations. *PLoS Biology* **4,** e138. ISSN: 1544-9173 (2006).

16. Fraser, J., Williamson, I., Bickmore, W. A. & Dostie, J. An Overview of Genome Organization and How We Got There: from FISH to Hi-C. *Microbiology and Molecular Biology Reviews* **79,** 347–372. ISSN: 1092-2172 (2015).

17. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326,** 289–293. ISSN: 0036-8075 (2009).

18. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502,** 59. ISSN: 1476-4687 (2013).

19. Chen, H. *et al.* Functional organization of the human 4D Nucleome. *Proceedings of the National Academy of Sciences* **112,** 8002–8007. ISSN: 0027-8424 (2015).

20. Nagano, T. *et al.* Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547,** 61. ISSN: 1476-4687 (2017).

21. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nature Reviews Genetics* **17,** nrg.2016.112. ISSN: 1471-0064 (2016).

22. Eagen, K. P. Principles of Chromosome Architecture Revealed by Hi-C. *Trends in Biochemical Sciences* **43.** ISSN: 0968-0004 (2018).

23. Paulsen, J. *et al.* Chrom3D: three-dimensional genome modeling from Hi-C and nuclear lamin-genome contacts. *Genome Biology* **18,** 21 (2017).

24. Rieber, L. & Mahony, S. miniMDS: 3D structural inference from high-resolution Hi-C data. *bioRxiv,* 122473 (2017).

25. Abbas, A. *et al.* Integrating Hi-C and FISH data for modeling of the 3D organization of chromosomes. *Nature Communications* **10,** 2049 (2019).

26. Caudai, C., Salerno, E., Zoppe, M., Merelli, I. & Tonazzini, A. ChromStruct 4: A Python Code to Estimate the Chromatin Structure from Hi-C Data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **PP,** 1–1. ISSN: 1545-5963 (2018).

27. Caudai, C., Salerno, E., Zoppe, M. & Tonazzini, A. Estimation of the Spatial Chromatin Structure Based on a Multiresolution Bead-Chain Model. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **PP,** 1–1. ISSN: 1545-5963 (2018).

28. Oluwadare, O., Zhang, Y. & Cheng, J. A maximum likelihood algorithm for reconstructing 3D structures of human chromosomes from chromosomal contact data. *BMC Genomics* **19,** 161 (2018).

29. Djekidel, M. N., Wang, M., Zhang, M. Q. & Gao, J. HiC-3DViewer: a new tool to visualize Hi-C data in 3D space. *Quantitative Biology* **5,** 183–190. ISSN: 2095-4689 (2017).

30. Kellerer, A. M. & Rossi, H. H. A generalized formulation of dual radiation action. *Radiation research* **178.** ISSN: 0033-7587 (1978).

31. Emami, B. *et al.* Tolerance of normal tissue to therapeutic irradiation. *International Journal of Radiation Oncology*Biology*Physics* **21,** 109–122. ISSN: 0360-3016 (1991).

32. Mesbahi, A. & Oladghaffari, M. An Overview on the Clinical Application of Radiobiological Modeling in Radiation Therapy of Cancer. *International Journal of Radiology & Radiation Therapy* **2** (2017).

33. Steel, G. G., McMillan, T. J. & Peacock, J. H. The 5Rs of Radiobiology. *International Journal of Radiation Biology* **56,** 1045–1048. ISSN: 0955-3002 (1989).

34. Henthorn, N. *et al.* Nanodosimetric Simulation of Direct Ion-Induced DNA Damage Using Different Chromatin Geometry Models. *Radiation Research* **188.** ISSN: 0033-7587 (2017).

35. McNamara, A. *et al.* Validation of the radiobiology toolkit TOPAS-nBio in simple DNA geometries. *Physica Medica* **33,** 207–215. ISSN: 1120-1797 (2017).

36. Friedland, W., Jacob, P. & Kundrát, P. Mechanistic simulation of radiation damage to DNA and its repair: on the track towards systems radiation biology modelling. *Radiation protection dosimetry* **143,** 542–8. ISSN: 0144-8420 (2010).

37. Friedland, W., Dingfelder, M., Kundrát, P. & Jacob, P. Track structures, DNA targets and radiation effects in the biophysical Monte Carlo simulation code PARTRAC. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **711,** 28–40. ISSN: 0027-5107 (2011).

38. Paganetti, H. Relative biological effectiveness (RBE) values for proton beam therapy. Variations as a function of biological endpoint, dose, and linear energy transfer. *Physics in Medicine and Biology* **59,** R419–R472. ISSN: 0031-9155 (2014).

39. Underwood, T. & Paganetti, H. Variable Proton Relative Biological Effectiveness: How Do We Move Forward? *International Journal of Radiation Oncology*Biology*Physics* **95,** 56–58. ISSN: 0360-3016 (2016).

40. Schuemann, J. *et al.* A New Standard DNA Damage (SDD) Data Format. *Radiation Research.* ISSN: 0033-7587 (2018).

41. Ingram, S. P. *et al.* Mechanistic modelling supports entwined rather than exclusively competitive DNA double-strand break repair pathway. *Scientific Reports* **9,** 6359. ISSN: 2045-2322 (2019).

42. McMahon, S. J., McNamara, A. L., Schuemann, J., Paganetti, H. & Prise, K. M. A general mechanistic model enables predictions of the biological effectiveness of different qualities of radiation. *Scientific Reports* **7,** 10790. ISSN: 2045-2322 (2017).

43. Friedland, W., Jacob, P. & Kundrát, P. Stochastic Simulation of DNA Double-Strand Break Repair by Non-homologous End Joining Based on Track Structure Calculations. *Radiation Research* **181,** 677–688. ISSN: 0033-7587 (2010).

44. INCERTI, S. *et al.* THE GEANT4-DNA PROJECT. *International Journal of Modeling, Simulation, and Scientific Computing* **01,** 157–178. ISSN: 1793-9623 (2010).

45. Schuemann, J. *et al.* TOPAS-nBio: An Extension to the TOPAS Simulation Toolkit for Cellular and Sub-cellular Radiobiology. *Radiation research* **191,** 125. ISSN: 0033-7587 (2019).

46. Wu, H., George, K., Kawata, T., Willingham, V. & Cucinotta, F. A. Comparison of F Ratios Generated from Interphase and Metaphase Chromosome Damage Induced by High Doses of Low- and High-LET Radiation. *Radiation Research* **155,** 57–62. https://doi.org/10.1667/0033-7587(2001)155[0057:COFRGF]2.0.CO;2 (2001).

47. Cornforth, M., Shuryak, I. & Loucas, B. Lethal and nonlethal chromosome aberrations by gamma rays and heavy ions: a cytogenetic perspective on dose fractionation in hadron radiotherapy. *Translational Cancer Research* **6,** S769–S778. ISSN: 2218-676X (2017).

48. Edens, L. J., White, K. H., Jevtic, P., Li, X. & Levy, D. L. Nuclear size regulation: from single cells to development and disease. *Trends in Cell Biology* **23,** 151–159. ISSN: 0962-8924 (2013).

49. Gillooly, J. F., Hein, A. & Damiani, R. Nuclear DNA Content Varies with Cell Size across Human Cell Types. *Cold Spring Harbor Perspectives in Biology* **7,** a019091 (2015).

50. Li, Y. *et al.* Moving Cell Boundaries Drive Nuclear Shaping during Cell Spreading. *Biophysical Journal* **109,** 670–686. ISSN: 0006-3495 (2015).

51. Bodgi, L. *et al.* Assessing Radiosensitivity of Bladder Cancer in vitro: A 2D vs. 3D Approach. *Frontiers in Oncology* **9,** 153 (2019).

52. Gomez-Roman, N., Stevenson, K., Gilmour, L., Hamilton, G. & Chalmers, A. J. A novel 3D human glioblastoma cell culture system for modeling drug and radiation responses. *Neuro-Oncology* **19,** 229–241. ISSN: 1522-8517 (2017).

53. Stevens, T. J. *et al.* 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544,** 59–64. ISSN: 0028-0836 (2017).

54. Ramani, V. *et al.* Sci-Hi-C: a single-cell Hi-C method for mapping 3D genome organization in large number of single cells. *Methods* **170,** 61–68. ISSN: 1046-2023 (2019).

55. Hufnagl, A. *et al.* The link between cell-cycle dependent radiosensitivity and repair pathways: A model based on the local, sister-chromatid conformation dependent switch between NHEJ and HR. *DNA Repair* **27,** 28–39. ISSN: 1568-7864 (2015).

56. Krietenstein, N. *et al.* Ultrastructural Details of Mammalian Chromosome Architecture. *Molecular Cell.* ISSN: 1097-2765 (2020).

57. Henthorn, N. *et al.* In Silico Non-Homologous End Joining Following Ion Induced DNA Double Strand Breaks Predicts That Repair Fidelity Depends on Break Density. *Scientific Reports* **8,** 2654 (2018).

58. Soutoglou, E. *et al.* Positional stability of single double-strand breaks in mammalian cells. *Nature Cell Biology* **9,** 675–682. ISSN: 1465-7392 (2007).

59. Lucas, J. S., Zhang, Y., Dudko, O. K. & Murre, C. 3D Trajectories Adopted by Coding and Regulatory DNA Elements: First-Passage Times for Genomic Interactions. *Cell* **158,** 339–352. ISSN: 0092-8674 (2014).

60. Sanders, J. T. *et al.* Radiation-Induced DNA Damage and Repair Effects on 3D Genome Organization. *bioRxiv,* 740704 (2019).

61. Yard, B. D. *et al.* A genetic basis for the variation in the vulnerability of cancer to DNA damage. *Nature Communications* **7,** 11428 (2016).

62. Taberlay, P. C. *et al.* Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. *Genome Research* **26,** 719–731. ISSN: 1088-9051 (2016).

63. Jia, R., Chai, P., Zhang, H. & Fan, X. Novel insights into chromosomal conformations in cancer. *Molecular Cancer* **16,** 173 (2017).

64. Díaz, N. *et al.* Chromatin conformation analysis of primary patient tissue using a low input Hi-C method. *Nature Communications* **9,** 4938 (2018).

65. Paulsen, J., Ali, T. M. L. & Collas, P. Computational 3D genome modeling using Chrom3D. *Nature Protocols* **13,** 1137. ISSN: 1750-2799 (2018).

66. Steensel, B. v. & Belmont, A. S. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169,** 780–791. ISSN: 0092-8674 (2017).

67. Lund, E., Oldenburg, A. R. & Collas, P. Enriched domain detector: a program for detection of wide genomic enrichment domains robust against local variations. *Nucleic Acids Research* **42,** e92–e92. ISSN: 0305-1048 (2014).

68. Rao, S. S. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159,** 1665–1680. ISSN: 0092-8674 (2014).

69. Henthorn, N. T. *et al.* Clinically relevant nanodosimetric simulation of DNA damage complexity from photons and protons. *RSC Advances* **9,** 6845–6858 (2019).

70. Pettersen, E. F. *et al.* UCSF Chimera – A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25,** 1605–1612. ISSN: 0192-8651 (2004).

71. Agostinelli, S. *et al.* Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506,** 250–303. ISSN: 0168-9002 (2003).

72. Meylan, S. *et al.* Simulation of early DNA damage after the irradiation of a fibroblast cell nucleus using Geant4-DNA. *Scientific Reports* **7,** 11923. ISSN: 2045-2322 (2017).

73. Friedland, W., Jacob, P., Bernhardt, P., Paretzke, H. G. & Dingfelder, M. Simulation of DNA damage after proton irradiation. *Radiation research* **159,** 401–10. ISSN: 0033-7587 (2003).

74. Sanche, L. Low-Energy Electron Damage to DNA and its Basic Constituents. *Physica Scripta* **68,** C108–C112. ISSN: 0031-8949 (2003).

75. Jafari-Mamaghani, M., Andersson, M. & Krieger, P. Spatial Point Pattern Analysis of Neurons Using Ripley's K-Function in 3D. *Frontiers in Neuroinformatics* **4,** 9 (2010).

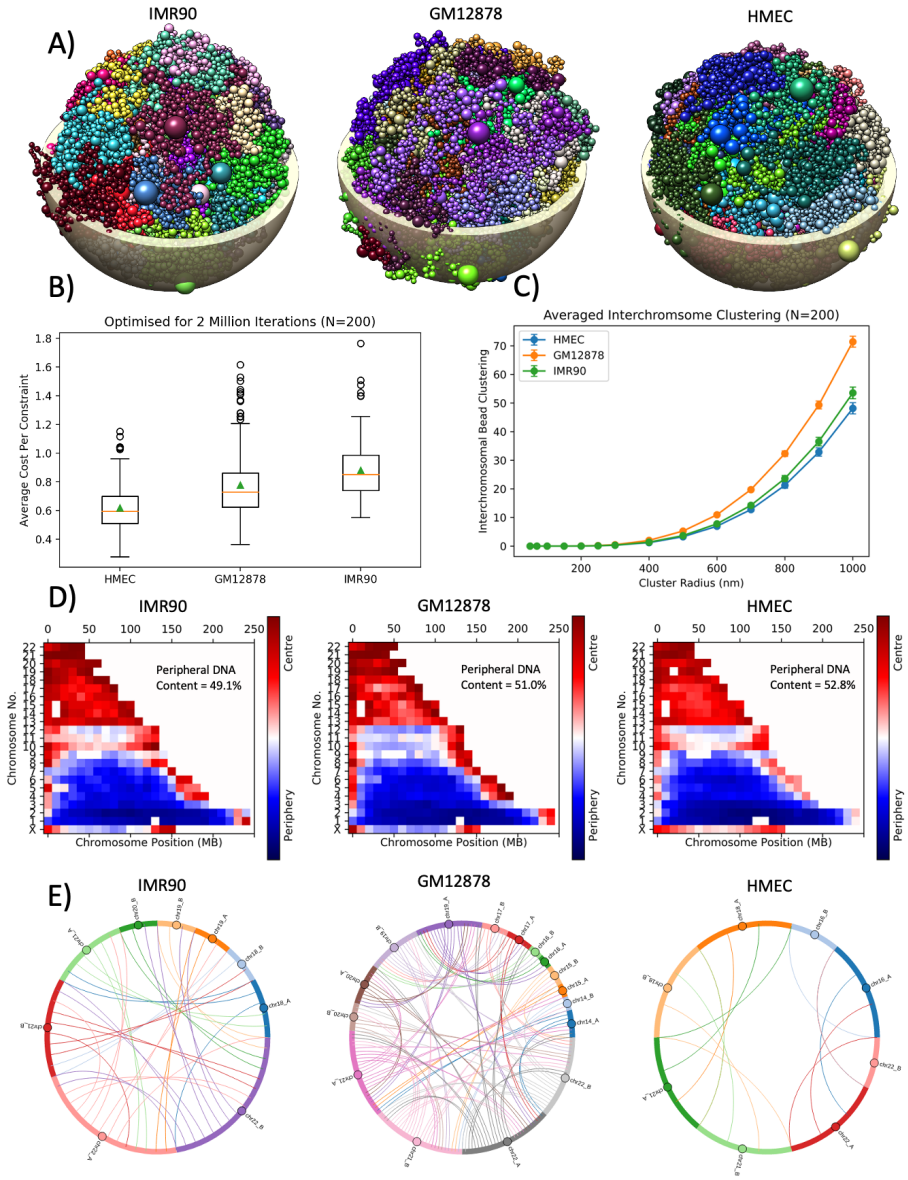

Figure 4.2: **Geometrical properties of the inferred 3D spatial chromatin arrangement.** **A)** rendered 3D example geometries for different cell-type Hi-C data. Each colour represents different chromosomes and the bead size represents the amount of DNA content. Rendered nuclear shell is to give an idea of scale for a 5 μm radius spherical nucleus. **B)** evaluation of G-NOME's ability to optimise the geometries of different Hi-C datasets. The cost function has been normalised to the total number of constraints which differs between cell types. Orange line represents the median and green triangle is the mean of the distribution. **C)** The resultant bead clustering for an increasing cluster inclusion radius. **D)** chromatin positioning within the nucleus for the different cell lines. Beads are scored based on being within the central or peripheral half of the nuclear sphere and is averaged over a 10 Mbp bin. The overall percentage of DNA content placed in the periphery is displayed within each plot. **E)** each line represents chromosomes that have interchromosomal beads within a 500 nm radius of one another. Lines are only displayed for the chromosomes that share proximal beads in at least 50% of the examined geometries. The data shown is averaged for 200 inferred geometries for each of the cell types. Error bars are the standard deviation of 200 geometries.

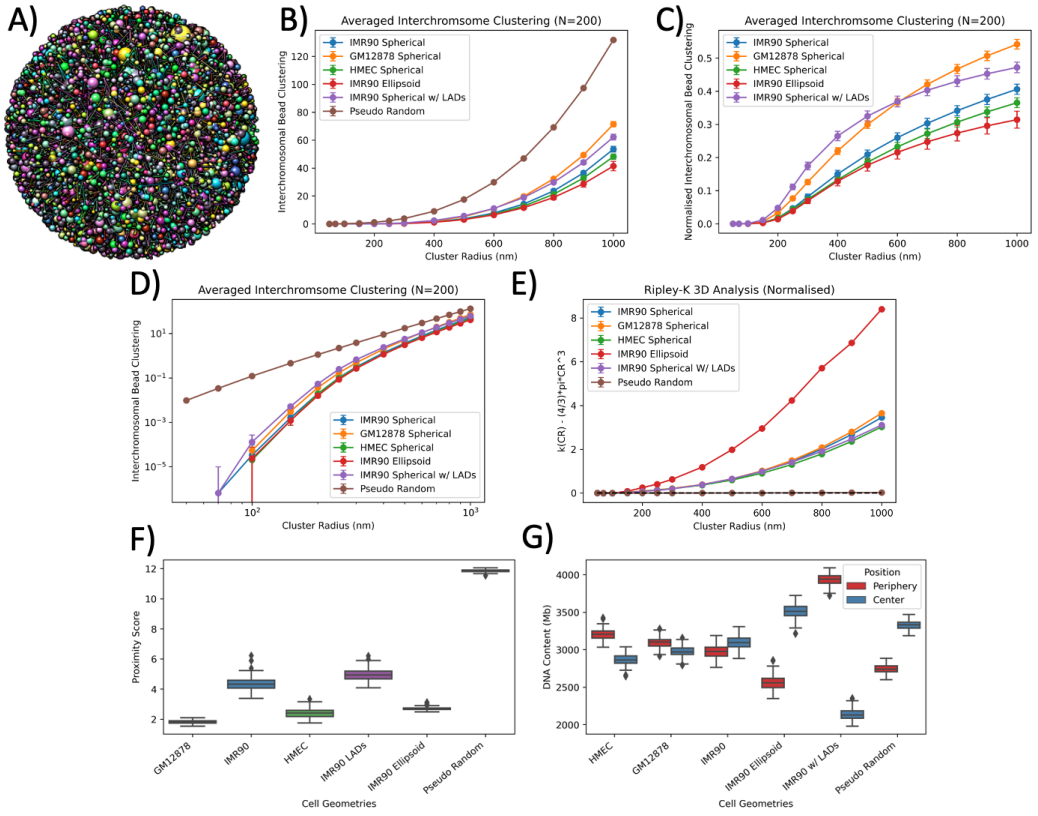

Figure 4.3: **Geometric comparison of solved Hi-C geometries and pseudo random geometries. A)** visualised example of a pseudo random geometry, beads are placed randomly within the restricted nuclear volume with no restriction of neighbouring beads. **B)** averaged interchromosomal bead clustering for all Hi-C cell-types and variants including the corresponding pseudo random values plotted on a linear-linear scale. **C)** normalised values of interchromosomal bead clustering for all Hi-C cell-types and variants using the corresponding pseudo random value as the normalisation parameter. **D)** same results as **B)** but on a log-log scale. **E)** Normalised 3D Ripley-K function (Eq. (4.1)). The normalisation results in complete spatial randomness giving a y-value=0 (shown by the dotted black line) for all cluster radii. Edge correction is applied for as $1/V_s$, where V_s is the fraction of overlapping volume of the cluster radius and the nuclear radius. All values are the average of 200 inferred geometries and error bars are the standard deviation. **F)** proximity score (lower value indicates a better optimisation of the contact constraints), which is a measure of the average euclidean distance constraint placed on TADs being proximal to one another from the Hi-C data. **G)** Spatial distribution of DNA content within the inferred geometries being place in either the central of peripheral half for the nuclear volume.

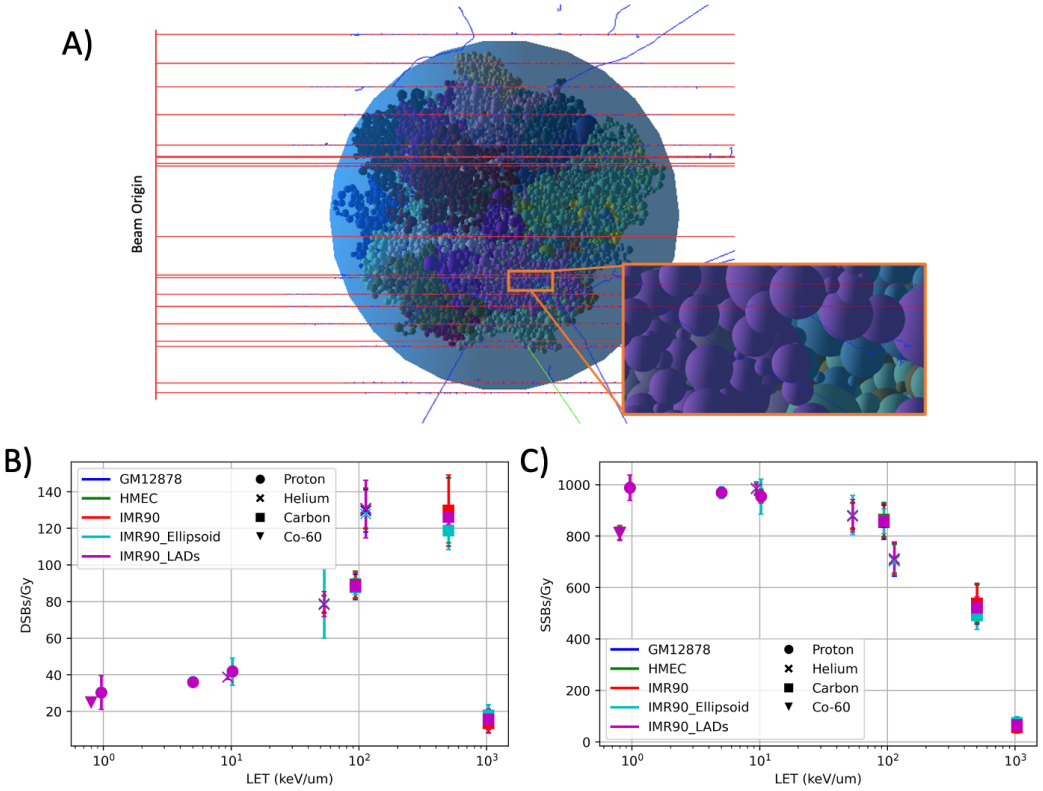

Figure 4.4: **Geant4 DNA damage simulation results. A)** visualisation of Hi-C geometry within Geant4-DNA being irradiated by a primary proton beam (red), generated secondary electrons (blue) and emergent gamma-rays from excitation (green). **B)** damage yields DSBs and **C)** damage yields SSBs per Gy of dose from the irradiation of Co-60 photons, protons, helium- and carbon-ions. Error bars are the standard deviation between the 200 geometries, where each geometry has 50 independent exposures.

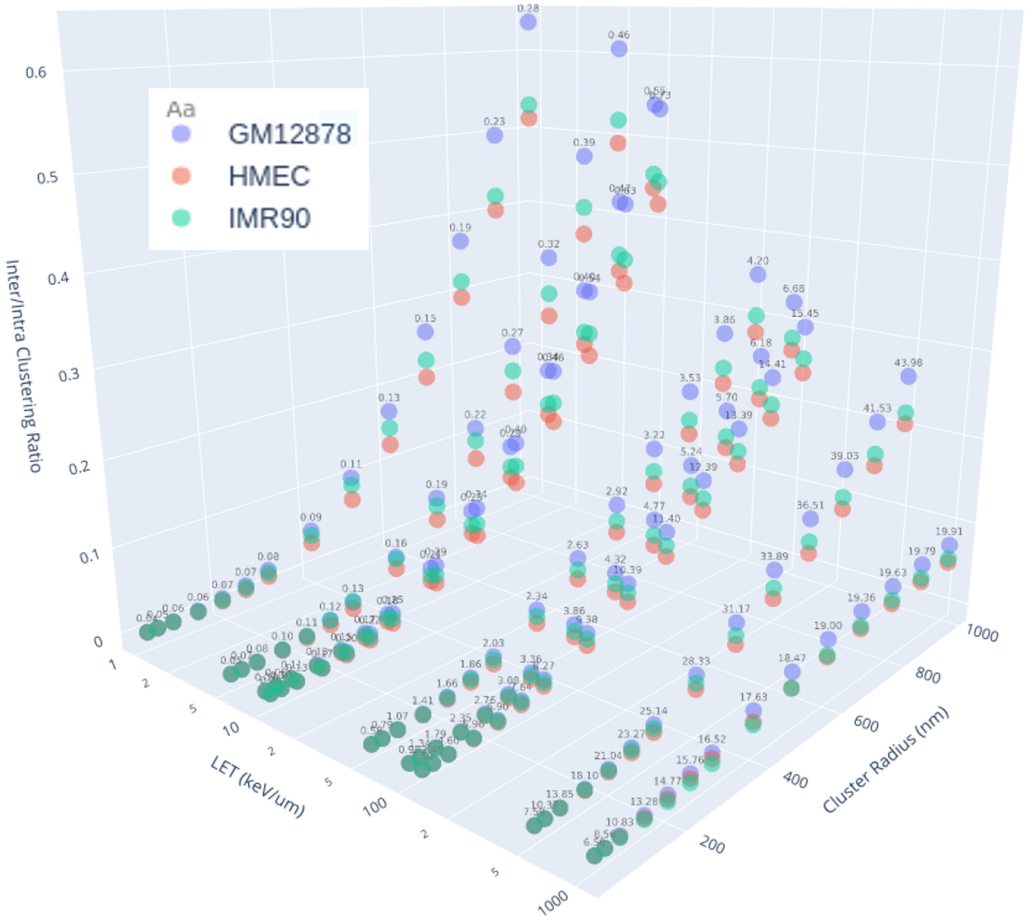

Figure 4.5: **Ratio of inter- and intra-chromosome DSB clustering (inter/intra CD Ratio) for a range of simulated cell-types, LET (keV/μm) and cluster radius (nm).** A lower and higher ratio indicates increased and decreased spatially clustered breaks respectively. Values of total DSB clustering are given as the floating numbers above points and are representative across cell-types. Interactive versions of this plot along with the corresponding total, inter- and intra-chromosomal DSB clustering graphs are available in S1-S8 Files. The 2D version of this plot for each fixed LET value is in S10 Fig.

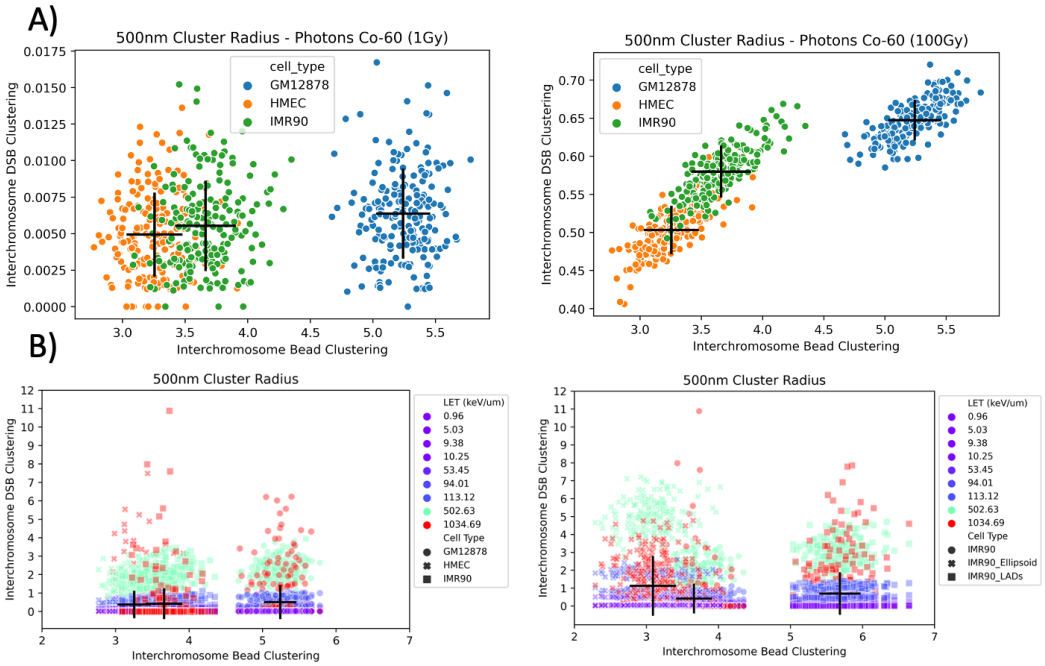

Figure 4.6: **Evaluation of interchromosomal geometric bead clustering and subsequent damage clustering. A)** relationship between interchromosomal bead clustering and interchromosomal DSB clustering for Co-60 photons irradiation at 1 Gy (left) and 100 Gy (right) at a 500 nm cluster radius. Black lines represent the standard deviation of each metric with the intersection corresponding to the mean value, each point represents a single G-NOME solved nucleus. **B)** same relationship of bead and DSB interchromosomal clustering, but for proton, helium- and carbon-ions of varying LET at cluster radii of 500nm with different cell-types (left) and cell variants (right).

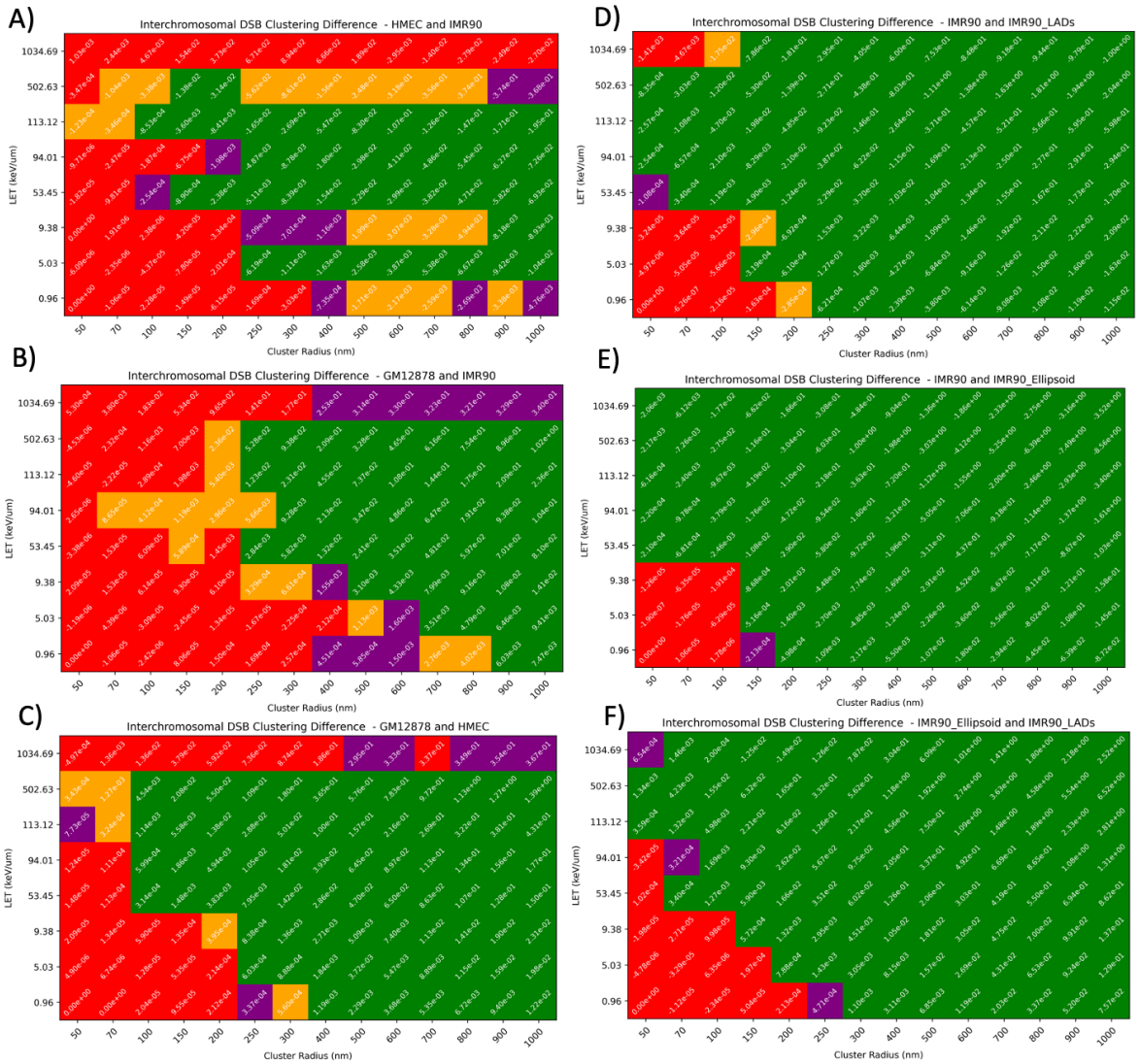

Figure 4.7: **Examination of statistically significant change in interchromosomal DSB clustering between cell-types (A - C)) and variants (D) - F)).** The values of the heat maps is the difference in the average interchromosomal clustering between the two cell types. The heat maps colouring shows the false discovery rate adjusted P-values from the Kolmogorov–Smirnov two-sided test for the full range of LET (keV/mum) and cluster radius (nm). The null hypothesis of the test is that two independent samples are drawn from the same continuous distribution. Colour codeing for adjusted P-values at varying thresholds: red ($P > 0.05$), purple ($0.05 > P > 0.01$), yellow ($0.01 > P > 0.001$) and green ($P < 0.001$). In this case, distributions with adjusted P-values < 0.05 are considered as statistically significantly different. Each of the tested distributions had 200 geometries per cell-type or variant group. Each value of interchromosomal DSB cluster density is an averaged result which comes from 50 independent 1Gy exposures at the listed LET and cluster radius values.

Chapter 5

A computational approach to quantifying miscounting of radiation-induced double-strand break immunofluorescent foci

Samuel P Ingram[1,2,*], John W Warmenhoven[1,3], Nicholas T Henthorn[1,3], A L Chadwick[1,3], E Santina[1,3], S J McMahon[4], J Schuemann[5], Norman F Kirkby[1,3], Ranald I Mackay[2,1], Karen J Kirkby[1,3], Michael J Merchant[1,3],

1 Division of Cancer Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Oxford Rd, Manchester M13 9PL, United Kingdom.
2 Christie Medical Physics and Engineering, The Christie NHS Foundation Trust, Wilmslow Rd, Manchester M20 4BX, United Kingdom.
3 The Christie NHS Foundation Trust, Manchester Academic Health Science Centre, Wilmslow Rd, Manchester M20 4BX United Kingdom.
4 Patrick G Johnston Centre for Cancer Research, Queens University Belfast, 97 Lisburn Rd, Belfast BT9 7AE, United Kingdom.
5 Massachusetts General Hospital & Harvard Medical School, Department of Radiation Oncology, 30 Fruit Street Boston, MA 02114, United States of America.

* samuel.ingram@postgrad.manchester.ac.uk

The paper is presented as the draft version as of 21st September 2021. Slight modifications to formatting for consistency.

5.1 abstract

Immunofluorescent tagging of DNA double-strand break (DSB) markers, such as γ-H2AX and other DSB repair proteins, are powerful tools in understanding biological consequences following irradiation. However, whilst the technique is widespread, there are many uncertainties related to its ability to resolve and reliably deduce the number of foci when counting using microscopy. We present a new tool for simulating radiation-induced foci and evaluate microscope performance within *in silico* immunofluorescent images. Simulations of the DSB distributions were generated using Monte Carlo track-structure simulation. For each DSB distribution, a corresponding DNA repair process was modelled and the un-repaired DSBs were recorded at several time points. Corresponding microscopy images for both a DSB and (γ-H2AX) fluorescent marker were generated and compared for different microscopes, radiation types and doses. Statistically significant differences in miscounting were found across most of the tested scenarios. These inconsistencies were propagated through to repair kinetics where there was a perceived change between radiation-types. These changes did not reflect the underlying repair rate and were caused by inconsistencies in foci counting. We conclude that these underlying uncertainties must be considered when analysing images of DNA damage markers to ensure differences observed are real and are not caused by non-systematic miscounting.

5.2 Introduction

The manner in which radiation can damage living cells has resulted in the scientific interest of the driving mechanisms of radiation-induced cytotoxicity. In early experiments, interactions of radiation within the nucleus were identified as the primary drivers for radiation-induced cell death [1]. The DNA was found to be the sensitive target of the nucleus, this is predominately driven by the way radiation imparts energy within the molecular structure of the DNA causing structural damage. This DNA damage can manifest itself in various forms, such as base lesions, single- and double-strand breaks in the sugar-phosphate backbone [2]. The type of DNA damage that is formed has varying corresponding cellular lethality, with DNA double-strand breaks (DSBs) being one of the most toxic [3]. Consequently this has caused a large amounts of interest in using DNA DSB induction as an indication of the cytotoxic characteristics of different radiation properties, such as dose, radiation particle and linear energy transfer (LET).

DNA damage response (DDR) signalling is a pivotal part of how adept cells are at preserving function following irradiation. It was discovered that the histone H2AX became phosphorylated (γ-H2AX) following the induction of DNA DSBs [4]. The γ-H2AX assay is a sensitive molecular marker for the building of an understanding of the initial DSB yields following irradiation and its subsequent DNA repair. This information has played a vital part in developing our current understanding of the mechanisms of DNA DSB formation [5] and how different radiation properties have different biological consequence [6–9]. Over time other key molecular targets, such as 53BP1, MDC1, ATM and proteins which form the MRN complex [10], have been sought out as alternative immunofluorescent markers of DNA DSB. Each molecular target has its role to play in the DDR, which can result in differences when inferring underlying mechanisms from them [11]. There has been an increased interest in using the DNA DSB markers in translational cancer research as a possible method of predicting patient-specific response and further optimising patient treatment [12–14]. However, whilst these novel approaches are enticing, with any interest in clinical application the limitations of the technique must be widely explored.

The developing understanding of the relationship between radiation and the cellular response has been driven by the applicability of radiotherapy in a large proportion of cancer treatment regimens. The efficacy of radiotherapy is dictated by the physical ability to sculpt the radiation dose to the tumour whilst sparing healthy tissue. The unavoidable radiation that is delivered to healthy tissue, specifically organs at risk (e.g. spinal cord), is what limits the maximum radiation dose that can be given to the tumour and therefore the tumour control probability from a treatment [15]. The quantitative measurement between the radiation that can be delivered to the tumour whilst being safe for the organs at risk is known as the therapeutic ratio. To utilise the experimental understanding of therapeutic ratio at a clinical level it is important to develop corresponding models which can further generalise predictions, to cover both the magnitude and spatial distributions of radiation doses delivered in treatment plans. However, these models are limited by the experimental uncertainties and

the ability to generalise to populations of patients which are inherently biologically diverse [16]. Therefore, efforts have continuously being made in the development of broader and/or more detailed models of radiation response to overcome these limitations and better define the therapeutic ratio, ensuring treatment plans are optimal. There are emerging attempts to model radiation response mechanistically, this increases the complexity but aims to fully capture the mechanisms at play increasing possibility of model generalisation. A subset of these developing models, include an explicit description of the induced DNA damage following irradiation [17–19] and some include aspects of the DDR (e.g. DNA repair) [20–22]. These models enable us to simulate the spatial distribution of DNA DSBs using Monte Carlo frameworks such as Geant4-DNA [23] and TOPAS-nBio [24] for a range of incident radiation setups. The information provided by mechanistic models can be coupled with our understanding of immunofluorescence and microscopy to start to explore the limitations of the experimental technique [25, 26]

In this study, we use radiation track-structure simulations [27] to produce representative spatial distributions of DNA DSBs and a bi-exponential repair model to predict the distribution at various time points. We then developed software called PyFoci to generate computational microscope images and look at the deviation in the foci detected and the known number of DSBs in the simulation. This software builds on previously developed models [25, 26, 28] through the addition of several mechanisms which improve it's likeness to the experimental setting. PyFoci is able to both emulate the fluorescent foci tagging being either direct (DSB markers) or indirect (γ-H2AX markers) to mark the spatial positions of the fluorescent antibodies. Indirect markers are controlled by explicit modelling of the chromatin conformation [29] and utilises recent insights of how H2AX becomes activated following DNA damage [30]. This is then convolved with the blurring caused by visualising sub-resolution fluorescent antibodies (approximately 10 nm) [31] for specific models of a microscope at varying magnification and numerical aperture (NA). The emulated microscope images can be analysed using standard foci counting techniques, allowing for the quantification of miscounting between the actual number of DSBs present within the simulation and number of DSB foci counted on the emulated microscope image. The amount of miscounting has been evaluated for a range of radiation types, including linear energy transfers (LETs) and doses to highlight the difficulties of comparing experimental DSB fluorescence foci across different radiation properties. We show that the level of miscounting significantly changes in the majority of tested radiation types and doses, which subsequently leads to perceived changes in repair kinetics, even when no such differences are present in the simulations. Therefore, we propose PyFoci as a tool for experimentalists to enhance their studies by quantifying these often omitted uncertainties and reinforce caution when trying to make mechanistic conclusions based on fluorescent foci imaging. We also conclude that this approach should be used to adequately compare simulated levels of DNA double-strand break damage and the subsequent repair in mechanistic models with experimental data. The creation of a computational microscope image can be seen as an intermediate step to incorporate the deviation between absolute DNA double-strand breaks and the number which would be represented via a florescent foci technique.

5.3 Materials and Methods

Modelling DNA Double-Strand Break Damage Distributions and DNA Repair Kinetics

To genrate a model of the distribution of the DNA damage radiation was simulated through a polymer bead model of the organised genome. The genome organisation is derived from modelling the topologically associated domain (TAD) contact probabilities of Hi-C data as a 3D polymer with each bead representing a single TAD. To get a spread of representative genome structures the Monte Carlo Markov-Chain model G-NOME [29] was used to produce 200 different IMR90 (normal lung fibroblast) geometries from published Hi-C data by Rao et al., [32] (GEO Accession GSE63525). Each geometry was solved as a flattened ellipsoid (11.8x11.8x1.0 μm radii) to be representative of cells which have been plated for microscopy [33].

To model proton irradiation, the track structure of the incident protons was simulated using the Monte Carlo toolkit Geant4 (v10.5.1) [34] with the default Geant4-DNA physics list [23]. Each track is simulated as a series of interaction limited steps, with each step accounting for the changes in energy and trajectory of both the primary and secondary particles, resulting in spatial information of energy deposition from the incident protons. To convert energy depositions into DNA strand breaks several conditions must be met. Firstly, the energy depositions must have occurred in the genome represented by the polymer beads. Secondly, a spatial sampling of 14.1% is applied to the bead to accounting for DNA sparsity within a TAD and reproduces DSB yields seen in our previous work [18]. Finally, an energy-based probability of break induction (0 at 5eV to 1 at 37.5eV) is applied for energy deposited in a backbone molecule, which is similar to other works [35] and incorporates the possibility of DNA damage occurring below the ionisation threshold of DNA [36, 37]. If all conditions are passed the energy deposition is accepted as a strand break and is randomly assigned to a strand of the double helix with equal probabilities. The equivalent position along the chromosome for the damage is randomly selected from the basepair range of the TAD the damage occurs in. Once all strand breaks are calculated from the simulated incident proton irradiation a clustering algorithm is used to distinguish which strand breaks are likely to form DSBs. The classification of DSB is given to strand breaks which are on the opposite strand and separated by 3.2 nm or less (equivalent to 10 bp).

To model photon irradiation, DSB induction is assumed to follow a Poisson distribution with an average of 30 DSBs/Gy. The damage is modelled through the same 200 Hi-C solved genome models as proton irradiation. Each DSB is assigned a chromosome at random, weighted by the chromosome size relative to the total genome size. The same process is applied to assigning a bead, which is weighted according to bead size. A random X, Y, Z position within the selected bead is also applied. All DNA damages are recorded in the Standard DNA Damage (SDD) format [38].

After DNA DSB induction we utilised a bi-exponential repair kinetics model which was applied across all radiation simulations to create a consistent repair rate (Figure 5.1). This was

uniformly applied so that only the difference caused by miscounting would be observable in the resultant repair kinetics. The bi-exponenial repair kinetics was a normalisation of the γ ray fit from Plante et al., [39] so it could be applied to any number of initial induced DSBs (Eq. (5.1)).

$$\lceil N(t) \rceil = N_0(a_1 e^{-t/\tau_1} + a_2 e^{-t/\tau_2}) \tag{5.1}$$

Where, $\lceil N(t) \rceil$ is the ceiling integer of the number of DSBs at a given time t hours and N_0 the initial number of DSBs. The four constants of fit are $a_1 = 0.711$, $a_2 = 0.289$, $\tau_1 = 1.54$ hours, $\tau_2 = 10$ hours.

Generating PyFoci Microscope Images

A schematic overview of the whole process to create the computational microscope images are shown in Figure 5.1. The spatial distribution of the damage as a function of time after irradiation derived from the bi-exponential repair model. These files can be read by PyFoci which maps the DNA DSBs to the same resolution (pixel spacing) of the microscope being used. The microscope point spread function (PSF) is derived experimentally through measurement of sub-resolution fluorescent beads of a known size, allowing for the PSF to be computed (distilled). The DSB positions are used as a point source, due to the typical size of antibody molecules (approximately 10nm) [31] being smaller than the resolution of the microscope. The DSBs are then convolved with the microscope's PSF to emulate the intrinsic blurring of a light source from microscopy in 3D. This includes the amount of out-of-plane fluorescence which is not blocked by the confocal pinhole. The light is then scored in a single 2D matrix at a set z-value to emulate a z-slice from a confocal microscope. The resultant matrix can then be viewed as a computational microscope image. These steps are more representative of using a DSB-marker which directly attaches to the repair proteins (e.g. Ku70/80 or DNA-PKcs) found at break ends [40], rather than indirect markers such as γ-H2AX or 53BP1.

To evaluate indirect markers a histone based model was deployed. The number H2AX histones required for placement was based on assuming 10% of H2A histones being the H2AX variant [41], each nucleosome having two copies of the H2A histone and each nucleosome being 146 basepairs [42]. The histones were randomly placed within the polymer beads (which represent TADs) which had DSBs present. Utilising the findings of Arnould et al., [30] it was assumed that activation of H2AX was restricted within the TAD the DSB belonged to (Figure S1). The strength of the H2AX activation was also taken from Arnould et al., [30] and was fitted using a Cauchy-Lorentz distribution (Eq. (5.2)) based on the Chip-Seq measurements of γ-H2AX read counts as a function of distance in Mbp from DSB

Microscope Parameters (XY pixel spacing, Z pixel spacing, Numerical Aperture)					
	10x	**20x**	**40x**	**63x**	**100x**
Airyscan	0.13, 1.14, 0.45	0.059, 0.3, 0.8	0.035, 0.15, 1.3	0.033, 0.12, 1.4	0.028, 0.1, 1.46
gSTED	—	0.065, 0.3, 0.95*	0.055, 0.20, 1.1	0.042, 0.13, 1.4	0.44, 0.13, 1.4
Lowlight	—	0.21, 0.5, 0.5	0.11, 0.3, 0.75	—	0.043, 0.1, 1.45
MultiPhoton	0.20, 1.0, 0.3	0.54, 0.25, 0.95**	0.058, 0.2, 0.85	0.044, 0.15, 1.2	—
Phenix	—	0.3, 0.8, 1.0	0.15, 0.5, 1.1	0.095, 0.4, 1.15	—
STED	—	0.075, 0.4, 0.75	0.045, 0.15, 1.1	0.034, 0.1, 1.4	0.035, 0.1, 1.4

Table 5.1: Microscope Parameters - the pixel spacing for XY and Z (μm) that the resultant PyFoci microscope image uses. Along with the numerical aperture used when measuring the microscope PSF. Values given for all size microscopes at their respective magnifications. *23x magnification not 20x. **25x magnification not 20x. The STED microscope also has a PSF at 25x which has the following values (0.06,0.25,0.95) for XY pixel spacing, Z pixel spacing and Numerical Aperture respectively.

(Figure S2).

$$Activation = 0.23 + \frac{0.38}{1 + 4(\frac{Distance}{0.45})^2} \tag{5.2}$$

As sub-TAD organisation is not present within the model, so the effective basepair distance was calculated by the euclidean distance between the break and histone, where the conversion is based on the diameter of TAD bead divided by the genomic content of the same bead. Each histone was given a value of activation based on its proximity to the DSB which would represent the intensity of fluorescence. Nearby breaks which share histones to activate were assumed to activate the histone in an additive manner. This was carried out for all DSBs and was visualised in the same manner as the DSB-marker method within PyFoci. This allows the production of γ-H2AX-marker based images to compare to the DSB-marker images.

The microscopy simulation of both marker types assumes perfect efficiency of the fluorescent antibody to attach to its molecular target. In total six microscopes were used each with several magnifications. The microscopes utilised were: Carl Zeiss LSM880 with Airyscan fast mode (Airyscan), Leica sp8 TCS inverted confocal with STED super-resolution mode (gSTED), Carl Zeiss Axiovert 200M (lowlight), Leica sp8 TCS upright confocal (MultiPhoton), Perkin Elmer Opera Phenix High Content Screening System (Phenix) and super resolution mode on gSTED system (STED). A PSF was measured for each microscope-magnification combination resulting in a total of 24 variations of microscope blurring. The PSFs are distilled and provided in the H5 hierarchical data format (*.h5) files which can be directly parsed into PyFoci. The pixel spacing and numerical aperture for each microscope combination are given in Table 5.1, with the full width half maximum (FWHM) of the central axis for each PSF is given in Table S1.

Automated Foci Counting on Generated Microscope Images

Computationally generated microscope images were generated for each of the two hundred solved nuclear geometries at sixteen radiation set-ups, twenty-three microscope-magnification

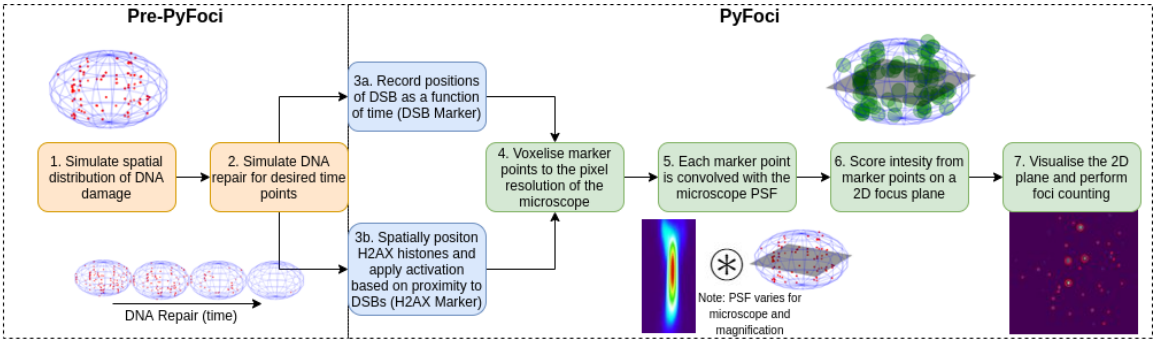

Figure 5.1: Schematic overview of creating computational microscope images using PyFoci. The generation of simulated DNA DSBs and repair (steps 1-2) is independent of PyFoci and can be interchanged with other models. Any model which can output DSB position as a function of time can use PyFoci to create computational microscope fluorescent foci images (for at least the DSB marker visualisation). The steps 3a and 3b represent the choice in visualisation approach, whilst DSB marker only requires information on the DSB spatial distribution, the γ-H2AX marker requires information on the chromosome geometry, namely the topologically associated domains the DSBs are created in. Steps 4-7 are the same regardless of the visualisation marker, the only difference is what biological object is being fluoresced (DSBs or γ-H2AX). Step 5 requires the microscopes PSF to be defined for the desired microscope and magnification under evaluation. Step 6 requires the user to define at what time point they wish to create the microscope image and at what z-axis slice. Step 7 creates the image and performs the foci counting on the produced image, this can be done with the in-built foci counter or can be exported to external foci counting software.

configurations and six time points (0 s, 15 minutes, 30 minutes, 2 hours, 6 hours, 24 hours) as a single z-slice centrally positioned in the cell nucleus. This resulted in a total of 441,600 computationally generated microscope images for both the DSB-marker and γ-H2AX-marker visualisation. Therefore, an automated foci counting approach was used to count the number of foci in all 883,200 images and compare the number of identified foci to the number of known breaks within the same slice. The automated foci detection was carried out using a Laplacian of Gaussian (LoG) technique found in the Python package scikit-learn (v0.23). The LoG technique relies on the user set minimum, maximum and number of sigma, along with a threshold value. The various sigma parameters control allowed size of detected foci and the threshold value determines the lower bound of the intensity required for detection. To be consistent in the foci detection the parameters were fixed for foci detection across all simulation set-ups within the same visualisation method (DSB-marker and γ-H2AX-Marker). Due to the intrinsic differences the is fluorescent component of the DSB-marker and γ-H2AX-marker images, there were two distinct sets of foci counting parameters for the visualisation methods. As the intensity values change drastically for different radiation parameters, the threshold was set as a percentage of the maximum intensity value of the computational microscope image. This maximum intensity value was calculated for each simulation set-up at the 15 minute time point and was set for all other time points, this was to ensure that the automated detection did not have a chance of increased foci detection at later time points. The parameters were optimised by matching the number of detected foci a subset of images from all the set-ups with a user manually counting foci. The LoG parameters used for the DSB-marker are as follows: $min\_sigma = 2, max\_sigma = 20,$ $num\_sigma = 1, threshold = 8\%$. The LoG parameters used for the γ-H2AX-marker are

as follows: $min\_sigma = 2$, $max\_sigma = 20$, $num\_sigma = 3$, $threshold = 4\%$.

5.4 Results

Comparison of Miscounting with DSB Markers

Through the simulation of the microscope image, it becomes possible to evaluate DNA DSB miscounting. This is possible as there is a known amount of simulated DSBs within a single confocal slice of the cell nucleus and this can be compared to the number of identified foci. As foci are commonly used as surrogates for identifying a DSB one can calculate the miscount as the difference between foci detected and the actual number of DSBs. Positive values of miscounting would indicate over-counting, most likely due to fluorescence of neighbouring slices being counted. Whereas, negative values of miscounting would indicate under-counting, which could be fluorescence from multiple DSBs not being distinguishable or the fluorescence being under the threshold value for the automated identification. The number of identified miscounts depends on the time point, radiation type, marker type and dose. To evaluate these relationships for the DSB marker visualisation a series of box-plots (Figure 5.2) compare the amount of miscounting across different time points, radiation types and doses.

As it is common in the literature to evaluate the differences in fluorescent foci between radiation types a series of Mann-Whitney tests were performed to identify a statistically significant difference in miscounting at each time point and dose. All tests were significant in simulations at 10 Gy with the difference between protons at 1.7 keV/μm and 7.15 keV/μm being not significant between 1 - 5 Gy. There are also some non-significant differences seen between low LET protons (keV/μm) and Cobalt-60 for 1 Gy at late ($>$6 hours) time points. The largest change between radiation types is with high-LET protons (27.95 keV/μm), there is a significant amount of under-counting, which increases with dose, that likely corresponds to DSBs becoming increasingly clustered and non-distinguishable. To get a perspective of the magnitude of these values the same graph has been presented as the percentage difference between foci detected and the number of actual DSBs in Figure S3, where the same distinct under-counting is seen for high-LET protons. A feature of examining percentage foci miscounting is that at later time points there is a tendency to see an increased proportion of over-counting due to the contribution of neighbouring slices outweighing the rate of actual DSBs within the slice being repaired. In some experiments researchers may be interested in how counted foci change as a function of dose, where each dose used is compared, this was also evaluated for statistically significant changes in miscounting (Figure S4). The majority of results demonstrated significant differences when attempting to compare different doses of the same radiation type with increasing foci miscount for increased dose. However, when compared to the percentage difference for dose comparison (Figure S5) it becomes apparent that whilst there is an increase in foci miscount for low LET radiations this does not translate to a distinct increase in percentage foci miscount.

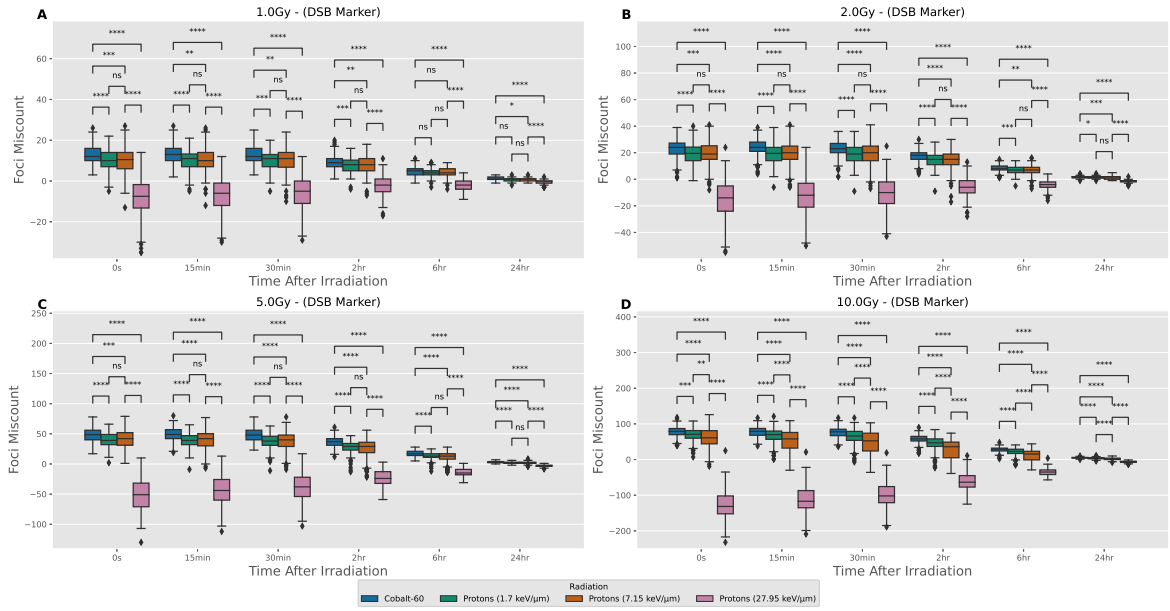

Figure 5.2: Distributions of miscounting between foci detected and the number of DSBs within the central microscope slice of the cell nucleus when using a **DSB marker of florescence**. Each panel is the same four radiation types and six time points being compared for different radiation doses, where A, B, C, D correspond to 1.0, 2.0, 5.0, 10.0 Gy respectively. Mann-Whitney test between different radiation types at each time point and dose to highlight statistically significant differences. P-values have been adjusted using the Bonferroni correction. Significance notation refers to the following thresholds: ns=P>0.05, *=P>0.01, **=P>1e-3, ***=P>1e-4, ****=P<1e-4. There are 200 samples for each box-plot. All microscope images have been emulated using the Airyscan 63x point spread function.

Comparison of Miscounting with γ-H2AX Markers

The γ-H2AX marker is evaluated in the same manner as the DSB marker for the same radiation types, time points and doses (Figure 5.3). There is a consistently high number of statistically significant differences in miscounting over the tested parameters. However, the amount of under-counting in high-LET protons is reduced at later time points, with several median values indicating there is over-counting of foci when compared to the DSB marker results. Whilst the amount of actual DSBs and the clustering is the same in both datasets the visualisation approach alters where the fluorescent light source is produced depending on the origin of the fluorescent signal. Within the γ-H2AX marker images, a single DSB results in multiple H2AX phosphorylation, each acting as a point of fluorescence, which in turn gives a larger area of fluorescence increasing the amount of contribution from breaks within neighbouring slices. Conversely, nearby DSBs within the same topologically associated domain can provide activation to the same H2AX histones producing slightly larger foci, which are non-distinguishable. The amount of foci miscount tends to show over-counting more frequently suggesting the light contribution from neighbouring slices being detected as foci is greater than the DSBs within the confocal slice being under-counted. With high-LET early time points centralising around a zero foci miscount it is thought that at this point the over-counting of foci fluorescence from neighbouring slice is of a similar magnitude to the under-counting of foci in the confocal slice which are not being counted as individual foci. This is supported when examining the percentage foci miscount, which

shows a similar centralisation around zero percentage foci miscount for high-LET early time points (Figure S6). When making dose comparisons (Figure S7) we see all scenarios which are significantly different. The corresponding analysis of the percentage foci miscount when comparing different doses suggest similar amounts of miscounting between doses (Figure S8).

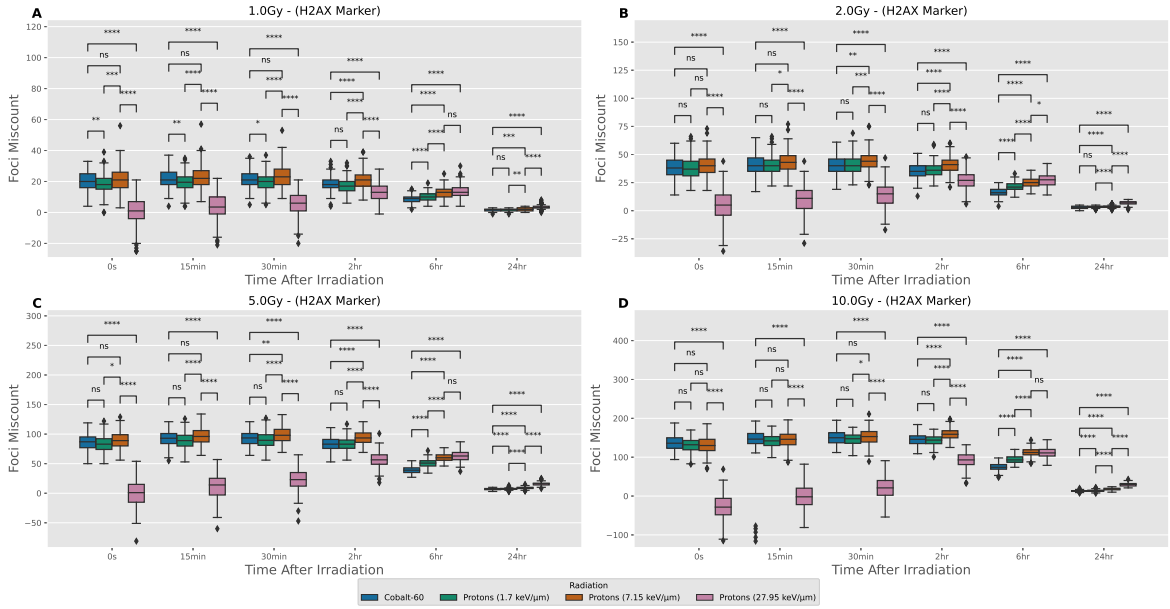

Figure 5.3: Distributions of miscounting between foci detected and the number of DSBs within the central microscope slice of the cell nucleus when using a γ**-H2AX marker of florescence**. Each panel is the same four radiation types and six time points being compared for different radiation doses, where A, B, C, D correspond to 1.0, 2.0, 5.0, 10.0 Gy respectively. Mann-Whitney test between different radiation types at each time point and dose to highlight statistically significant differences. P-values have been adjusted using the Bonferroni correction. Significance notation refers to the following thresholds: ns=P>0.05, *=P>0.01, **=P>1e-3, ***=P>1e-4, ****=P<1e-4. There are 200 samples for each box-plot. All microscope images have been emulated using the Airyscan 63x point spread function.

Implications for Repair Kinetics

Typically, DSB fluorescence foci would be analysed as a function of time to examine the repair kinetics. When comparing across radiation types it is common to normalise to the first time point to account for the differences in the total number of DSBs. The normalised comparison of the radiation types for the actual number of DSBs and the identified foci in the DSB marker and γ-H2AX marker images are presented in Figure 5.4 for a dose of 2 Gy and using the Airyscan 63x PSF. To evaluate the perceived effects of changes in repair kinetics that derive from miscounting alone the same repair rate was applied, in the form of the bi-exponential equation (Eq. (5.1)), to all radiation types. The differences within the DSB and γ-H2AX marker visualisations are given in Figure 5.4B and Figure 5.4C respectively. To better correspond to experimental set-ups the time normalisation point was chosen to be at 15 minutes as this would allow time for the DNA damage response to recruit necessary proteins that would be observable in fluorescence experiments. The differences in the visualised repair kinetics are due to varying miscounting at subsequent time points, this results

in in the identical repair kinetics from Figure 5.4A being observed as different when visualised via a DSB or γ-H2AX marker. The largest miscount for the DSB marker (Figure 5.2) and γ-H2AX marker (Figure 5.3) was seen for high-LET proton and this is propagated to the largest observed effects. When using a DSB marker (Figure 5.4B) high-LET protons DSBs were increasingly under-counted (Figure S3) which results in the perceived quickening of repair kinetics. Conversely, when using an γ-H2AX marker (Figure 5.4C) DSBs are initially under-counted due to overlaps, but become over-counted due to signal from other layers at later time points (Figure S6) resulting in the perceived slow down of repair kinetics. However, it should be acknowledged that increasingly research publications are avoiding normalisation as it is highly sensitive to deviations in the normalisation time point. The same plot without normalisation can be seen in Figure S9, whilst there are differences in the initial number of DSBs seen within the foci detection, these differences appear reduced when compared to the actual values. The overall, perceived faster and slower repair of high-LET protons is still visible for the DSB and γ-H2AX marker respectively.

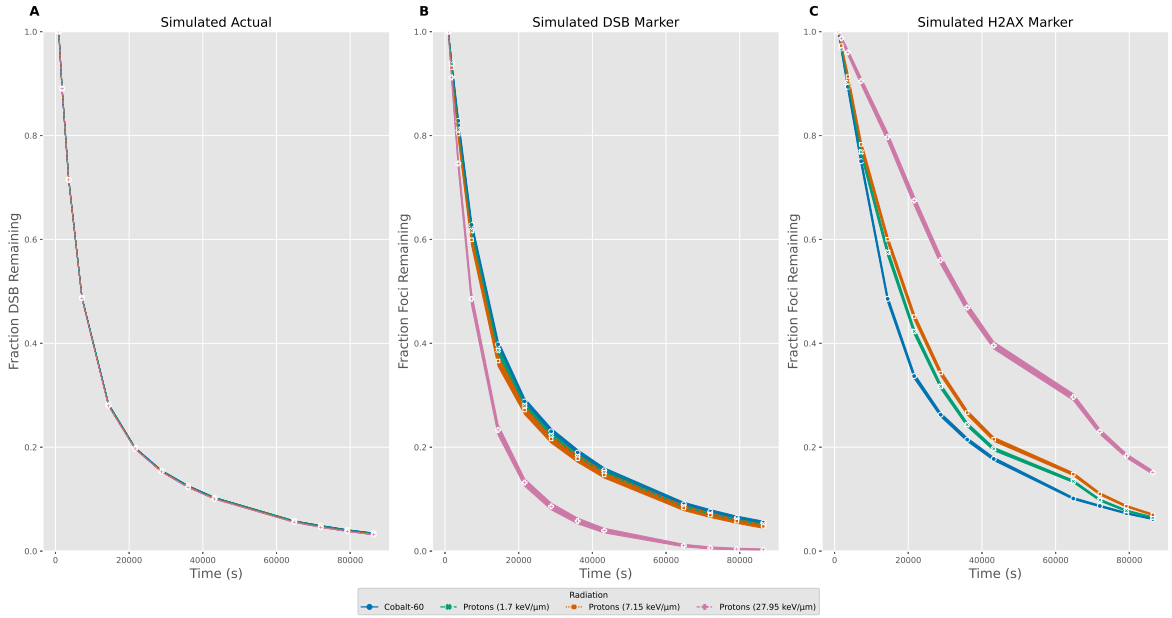

Figure 5.4: Repair kinetics comparison between virtual cells irradiated with four different incident radiation conditions. The fraction of DSB/Foci remaining have been normalised to the 15 minute time point to allow comparison of repair kinetics regardless of the total number of DSB breaks induced. A) shows the repair kinetics of the actual number of DSBs within the evaluated microscope slice. The repair kinetics applied using the bi-exponential repair model is kept constant across all radiation types to separate out the impact of miscounting alone. B) the repair kinetics when calculated from the foci detected in the DSB marker microscope images. C) the repair kinetics when calculated from the foci detected in the γ-H2AX marker microscope images. All microscope images have been emulated using the Airyscan 63x and the results shown are for 2 Gy of radiation.

Effects of Microscope Magnification

The emulation of the microscope visualisation is achieved by convolution with the experimentally derived PSF. This PSF is unique to each microscope and each magnification. The effects of the magnification on percentage foci miscount was evaluated for the Airyscan microscope for both the DSB and γ-H2AX marker visualisations at 2 Gy and 15 minutes

post-irradiation time point (Figure 5.5). To help give perspective for the percentage foci miscount the actual number of DSBs within the evaluated slice are provided within each box-plot as the white numeric value. The increased number of actual DSBs within a slice is a respective measure of how the larger z-slice resolution encompasses a larger volume of measurement. It is unlikely that different magnifications would be used in a single experiment, but in the event of comparing between experiments where different magnifications have been used, the quantification of miscounting may be of use. The percentage miscount is largely preserved across magnifications greater than 10x, with a significant undercounting at 10x. The Z-axis pixel spacing for 10x is 1.14 μm (Table 5.1) which is a considerable proportion of the 2 μm thickness of the simulated flattened cell nucleus causing significant spatial averaging increasing the likelihood of not distinguishing DSBs. Interestingly, the percentage foci miscount is similar between the other magnifications suggesting that normalised levels of foci may be comparable between different experiments where the same microscope, radiation and dose has been used, but different magnifications. To ensure this was not a feature of just the 15 minute time point the same analysis was run on the distribution of all time points combined (Figure S10) where the overall trend persists, but the spread in the distribution increases due to the DSB repair giving a wider range of DSBs to being visualised for each box-plot.

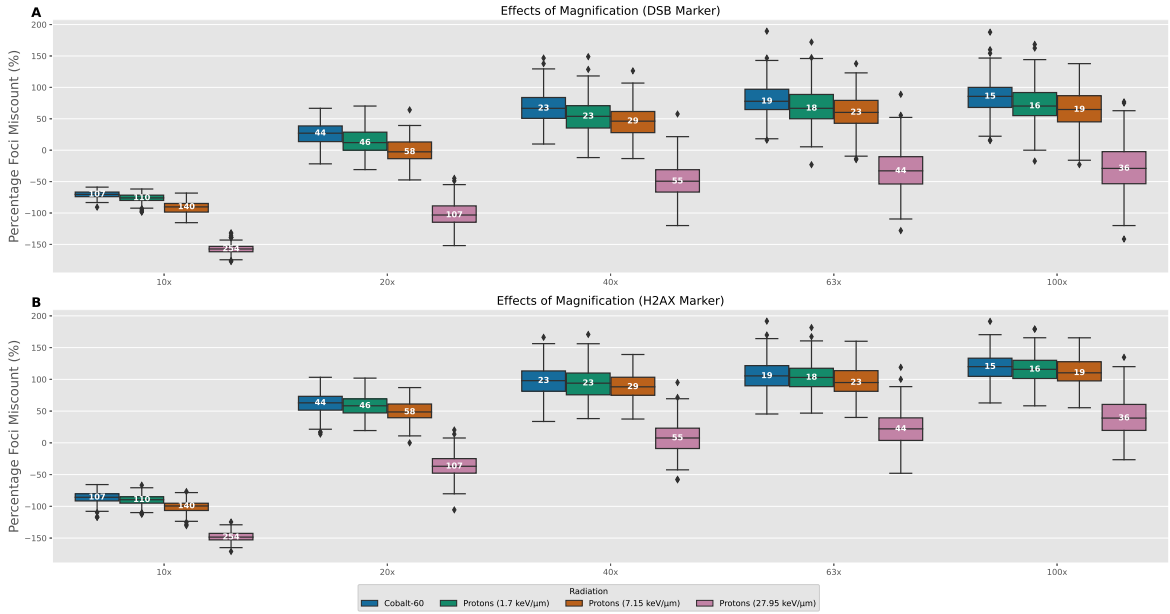

Figure 5.5: Comparison of percentage miscounting at different Airyscan magnifications. The value shown in white within each box-plot is the actual number of DSBs within the evaluated microscope slice. This aids in giving perspective on the amount of miscounting each percentage corresponds to. A) are the results when using a DSB marker. B) are the results when using an γ-H2AX marker. All microscope images are at 2 Gy dose and for the 15 minute time point. There are 200 samples for each box-plot.

Effects of Microscope Resolution

To evaluate the effects of different microscopes, including the various magnifications, the average percentage foci miscount was compared against voxel size across all twenty-three microscope/magnification configurations for 2 Gy dose (Figure 5.6). The voxel size is cal-

culated as the X*Y*Z pixel spacing which can be found in Table. 5.1, small voxel sizes are typical for higher resolution microscopes at high magnification. There appears to be a negative relationship between voxel size and percentage foci miscount, with a decreasing value with increasing voxel size. The change in percentage foci miscounting indicates optimal foci counting relies on a balance between the tendency to over-count at small voxel sizes and under-count at larger voxel sizes. Over-counting is due to neighbouring slices interference being larger at higher resolutions and the under-counting is the result of increased spatial averaging and a smaller influence of neighbouring slices.

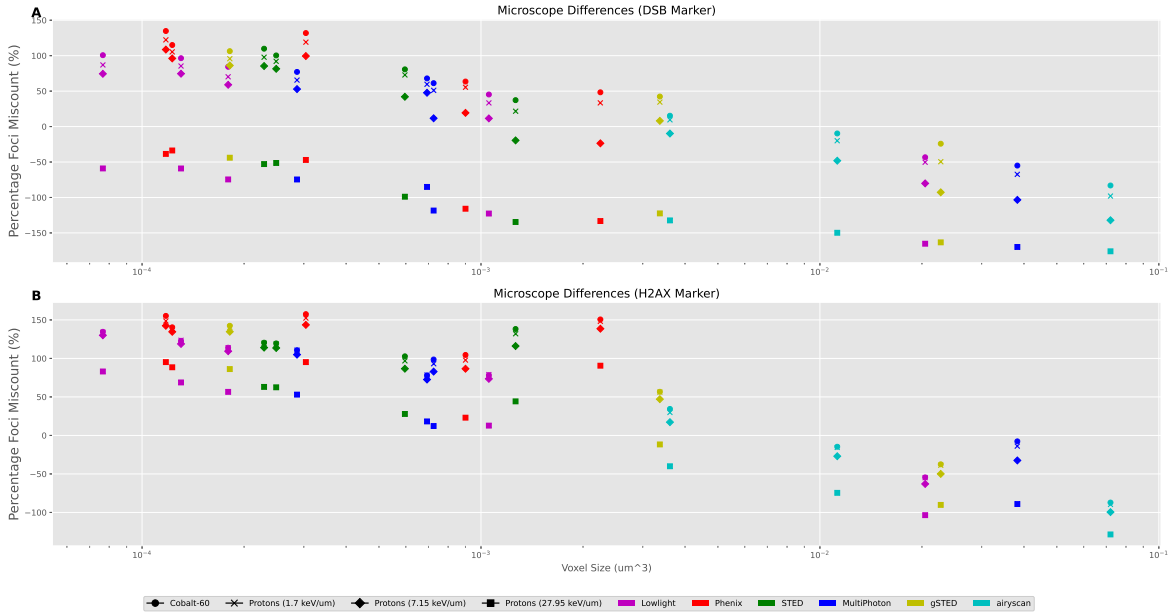

Figure 5.6: Comparison of percentage miscounting across all available microscope and magnifications. Voxel size has been used as the descriptor of each microscope and is presented on a log axis. A) are the results when using a DSB marker. B) are the results when using an γ-H2AX marker. Error bars are given as the standard error in the mean for the 1200 microscope images per point (200 cells and 6 time points). All microscope images are at 2 Gy dose.

Impact of Deconvolution on Miscounting

Emulating confocal microscopy allows for a single z-slice to be chosen whilst reducing out-of-plane fluorescence. However, as we have demonstrated, whilst the "intensity" may not be focused, with much of it being filtered out, there is still a contribution to the in-focus plane which can lead to foci over-counting. To evaluate the magnitude of this problem we emulate the changes in foci counting if the fluorescent points are not convolved with the microscope PSF. This is analogous to what would be seen if you were able to perform a perfect deconvolution on the microscope image, essentially mitigating the microscope blurring effect. The impact of deconvolution can be seen in Figure 5.7, where we have compared the foci detected for the actual system, the DSB and γ-H2AX marker visualisations and the DSB and γ-H2AX marker with deconvolution visualisations. This was compared for the Airyscan 63x microscope at the 30 minutes (Figure 5.7A) and 24 hours (Figure 5.7B) time points for 1 Gy of dose. We see that at the 30 minute time point of the high-LET protons is usually under-counted in all but the γ-H2AX marker. For the other radiation types tested

using a deconvolved DSB or γ-H2AX marker gave better agreement with the actual number of DSBs. This remains the case at 24-hours with the deconvolved versions of visualisation giving better agreement than their counterparts which include the convolved microscope PSF. Furthermore, we see that at 24-hours the overall best agreement is seen when using a DSB marker that has been perfectly deconvolved for all radiation types. This is due to the DSB marker being a direct indicator of DSBs and the reduced number of DSBs at 24 hours reduces the likelihood of breaks clustering together into a single foci. These results help confirm that the fluorescence from neighbouring slices are a major driver of miscounting in confocal imaging.

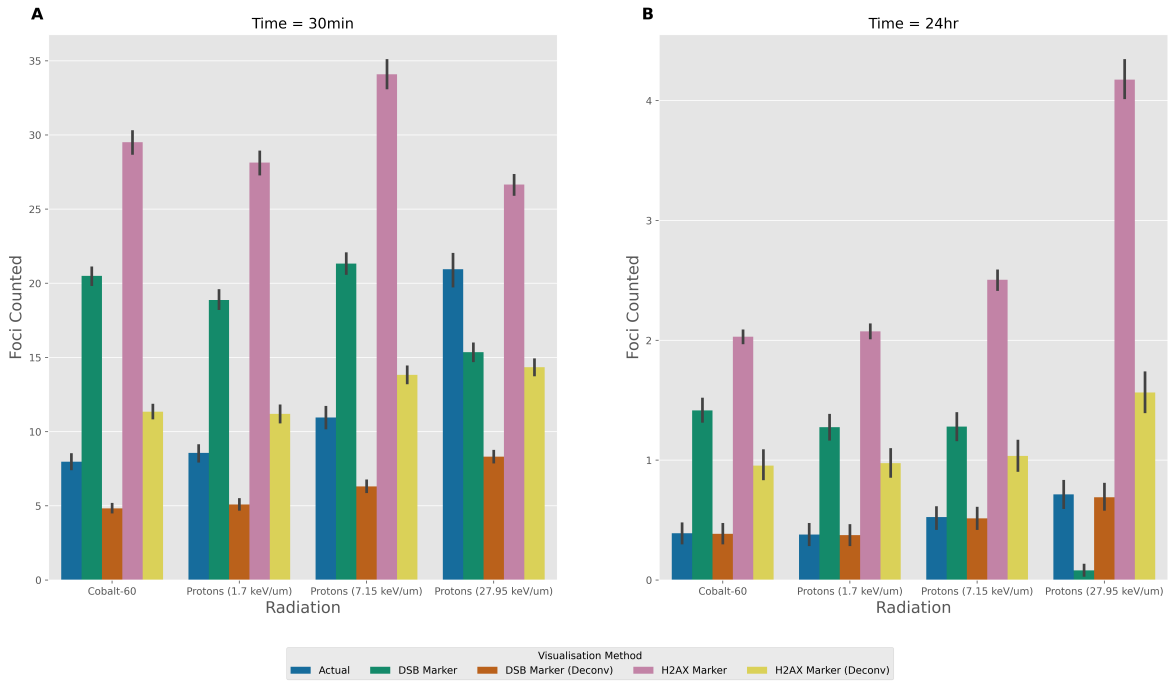

Figure 5.7: Comparison of visualisation methods including deconvolution. This compares the number of foci detected for the four radiation types delivering 1 Gy at two time points. Where A and B is 30 minutes and 24 hours post-irradiation respectively. The actual value is given by the blue bar which is the known simulated number of DSBs within the evaluated z-slice. Whereas, the other bars represent the foci detected from the corresponding visualisation method. Marker visualisation which has been deconvolved represent perfect deconvolution as the image is created before the microscope PSF is applied. Each bar is the averaged result of 200 independent samples and the error bars are the 95% confidence intervals. All data is for an emulated Airyscan 63x microscope.

Underlying drivers of Miscounting

It has been identified that both marker type and visualisation can play a significant role in the amount of possible miscounting and the subsequent repair kinetics you can arrive at given a particular radiation set-up. Most of the differences between radiation types are being driven by the spatial distribution of the DSBs and more specifically their proximity to one another. Therefore, we calculate the clustering of the DSBs at all the time points, doses and radiation types. This clustering metric allows us to unify over all of these different parameters and compare the systems based on the DSB spatial distribution alone. The clustering is a measure of the average number of DSBs within proximity to each DSB of the sys-

tem. This proximity has been defined as 200 nm in this study, but similar trends are seen at both 100 nm (Figure S11) and 500 nm (Figure S12) . We then compare the amount of miscounting for different clustering values and look for an overall pattern (Figure 5.8). We see that across all visualisation approaches an increased clustering value equates to increasing levels of under-counting. These effects are reduced in the γ-H2AX marker (Figure 5.8B) where the under-counting due to clustering appears to balance with the increased contribution of fluorescence from neighbouring slice being identified as foci. This can be confirmed by the corresponding deconvolved γ-H2AX marker system (Figure 5.8D) where the contribution from neighbouring slices are reduced and the pattern of increased under-counting at higher clustering values becomes apparent again. The effects in the DSB Marker is similar, but with a smaller contribution of neighbouring fluorescence being counted as foci (Figure 5.8A), but a still noticeable increase in under-counting in the corresponding deconvolved system (Figure 5.8C)

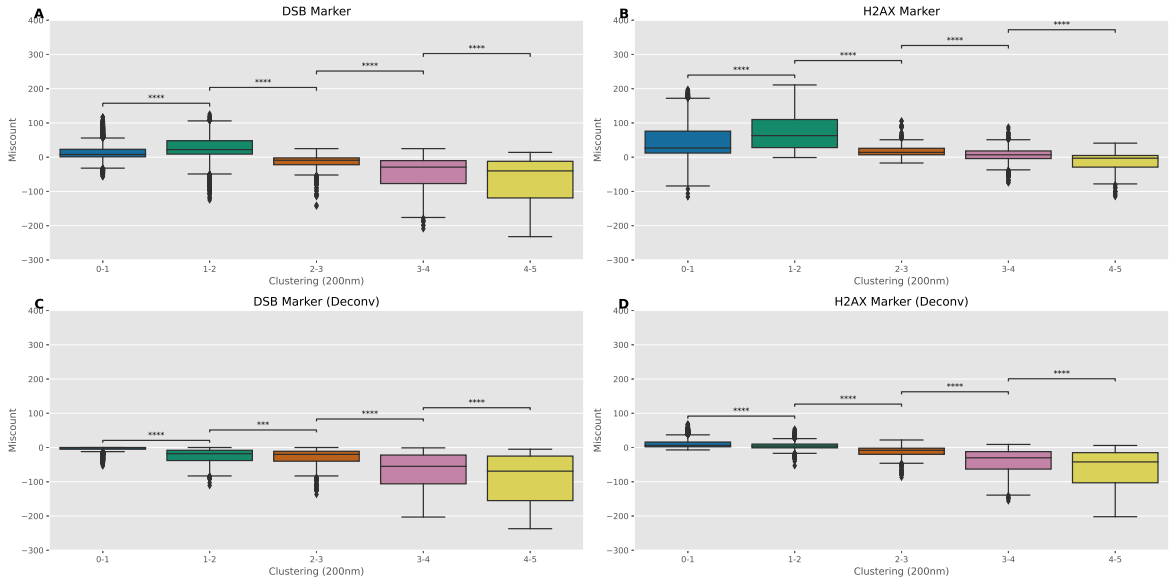

Figure 5.8: Categorising the magnitude of miscounting as a function of DSB clustering. The values of clustering are calculated based on the known amount of DSBs within the cell nucleus at a given time point. Clustering equates to the average number of DSBs within proximity to each DSB within the simulated cell nucleus. Where proximity for this study is characterised as 200 nm. Clustering is increased for early time points, higher dose, higher LET and is decreased for the opposite. Therefore, the clustering combines time, dose and radiation type parameters into a single metric. Panel A and B are for the DSB and γ-H2AX markers respectively. Panel C and D are for the DSB and γ-H2AX markers with perfect deconvolution respectively. Mann-Whitney test between categorised clustering values to highlight statistically significant differences. P-values have been adjusted using the Bonferroni correction. Significance notation refers to the following thresholds: ns=P>0.05, *=P>0.01, **=P>1e-3, ***=P>1e-4, ****=P<1e-4. All microscope images have been emulated using the Airyscan 63x point spread function.

5.5 Discussion

In this study, the impact on foci miscounting has been evaluated for a range of radiations, doses, microscopes and magnifications. Through the modelling of both the DNA DSB positioning and the emulation of viewing through a microscope, we have been able to leverage a unique position of being able to compare the detected foci to the simulated ground

truth. This provides insight into the ability to resolve DNA DSBs when using a fluorescence marker which is subjected to a microscopes point spread function. It has been shown that even under ideal simulated experimental conditions there are significant differences in the amount of miscounting you should expect between different radiations (Co-60 and protons at varying LET), shown in Figure 5.2-5.3. This can result in the perceived change of repair kinetics, even when these are simulated as constants (Figure 5.4). This brings into question how best to evaluate changes between test groups of experimental work, given that these differences in miscounting are usually omitted. This work has highlighted that the main sources of foci miscount arrive from either over-counting DSBs from neighbouring slices or under-counting DSBs which cannot be distinguished from one another. Whilst counting breaks from other slices is not inherently unwanted as this damage does exist within the cell, the rate at which this happens varies based on the radiation quality leading to a non-systematic experimental set-up uncertainty. For example, while it is widely thought that high-LET protons may cause additional complex DNA damage which requires increased time for DNA repair [6, 7, 39] we show the increased under-counting in γ-H2AX simulated images can lead to visual perception of slower repair kinetics caused by miscounting alone. It becomes important to think of how we might disentangle changes due to the uncertainty of the technique and the underlying biology.

This work has focused on confocal microscopy and viewing a single z-slice. However, we acknowledge the increased amount of multi-slice imaging and 3D foci detection [43–45], especially when using super-resolution microscopy. The developed PyFoci software is well poised to explore the effects of foci miscounting when analysing 3D data sets in subsequent work. The variation in foci detection is increasingly difficult to replicate due to the diverse available methodologies. Therefore, this work attempted to perform a widely available approach with the Laplacian of Gaussian technique prioritising consistency over optimisation. This consistency is at the backbone of the evaluations we have made, but we acknowledge that counting accuracy may be improved if the foci counting technique was optimised specifically for an individual setup (microscope, magnification, radiation type and dose). To enable this work to be more representative of an individual's experimental approach we have provided the DSB and γ-H2AX marker distributions which can be visualised using the PyFoci software. This allows the same images (including 3D stacks) to be evaluated using a foci detection system better matched to ones own experiments.

In modelling the phosphorylation of H2AX some simplifications were made. This includes a spatially linear spread of the activation region from a DSB, whereas it is believed that phosphorylation is spread through physical contact rather than propagation along the chromosome [46]. However, without sub-TAD resolution geometry, it is not possible to incorporate these contact points at present. The amount of H2AX is known to be cell-type dependent [4], in this study we use a constant of 10% of H2A histones being the variant H2AX. If this work was to be expanded to measure changes due to cell-type we would recommend that the H2AX variant percentage is altered to best capture any cell-type specific effects [47]. Whilst not included in this study, it has been suggested that DSBs inflicted at the TAD

boundary may result in a multi-TAD activation [46], as these regions make up a small proportion of the genome, making the chances of damage induction at these regions small, this effect has not been included. Furthermore, the model currently assumes a stationary break end, where in fact the break ends are known to be mobile, this may potentially increase congregation of breaks ends increasing the difficulty to distinctly identify multiple DSBs at a single site. This is confounded by other drivers of motions, such as cell cycle and chromatin condensation which require further geometric modelling developments before they can be accounted for. Finally, although γ-H2AX marker has been explicitly detailed here with the H2AX histones used as points of fluorescence, it has also been identified [30] that 53BP1 and MDC1 similarly spread across the TAD. Therefore, results for the γ-H2AX marker may be a good approximation for these markers but would benefit from further evaluation of how these markers differ to γ-H2AX.

The PyFoci software can be readily applied to any *in silico* model of DNA damage and repair which can output DNA DSB positions as a function of time. Through the modelling of the microscopy and the inherent uncertainties which come from the technique, it becomes apparent that modellers may prefer to generate these computational microscope images and perform foci counting when trying to compare their models to the experimental data. It is common for modellers to evaluate their DNA repair systems which are representative of the absolute system to experimental data [22, 48, 49], but this work has identified that the simulated system should be distorted in the same manner as the experimental fluorescence imaging is a distortion on the actual biology. This would allow for better matching between the simulated DNA repair from computational foci imaging to the corresponding experimental foci imaging.

5.6 Conclusion

This work has captured some of the underlying uncertainties related to using fluorescence foci microscopy techniques for varying radiation exposures. These uncertainties appear to lead to discernible differences in the resultant information gained from these experimental techniques and vary significantly between different forms of radiation. This variance makes it increasingly difficult to disentangle experimentally observed changes in repair kinetics from the underlying changes in foci miscounting. Therefore, we advise the use of computational microscopy with known simulated DNA DSB distributions to help investigate the variability in miscounting across the desired experimental tests. By taking into consideration these changes in miscounting, we believe that changes in the experimental results could be more rigorously evaluated.

5.7 Acknowledgements

The authors would like to acknowledge the help of Kang Zeng and Steven Bagley from the Cancer Research UK Manchester Institute for obtaining the microscope point spread functions and their insight into microscopy. The authors would like to acknowledge the assistance given by Research IT and the use of the Computational Shared Facility at The University of Manchester.

Conflict of interest statement.

None declared.

5.8 Funding

This work was supported by the NIHR Manchester Biomedical Research Centre [BRC-121520007 to K.J.K, R.I.M, N.F.K and M.J.M]; the STFC Global Challenge Network+ in Advanced Radiotherapy [ST/N002423/1 to S.P.I, M.J.M, K.J.K and R.I.M]; the European Union Horizon 2020 Research and Innovation [730983 – INSPIRE to S.P.I, N.F.K, K.J.K, R.I.M, M.J.M, N.T.H and J.W.W]; the United Kingdom Engineering and Physical Science Research Council [EP/S024344/1 – BioProton to S.P.I, K.J.K, R.I.M, N.F.K and M.J.M];a UK Research and Innovation Future Leaders Fellowship [MR/T021721/1 to S.J.M]; the National Institute of Health (NIH)/National Cancer Institute (NCI) [R01 CA187003 - TOPAS-nBio to J.S]; The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

5.9 Data Availability

All datasets are available at https://figshare.com/s/ae99c2d5888604e16eec. The PyFoci source code is available at https://gitlab.com/PRECISE-RT/releases/pyfoci.

References

1. Warters, R., Hofer, K., Harris, C. & Smith, J. Radionuclide toxicity in cultured mammalian cells: elucidation of the primary site of radiation damage. *Current topics in radiation research quarterly* **12,** 389–407. ISSN: 0011-3964 (1978).
2. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362,** 709–715. ISSN: 0028-0836 (1993).

3. Ward, J. DNA Damage Produced by Ionizing Radiation in Mammalian Cells: Identities, Mechanisms of Formation, and Reparability. *Progress in Nucleic Acid Research and Molecular Biology* **35,** 95–125. ISSN: 0079-6603 (1988).

4. Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139. *Journal of Biological Chemistry* **273,** 5858–5868. ISSN: 0021-9258 (1998).

5. Podhorecka, M., Skladanowski, A. & Bozko, P. H2AX Phosphorylation: Its Role in DNA Damage Response and Cancer Therapy. *Journal of Nucleic Acids* **2010,** 920161. ISSN: 0000-0004 (2010).

6. Wu, J. *et al.* Repair rates of DNA double-strand breaks under different doses of proton and γ-ray irradiation. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **276.** #RepairKinetics, 1–6. ISSN: 0168-583X (2012).

7. Chaudhary, P. *et al.* Variations in the Processing of DNA Double-Strand Breaks Along 60-MeV Therapeutic Proton Beams. *International Journal of Radiation Oncology*Biology*Physics* **95,** 86–94. ISSN: 0360-3016 (2016).

8. Nakajima, N. I. *et al.* Visualisation of γH2AX Foci Caused by Heavy Ion Particle Traversal; Distinction between Core Track versus Non-Track Damage. *PLoS ONE* **8,** e70107 (2013).

9. Szymonowicz, K. *et al.* Proton Irradiation Increases the Necessity for Homologous Recombination Repair Along with the Indispensability of Non-Homologous End Joining. *Cells* **9,** 889 (2020).

10. Tobias, F. *et al.* Spatiotemporal Dynamics of Early DNA Damage Response Proteins on Complex DNA Lesions. *PLoS ONE* **8,** e57953 (2013).

11. Nakamura, A. J., Rao, V. A., Pommier, Y. & Bonner, W. M. The complexity of phosphorylated H2AX foci formation and DNA repair assembly at DNA double-strand breaks. *Cell Cycle* **9,** 389–397. ISSN: 1538-4101 (2010).

12. Ivashkevich, A., Redon, C. E., Nakamura, A. J., Martin, R. F. & Martin, O. A. Use of the γ-H2AX assay to monitor DNA damage and repair in translational cancer research. *Cancer Letters* **327,** 123–133. ISSN: 0304-3835 (2012).

13. Li, P. *et al.* Correlation of dynamic changes in γ-H2AX expression in peripheral blood lymphocytes from head and neck cancer patients with radiation-induced oral mucositis. *Radiation Oncology* **8,** 155 (2013).

14. Bourton, E. C., Plowman, P. N., Smith, D., Arlett, C. F. & Parris, C. N. Prolonged expression of the γ-H2AX DNA repair biomarker correlates with excess acute and chronic toxicity from radiotherapy treatment. *International Journal of Cancer* **129,** 2928–2934. ISSN: 1097-0215 (2011).

15. Bentzen, S. M. *et al.* Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): An Introduction to the Scientific Issues. *International Journal of Radiation Oncology*Biology*Physics* **76,** S3–S9. ISSN: 0360-3016 (2010).

16. Barnett, G. C. *et al.* Normal tissue reactions to radiotherapy: towards tailoring treatment dose by genotype. *Nature Reviews Cancer* **9,** nrc2587. ISSN: 1474-1768 (2009).

17. Friedland, W., Bernhardt, P., Jacob, P., Paretzke, H. & Dingfelder, M. Simulation of DNA damage after proton and low LET irradiation. *Radiation protection dosimetry* **99,** 99–102. ISSN: 0144-8420 (2002).

18. Henthorn, N. T. *et al.* Clinically relevant nanodosimetric simulation of DNA damage complexity from photons and protons. *RSC Advances* **9,** 6845–6858 (2019).

19. Ballarini, F. From DNA Radiation Damage to Cell Death: Theoretical Approaches. *Journal of Nucleic Acids* **2010,** 350608. ISSN: 2090-0201 (2010).

20. Friedland, W., Jacob, P. & Kundrát, P. Mechanistic simulation of radiation damage to DNA and its repair: on the track towards systems radiation biology modelling. *Radiation protection dosimetry* **143,** 542–8. ISSN: 0144-8420 (2010).

21. McMahon, S. J., McNamara, A. L., Schuemann, J., Paganetti, H. & Prise, K. M. A general mechanistic model enables predictions of the biological effectiveness of different qualities of radiation. *Scientific Reports* **7,** 10790. ISSN: 2045-2322 (2017).

22. Ingram, S. P. *et al.* Mechanistic modelling supports entwined rather than exclusively competitive DNA double-strand break repair pathway. *Scientific Reports* **9,** 6359. ISSN: 2045-2322 (2019).

23. INCERTI, S. *et al.* THE GEANT4-DNA PROJECT. *International Journal of Modeling, Simulation, and Scientific Computing* **01,** 157–178. ISSN: 1793-9623 (2010).

24. Schuemann, J. *et al.* TOPAS-nBio: An Extension to the TOPAS Simulation Toolkit for Cellular and Sub-cellular Radiobiology. *Radiation research* **191,** 125. ISSN: 0033-7587 (2019).

25. Barbieri, S. *et al.* MODELLING γ-H2AX FOCI INDUCTION TO MIMIC LIMITATIONS IN THE SCORING TECHNIQUE. *Radiation protection dosimetry.* ISSN: 0144-8420 (2018).

26. Rabus, H., Barbieri, S., Baiocco, G., Ottolenghi, A. & Giesen, U. INVESTIGATION INTO THE PROBABILITY FOR MISCOUNTING IN FOCI-BASED ASSAYS. *Radiation protection dosimetry.* ISSN: 0144-8420 (2018).

27. Hill, M. Radiation Track Structure: How the Spatial Distribution of Energy Deposition Drives Biological Response. *Clinical Oncology* **32,** 75–83. ISSN: 0936-6555 (2020).

28. Tommasino, F. & Durante, M. Proton Radiobiology. *Cancers* **7,** 353–381 (2015).

29. Ingram, S. P. *et al.* Hi-C implementation of genome structure for in silico models of radiation-induced DNA damage. *PLOS Computational Biology* **16,** e1008476. ISSN: 1553-734X (2020).

30. Arnould, C. *et al.* Loop extrusion as a mechanism for formation of DNA damage repair foci. *Nature,* 1–6. ISSN: 0028-0836 (2021).

31. Reth, M. Matching cellular dimensions with molecular sizes. *Nature Immunology* **14,** 765. ISSN: 1529-2916 (2013).

32. Rao, S. S. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159,** 1665–1680. ISSN: 0092-8674 (2014).

33. Li, Y. *et al.* Moving Cell Boundaries Drive Nuclear Shaping during Cell Spreading. *Biophysical Journal* **109,** 670–686. ISSN: 0006-3495 (2015).

34. Agostinelli, S. *et al.* Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506,** 250–303. ISSN: 0168-9002 (2003).

35. Friedland, W., Dingfelder, M., Kundrát, P. & Jacob, P. Track structures, DNA targets and radiation effects in the biophysical Monte Carlo simulation code PARTRAC. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **711,** 28–40. ISSN: 0027-5107 (2011).

36. Sanche, L. Low-Energy Electron Damage to DNA and its Basic Constituents. *Physica Scripta* **68,** C108–C112. ISSN: 0031-8949 (2003).

37. PRISE, K. M. *et al.* A review of dsb induction data for varying quality radiations. *International Journal of Radiation Biology* **74,** 173–184. ISSN: 0955-3002 (2009).

38. Schuemann, J. *et al.* A New Standard DNA Damage (SDD) Data Format. *Radiation Research.* ISSN: 0033-7587 (2018).

39. Plante, I., Slaba, T., Shavers, Z. & Hada, M. A Bi-Exponential Repair Algorithm for Radiation-Induced Double-Strand Breaks: Application to Simulation of Chromosome Aberrations. *Genes* **10,** 936 (2019).

40. Abdisalaam, S., Davis, A. J., Chen, D. J. & Alexandrakis, G. Scanning fluorescence correlation spectroscopy techniques to quantify the kinetics of DNA double strand break repair proteins after γ-irradiation and bleomycin treatment. *Nucleic Acids Research* **42,** e5–e5. ISSN: 0305-1048 (2014).

41. Yuan, J., Adamski, R. & Chen, J. Focus on histone variant H2AX: To be or not to be. *FEBS Letters* **584,** 3717–3724. ISSN: 0014-5793 (2010).

42. Zlatanova, J., Leuba, S. H. & Holde, K. v. Chromatin Fiber Structure: Morphology, Molecular Determinants, Structural Transitions. *Biophysical Journal* **74,** 2554–2566. ISSN: 0006-3495 (1998).

43. Hagiwara, Y. *et al.* 3D-structured illumination microscopy reveals clustered DNA double-strand break formation in widespread γH2AX foci after high LET heavy-ion particle radiation. *Oncotarget* **8,** 109370–109381. ISSN: 1949-2553 (2017).

44. Ruprecht, N., Hungerbühler, M. N., Böhm, I. B. & Heverhagen, J. T. Improved identification of DNA double strand breaks: γ-H2AX-epitope visualization by confocal microscopy and 3D reconstructed images. *Radiation and Environmental Biophysics* **58,** 295–302. ISSN: 0301-634X (2019).

45. Memmel, S. *et al.* FocAn: automated 3D analysis of DNA repair foci in image stacks acquired by confocal fluorescence microscopy. *BMC Bioinformatics* **21,** 27 (2020).

46. Collins, P. L. *et al.* DNA double-strand breaks induce H2Ax phosphorylation domains in a contact-dependent manner. *Nature Communications* **11,** 3158 (2020).

47. Bobkova, E. *et al.* Recruitment of 53BP1 Proteins for DNA Repair and Persistence of Repair Clusters Differ for Cell Types as Detected by Single Molecule Localization Microscopy. *International Journal of Molecular Sciences* **19,** 3713 (2018).

48. McMahon, S. J., Schuemann, J., Paganetti, H. & Prise, K. M. Mechanistic Modelling of DNA Repair and Cellular Survival Following Radiation-Induced DNA Damage. *Scientific Reports* **6,** 33290. ISSN: 2045-2322 (2016).

49. Warmenhoven, J. W. *et al.* Insights into the non-homologous end joining pathway and double strand break end mobility provided by mechanistic in silico modelling. *DNA Repair* **85,** 102743. ISSN: 1568-7864 (2020).

Chapter 6

Final Discussion

6.1 Advances to Manchester Mechanistic Models

The work carried out in this thesis has been able to elevate different sections of the overarching Manchester Mechanistic Model collection. To detail this more clearly a schematic of the Manchester Models have been made with the additions positioned at various points along the process (Fig. 6.1). All additions made from the thesis have been made open-source either directly (https://gitlab.com/PRECISE-RT/releases) or through our collaboration with TOPAS-nBio. These additions take steps towards better modelling of cellular radio-sensitivity through cell-type-specific geometries and re-distribution through the addition and interfacing of HR with C-NHEJ the most prominent pathways for DSB repair during G2 of the cell cycle. PyFoci is an important tool in the DNA repair portion of the Manchester Models which aids in the fitting and evaluating DaMaRiS against fluorescent foci experimental data by bringing the simulated result closer to the experimental process.

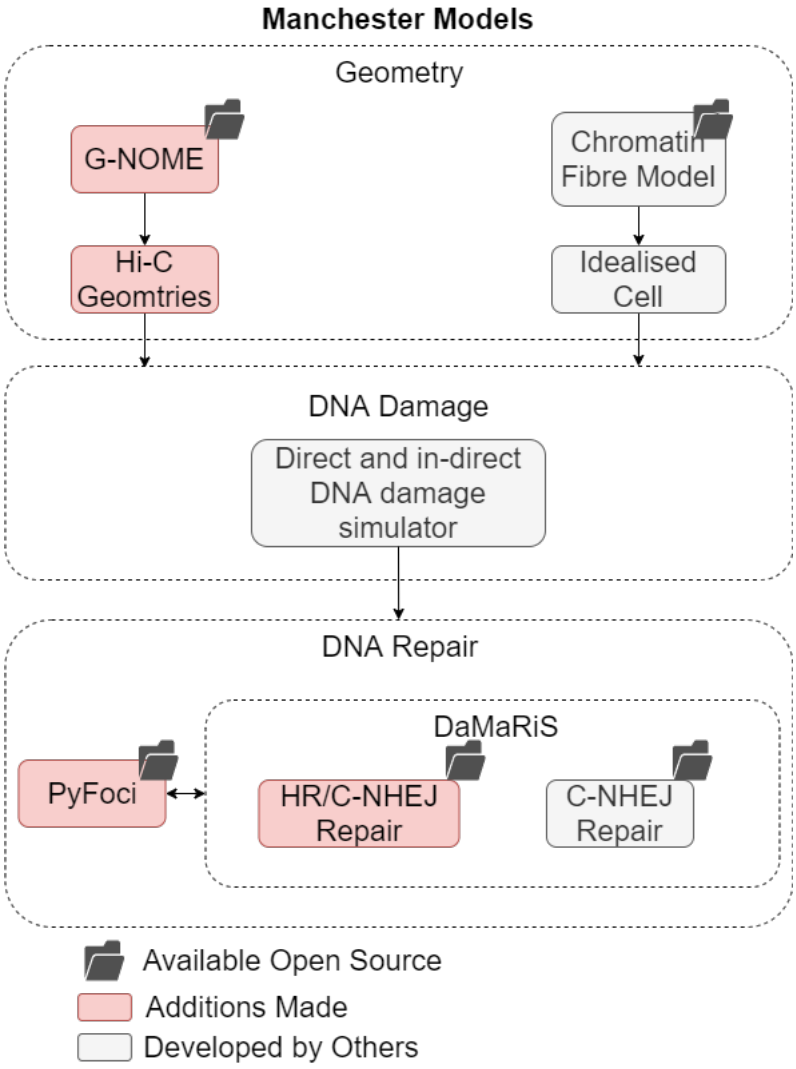

Figure 6.1: Overview of the additions made to the Manchester Mechanistic Models for radiobiological modelling from this thesis. Red boxes indicate direct additions and grey boxes indicate work developed by others. The open folder icon shows which parts of the models are open-source.

6.2 Scientific Insights

6.2.1 Importance of the Break End

The work carried out in Chapter 2 demonstrated how proteins should not be thought of as single-minded players explicit to one repair pathway. This is an important distinction to consider as traditional descriptions of repair pathways often refer to proteins "belonging" to particular pathways. As a result this is a detrimental notion when attempting to build realistic models which are striving to capture the true nature of DNA repair. A more productive way of thinking of protein recruitment can be achieved by concentrating on the current state of the DNA break end and what proteins are applicable for recruitment. This allows for a more natural way of modelling and avoids prescriptive assumptions of proteins acting beyond their capabilities.

This type of modelling is beneficial as rules can be applied based on the physical state of the break end rather than just an approved set of recruitment options. This ensures that repair pathways are not artificially removed from the system which can be especially impactful when trying to model changes in protein availability. Without this approach, the user would need to develop models for each perturbation (e.g. protein deficiencies) they would want to introduce to the system. This includes explicit modelling of backup repair pathways on an unsuccessful attempt of repair, whereas if the break end simply reports its alterations other proteins which are now applicable can prompt the repair of alternative pathways without the need of direct intervention. Furthermore, this promotes the ability of co-recruitment of proteins which can alter the break ends in multiple ways and can arrive at the natural promotion of a particular backup pathway.

The DaMaRiS framework has been designed with the importance of the break end in mind. Models use a generic break end object which has their identity modified depending on what proteins are loaded. In conjunction with user-defined pathway files, can enable modellers to assess various ways of how proteins may recruit and inhibit other proteins. Future development of the DaMaRiS project will likely focus on increasing the information which is stored in the break end objects. An example of this would be to store information such as single-strand overhang length and base sequence which are important for DNA repair and are currently inferred in models. Through explicit modelling of these features of the DNA break end, it increases the realistic capture of how proteins interact and perform a successful repair and would allow for further generalisation of the pathway files.

6.2.2 Scales of Geometry

The work carried out in Chapter 3 expands to the available methodologies to create computationally viable geometric depictions of the genome. Prior to this work, geometric models where either inferred from imaging data or created using mathematically repeated sequences. Neither of these approaches provided a systematic way of accounting for cell type

specific geometric differences or intra-cell type geometric differences across a population. Through the incorporation of biological information, Hi-C allows us to better capture how and why different cell types have specific radiosensitivity. Furthermore, as Hi-C has demonstrated the geometric patterns observed from within the genome are largely functionally driven; this could potentially shed some insight on functionally irregular cells, such as cancer cells and senescent cells, both of which are clinically interesting. Finally, our simulations of Hi-C geometries being irradiated has shown how alterations in the nucleus shape may inevitably alter the DNA damage distribution which is a potential layer of inconsistency between radioresponse carried out *in vitro* and clinically.

The Hi-C defined cell-type-specific chromatin conformation at the resolution of TADs. These typically range from between kbps to Mbps in size and are therefore large when compared to the scale at which DNA is damaged by incident radiation. At present, there is no method to build the geometry within TADs using experimentally inferred information, but there are attempts to model the inner-TAD geometry using a voxelised packing approach. Whilst this allows for increased detail for the DNA damage you can score it is also devoid of any cell-specific inner-TAD structuring. It would be advantageous to expand the Hi-C analysis to incorporate higher resolution structures and possibly capturing differences in the amount of heterochromatin and euchromatin directly from the data. One such approach to doing this may be to expand the G-NOME solver for inputs of Micro-C data which goes down to nucleosome resolution. However, with the availability of Micro-C data being extremely sparse and costly to perform this type of development, at present, might not have significant application.

A significant amount of geometric modelling for cell structures typically revolve around the structure of the nucleus as DNA is a major target for radiation. However, as our understanding of radiobiology grows, so do the targets of interest for radiation. Structures such as mitochondria and the cell membrane are interesting targets to model for radiation damage. Therefore, as we continue our pursuit of complete mechanistic models, including cell-type specific responses, it will become increasingly important to build up a model of the whole cell. Platforms such as Geant4-DNA and TOPAS-nbio are incredibly useful for this pursuit as it enables researchers from a wider remit of interest to collaborate and share their geometries which could potentially be used in conjunction to speed up this development.

6.2.3 Experimental Skepticism

In any type of modelling, it can be of great benefit to have a ground truth to reference, the work in Chapter 5 demonstrates how you must take care when considering experimental data as that truth. Radiobiological modelling uses experimental data for both fitting behaviours and benchmarking their results. A common issue when doing this is that there is a large amount of uncertainty typically seen across biological experiments. This can make it difficult to correctly identify patterns and assess how well a model should mimic the results seen. An approach to dealing with this, as seen in Chapter 5, is to devise an independent

model which attempts to capture the sensitivity of an experimental technique and quantify levels of uncertainty. This is analogous to how the detectors in Physics experiments are typically simulated to understand the magnitude of response required to prove or disprove a theoretical hypothesis. The narrowing down of these uncertainties will help ensure the models predict response differences akin to the sensitivity the experimental data can provide and reduce the chance of over-fitting to experimental error.

The concept of over-fitting in modelling is well discussed but typically more often from the point of view of model design, whereby you want to minimise the number of parameters that don't have mechanistic meaning in order to minimise "fudge factors". The work carried out in Chapter 5 demonstrates that radiobiological models need to consider the sensitivity of their parameters and if the experimental data they use can be trusted for the level of accuracy they attempt to derive. To do this it may be advantageous to either emulate the same experimental uncertainties within the model or to attempt to correct the experimental data error for non-accounted uncertainties. The PyFoci model is specific for immunofluorescent foci data, a common data source for radiobiological modellers, but the concept can be expanded for other detectors commonly used in radiobiology experiments.

The difficulty of relying on experimental data is a substantial problem for radiobiological models. This is most prominent for those who are attempting to provide clinical insight. Using approaches such as PyFoci may aid in avoiding some of the pitfalls of dealing with uncertain data, but to minimise these issues experiments should be designed around ensuring robustness. This can be extremely onerous for the experimentalist as it requires significant modelling of all the equipment being used and a large number of experimental repeats. This goes beyond what is normally required for mechanistic exploration in biology experiments and requires specific experiments designed around building models. It is in these areas of collaboration between experimentalists and modellers where significant progress can be made in ensuring robust generalisable data can be obtained.

6.3 Modelling Limitations

6.3.1 Protein Overload

The work carried out in Chapter 2 demonstrated how mechanistic modelling of DNA repair recruitment can capture some of the complexities of how repair pathways interact with one another. Protein-level modelling is highly advantageous due to them being able to broaden predictions to both increases and decreases in protein availability. This is a move towards building clinically robust models which can deal with real-world data where patient cohorts have a variety of repair capabilities. However, this increased functionality comes at the cost of requiring a detailed understanding of how repair proteins interlace to form their respective functional repair pathways. Furthermore, once a pathway is understood conceptually, something which has not been fully realised for most repair pathways, you require experimental data to match the behaviour of the "key" proteins within the model. The concept of

only modelling “key” proteins is beneficial in the fact that you reduce the complexity of the model, but it is prone to the subjectivity of the researcher and as repair pathways become increasingly understood what qualifies as “key” proteins can change. This requirement of defining which proteins to model, ensuring the theoretical mechanisms are established and making sure there is strong experimental data to fit against outlines the difficulties which surround this mechanistic modelling and poses as a limitation.

Once the selection of proteins have been identified the next hurdle to account for is that some proteins have a range of functionalities. A key example of this is the protein kinase ATM which is indubitably very important following irradiation of DNA as it helps orchestrates the response. Therefore, it is only logical that ATM would be classified as a key protein to model mechanistically, the difficulty arises when you attempt to assess how, when and why ATM is used following irradiation. This is because ATM plays multiple roles in DNA damage sensing, DNA repair protein recruitment and cellular homeostasis. The range of targets ATM interacts with makes it difficult to incorporate all of its mechanisms without then subsequently modelling a large array of other proteins. The specificity of including protein-protein interactions does in turn limit the inclusion of proteins that interact with vast numbers of other proteins. One method to resolve this issue would be to partially model the functions of ATM or non-explicitly model ATM as a concentration rather than a molecular object. These approaches are likely the only reasonable way of incorporating key orchestrator proteins without changing the modelling philosophy.

6.3.2 System Dynamism

The motion of the DNA is modelled by sub-diffusion by DaMaRiS and has a large impact on the kinetics and fidelity of repair. Within the DNA repair model, the system removes the physical modelling of DNA structure and treats everything as objects in space confined to the nucleus. However, in reality, it is the physical structure of the DNA and tethering protein complexes that likely reduces both the magnitude and extent of the DNA end motion. It remains unclear if there is strong enough theoretical and/or experimental evidence to be able to accurately model motion without the inclusion of the physical structure. This is due to the mechanisms of motion of the DSB break end within the cell nucleus depending on both micro-movement of the end and macro-movement of the larger chromatin structure. This remains an outstanding part of the model which is not simple to overcome, requiring tracked imaging data at multiple resolutions.

Beyond the motion of the DNA which is a response to damaged DNA, there is that which occurs from the cell progressing through the cell cycle. The progression through the cell cycle has a large impact on the radio-sensitivity, not only does it alter the available repair pathways, but there are morphological and metabolic changes that will alter the damage response. Furthermore, it is known that the progression through the cell cycle is safeguarded by checkpoints which can inhibit progression if a sufficient amount of DNA damage is present. Finally, the cell cycle progression to mitosis is critical to capture possible mitotic catastro-

phe events. Whilst, the DaMaRiS framework does currently have ways of assessing these checkpoints, the combination of controlling them along with the up- and down-regulation of different repair pathways is not currently complete and is no small feat. However, when complete it may be possible to track not only events of mitotic catastrophe but also successful divisions. The tracking of generations of cells is an interesting topic when starting to examine how misrepair events could potentially lead to secondary cancers.

At present proteins within DaMaRiS are not explicitly modelled, but are modelled through recruitment time constants. This is a beneficial way of modelling proteins as it reduces the number of computational objects the simulation has to track, reducing compute time. It also offers the ability to emulate increases and decreases in recruitment efficiency through the changing of the single time constant. Where this approach becomes limited is that there is currently no consideration of protein depletion. Whilst this is unlikely to be problematic for proteins that are ubiquitous in the nucleus (e.g. ku70/80) it may pose inaccuracies for proteins that have been reported to deplete at moderate doses (e.g. Rad51). To capture this mechanism it would be beneficial to extend the DaMaRiS framework to include a depletion factor that can trigger if there are many break end objects of a given state. Alternatively, this could be deployed single object representation of the population of a given protein which can have its availability altered based on how frequent it is recruited and produced. As with most new features suggested the limiting factor would be if there is enough experimental data to fit against for these added details.

6.3.3 Multi-Scale Modelling

The design of mechanistic modelling which can capture how damage is formed on a molecule of the DNA ($\sim 1e^{-15}$ m) through to the response in patients (~ 1 m) is vulnerable to a variety of multi-scale modelling issues. Multi-scale models refer to the running of multiple models at varying scales simultaneously, which can interact with one another to describe a system as a whole. Typically each model will be designated a scale model, incorporating the relevant physical laws which govern that particular sub-system. For example, a DNA damage model using radiation transport physics and a DNA repair model using protein dynamics. The use of multi-scale models is to attempt to overcome the inaccuracies of macro-scale models whilst spanning beyond the limited scope that micro-scale models can offer. Through the combination of multiple models, it is hoped one might achieve high detail macro-scale modelling. However, in practice using multi-scale models requires the developer to compromise between accuracy and efficiency for each scale of the model to best fit the needs of the overall endpoint they wish to capture. Furthermore, there can the difficulties in interfacing models, especially if independently developed, which outlines the need for wider consensus on model reporting standards, for example the Standard DNA Damage (SDD) data format [1] is an example of good practice.

The issue with radiobiological modelling for radiotherapy specific endpoints is that it is not always clear on what parts of the micro-scale model are needed to achieve high accuracy

whilst attempting to minimise the computational overhead of irrelevant information. This is somewhat driven by the system being sequentially driven, in that the response at one scale is a derivative of the prior-scale model. This uncertainty around the theoretical understanding of how molecular damage at the DNA results in a particular clinical endpoint limits the ability to encapsulate micro-scale models with pre-computed relationships. Therefore, to evidence these types of models for clinical use it is becoming increasingly pertinent to focus on particularly well-understood endpoints which have clear pathogenesis. This will act as an exemplar multi-scale mechanistic models which will bolster further development of the composite micro-scale models.

6.4 Future Direction

6.4.1 Validation and Robustness

Through using a mechanistic modelling approach, in theory, each step of the model can be validated by experimental measurements. This allows for the segmentation of single or multiple mechanisms which can be tested and validated. Whilst this level of testing can be useful to ensure correct model characterisation, it also brings a large overhead of work required when compared to non-mechanistic models. Therefore the strategic design of experiments is required with a preference for multiple experimental outputs which can be used for validation of several parts of the model at once. This requires careful consideration of discerning which areas in the literature are well understood and which are ambiguous.

Model validation is only one side of a successful model. As discussed in Chapter 5 the use of experimental data is limited by how accurate the experimental technique is. This means it becomes possible to validate your model, but for the model not to be very robust with large error bands. This is not a good metric of models which are intended to inform clinical decision making. Therefore, measuring model robustness is an incredibly important step of model application. One such approach to doing this is to calculate an uncertainty budget for each parameter of the model. The overall uncertainty of the model is a composite of all the uncertainties together and provides a good assessment of how robust the model is in its predictions. The measured robustness can then be accurately represented when providing predictions at a clinical level.

As the DaMaRiS framework is developed it becomes increasingly important in how the model is optimised to best represent the chosen available experimental data. This process should be automated as doing this manually is extremely time-consuming. Therefore, there would be a benefit in developing a generalised optimisation process, something which is currently missing from DaMaRiS. Automated optimisation of DaMaRiS fitting will allow for more time to be focused on mechanistic development rather than tuning. Furthermore, optimisation can remove areas of subjectivity, which whilst not innately bad, can cause unwanted model bias. Finally, this would also allow for robustness to be measured during the optimisation step, which could potentially result in more robust models being found.

6.4.2 Mechanistic Modelling for Tissues

The current developments in the Manchester Models have been around building a detailed picture of the DNA damage and subsequent DNA repair. At present, the DaMaRiS model reports up to repair fidelity and does not include mechanisms of cell fate in the model. This is an important final step of the model for multiple reasons; firstly, cell death is a widespread metric of radiobiological models and quantifying this in the DaMaRiS project will allow cross-comparison; Secondly, cell fate is important for subsequent scales of the models, namely, how do the effects in a single cell start to manifest at the tissue scale. I believe this is the next major step for the Manchester Models to extend the predictions being made to clinically observable effects. However, making the transition from single-cell simulation to tissue simulation is incredibly complex from both a theoretical and computational standpoint.

From a theoretical standpoint, it becomes important that we define which metrics are essential for the expansion to the tissue scale. The benefits of the mechanistic approach of the DaMaRiS model is that multiple cell fate endpoints can be modelled and transferred to this state, but in turn this then requires a detailed understanding of how this impacts the tissue level effect. The pathogenesis of radiation-induced tissue injury spans a variety of biological processes which can be specific to the tissue endpoint in question. To create a model which can mechanistically model the pathogenesis you need to have ways of considering inflammation, cell depletion, fibrogenesis and senescence. Each of these topics is incredibly detailed and have various implications at a wide range of time scales.

From a computational standpoint, given the breadth of what tissue-level modelling wants to achieve it is important to consider the architecture of the modelling at this scale. As these endpoints need to consider a vast number of cells, with intercellular communication and shifting environmental conditions, typically these types of models gravitate towards system modelling approaches over discrete physical object models. However, a good exception to this is the use of agent-based models which have the benefit of capturing physical single object behaviour whilst focusing on the overall system response. Whilst, there is no one correct modelling approach, given the nature of how the Manchester Models currently depict DNA damage and repair an agent-based approach would allow for the continued concept of using discrete physical objects. To enable such a modelling technique encapsulation and correlation of the DNA damage and repair models would be required as it is unfeasible to model the independent response of millions of cells. Whilst this may be achieved statistically, it may be of interest to explore how machine learning models (e.g. generative adversarial networks) capture complex relationships to examine if it could potentially improve encapsulation but may come at the cost of model transparency.

6.4.3 Clinical Utility

The Manchester Models are some of the most detailed depictions of radiation-induced DNA damage and repair, and as such, they make for excellent scientific tools to conduct radiobi-

ology research. However, the aim of the models is to provide a form of clinical utility in the hope to further optimise the use of proton radiotherapy. This is difficult as whilst detailed models are potentially more impactful due to the increased insight they may be able to offer they also require longer development periods and more points of validation. Regardless, the ability to offer clinical utility will remain a clear path of development for the Manchester Models which will look to answer questions posed by clinical staff.

As the Manchester Models are approaching prediction of cell fate, with tissue-scale modelling likely to come next it is increasingly clear that selecting specific clinical endpoints that are relevant to proton therapy is the best way to tackle the breadth of required development. These clinical "use cases" will help provide the real-world endpoint data that the developed models need to predict. It is important these "use cases" are relevant to our Clinical partners, validated and robust. Furthermore, careful consideration must be given when implementing any model clinically as it should be fully realised that any potential benefit predicted will need to be balanced with likely degradation to other aspects of the treatment plan. Implementation can be often underrepresented in research and can be difficult to carry out correctly, it requires researchers and clinical staff to communicate clearly in a common language, to ensure the limitations of the model are understood so that any actions of the clinical staff are correctly informed.

Chapter 7

Conclusions

Radiotherapy is a potent tool for the treatment of cancer. The availability of proton therapy is growing quickly around the world, making it increasingly pertinent to understand and subsequently optimise the biology of the modality. *In silico* radiobiological models are useful tools to attempt to do this as they can have roles in both scientific evaluation and clinical decision making. In this thesis, several models have been developed to improve the detail at which biological processes can be predicted. Whilst, the tools at present have been designed for scientific evaluation each model has latent potential for impact at a clinical level.

The work of this thesis starts with the addition of the pathway choice between NHEJ and HR the two most dominant types of DNA DSB repair. This work outlined two main areas of consideration when moving to model cell fate: 1) the distribution of the DNA break ends is an important metric for misrepair and 2) the model accuracy is limited by the experimental data it is based on. These areas resulted in a broad journey across the cellular radiobiological landscape where the subsequent models produced allows for not only better model detail but also the ability to make hypotheses and outline key experimental questions.

The DaMaRiS model can now capture the key steps of protein recruitment and inhibition that guides the repair of DNA DSB ends during the G2 cell phase. It was shown that NHEJ and HR do not operate exclusively competitively but are better modelled as complementary resulting in the "entwined" pathway approach. To better incorporate how DNA DSBs are distributed within the cell nucleus the Manchester Models required an accurate depiction of how DNA was distributed. This was found in the form of Hi-C data and the developed G-NOME model allows for the interpretation of this data to realistic chromosome conformations ready for DNA damage simulations. This brought with it cell-type-specific conformation changes and the work in this thesis demonstrated that these alone can lead to statistically significant changes in the damage distribution. The final aspect of this thesis was the development of the PyFoci model, which leverages both the models of *In silico* DNA damage and Hi-C conformation to predict the amount of miscounting in immunofluorescent foci counting. This miscounting changes significantly between different radiation parameters which can result in perceived biological differences. This work contributes not only to the modellers looking to evaluate the accuracy of their "ground truth", but also experimentalists looking to ensure the robustness of their data. In combination, it is hoped that these models, all of which are open-sourced, can be utilised by the community and continue to develop our understanding of these pertinent topics.

Appendices

Appendix A

Supplementary Material for Ch. 3 - Mechanistic modelling supports entwined rather than exclusively competitive DNA double-strand break repair pathway

Protein Recruitment Kinetics

Accurate protein recruitment kinetics are essential for the total repair timings. This section details the results of fitting recruitment time constants in the *in silico* model and how they compare to the recruitment kinetics seen in literature from fluorescence foci imaging. It should be noted that no absolute quantification efforts have been made. Instead, the relative build-up of foci through to levels of a plateau as a function of time have been used. There is a comparison of the changes in simulated recruitment kinetics throughout the tested repair choice scenarios for Ku, DNA-PKcs and CtIP.

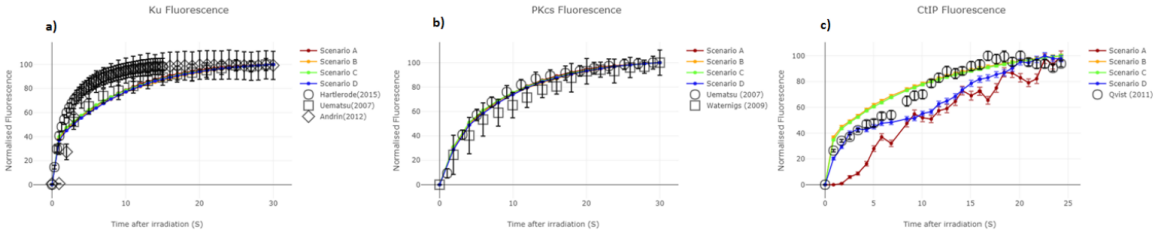

Figure A.1: protein recruitment kinetics scenario comparison - a) shown insensitivity of the recruitment kinetics of Ku70/80 for the tested scenarios. b) shown insensitivity of the recruitment kinetics of DNA-PKcs for the tested scenarios. c) shown variation in the recruitment kinetics of CtIP for the tested scenarios. All error bars are $\pm$SEM from either the reported experimental data or from 200 repeated simulations each with their own independent exposures on different cells.

Flexibility of scenarios (Variations in τ_{RR})

Throughout the repair choice scenario testing the time constant for the progression between DNA end that has been resected and DNA end repaired (τ_{RR} has been allowed to vary. This allowed variable leads to a spectrum of possible repair kinetics. T he allowed variation for each cell system tested is shown in Figure S2, Figure S3, Figure S4 and Figure S5 for Scenarios A, B, C and D respectively. The possible variation is shown by varying τ_{RR} for multiple simulations between 100s and 100,000s.

Variations of XLF-deficiency modelling

This study models XLF-deficiency as the failure of synapsis stabilisation between two DNA-PK complexes. An alternative model of XLF-deficiency was explored where the deficiency resulted in the inability to ligate DNA ends, similar to the modelled Lig4-deficient system. This alternative approach was evaluated for a range of synapsis stabilisation impairments, where the time constant applied to synapsis stabilisation was increased (Figure S6). The simulated results for the XLF deficiency leading to the inability to ligate with varying levels of synapsis stabilisation impairment is shown in Figure S7.

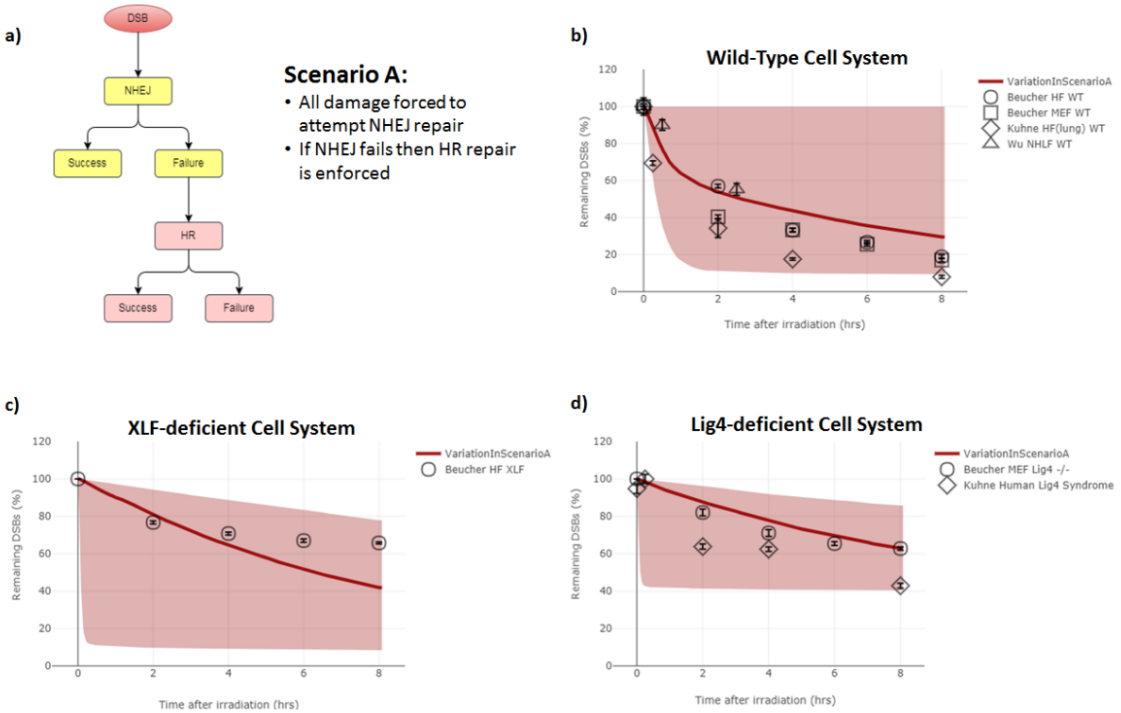

Figure A.2: Flexibility in Scenario A - the simulated repair kinetics displayed in Figure 3 is shown as the emboldened red line. The coverage of red colour represents the possible repair kinetics when varying τ_{RR} for multiple simulations between 100s and 100,000s. b), c) and d) represent the Wild-Type, XLF-Deficient and Lig4-Deficient cell systems respectively, with a) being a simplified schematic of the tested repair choice scenario. The available variation of Scenario A results shows that the majority of data-points can be mimicked though this would require multiple τ_{RR} time points to be used. Whilst the use of multiple τ_{RR} time points would not be used, the understanding that you can achieve the flexibility to mimic experimental values gives merit to Scenario A's repair choice representation.

Detailed diagrams of scenarios

The repair choice scenarios modelled have been optimised to fit with protein recruitment kinetics where possible. The diagrams below are schematic depictions of the simulated repair pathways including the time constants used. For the different cell systems modelled (Wild-Type, XLF-deficient and Lig4-deficient) the time constants remain fixed for each scenario with only set progression points removed, as shown in Figure 2. Protein names listed within the boxes represents the stages at which it is believed the protein is still present and therefore would be counted in the simulated recruitment kinetics graphs (Figure S1). Boxes in grey represents stages of the model which were not explicitly modelled. With dotted lines representing the time constant which encompasses non-explicitly modelled steps. Yellow boxes represent Non-Homologous End Joining steps, red boxes represent Homologous Recombination steps and blue boxes represent DNA end processing steps. The "Exposed DSB End" with the red lining represents the starting point of the repair simulation as all DNA ends are placed here at the start.

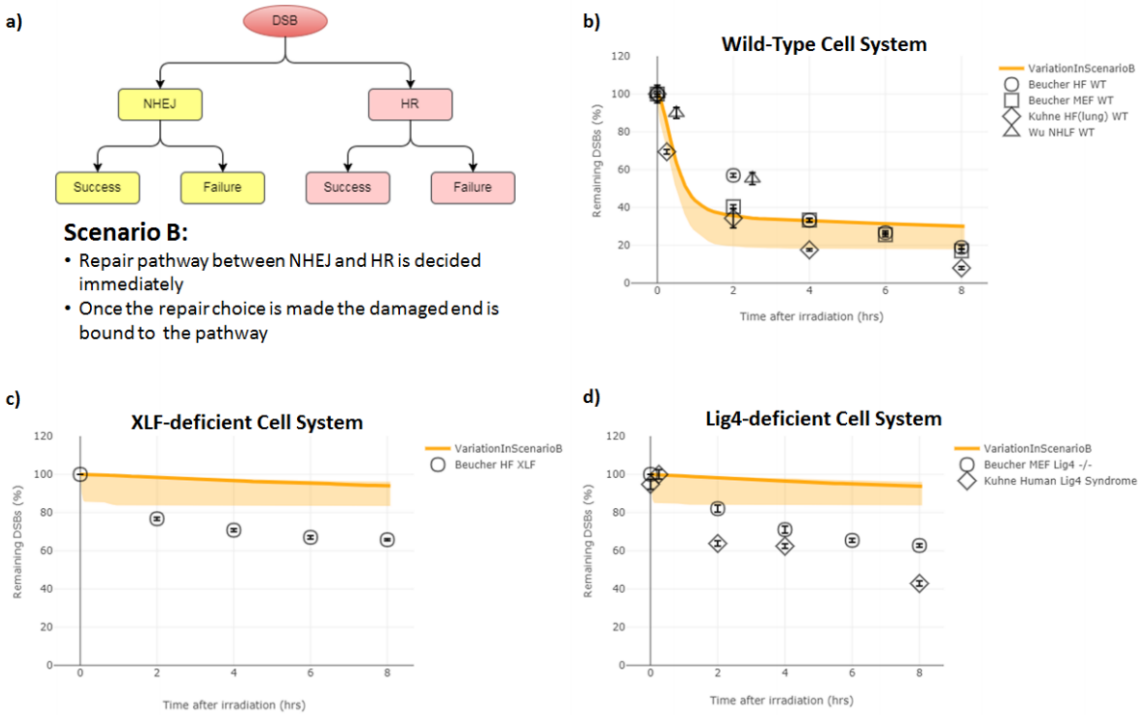

Figure A.3: Flexibility in Scenario B - the simulated repair kinetics displayed in Figure 3 is shown as the emboldened orange line. The coverage of orange colour represents the possible repair kinetics when varying τ_{RR} for multiple simulations between 100s and 100,000s. b), c) and d) represent the Wild-Type, XLF-Deficient and Lig4-Deficient cell systems respectively, with a) being a simplified schematic of the tested repair choice scenario. The shown variation in all cell systems demonstrates that Scenario B is unable to mimic both the values and shape of the experimental data. This supports the argument of a purely competitive system with no repair pathway cross-talk being unrepresentative of repair choice.

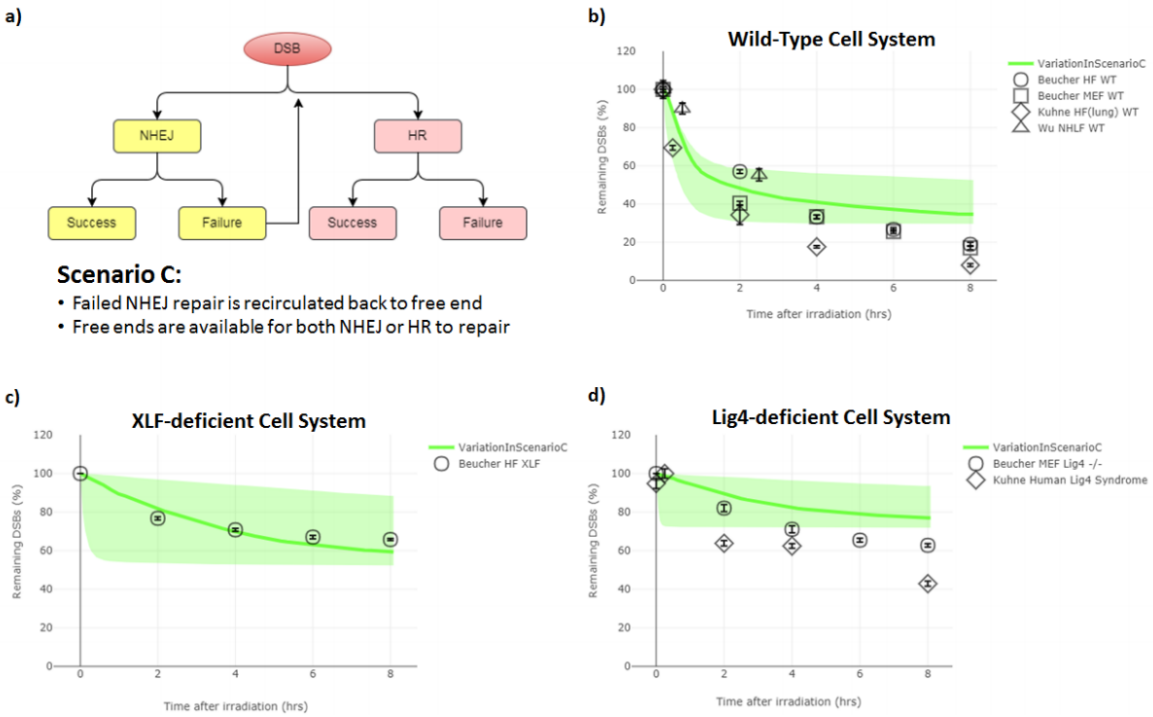

Figure A.4: Flexibility in Scenario C - the simulated repair kinetics displayed in Figure 3 is shown as the emboldened green line. The coverage of green colour represents the possible repair kinetics when varying τ_{RR} for multiple simulations between 100s and 100,000s. b), c) and d) represent the Wild-Type, XLF-Deficient and Lig4-Deficient cell systems respectively, with a) being a simplified schematic of the tested repair choice scenario. Whilst the repair kinetics of the XLF-deficient cell system c) can be well represented by Scenario C, the possible variation achievable in the WT a) and Lig4-deficient d) cell systems show that repair kinetics at 6 and 8 hours cannot be mimicked. This failure supports the argument that a purely competitive system with cross-talk through re-competition is not representative of repair choice.

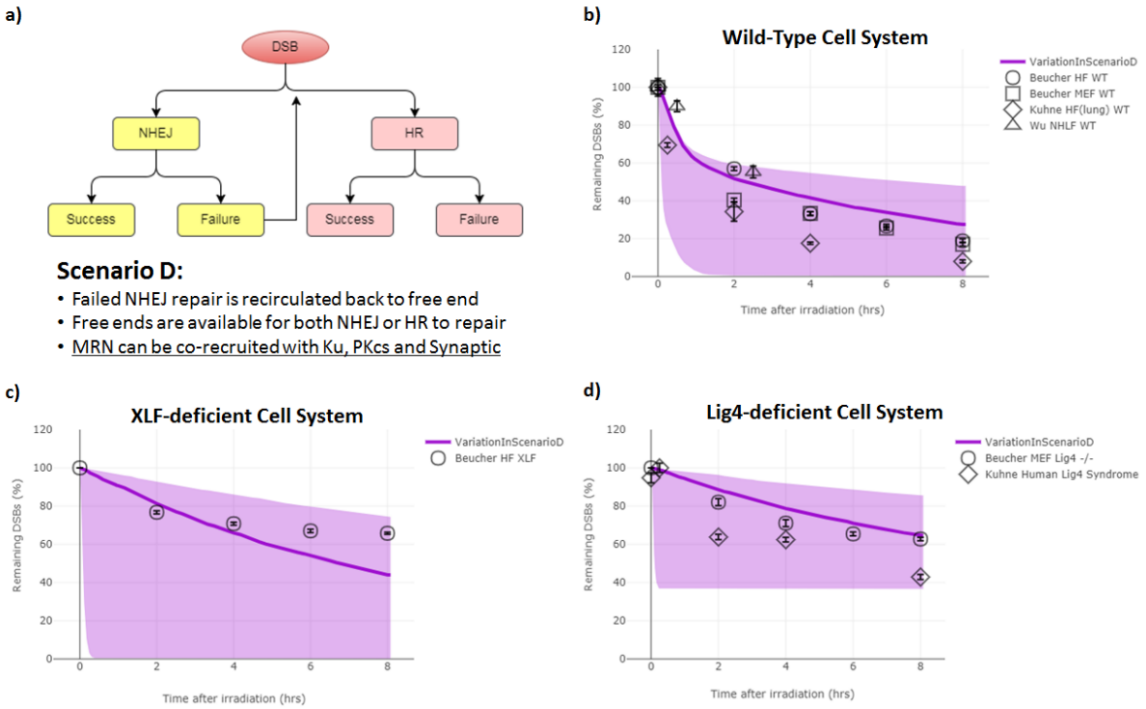

Figure A.5: Flexibility in Scenario D - the simulated repair kinetics displayed in Figure 3 is shown as the emboldened purple line. The coverage of purple colour represents the possible repair kinetics when varying $\tau_{RR}$for multiple simulations between 100s and 100,000s. b), c) and d) represent the Wild-Type, XLF-Deficient and Lig4-Deficient cell systems respectively, with a) being a simplified schematic of the tested repair choice scenario. Similarly to Scenario A the achievable variation shown in Scenario D shows it can mimic the majority of experimental data points. This flexibility gives merit to the scenario and a system which utilises non-directed competition along with co-localisation of repair proteins to represent the repair choice.

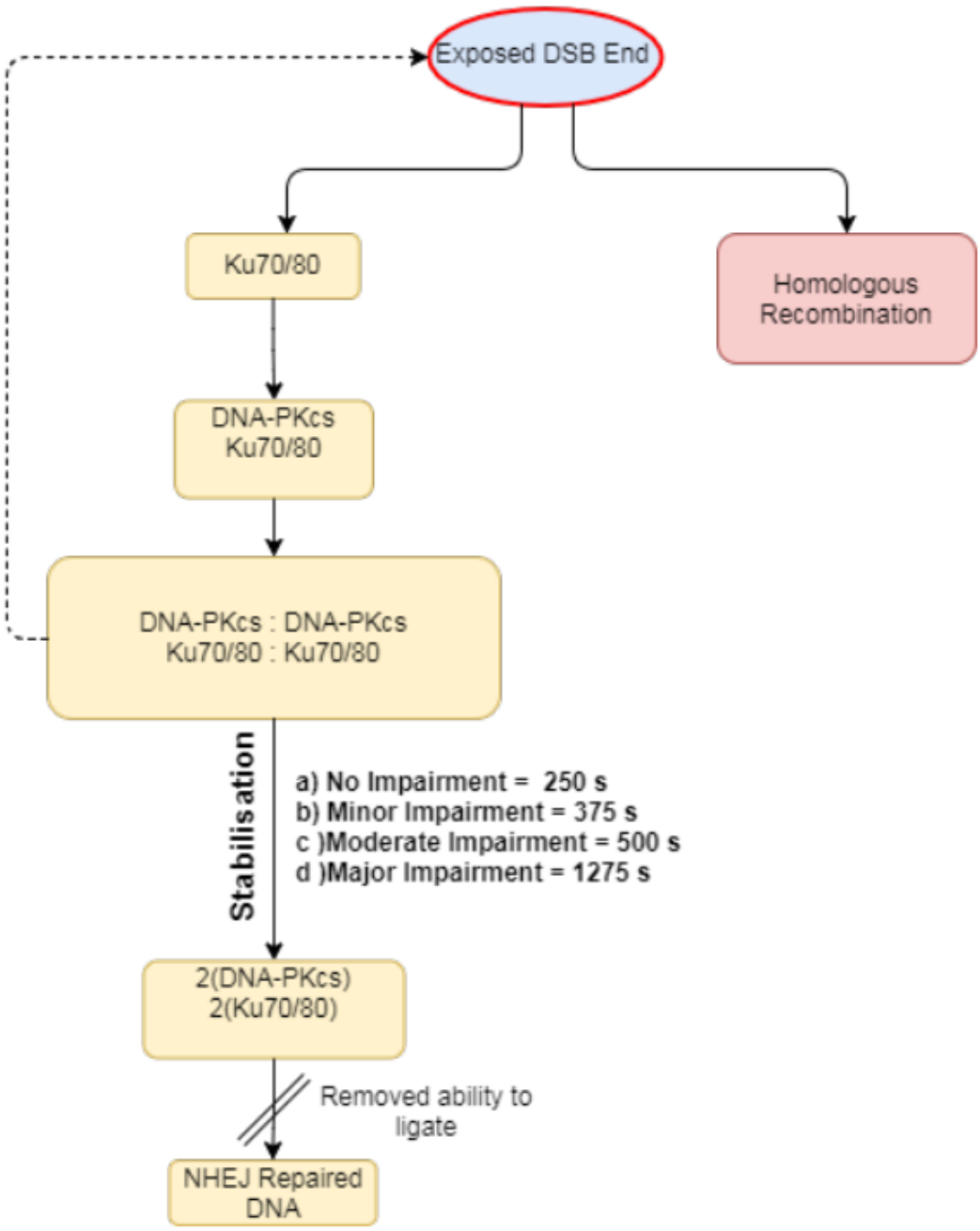

Figure A.6: Schematic of variations in modelling XLF-deficiency - diagram depicting the alternative approach to modelling XLF-deficiency where ligation is made impossible and the stabilisation of the synapsis complex is impaired taking longer to complete. The level of impairment ranges from no impairment (250 s), minor impairment (375 s), moderate impairment (500 s) and major impairment (1275 s). The progression to "Homologous Recombination" varies in the same manner as to what is shown in Figure 2 for each repair choice scenario.

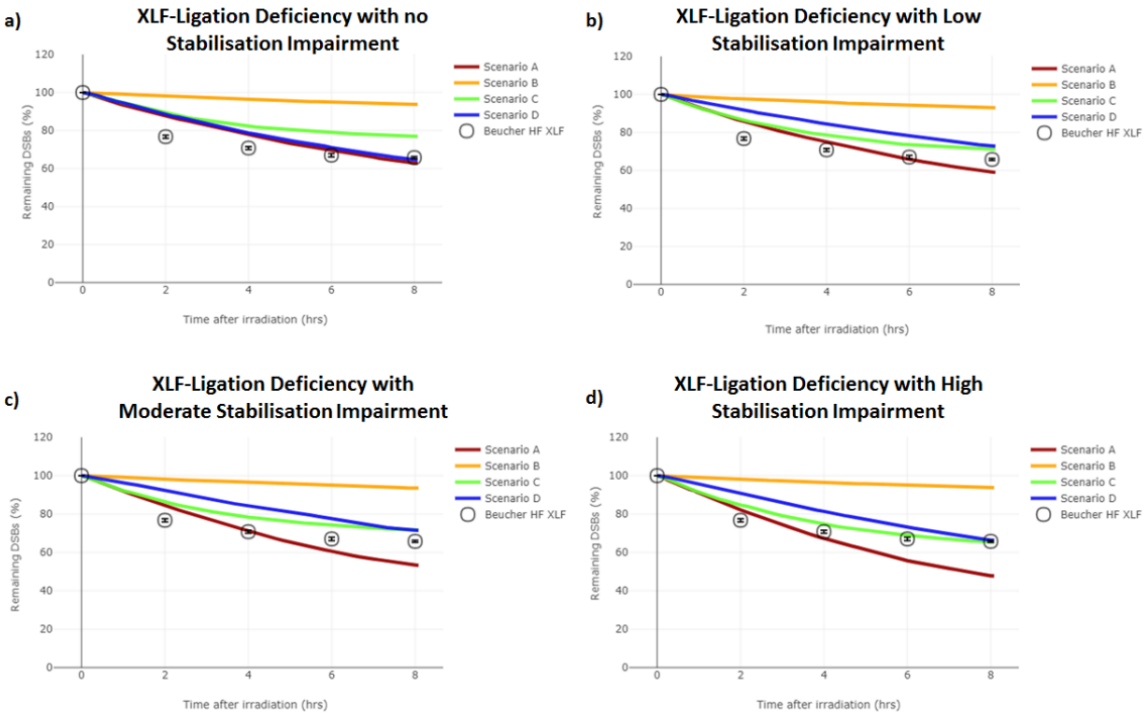

Figure A.7: Simulated results of variations in modelling XLF deficiency - results are shown for each scenario with the same optimised time constants used in Figure 3 with the variation shown in Figure S6. The exact time constants used for each repair choice scenario is shown in Figure S8-S11. a) no stabilisation impairment present, results are the same as the simulated Lig4-deficient system in Figure 3, but have been plotted against experimental XLF-deficient data. b) minor stabilisation impairment, increasing the the stabilisation time constant by 1.5 times. c) moderate stabilisation impairment, increasing the stabilisation time constant by 2.0 times. d) major stabilisation impairment, increasing the stabilisation time constant by 5.0 times. The error in the simulated data is the $\pm$SEM is displayed as the line width for 50 repeated simulations each with their own independent exposures on different cells.

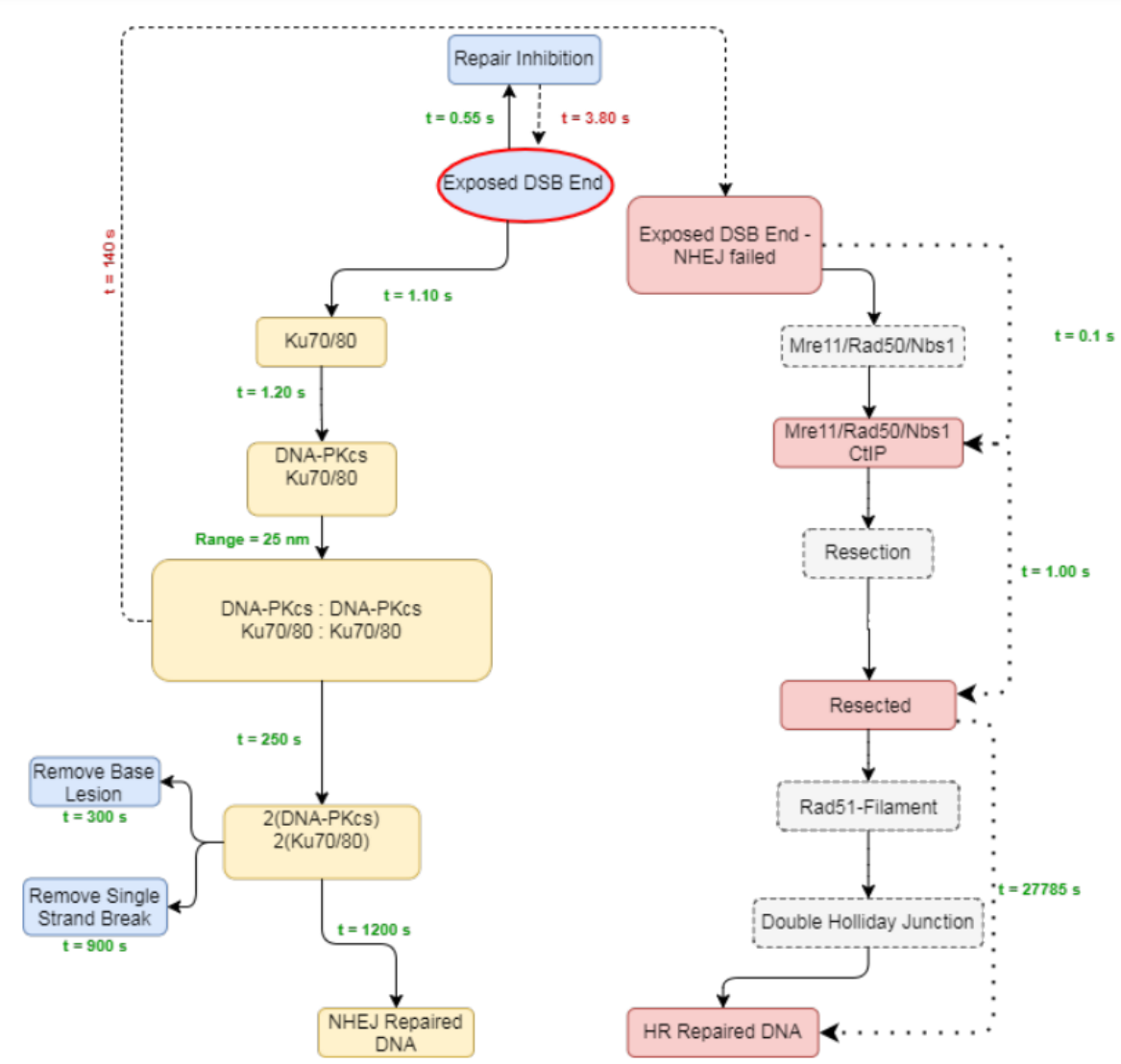

Figure A.8: Detailed schematic of Scenario A - diagram shows the time constants used for modelling Scenario A with the DaMaRiS framework (v0.3). Progression time constants are in green text and regression time constants are in red (with a dashed line).

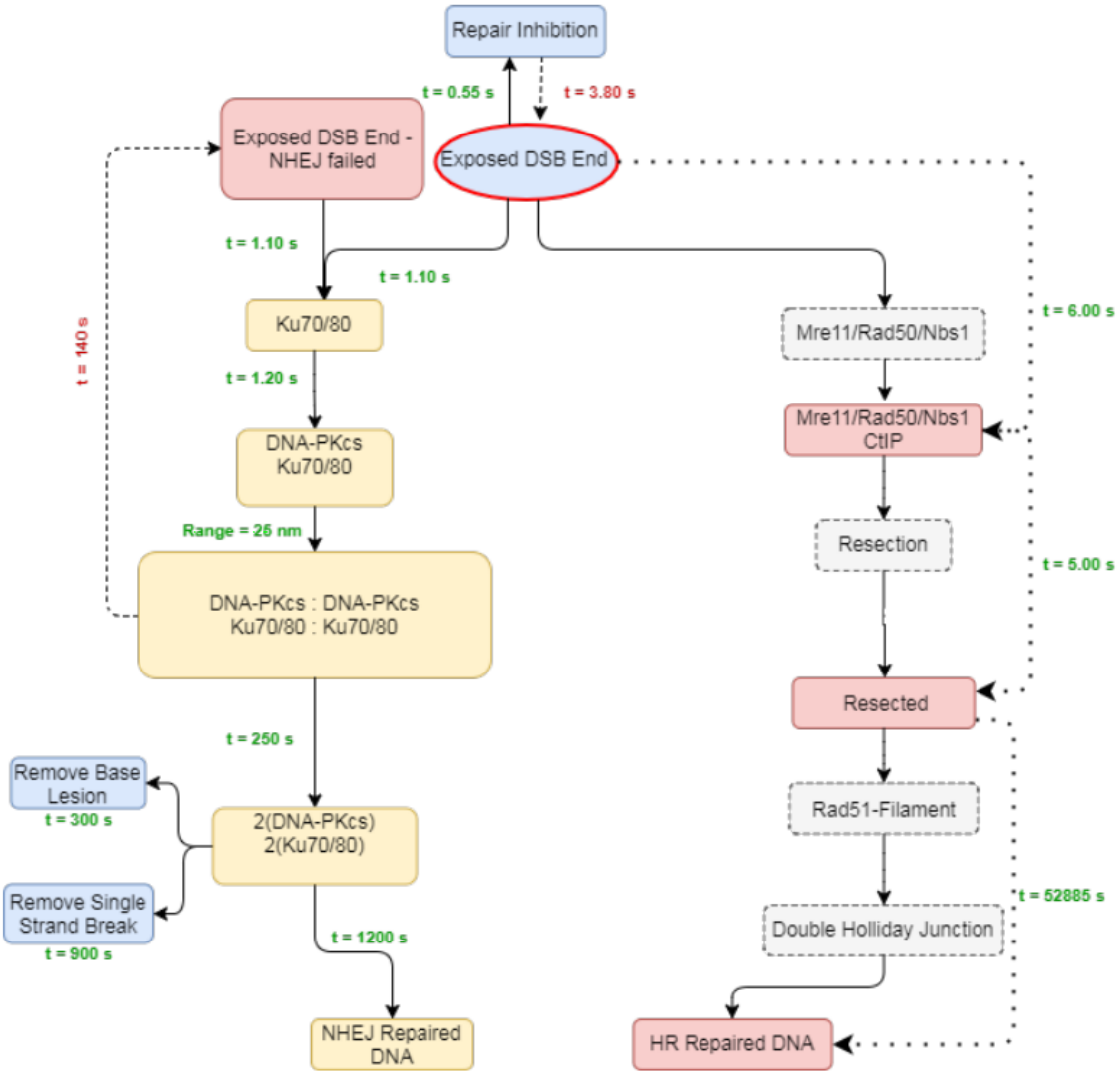

Figure A.9: Detailed schematic of Scenario B - diagram shows the time constants used for modelling Scenario A with the DaMaRiS framework (v0.3). Progression time constants are in green text and regression time constants are in red (with a dashed line).

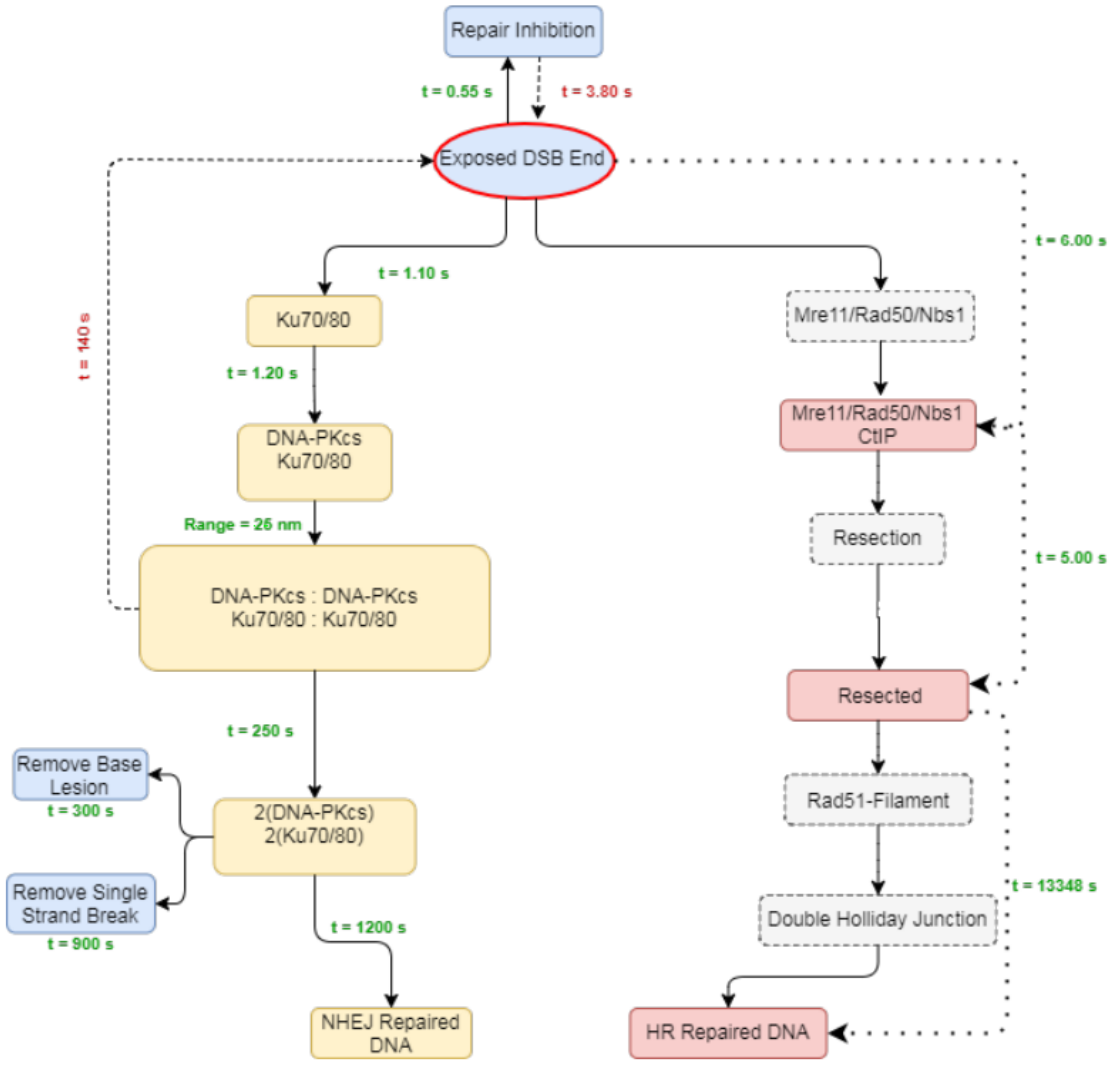

Figure A.10: Detailed schematic of Scenario C - diagram shows the time constants used for modelling Scenario A with the DaMaRiS framework (v0.3). Progression time constants are in green text and regression time constants are in red (with a dashed line).

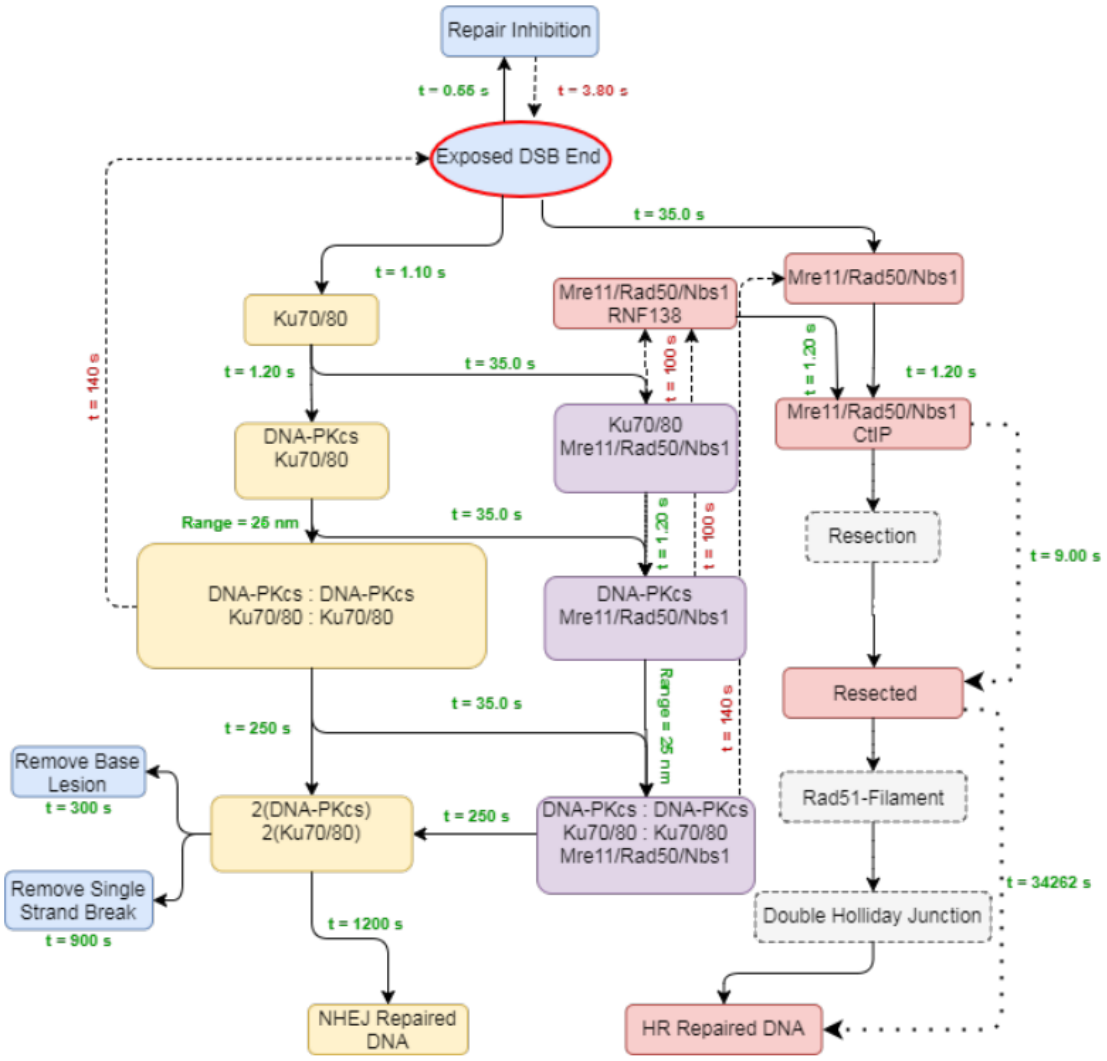

Figure A.11: Detailed schematic of Scenario D - diagram shows the time constants used for modelling Scenario A with the DaMaRiS framework (v0.3). Progression time constants are in green text and regression time constants are in red (with a dashed line).

Appendix B

Supplementary Material for Ch. 4 - Hi-C implementation of genome structure for *in silico* models of radiation-induced DNA damage

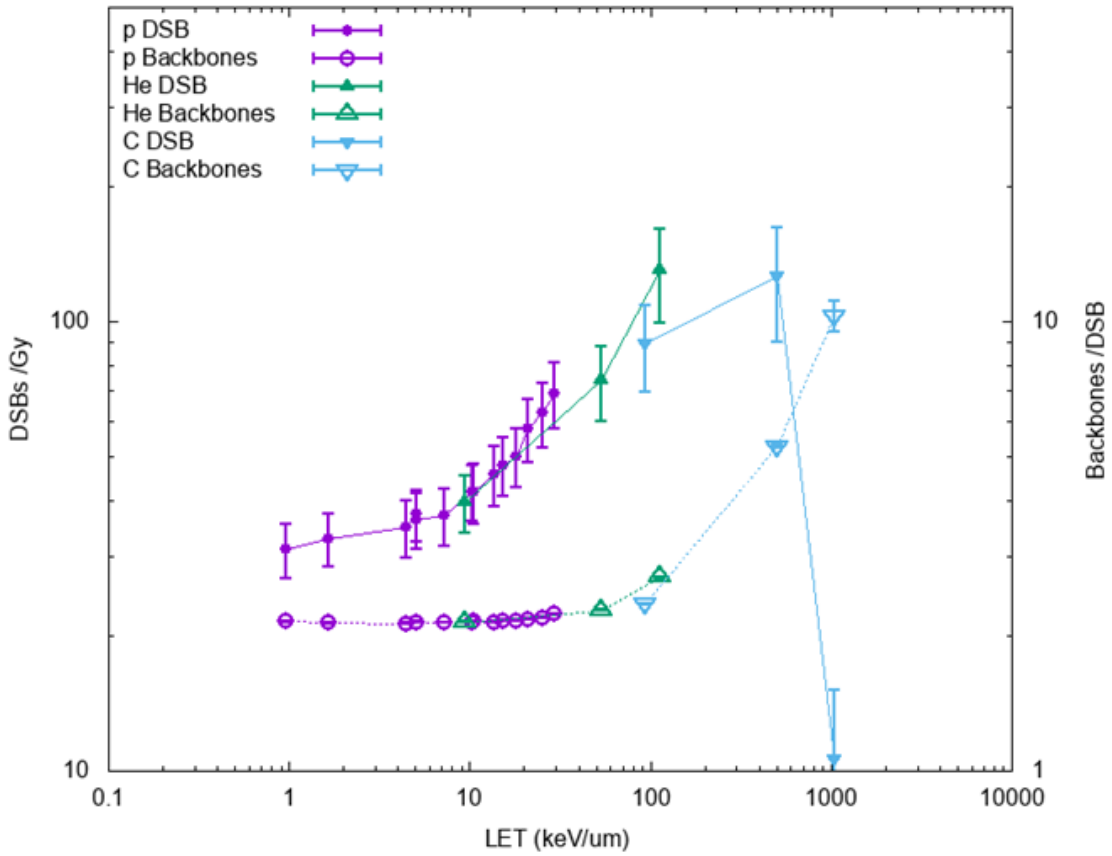

Figure B.1: **DSB Damage Complexity.** Average number of DSBs/Gy (filled symbols) and backbones per DSBs (empty symbols) for a range of LET values across different particle types. Error bars are displayed as the standard error of the mean for 100 repeats.

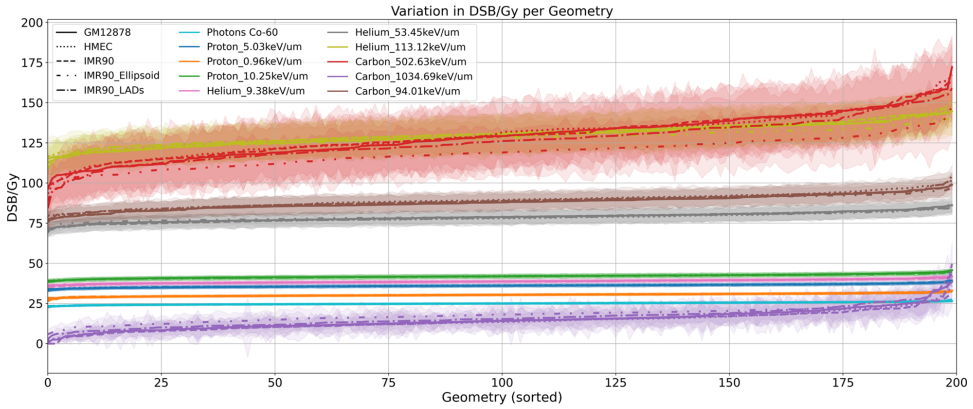

Figure B.2: **DSB Yield Variation Across Generated Geometries.** Yields of DSB per Gy of dose for the each of the 200 geometries created. Results have been sorted from smallest to largest yields to allow for easier interpretation. Different cell-types are shown as different line types with each radiation quality presented as a different colour. Errors are the transparent area around the line and are the standard error in the mean for 50 independent exposures per geometry.

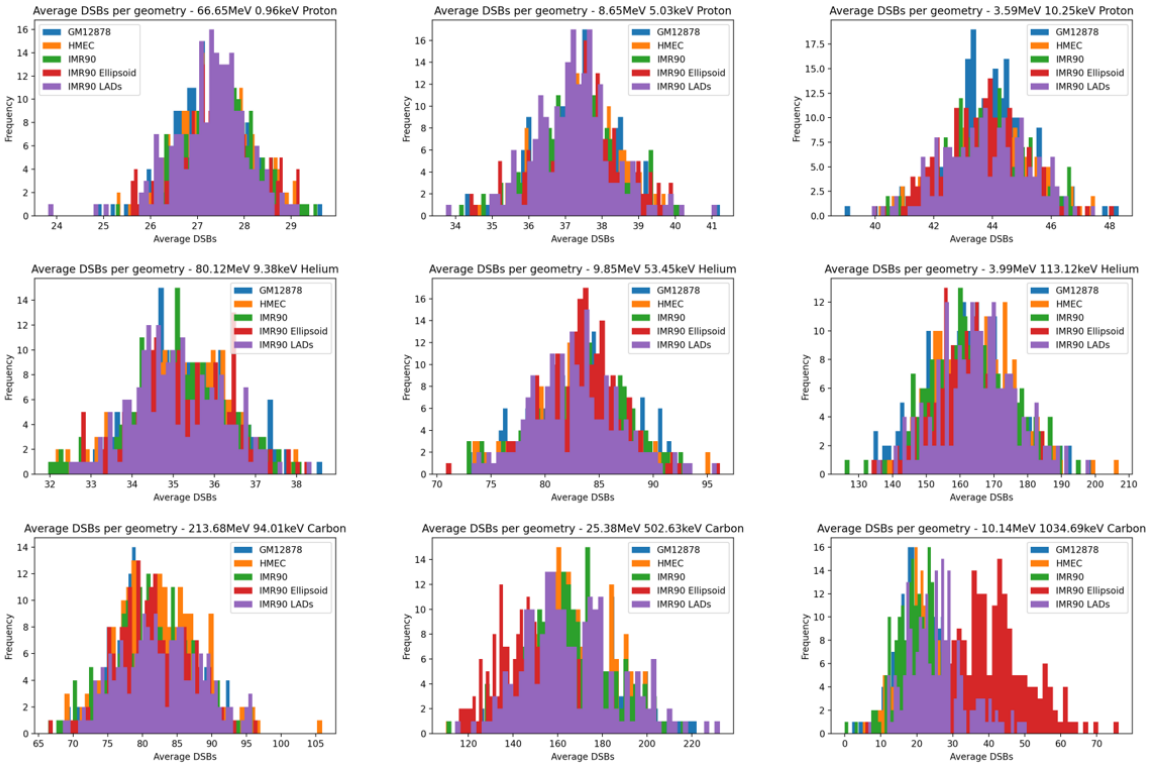

Figure B.3: **DSB Yield Distribution.** Double-strand break yield histograms for 200 geometries of each cell-type and variant.

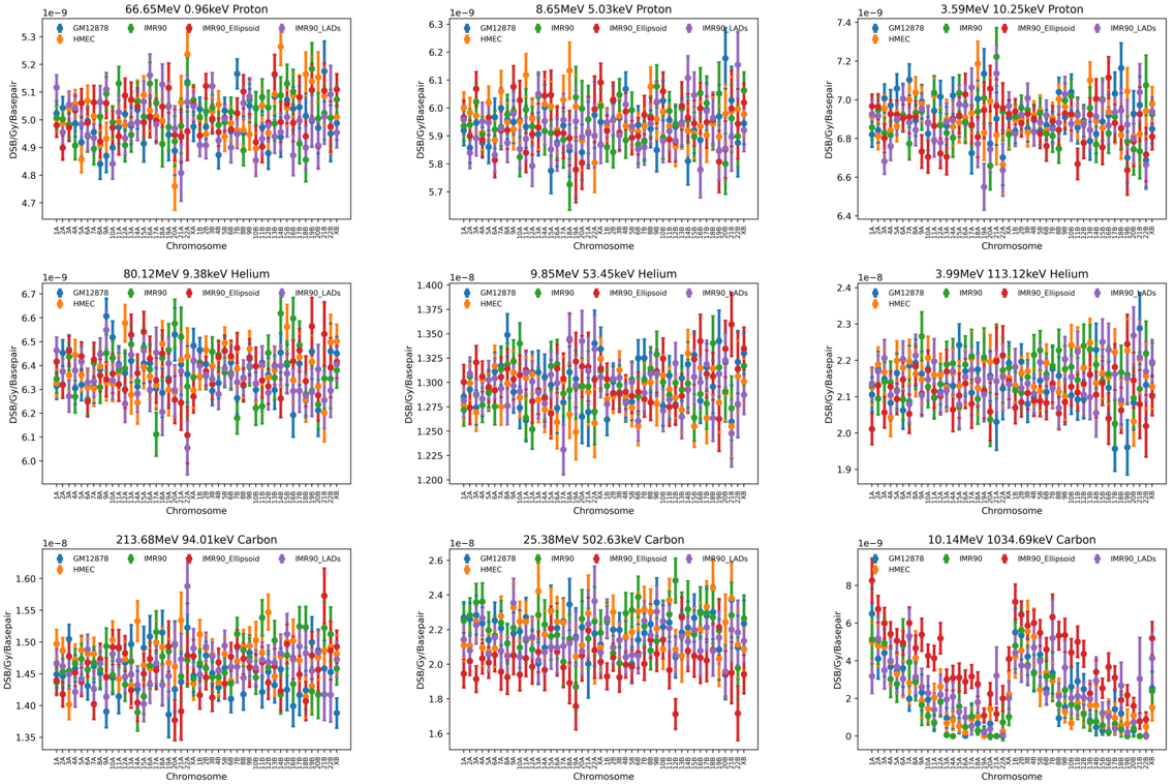

Figure B.4: **DSB Yield Chromosome Distribution.** Double-strand break per 1Gy of dose per DNA basepair on each of the modelled 46 chromosomes for each cell-type and variant. Error bars are displayed as the standard error of the mean for 200 geometries for each cell-type and variant with each geometry having 50 independent exposures.

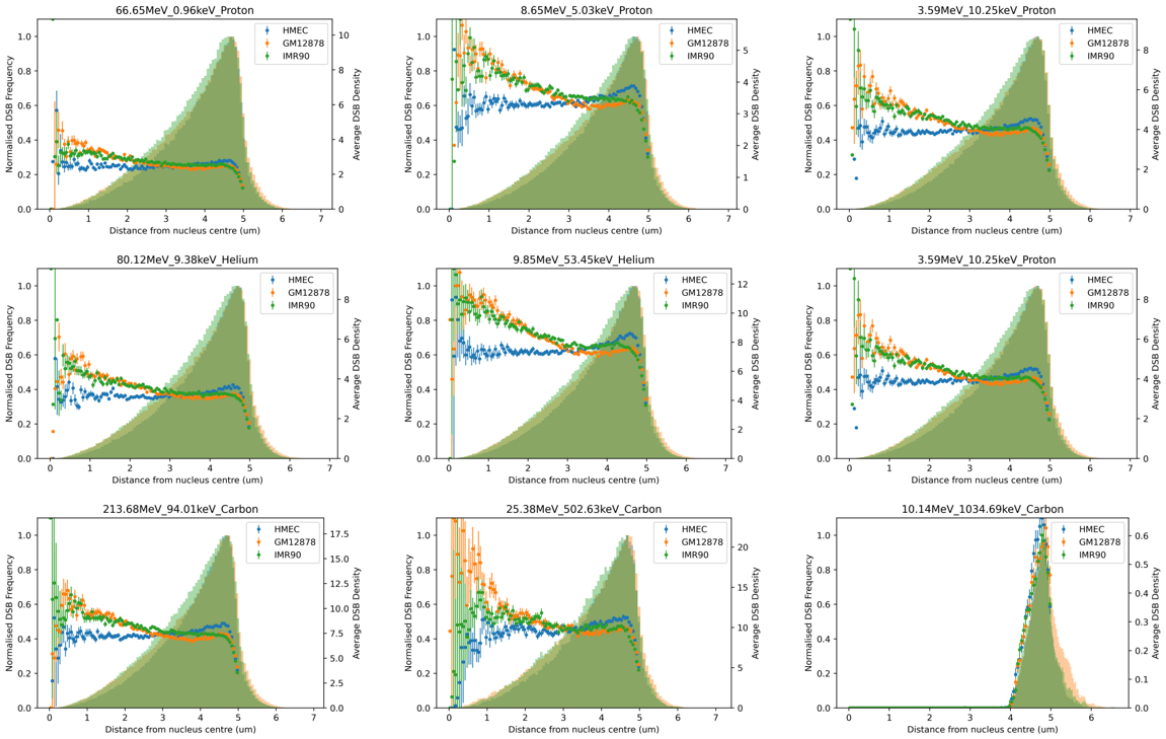

Figure B.5: **Spatial distribution of DNA DSB yields for different cell types.** Dual axis plot – left y-axis shows the histogram plot of the Normalised DSB frequency and right y-axis is the corresponding average DSB density for the same x-axis bin per geometry. Both are given as a function of distance from the nucleus centre. The cell types are all solved for a spherical nucleus and do not include LADs. The DSB frequency was normalised to the maximum number of DSBs within any bin for a given cell type. DSB density is calculated as the average number of DSBs per geometry (N=200) within a bin divided by the volume (um3) of the spherical shell of the bin. Error bars in the DSB density are the standard error in the mean for all 200 geometries for each cell type. Title shows "energy_LET_particle" for each subplot.

]

Particle	Energy (MeV)	LET (keV/um)	Range (um)
Proton	66.65	0.96	>20
Proton	8.65	5.03	>20
Proton	3.59	10.25	>20
Helium	80.12	9.38	19.9
Helium	9.85	53.45	12.5
Helium	3.99	113.12	10.0
Carbon	213.68	94.01	>20
Carbon	25.38	502.63	>20
Carbon	10.14	1034.69	5.86

Table B.1: **Incident Particle Range.** Examination of particle range travelling through 20um of water. Simulation was carried out for each energy with the furthest depth of each particle scored and averaged for 100 single-particle transversals. Particles which transverse beyond the 20um of water are simply signified as having ranges beyond 20um.

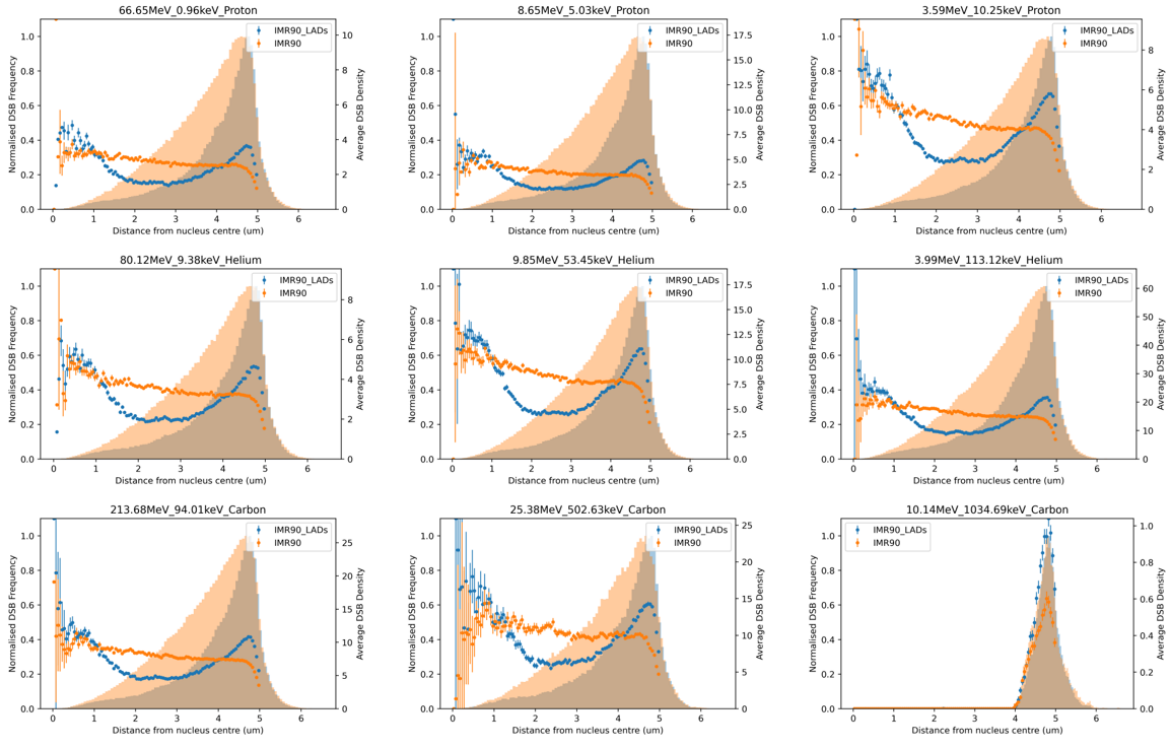

Figure B.6: **Spatial distribution of DNA DSB yields for the addition of LADs.** Dual axis plot – left y-axis shows the histogram plot of the Normalised DSB frequency and right y-axis is the corresponding average DSB density for the same x-axis bin per exposure. Both are given as a function of distance from the nucleus centre. Comparison between IMR90 with and without LADs constraints for a spherical nucleus. The DSB frequency was normalised to the maximum number of DSBs within any bin for a given cell variant. DSB density is calculated as the average number of DSBs per geometry (N=200) within a bin divided by the volume (um3) of the spherical shell of the bin. Error bars in the DSB density are the standard error in the mean for all 200 geometries for each cell variant. Title shows "energy_LET_particle" for each subplot.

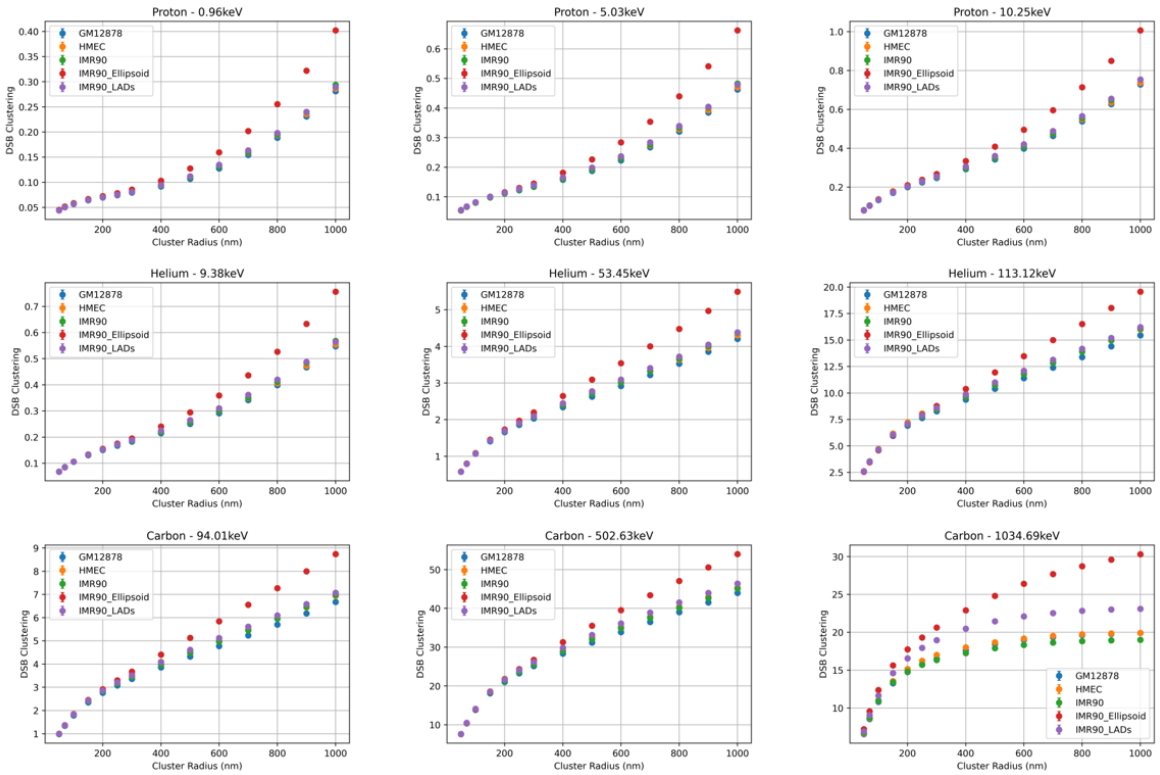

Figure B.7: **DSB Clustering per Radiation Quality Plots.** Double-strand break clustering as a function of the cluster radius for all cell-types and variants. Error bars are displayed as the standard error of the mean for 200 geometries for each cell-type and variant with each geometry having 50 independent exposures.

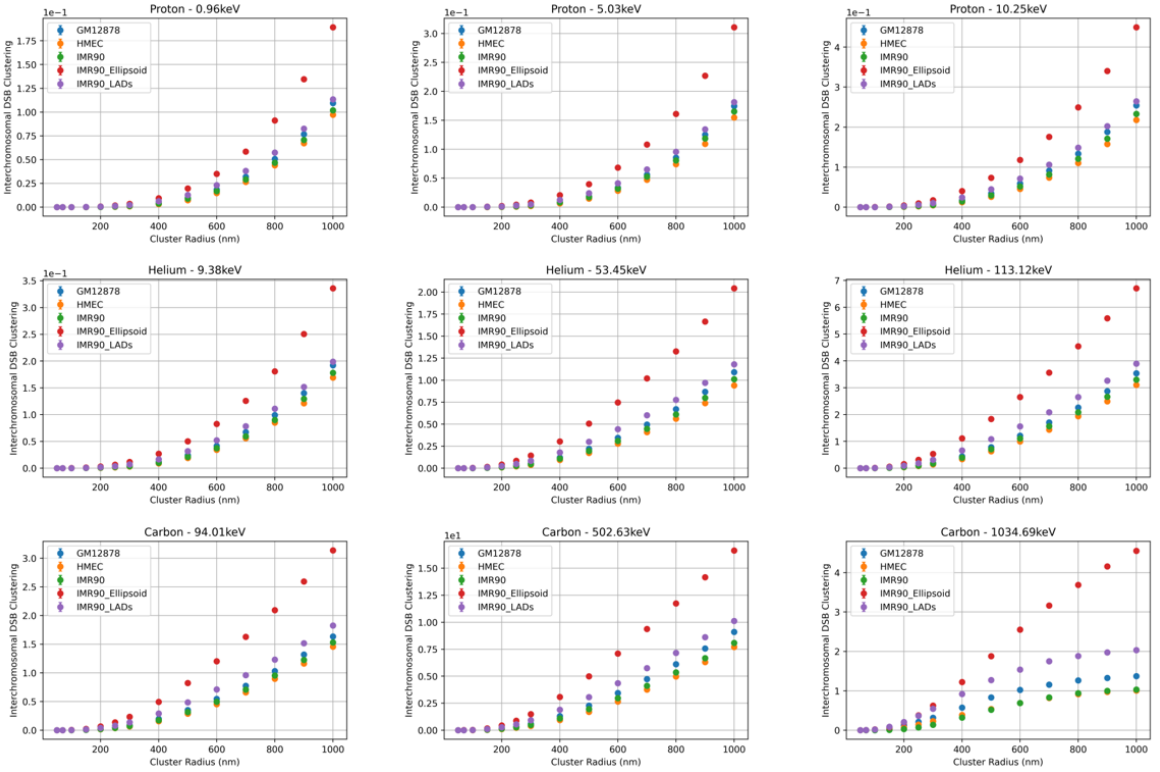

Figure B.8: **DSB Interchromosomal Clustering per Radiation Quality Plots.** Double-strand break inter-chromosomal clustering as a function of the cluster radius for all cell-types and variants. Error bars are displayed as the standard error of the mean for 200 geometries for each cell-type and variant with each geometry having 50 independent exposures.

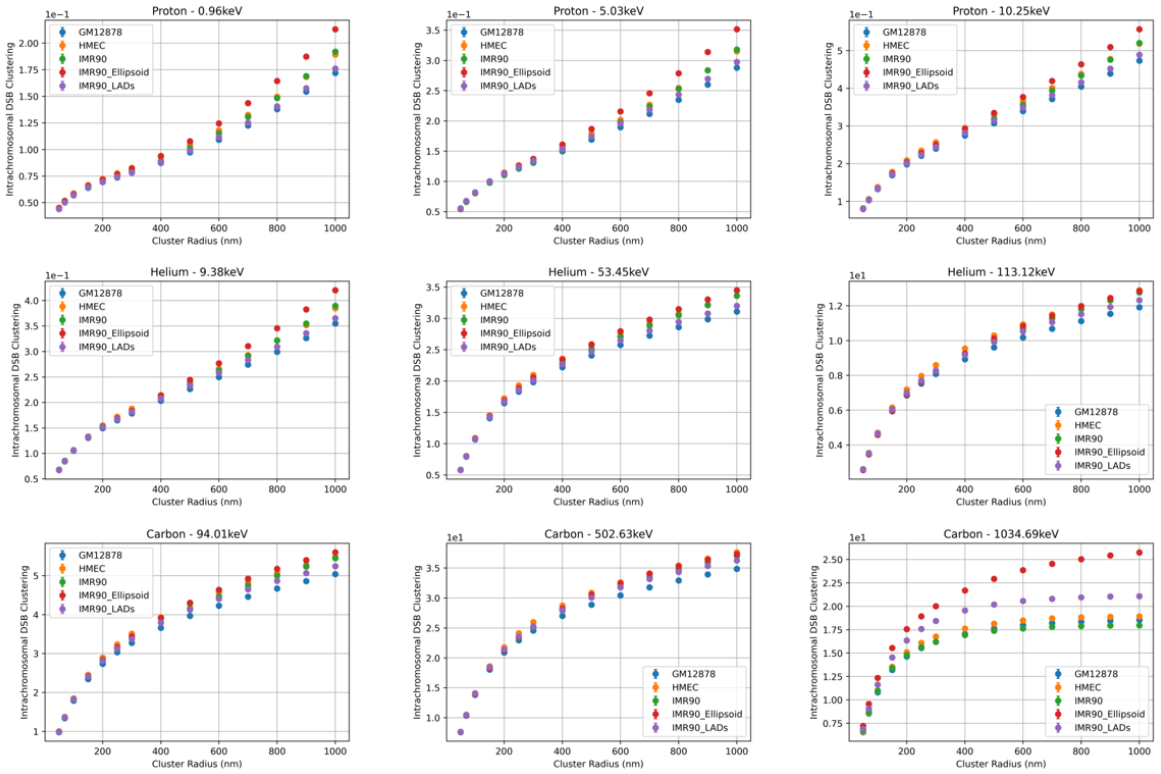

Figure B.9: **DSB Intrachromosomal Clustering per Radiation Quality Plots.** Double-strand break intra-chromosomal clustering as a function of the cluster radius for all cell-types and variants. Error bars are displayed as the standard error of the mean for 200 geometries for each cell-type and variant with each geometry having 50 independent exposures.

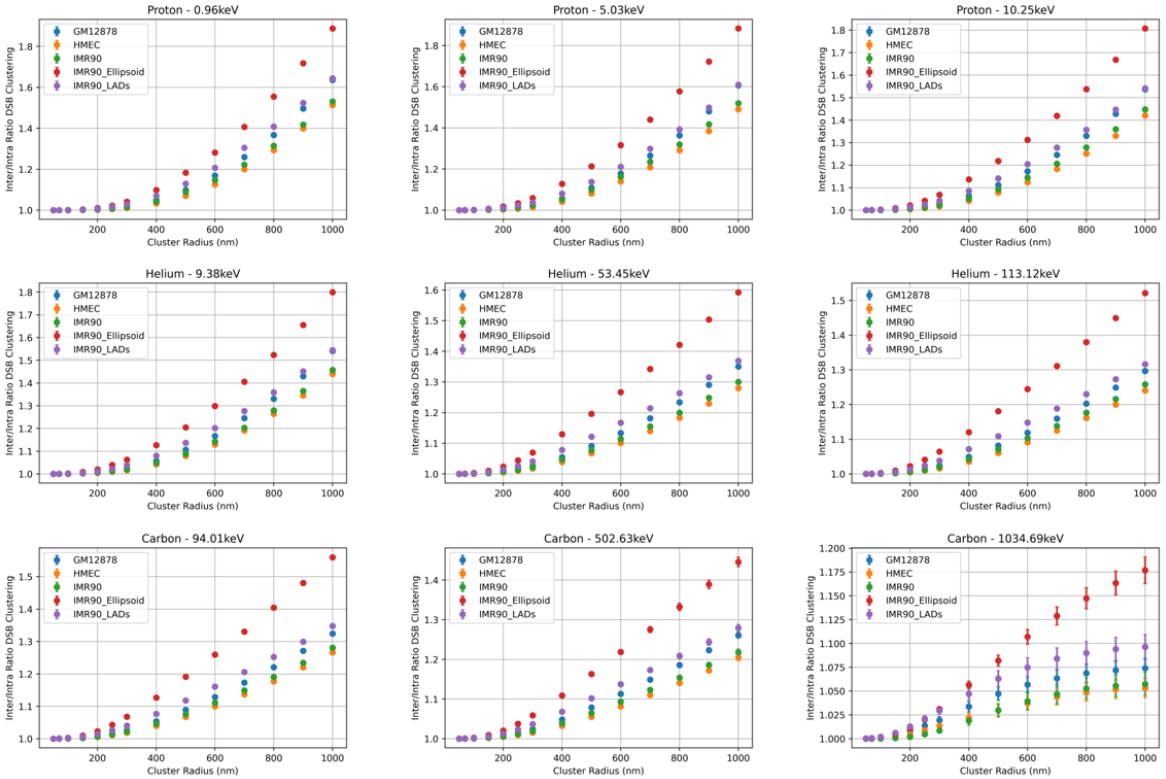

Figure B.10: **DSB Inter/Intra Clustering per Radiation Quality Plots.** Double-strand break inter/intra chromosomal clustering as a function of the cluster radius for all cell-types and variants.

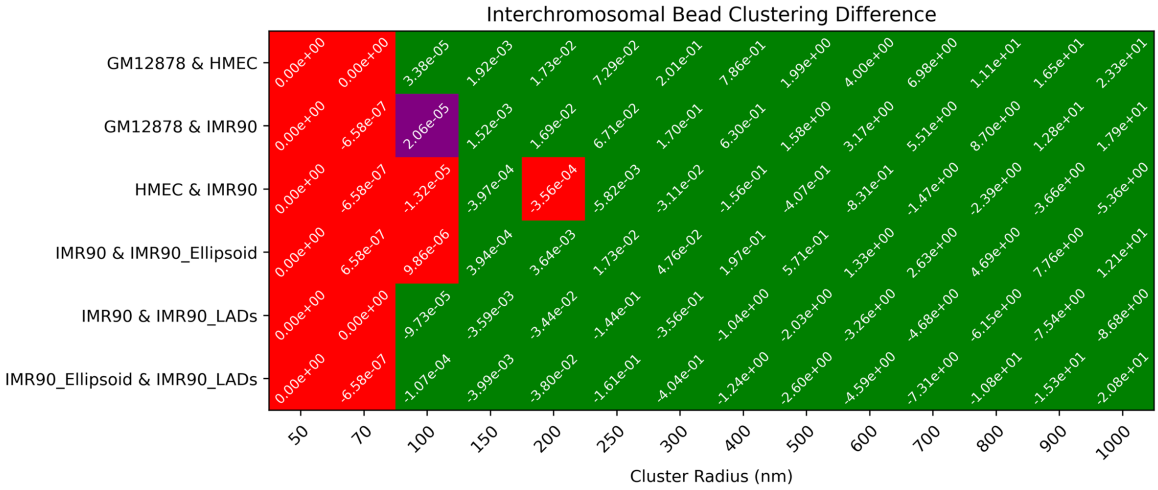

Figure B.11: **Differences between Interchromsomal Bead Clustering of Different Cell Lines.** False discovery rate adjusted P-values from a 2-sided Kolmogorov-Smirnov test on the interchromosomal bead clustering values for the different cell-types and variants. Colour coding for adjusted P-values at varying thresholds: red (P > 0.05), purple (0.05 > P > 0.01), yellow (0.01 > P > 0.001) and green (P < 0.001). In this case, distributions with adjusted P-values < 0.05 will be considered as having significant statistical difference to one another. Each of the tested distributions had 200 geometries per cell-type or variant group.

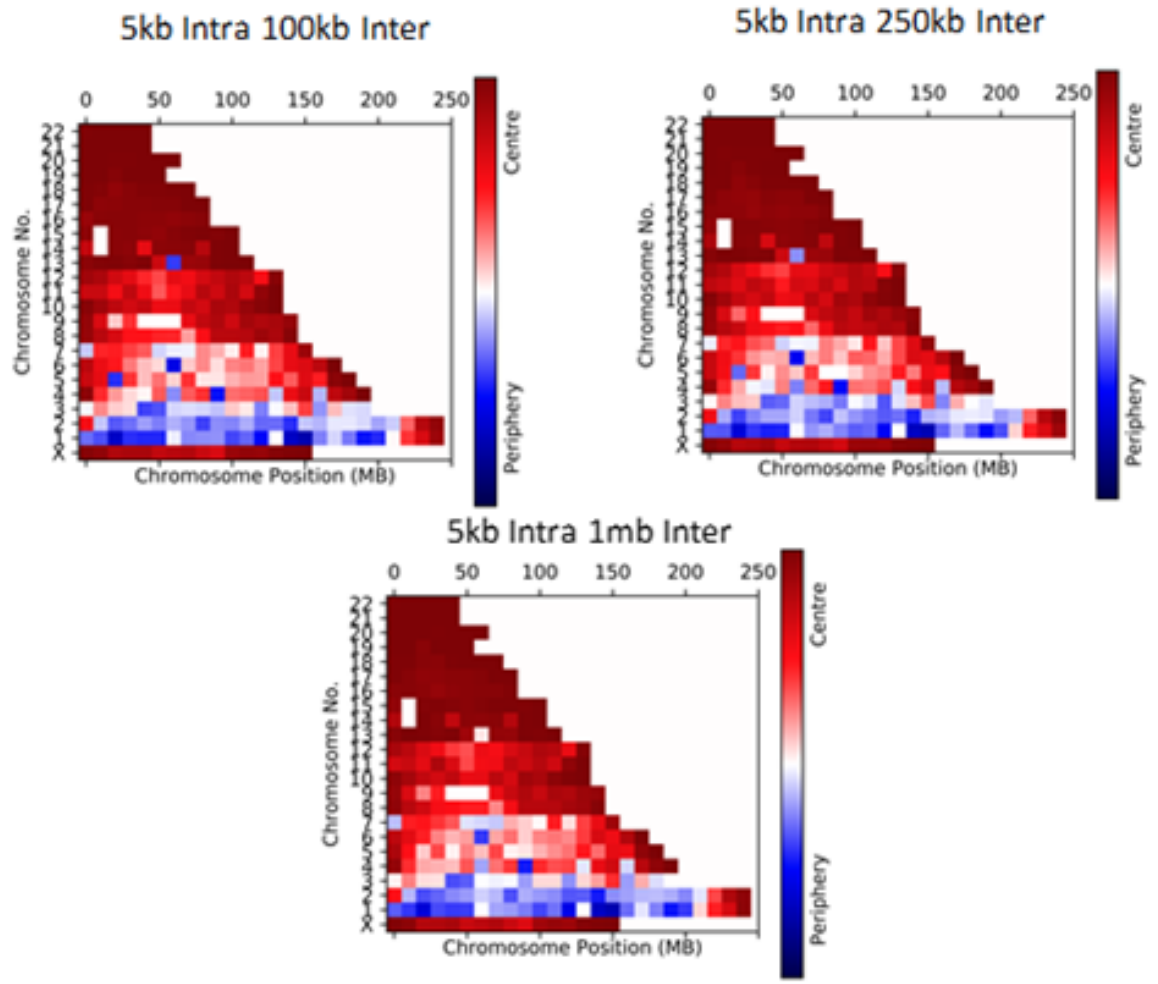

Figure B.12: **Nuclear Positioning of Beads at Varying Hi-C Contact Resolutions.** Bead positioning between periphery and central locations for a range of different interchromosomal contact resolutions at the finest available intrachromosomal contact resolution. Each category consists of 200 geometries created from the corresponding gtrack file created from using different analysis resolutions.

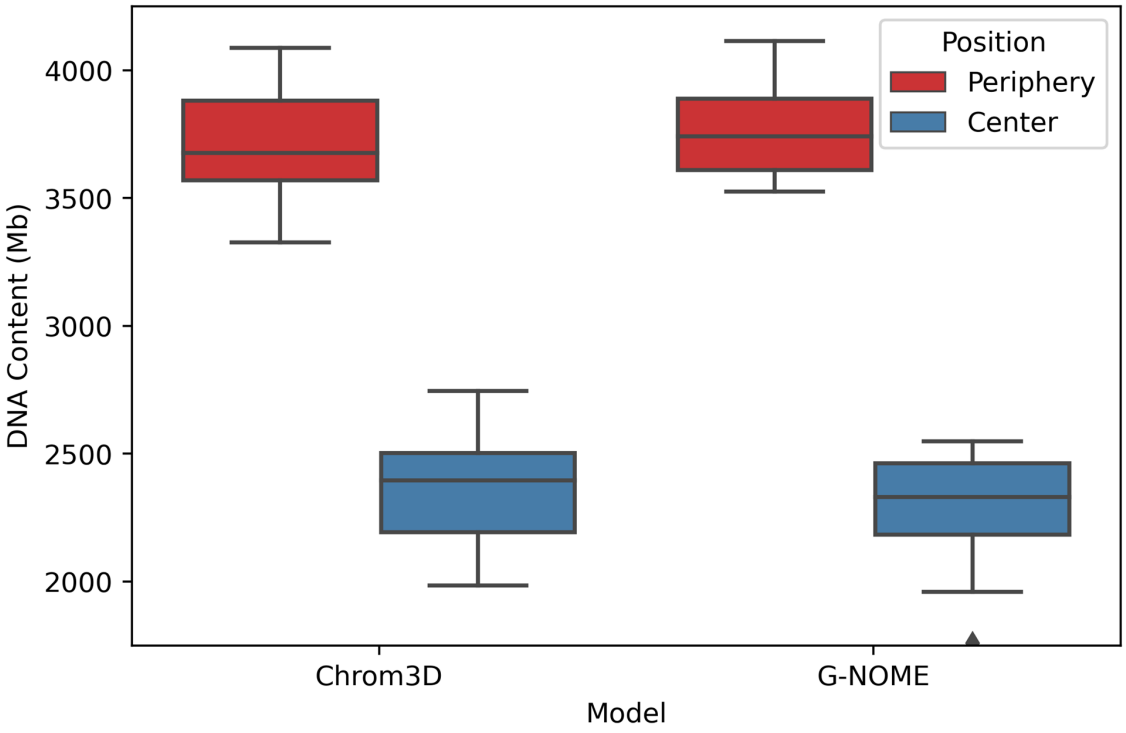

Figure B.13: **DNA content position model comparison.** Box plots of the DNA content positioned either in the peripheral half or central half of the cell nucleus volume. These results are for 50 geometries from G-NOME and 50 geometries from Chrom3D (v1.0.2). In both models the same input IMR90 noLADs gtrack file was optimised for 1 million iterations, 5-micron nuclear radius and 0.15 occupancy volume.

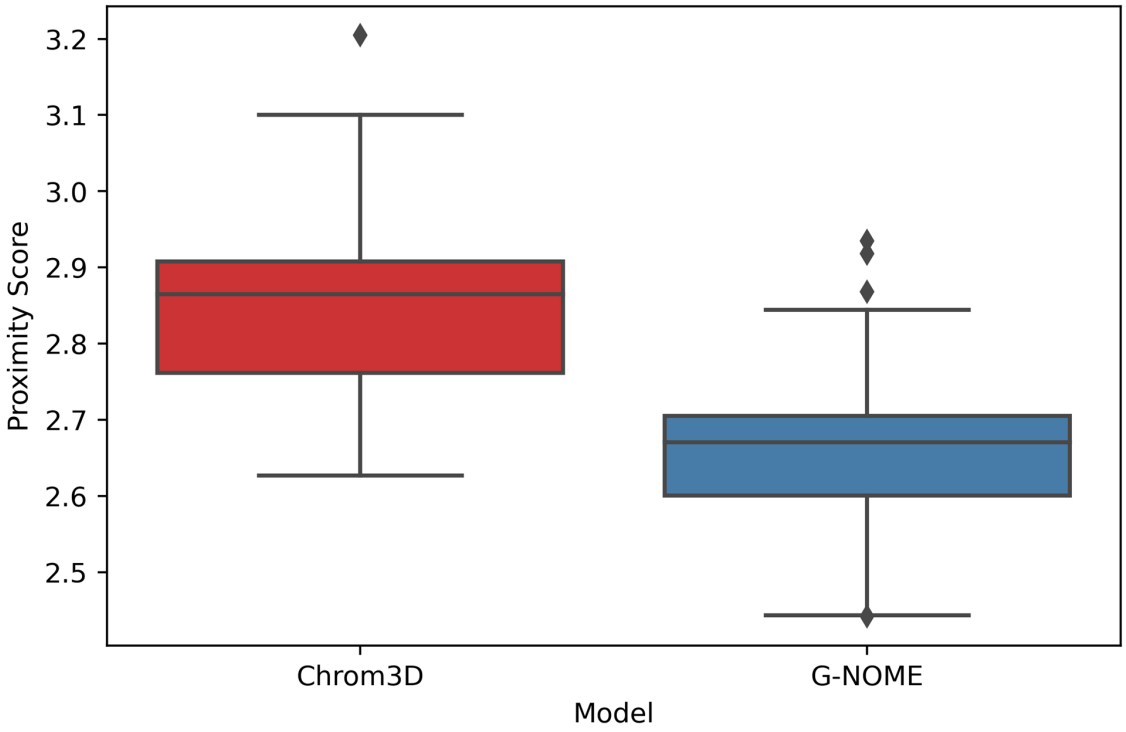

Figure B.14: **Proximity score model comparison.** Box plots of the proximity scores which is the average Euclidean distance between TADs which have a constraint to be proximal other TADs (lower value indicates a better optimisation of the contact constraints). To put these differences into perspective for a randomly distributed geometry where the proximity score is approximately 12. These results are for 50 geometries from G-NOME and 50 geometries from Chrom3D (v1.0.2). In both models the same input IMR90 noLADs gtrack file was optimised for 1 million iterations, 5-micron nuclear radius and 0.15 occupancy volume.

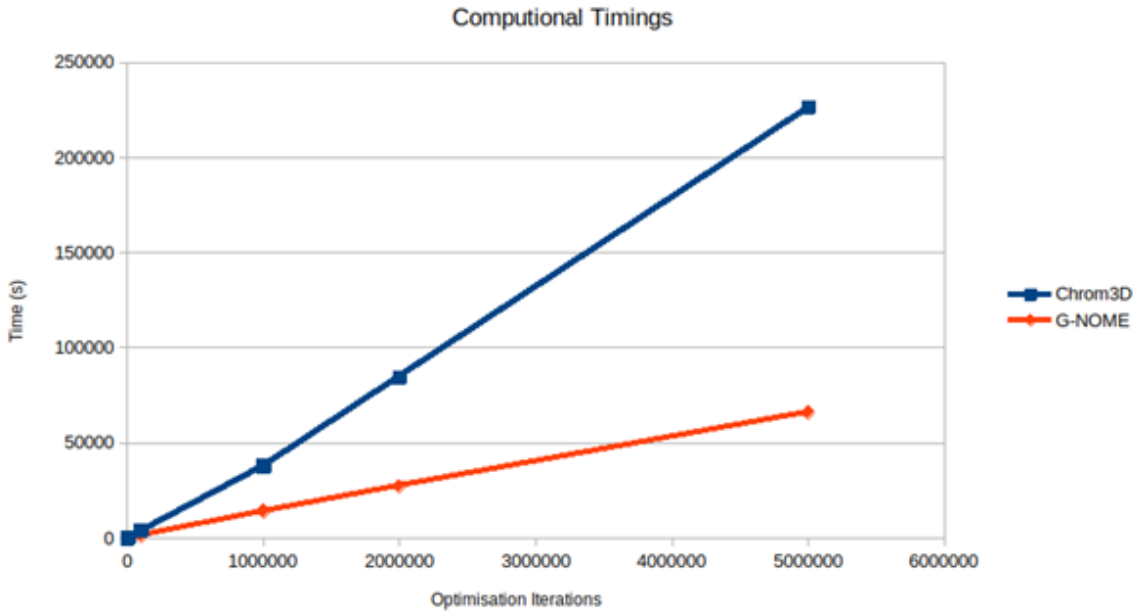

Figure B.15: **Nominal evaluation of speed performance.** Timing performance for a nominal single IMR90 spherical cell 3D geometry generation using both G-NOME and Chrom3D (v1.0.2). In this case for 2 million iterations (the number used for the evaluation of different cell types) is 23.5 hours in Chrom3D and 7.6 hours in G-NOME.

Appendix C

Supplementary Material for Ch. 5 - A computational approach to quantifying miscounting of radiation-induced double-strand break immunofluorescent foci

PSF central-axis full width half maximum in μm (X,Y,Z axis)					
	10x	**20x**	**40x**	**63x**	**100x**
Airyscan	0.72, 0.54, 4.97	0.37, 0.28, 1.7	0.26, 0.18, 0.78	0.25, 0.19, 0.68	0.25, 0.19, 0.61
gSTED	–	0.52, 0.51, 1.29*	0.32, 0.5, 0.97	0.2, 0.22, 1.01	0.2, 0.2, 0.84
Lowlight	–	0.72, 0.69, 3.68	0.49, 0.43, 1.97	–	0.27, 0.24, 0.74
MultiPhoton	1.0, 0.94, 9.9	0.41, 0.43, 2.91**	0.32, 0.34, 0.98	0.23, 0.31, 0.62	–
Phenix	–	1.12, 1.14, 3.37	0.56, 0.62, 1.59	0.4, 0.41, 1.42	–
STED	–	0.54, 0.45, 9.34	0.26, 0.32, 2.25	0.2, 0.29, 1.32	0.2, 0.2, 0.77

Table C.1: The FWHM of the PSF intensity along each central axis. *23x magnification not 20x. **25x magnification not 20x. The STED microscope also has a PSF at 25x which has the following values (0.35,0.38,2.85) for X,Y,Z axis respectively.

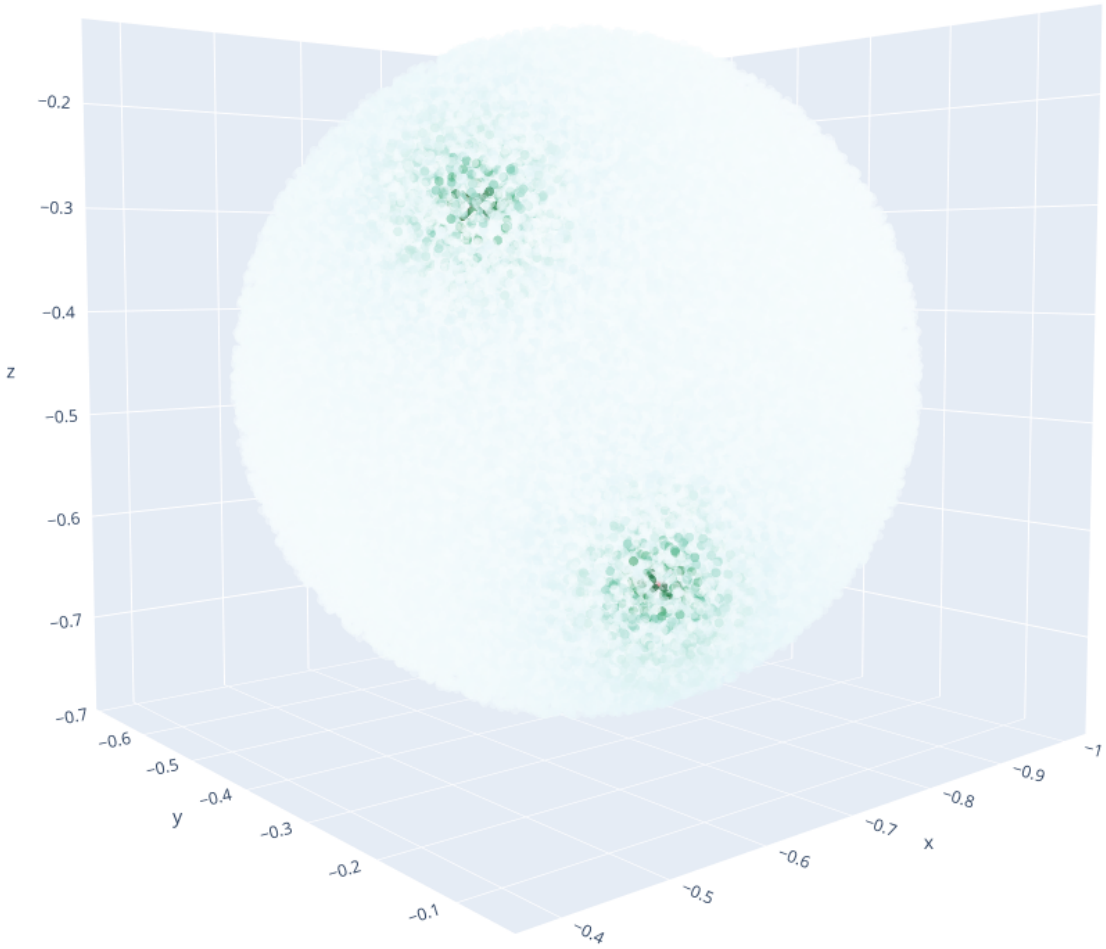

Figure C.1: Activation spread of the H2AX histones surrounding two DSBs (red crosses) for a single topological associated domain (TAD). Each small sphere represents an H2AX histone where no activation is displayed by the sphere being white and activation is displayed as an increasing green colour. Activation values are calculated based on Chip-Seq measurements Arnould et al., fitted using a Cauchy-Lorentz distribution. The positions of these histones along with the activation value is used by PyFoci in the creation of γ-H2AX marker computational microscope images.

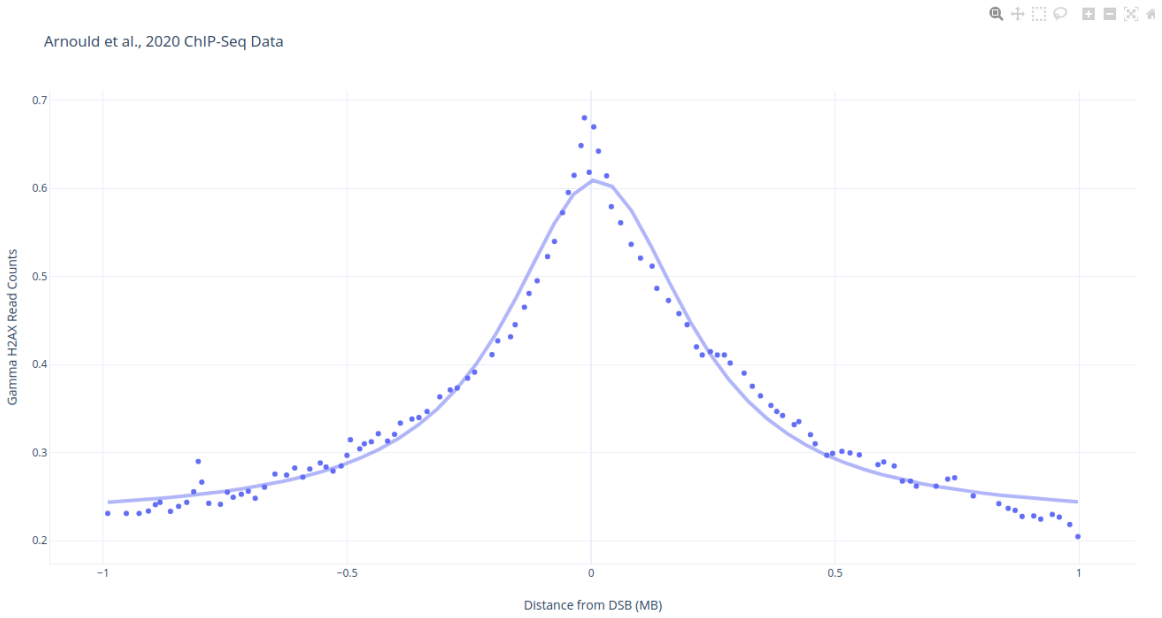

Figure C.2: The Cauchy-Lorentz distribution fit (blue-solid line) for γ-H2AX read counts based on the Chip-Seq measurements (blue dots) by Arnould et al. This work uses the γ-H2AX read counts as a surrogate for histone activation.

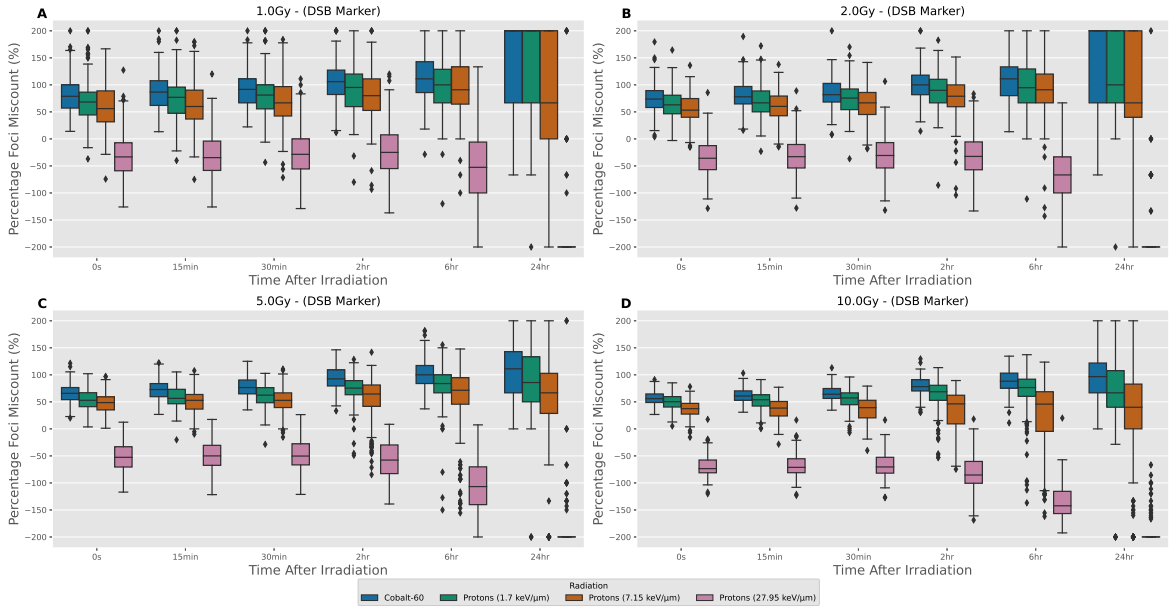

Figure C.3: Distributions of the percentage difference between foci detected and the number of DSBs within the central microscope slice of the cell nucleus when using a **DSB marker of florescence**. The percentage difference is calculated as the $(counted - simulated/((counted + simulated) * 0.5) * 100$. Values of -200% or +200% were checked and correspond to either a foci not being detected due to a low threshold value or at least one break detected when no breaks are present within confocal slice respectively. Each panel is the same four radiation types and six time points being compared for different radiation doses, where A, B, C, D correspond to 1.0, 2.0, 5.0, 10.0 Gy respectively. All microscope images have been emulated using the Airyscan 63x point spread function.

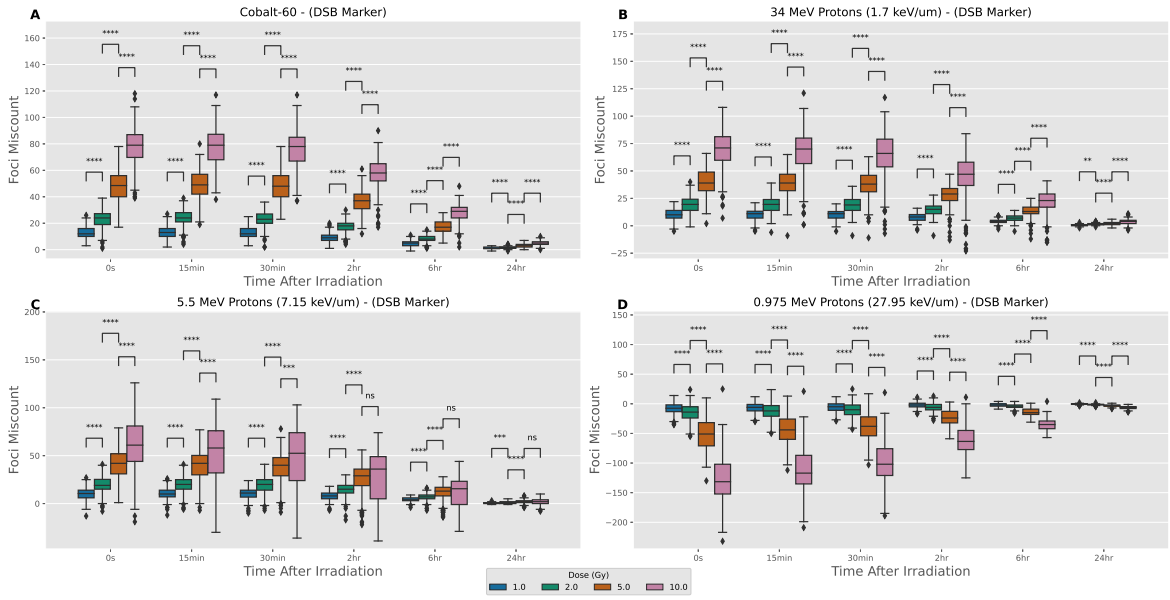

Figure C.4: Distributions of miscounting between foci detected and the number of DSBs within the central microscope slice of the cell nucleus when using a **DSB marker of florescence**. Each panel is the same four doses and six time points being compared for different radiation types, where A, B, C, D correspond to Colbalt-60, 34 MeV Protons, 5.5 MeV Protons, 0.975 MeV protons respectively. Mann-Whitney test between different radiation types at each time point and dose to highlight statistically significant differences. P-values have been adjusted using the Bonferroni correction. Significance notation refers to the following thresholds: ns=P>0.05, *=P>0.01, **=P>1e-3, ***=P>1e-4, ****=P<1e-4. There are 200 samples for each box-plot. All microscope images have been emulated using the Airyscan 63x point spread function.

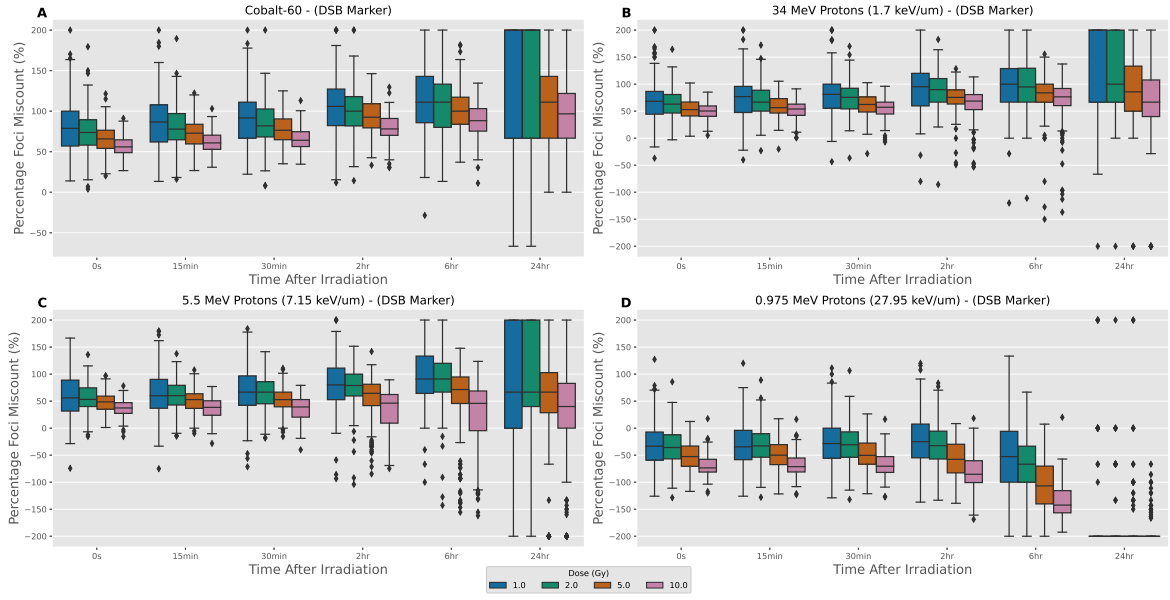

Figure C.5: Distributions of the percentage difference between foci detected and the number of DSBs within the central microscope slice of the cell nucleus when using a **DSB marker of florescence**. The percentage difference is calculated as the $(counted - simulated/((counted + simulated) * 0.5) * 100$. Values of -200% or +200% were checked and correspond to either a foci not being detected due to a low threshold value or at least one break detected when no breaks are present within confocal slice respectively. Each panel is the same four doses and six time points being compared for different radiation types, where A, B, C, D correspond to Colbalt-60, 34 MeV Protons, 5.5 MeV Protons, 0.975 MeV protons respectively. All microscope images have been emulated using the Airyscan 63x point spread function.

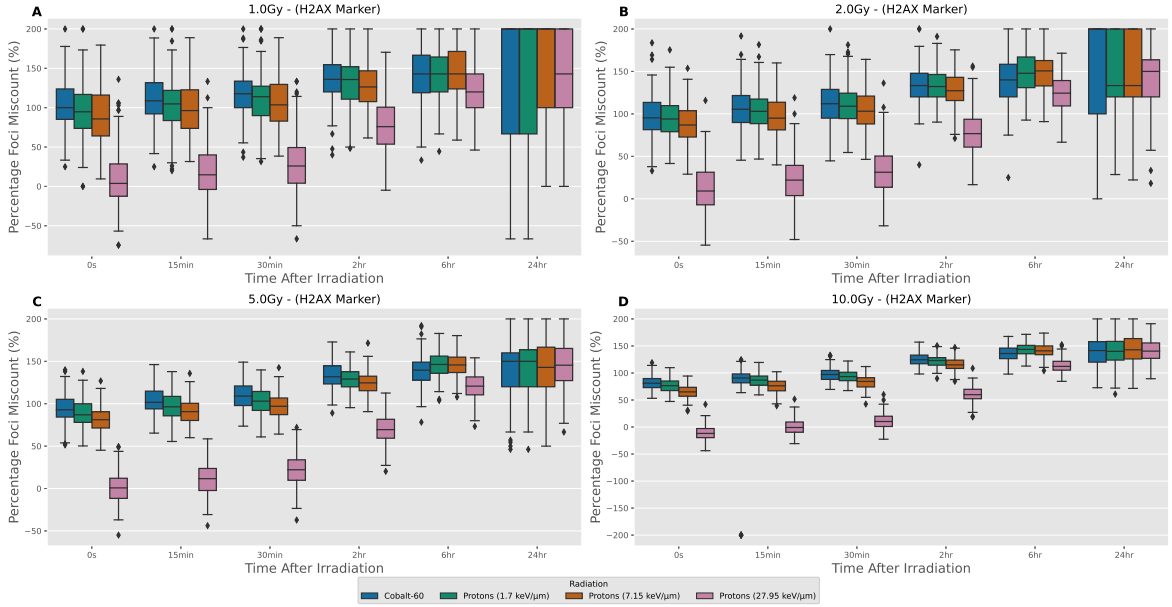

Figure C.6: Distributions of the percentage difference between foci detected and the number of DSBs within the central microscope slice of the cell nucleus when using a γ**-H2AX marker of florescence**. The percentage difference is calculated as the $(counted - simulated/((counted + simulated) * 0.5) * 100$. Values of -200% or +200% were checked and correspond to either a foci not being detected due to a low threshold value or at least one break detected when no breaks are present within confocal slice respectively. Each panel is the same four radiation types and six time points being compared for different radiation doses, where A, B, C, D correspond to 1.0, 2.0, 5.0, 10.0 Gy respectively. All microscope images have been emulated using the Airyscan 63x point spread function.

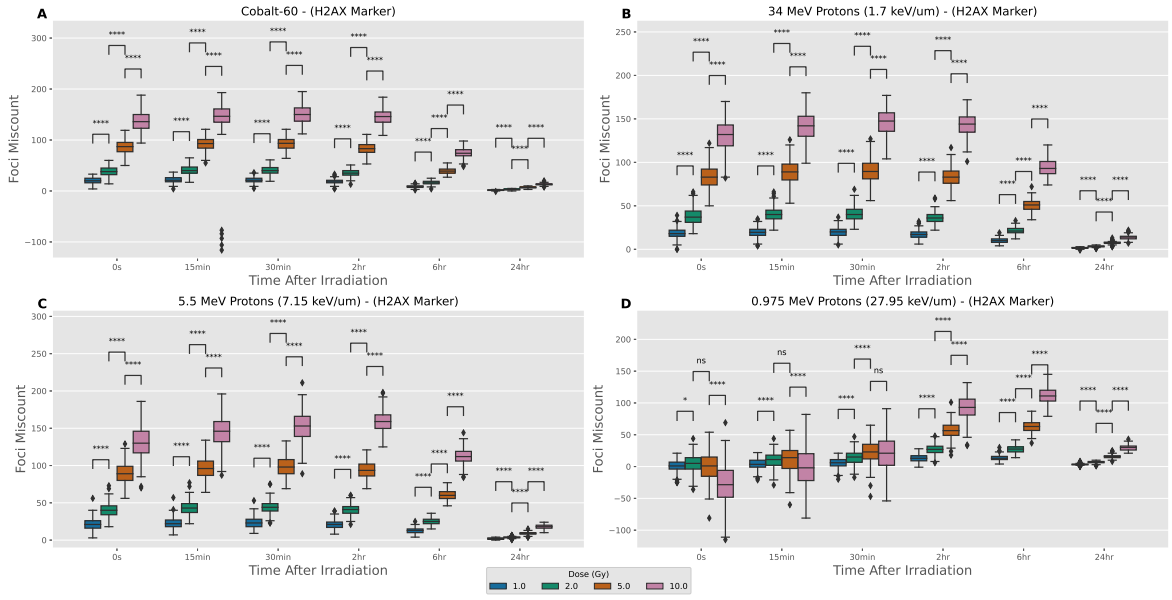

Figure C.7: Distributions of miscounting between foci detected and the number of DSBs within the central microscope slice of the cell nucleus when using a γ**-H2AX marker of florescence**. Each panel is the same four doses and six time points being compared for different radiation types, where A, B, C, D correspond to Colbalt-60, 34 MeV Protons, 5.5 MeV Protons, 0.975 MeV protons respectively. Mann-Whitney test between different radiation types at each time point and dose to highlight statistically significant differences. P-values have been adjusted using the Bonferroni correction. Significance notation refers to the following thresholds: ns=P>0.05, *=P>0.01, **=P>1e-3, ***=P>1e-4, ****=P<1e-4. There are 200 samples for each box-plot. All microscope images have been emulated using the Airyscan 63x point spread function.

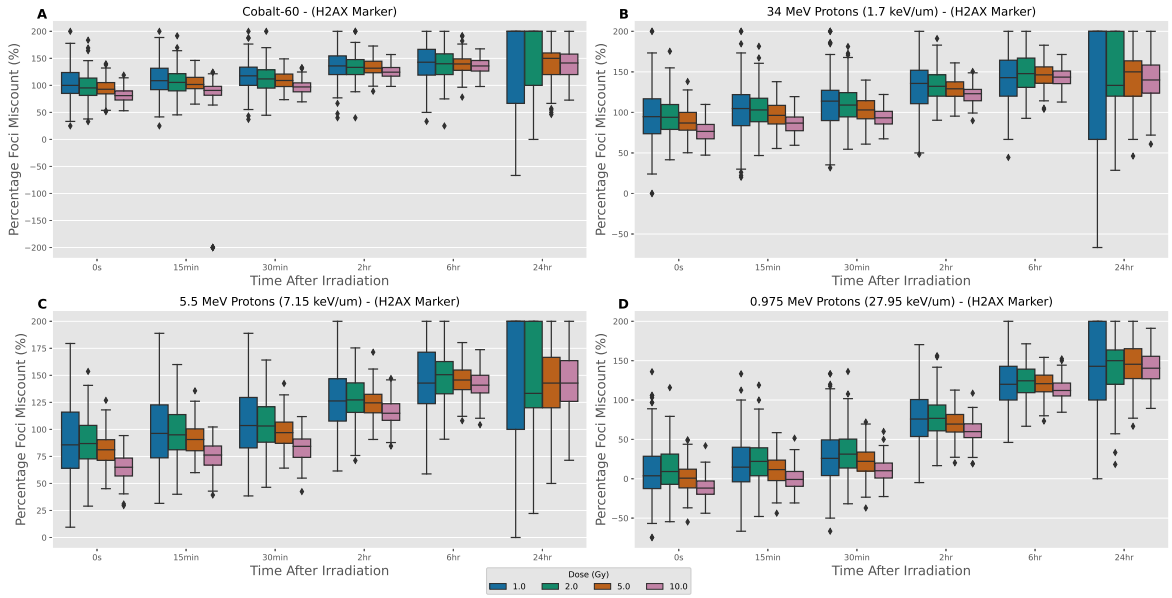

Figure C.8: Distributions of the percentage difference between foci detected and the number of DSBs within the central microscope slice of the cell nucleus when using a **γ-H2AX marker of florescence**. The percentage difference is calculated as the $(counted - simulated/((counted + simulated) * 0.5) * 100$. Values of -200% or +200% were checked and correspond to either a foci not being detected due to a low threshold value or at least one break detected when no breaks are present within confocal slice respectively. Each panel is the same four doses and six time points being compared for different radiation types, where A, B, C, D correspond to Colbalt-60, 34 MeV Protons, 5.5 MeV Protons, 0.975 MeV protons respectively. All microscope images have been emulated using the Airyscan 63x point spread function.

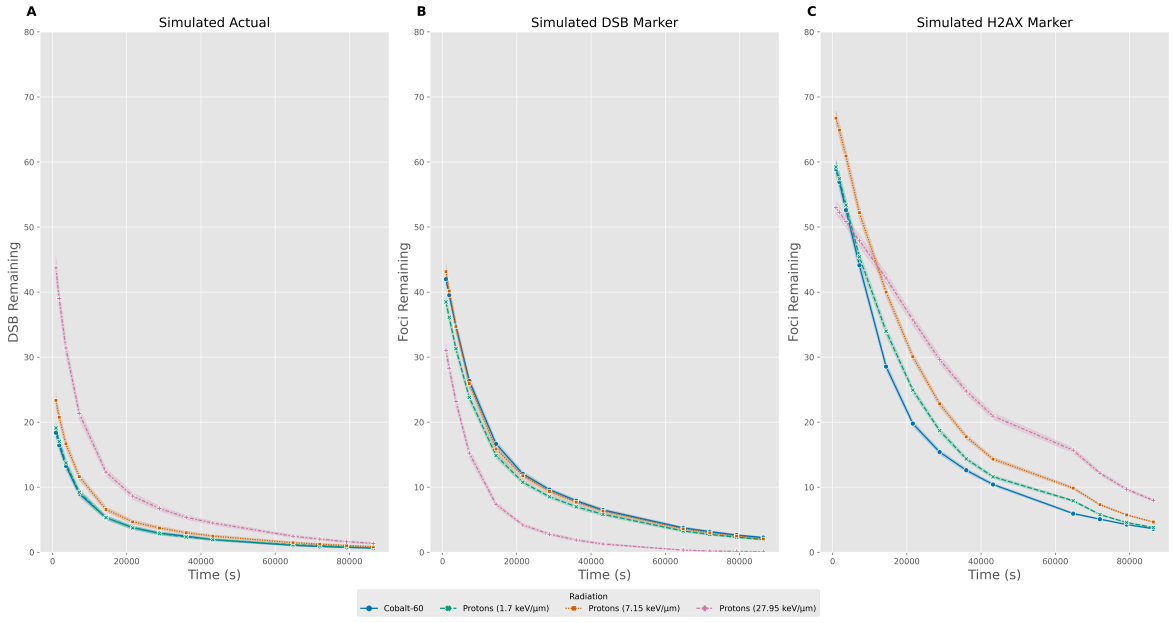

Figure C.9: Repair kinetics comparison between cells irradiated with four different incident radiation conditions. A) shows the repair kinetics of the actual number of DSBs within the evaluated microscope slice. The repair kinetics applied is kept constant across all radiation types to separate the impact of miscounting alone. B) the repair kinetics when calculated from the foci detected in the DSB marker microscope images. C) the repair kinetics when calculated from the foci detected in the γ-H2AX marker microscope images. All microscope images have been emulated using the as the Airyscan 63x and the results shown are for 2Gy of radiation.

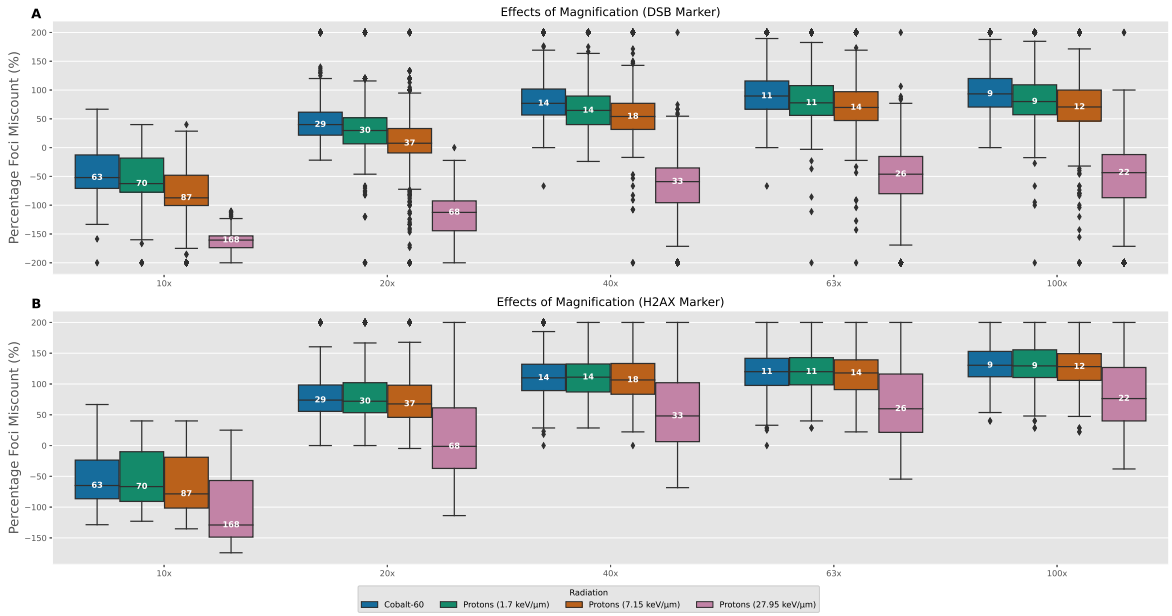

Figure C.10: Comparison of percentage miscounting at different Airyscan magnifications. The value shown in white within each box-plot is the actual number of DSBs within the evaluated microscope slice to give perspective on the amount of miscounting each percentage corresponds to. A) are the results when using a DSB marker. B) are the results when using an γ-H2AX marker. All microscope images are at 2 Gy dose and for the all 6 time points (0s, 15mins, 30 mins, 2hrs, 6hrs and 24hrs). There are 1200 samples for each box-plot.

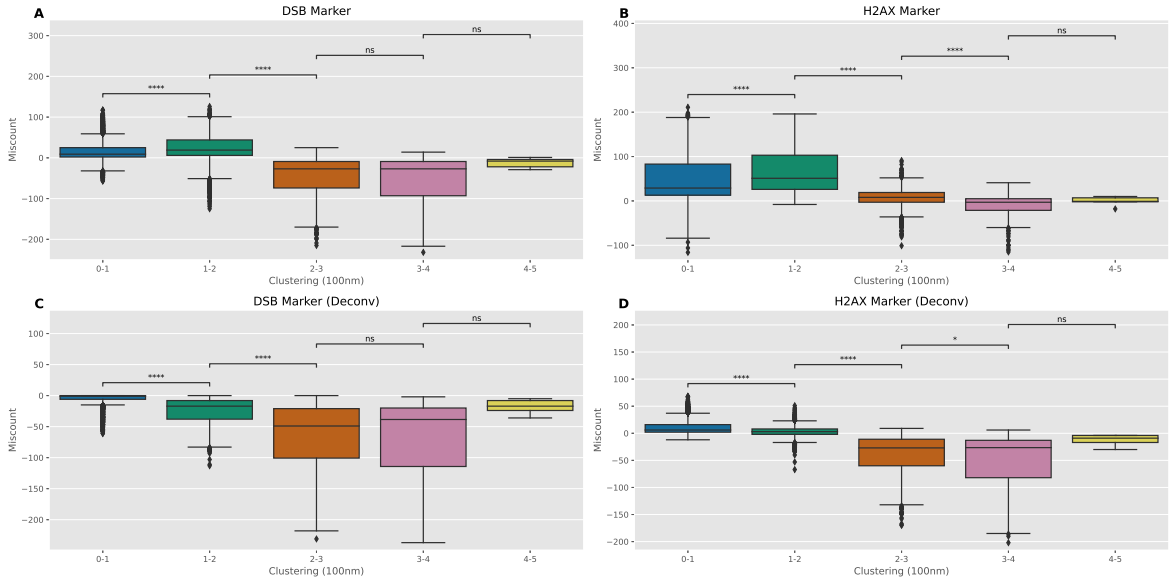

Figure C.11: Categorising the magnitude of miscounting as a function of DSB clustering. The values of clustering are calculated based on the known amount of DSBs within the cell nucleus at a given time point. Clustering equates to the average number of DSBs within proximity to each DSB within the simulated cell nucleus. Where proximity for this study is characterised as 100 nm. Clustering is increased for early time points, higher dose, higher LET and is decreased for the opposite. Therefore, the clustering combines time, dose and radiation type parameters into a single metric. Panel A and B are for the DSB and γ-H2AX markers respectively. Panel C and D are for the DSB and γ-H2AX markers with perfect deconvolution respectively. Mann-Whitney test between categorised clustering values to highlight statistically significant differences. P-values have been adjusted using the Bonferroni correction. Significance notation refers to the following thresholds: ns=P>0.05, *=P>0.01, **=P>1e-3, ***=P>1e-4, ****=P<1e-4. All microscope images have been emulated using the Airyscan 63x point spread function.

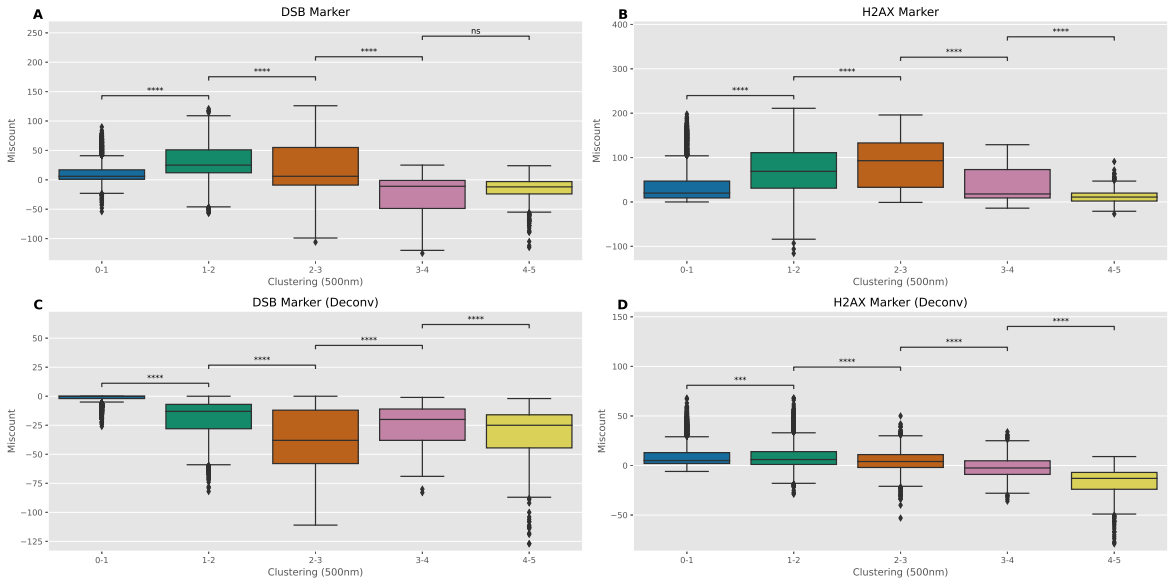

Figure C.12: Categorising the magnitude of miscounting as a function of DSB clustering. The values of clustering are calculated based on the known amount of DSBs within the cell nucleus at a given time point. Clustering equates to the average number of DSBs within proximity to each DSB within the simulated cell nucleus. Where proximity for this study is characterised as 500 nm. Clustering is increased for early time points, higher dose, higher LET and is decreased for the opposite. Therefore, the clustering combines time, dose and radiation type parameters into a single metric. Panel A and B are for the DSB and γ-H2AX markers respectively. Panel C and D are for the DSB and γ-H2AX markers with perfect deconvolution respectively. Mann-Whitney test between categorised clustering values to highlight statistically significant differences. P-values have been adjusted using the Bonferroni correction. Significance notation refers to the following thresholds: ns=P>0.05, *=P>0.01, **=P>1e-3, ***=P>1e-4, ****=P<1e-4. All microscope images have been emulated using the Airyscan 63x point spread function.