

**Determining the molecular and pathological responses
of neoadjuvant radiotherapy on soft tissue sarcomas**

Steven William George Nottley

University College London
and
The Francis Crick Institute

PhD Supervisors: Dr Nischalan Pillay and Prof Peter Van Loo

A thesis submitted for the degree of
Doctor of Philosophy
University College London

February 2025

Declaration

I Steven Nottley confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Soft tissue sarcomas (STS) are a heterogeneous group of mesenchymal malignancies, often treated with neoadjuvant radiotherapy (RT) to improve local control. However, the genomic impact of RT on STS remains poorly understood. This thesis investigates the mutational and transcriptomic responses of STS to RT using whole-exome sequencing (WES), RNA sequencing (RNAseq), and the highly sensitive NanoSeq technology.

Initial WES analyses of pre- and post-RT tumour samples revealed a low tumour mutational burden (TMB) with no significant increase in single nucleotide variants (SNVs) or small insertions and deletions (indels) post-treatment. Given the known role of RT in inducing DNA damage, I hypothesised that conventional WES might lack the sensitivity to detect low-frequency mutations. To address this, I employed NanoSeq, a duplex sequencing approach capable of detecting rare mutations with unprecedented accuracy. Analysis of paired pre- and post-RT samples using NanoSeq demonstrated a significant increase in indels and a shift toward microhomology-mediated end joining, suggesting a mutational footprint of RT previously undetectable with standard sequencing.

Transcriptomic analysis revealed differentially expressed genes and pathways when comparing pre- and post-RT samples, shedding light on the molecular response to RT and identifying potential biomarkers of disease. A machine learning model trained on gene expression data successfully distinguished patients with favourable vs. poor post-RT outcomes.

These findings provide novel insights into the genomic and transcriptomic effects of RT on STS. By leveraging high-resolution sequencing technologies, this work enhances our understanding of RT-induced mutagenesis and lays the foundation for improved patient stratification based on molecular response. This research identifies potential biomarkers of disease progression and therapeutic targets, which, with further research and validation, could inform both post-RT surveillance strategies and the development of adjuvant treatment approaches in clinical practice.

Impact Statement

Radiotherapy is a widely used treatment for soft tissue sarcomas (STS), yet its long-term effects on the tumour genome and transcriptome remain incompletely understood. This thesis applies cutting-edge sequencing methodologies to investigate how neoadjuvant RT reshapes the molecular landscape of STS, with significant implications for both clinical practice and future research.

From an academic perspective, this work advances our understanding of the mutational consequences of RT. The application of NanoSeq—a highly sensitive duplex sequencing approach—revealed an increased burden of low-frequency mutations, particularly small insertions and deletions, which were undetectable using conventional whole-exome sequencing (WES). This demonstrates that traditional sequencing approaches may underestimate the extent of genomic alterations caused by treatment. The observed shift toward microhomology-mediated repair mechanisms aligns with known mutational processes induced by DNA damage, providing new avenues for investigating RT resistance mechanisms.

Beyond academia, this research has potential clinical implications. The identification of differentially expressed genes and pathway alterations post-RT offers valuable biomarkers that could predict patient outcomes. A machine learning model trained on gene expression data successfully stratified patients based on their likelihood of disease progression, highlighting the feasibility of integrating molecular profiling into clinical decision-making. These findings could contribute to the development of biomarker-driven treatment strategies, allowing clinicians to personalise post-RT surveillance and therapeutic interventions.

At a broader level, this thesis contributes to the growing field of radiogenomics, with potential applications in other tumour types treated with RT. By refining our understanding of RT-induced molecular changes, this research supports efforts to mitigate therapy-associated risks, optimise patient monitoring, and inform the design of novel therapeutic strategies that leverage molecular vulnerabilities induced by RT.

In summary, this work bridges the gap between genomic research and clinical application, providing a molecular framework for understanding RT response in STS. Through improved detection of RT-induced mutations and identification of transcriptomic biomarkers that could be used for patient risk stratification, this research has the potential to influence both future studies and clinical practice, ultimately improving outcomes for STS patients.

Acknowledgements

This PhD would not have been possible without the support of many people, to whom I am deeply grateful.

Firstly, to my wonderful supervisors, Nischalan Pillay and Peter Van Loo, for conceiving this project and entrusting me with it. Nischalan, thank you for your time, care, and mentorship. You have always gone above and beyond in the support you have given me - both academically and personally - for which I will be forever grateful. I have grown as a scientist, and I am certain I will be a better pathologist because of this experience and training. Peter, thank you for your encouragement, insightful advice, and invaluable support in developing both the ideas for this project and my bioinformatics skills.

I would also like to thank the members of my thesis committee: Jasmin Fisher, Maria Hawkins, and Mahbubl Ahmed for your input and encouragement throughout this journey.

To the members of the Pillay and Van Loo labs, you have made this experience truly memorable. Thank you for all your support and help over the years: Amy Bowes, Sara Waise, Chris Steele, Akanksha Farswan, Christie Davies, Tom Butters, Simon Haeffliger, Nana Mensah, Tom Lesluyes, Haixi Yan, Toby Baker, Carla Castignani, Dolapo, Shadi Hames-Fathi, Maxime Tarabichi, and Jonas Demeulemeester.

To the Royal National Orthopaedic Hospital Pathology Department team, especially Adrienne Flannagan, Fernanda Amary, Roberto Tirabosco, and Hongtao Ye.

I am very grateful to Iñigo Martincorena and Federico Abascal for their generosity in applying their NanoSeq technique to patients in this cohort and for the informative discussions we had regarding data analysis.

A special thanks to Cancer Research UK for funding my Clinical Research Training Fellowship.

I am deeply grateful to the patients who participated in this study - without whom this work would not have been possible.

To my friends, who have been a constant throughout this journey: Elissa Franceschi, Katie Bend, Beccie Haigh, Ross Tate, Daniel Conway, Daniel Girling, Sunny Sharma, James Geer, Max Tobin, Dillon Steele, Brad Coan, Ajay Chamba, Jamie Ramacciotti, Julia Dray, Ben Ifrah, Dan Davies, Alex Shavick, Jonathan Gareze, and Anna KT Porko. You all bring me so much joy and have been incredibly supportive throughout this process.

Thank you to my family, especially my parents, for all their support. A particular thanks to my Aunt Judy for her incredible support during the writing of this thesis - I am forever grateful.

Finally, this PhD would not have come about if not for my brilliantly talented fiancé, Felipe Rodrigues, who inspired and encouraged me to embark on this journey with him. Your unwavering belief in me and your continuous love and support through all the challenges have helped me overcome every hurdle. The submission of this thesis falls on the seventh anniversary of the day we met - you have changed my whole perspective on life, and I thank you for all the motivation and love.

Dedication

This research has been particularly motivated by the tragic loss of two dear friends to sarcoma.

Edward Showler was my classmate in medical school, and we graduated together in 2013. He sadly passed away from Clear Cell Sarcoma in 2017 at the age of 28. In his memory, his parents, John and Ellie Showler, established a foundation that has since raised nearly a million pounds for Clear Cell Sarcoma research. Much of this funding has directly supported work in the Pillay lab, where I have seen firsthand how their generosity has advanced our understanding of his disease and is contributing towards finding potential treatments. Ed would be very proud.

Jack Holmes passed away in 2024 at the age of 34 from CIC-DUX4 sarcoma. Following his diagnosis in 2023, he was determined to make a difference by raising funds for sarcoma research. Together, we co-founded Transcend Sarcoma, a charity dedicated to funding research into translocation-associated sarcomas. This would not have been possible without the incredible dedication of his wife, Rachel Kapo, his brother, Mark Holmes, and his friends, Harry Fothergill and Assad Rashid.

To Ed and Jack this thesis is dedicated to your memory.

www.edwardshowlerfoundation.com

www.transcendsarcoma.com

Table of Contents

Declaration..	2
Abstract	3
Impact Statement	4
Acknowledgements	6
Dedication...	8
Table of Contents	9
Table of figures	13
List of tables	16
Abbreviations	17
 Chapter 1.Introduction	 19
1.1 Introduction to soft tissue sarcomas	19
1.2 The challenge to provide a timely and accurate diagnosis	23
1.3 The genomics of soft tissue sarcomas	26
1.3.1 Translocation associated sarcomas	26
1.3.2 Complex genomic sarcomas	28
1.4 Risk factors for the development of soft tissue sarcomas	32
1.4.1 Inherited cancer predisposition syndromes	32
1.4.2 Viral associations	33
1.4.3 Radiation induced sarcomas	34
1.5 The treatment of soft tissue sarcomas	38
1.5.1 Surgery	38
1.5.2 Chemotherapy	39
1.5.3 The history of radiotherapy	41
1.5.4 Preoperative vs. postoperative radiotherapy	44
1.5.5 Histology-specific considerations in radiotherapy	45
1.5.6 Current UK radiotherapy guidelines	46
1.5.7 Side effects of radiotherapy	47
1.6 Prognostic factors in soft tissue sarcoma	49
1.6.1 Histological grade	49
1.6.2 Tumour size	51
1.6.3 Anatomical site	52
1.6.4 Stage at diagnosis	52
1.6.5 Resection margin status	55
1.6.6 Patient age, performance status, and socioeconomic background	56
1.7 The biological effects of radiotherapy	57
1.7.1 Mutational signatures associated with radiotherapy	57
1.7.2 Radiation induced mutational signatures	58
1.7.3 Radiation-induced APOBEC mutagenesis	59
1.7.4 ID8 and DNA Repair pathways	59
1.8 The pathological assessment of the response to radiotherapy	61
1.8.1 Histopathological features of radiotherapy response	62
1.8.2 Correlation with radiological response criteria	62
1.8.3 Lack of consensus in pathological reporting	63
1.8.4 Implications for tissue processing and molecular analysis	63

1.8.5 Conclusion	63
1.9 Rationale and objectives of this thesis	65
Chapter 2. Materials & Methods.....	68
2.1 Ethics approval and patient sample selection.....	68
2.1.1 Radiotherapy treatment.....	70
2.1.2 Defining disease progression.....	70
2.2 Histological assessment of samples	74
2.3 Nuclei acid extraction.....	74
2.4 Sequencing protocols	74
2.5 Alignment, variant calling, and annotation	76
2.5.1 Computational Resources	76
2.5.2 Alignment and pre-processing	76
2.5.3 Variant calling	77
2.5.4 Panel of Normals (PON) generation	77
2.5.5 Functional annotation	78
2.5.6 Quality control and filtering.....	78
2.5.7 Consensus and manual review	79
2.5.8 Final filtering criteria	79
2.5.9 REVEL score annotation	80
2.6 COSMIC cancer genes.....	81
2.7 Copy number analysis	81
2.7.1 ASCAT	81
2.7.2 Copy Number analysis methodology	82
2.8 Nanoseq	84
2.9 Mutational signature analysis	84
2.10 Processing of RNAseq data	85
2.11 Differential Gene Expression analysis	85
2.12 Gene Set Enrichment Analysis.....	86
2.13 Gene Ontology analysis	86
2.14 PROGENy pathway activity analysis	86
2.15 Immune cell inference with xCell.....	87
2.16 Gene expression-based modelling of disease progression	87
2.17 Data management and storage.....	88
Chapter 3. Genomic responses to neoadjuvant radiotherapy in soft tissue sarcomas.	89
3.1 Introduction.....	89
3.2 The mutational landscape of the London Sarcoma Service cohort.....	91
3.2.1 Tumour mutational burden varies across and within different subtypes	91
3.2.2 Identification of somatic mutations within known cancer genes	94
3.2.3 Summary.....	107
3.3 Comparative analysis of somatic mutation burden in pre- and post-radiotherapy sarcoma samples.	108
3.3.1 Unpaired comparison of all pre- and post-radiotherapy samples. ...	110
3.3.2 Unpaired comparison of pre- and post-radiotherapy samples within sarcoma subtypes	111

3.3.3 Paired analysis of patients with matching pre- and post-radiotherapy samples.....	112
3.3.4 Assessment of the Indel-to-SNV and the Deletion-to-Insertion ratio.....	114
3.3.5 Case study of somatic mutation dynamics in patient 58: a longitudinal analysis across pre-, post-radiotherapy, and recurrence stages.	117
3.3.6 Summary.....	120
3.4 Comparison of copy number alterations pre- and post- radiotherapy....	121
3.4.1 Assessing Copy Number Alterations: methods and metrics.	122
3.4.2 Comparison of copy number alteration metrics across all pre- and post-radiotherapy samples.....	124
3.4.3 Subtype specific analysis of Copy Number Alteration metrics in pre- and post-radiotherapy samples.....	127
3.4.4 Comparative analysis of Copy Number Alteration metrics in paired pre- and post-radiotherapy samples.	129
3.4.5 Summary.....	131
3.5 Mutational signature analysis	132
3.5.1 Analysis of Single base substitution (SBS) signatures.....	132
3.5.2 Analysis of Indel (ID) signatures	137
3.5.3 Identification of mismatch repair deficiency	143
3.5.4 Summary.....	146
3.6 Using NanoSeq to more accurately interrogate the genomic response to radiotherapy.	147
3.6.1 Comparison of the frequency of somatic mutations pre- and post-radiotherapy.....	149
3.6.2 Assessment of the Indel-to-SNV and the Deletion-to-Insertion ratios.	151
3.6.3 Radiotherapy-induced shift towards microhomology-mediated DNA repair.....	153
3.6.4 Mutational signature analysis of NanoSeq cohort	155
3.6.5 Summary.....	159
3.7 Discussion of chapter 3	160
Chapter 4. The transcriptomic response to neoadjuvant radiotherapy. ...	163
4.1 Introduction.....	163
4.2 Outline of dataset.....	165
4.2.1 Summary of dataset	168
4.3 Exploratory analysis of transcriptomic data quality and clustering trends.	169
4.3.1 Multiregional sampling.....	173
4.3.2 Summary of transcriptome clustering analyses.....	176
4.4 Differential gene expression analysis of pre- and post- radiotherapy samples.	177
4.5 Differential gene expression analysis of pre- and post- radiotherapy samples by histological subtype	180
4.5.1 Shared differentially expressed genes between subtypes.....	182
4.5.2 Summary.....	187
4.6 Gene Set Enrichment Analysis (GSEA) of pre- vs post-radiotherapy Samples	188

4.7 Gene Ontology analysis.....	190
4.8 PROGENy analysis.....	194
4.8.1 Subtype specific analysis	196
4.8.2 Comparison to TCGA sarcoma samples	198
4.8.3 Summary.....	200
4.9 Cellular composition analysis using xCell.....	201
4.10 Discussion of chapter 4	205
Chapter 5. Transcriptomic determinants of radiotherapy response in soft tissue sarcomas.....	208
5.1 Introduction.....	208
5.2 Clinical characteristics of post-radiotherapy responders vs progressors	211
5.3 Differential Gene Expression and Gene Set Enrichment Analysis.....	213
5.4 PROGENy analysis.....	220
5.5 Cellular composition analysis	222
5.6 Modelling disease progression in soft tissue sarcomas following radiotherapy.	224
5.6.1 External validation of model in the TCGA Sarcoma cohort.....	233
5.7 Discussion of chapter 5	237
Chapter 6. Discussion.....	241
6.1 Genomic alterations and mutational signatures following radiotherapy.....	242
6.1.1 The mutational landscape of the LSS cohort.....	242
6.1.2 Somatic mutation burden pre- and post-radiotherapy	245
6.1.3 Critical appraisal of methods and limitations	249
6.2 Transcriptomic responses to neoadjuvant radiotherapy.....	251
6.2.1 Transcriptomic clustering trends	251
6.2.2 Differential Gene Expression Analysis.....	253
6.2.3 Subtype-specific transcriptomic changes post-radiotherapy.....	259
6.2.4 Genes with variable expression patterns following radiotherapy.....	262
6.2.5 Pathway enrichment analysis highlights immune activation post-radiotherapy	266
6.2.6 PROGENy analysis identifies androgen signalling as a potential target for adjuvant therapy.....	268
6.3 Transcriptomic determinants of radiotherapy resistance.....	270
6.3.1 Clinical characteristics of responders vs. progressors	270
6.3.2 Differential gene expression analysis identifies potential biomarkers of disease progression.	271
6.4 Summary and future directions.....	273
Chapter 7. Appendix.....	275
7.1 Publications.....	275
7.2 Statement of contributions.....	276
Reference List	277

Table of figures

Figure 1.1. A timeline of the history of radiotherapy	42
Figure 2.1. Consort diagram of patient inclusion and sample quality control flowchart	70
Figure 2.2. Overview of the London Sarcoma Service patient cohort, clinical features, and sequencing data availability	72
Figure 2.3. Histological assessment of tumour samples	75
Figure 3.1. Distribution of Tumour Mutational Burden by sarcoma subtype	92
Figure 3.2 Oncoplot of somatic mutations in known cancer genes.	95
Figure 3.3. Comparison of the frequency of SNVs and indels in pre- and post- radiotherapy sarcoma samples.	110
Figure 3.4. Comparison of the frequency of SNVs and indels pre- and post- radiotherapy across different histological subtypes.	111
Figure 3.5. Comparison of the frequency of SNVs and indels in paired pre- and post-radiotherapy sarcoma samples.	113
Figure 3.6. Comparison of Deletion-to-Insertion and Indel-to-SNV Ratios	116
Figure 3.7 Timeline of treatment and disease progression for patient 58	117
Figure 3.8. Temporal evolution of somatic mutations in patient 58's sarcoma specimens across treatment stages.	119
Figure 3.9 Comparison of copy number alteration metrics in pre- and post- radiotherapy samples	125
Figure 3.10. Comparison of copy number alteration metrics in pre- and post- radiotherapy myxofibrosarcoma samples	128
Figure 3.11. Comparison of copy number alteration metrics in paired pre- and post- radiotherapy samples	130
Figure 3.12. Distribution of SBS mutational signatures across pre-radiotherapy, post-radiotherapy, recurrence, and metastasis samples in WES data	133
Figure 3.13. Distribution of SBS mutational signatures across pre-radiotherapy, post-radiotherapy, and metastasis samples in WGS data.	134
Figure 3.14. Distribution of SBS mutational signatures across sarcoma subtypes.	136

Figure 3.15. Distribution of indel mutational signatures across pre-radiotherapy, post-radiotherapy, recurrence, and metastasis samples in WES data.....	139
Figure 3.16. Distribution of indel mutational signatures across pre-radiotherapy, post-radiotherapy, and metastasis samples in WGS data.	140
Figure 3.17. Distribution of Indel Mutational Signatures across sarcoma subtypes.	142
Figure 3.18. Expression of mismatch repair genes.....	144
Figure 3.19. Indels and SNVs per Cell Pre- and Post-Radiotherapy.	150
Figure 3.20. Comparison of Indel-to-SNV and Deletion-to-Insertion ratios	152
Figure 3.21. Proportion of microhomology Indels to total Indels pre- and post-radiotherapy.	154
Figure 3.22. Single base substitution signatures pre- and post-radiotherapy samples across the NanoSeq cohort.....	156
Figure 3.23. Indel (ID) mutational signatures in pre- and post-radiotherapy samples across the NanoSeq cohort.....	158
Figure 4.1. PCA plot demonstrating histological subtypes.....	170
Figure 4.2 UMAP demonstrating clustering of samples according to histological subtype.	171
Figure 4.3. PCA plot demonstrating the radiotherapy status of the samples.	172
Figure 4.4. PCA plot demonstrating the effect of radiotherapy in patients with multi-region sampling of resection specimens.	174
Figure 4.5. PCA plot demonstrating the effect of sequencing batch.	175
Figure 4.6. Volcano plot of differential gene expression following radiotherapy...	178
Figure 4.7. Top differentially expressed genes following radiotherapy.	179
Figure 4.8. Upset plot displaying unique and shared differentially expressed genes following radiotherapy across histological subtypes.	183
Figure 4.9. Heatmap of differentially expressed genes observed in at least three sarcoma subtypes following radiotherapy.....	186
Figure 4.10. Heatmap of enriched Hallmark pathways in post-radiotherapy Samples.....	189
Figure 4.11. Top 30 Upregulated GO Biological Processes following radiotherapy.	191
Figure 4.12. Top 30 Downregulated GO Biological Processes following radiotherapy.....	193

Figure 4.13. PROGENy pathway activity scores pre- and post-radiotherapy.....	195
Figure 4.14. PROGENy pathway scores pre- and post-radiotherapy	197
Figure 4.15. Comparison of PROGENy scores in myxofibrosarcomas between the TCGA and the radiotherapy datasets	199
Figure 4.16. Changes in cellular composition identified by xCell analysis following radiotherapy	202
Figure 5.1. Volcano plot of differential gene expression between progressors and responders following radiotherapy.....	214
Figure 5.2. Top differentially expressed genes in patients with subsequent disease progression.	215
Figure 5.3. Enriched Hallmark pathways in progressors vs. responders following radiotherapy.	218
Figure 5.4. PROGENy pathway activity scores for responders vs. progressors. .	221
Figure 5.5. Changes in cellular composition identified by xCell analysis in progressors following radiotherapy.....	223
Figure 5.6. LASSO Cross-validation plot.....	227
Figure 5.7. Confusion matrix evaluating the model on the testing data	230
Figure 5.8. ROC curve for the post-RT progression classifier on the held-out test set.....	232
Figure 5.9. Kaplan–Meier overall survival by RF score	235

List of tables

Table 1. Soft tissue sarcomas subtypes included in this study	21
Table 2. FNCLCC grading system.....	50
Table 3.1. High pathogenicity missense mutations.....	98
Table 3.2. Moderate pathogenicity missense mutations	99
Table 3.3. COSMIC cancer genes mutated in multiple patients across cohort. ...	103
Table 3.4. COSMIC cancer genes mutated in each sarcoma subtype	105
Table 4.1 Clinical characteristics of all patient samples.....	166
Table 4.2. Clinical characteristics of paired patient samples	167
Table 4.3. Numbers of differentially expressed genes following radiotherapy broken down by histological subtype.....	181
Table 4.4. Differentially expressed genes following radiotherapy that are shared between multiple soft tissue sarcoma subtypes.	184
Table 5.1. Summary of histological subtypes and disease progression status in post-radiotherapy cohort.	212
Table 5.2. Performance metrics of the Random Forest progression prediction model.....	231

Abbreviations

ASPS	Alveolar Soft Part Sarcoma
CCS	Clear Cell Sarcoma
CNA	Copy Number Alteration
CNV	Copy Number Variation
COSMIC	Catalogue of Somatic Mutations in Cancer
ddLPS	Dedifferentiated Liposarcoma
EMC	Extraskeletal Myxoid Chondrosarcoma
FFPE	Formalin-Fixed Paraffin-Embedded
FDR	False Discovery Rate
GSEA	Gene Set Enrichment Analysis
ID	Insertion and Deletion Mutational Signature
IMRT	Intensity-Modulated Radiotherapy
INDEL	Insertion/Deletion
LASSO	Least Absolute Shrinkage and Selection Operator
MEC	Myoepithelial Carcinoma
MFS	Myxofibrosarcoma
MMEJ	Microhomology-Mediated End Joining
mLPS	Myxoid Liposarcoma
NMEJ	Non-Microhomology-Mediated End Joining
NGS	Next-Generation Sequencing
PCA	Principal Component Analysis
pLMS	Pleomorphic Leiomyosarcoma
pLPS	Pleomorphic Liposarcoma
PROGENy	Pathway Activity Inference Tool
ROC	Receiver Operating Characteristic
RT	Radiotherapy
SBS	Single Base Substitution Mutational Signature
SNV	Single Nucleotide Variant
SpCS	Spindle Cell Sarcoma
SS	Synovial Sarcoma
TCGA	The Cancer Genome Atlas

TMB	Tumour Mutational Burden
UMAP	Uniform Manifold Approximation and Projection
UPS	Undifferentiated Pleomorphic Sarcoma
VAF	Variant Allele Frequency
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

Chapter 1. Introduction

1.1 Introduction to soft tissue sarcomas

Soft tissue sarcomas (STS) are rare malignant tumours arising from mesenchymal tissues, including muscle, fat, blood vessels, and connective tissue. They account for approximately 1% of all adult malignancies and have an annual incidence of 5 per 100,000 individuals globally (Sbaraglia, Bellan *et al.* 2021). These tumours represent a highly heterogeneous group, with over 80 histological subtypes classified by the World Health Organization based on a combination of morphological, immunohistochemical, and molecular characteristics (WHO 2020). Additionally individual subtypes tend to have a distinct pattern of presentation including the anatomical site, age distribution, behaviour, response to treatment, and prognosis (Hayes, Nixon *et al.* 2024).

STS can arise anywhere in the body, but approximately 50% occur in the extremities, particularly the thighs. Other common sites include the retroperitoneum (30%) and the trunk or head and neck (15%) (Sbaraglia and Dei Tos 2019). The diversity in anatomical presentation correlates with specific histological subtypes, such as myxoid liposarcoma, which typically affects the thighs of younger adults, and myxofibrosarcoma, commonly found in the elderly and located superficially (above the fascia) (Sbaraglia and Dei Tos 2019).

Given their origin from mesenchymal tissues, sarcomas are fundamentally distinct from carcinomas, which originate from epithelial cells and represent some of the most common cancers worldwide. Their biological behaviour is closely linked to their putative origin from mesenchymal stem cells (MSCs), multipotent progenitor cells responsible for generating connective tissues such as bone, cartilage, fat, and muscle. MSCs are not only essential for tissue differentiation and repair but are also implicated in tumour initiation and progression. Their inherent properties, including extensive self-renewal, plasticity, and immunomodulatory functions, are thought to contribute to the aggressive and heterogeneous nature of sarcomas (Rodriguez, Rubio *et al.* 2012, O'Donnell III, Muñoz *et al.* 2025).

In addition to MSCs, cancer stem cells (CSCs) play a pivotal role in sarcomagenesis. Sharing many characteristics with MSCs, CSCs are a subpopulation of tumour cells that exhibit resistance to conventional therapies, driving tumour heterogeneity, recurrence, and metastasis. These characteristics are particularly relevant in high-grade sarcomas, where therapeutic challenges are pronounced (O'Donnell III, Muñoz *et al.* 2025).

Understanding the heterogeneity of soft tissue sarcomas is critical for optimising diagnosis, management, and research. To provide a foundation for the subsequent discussion, Table 1 summarises the key features, common anatomical locations, and prognosis of the sarcoma subtypes examined in this study.

Table 1. Soft tissue sarcomas subtypes included in this study

Diagnosis	Key Features	Common Locations	Prognosis
Alveolar soft part sarcoma (Paoluzzi and Maki 2019)	Characteristic ASPL::TFE3 fusion; often slow-growing	Extremities, trunk	Poor, as metastases are common at diagnosis
Clear cell sarcoma (Ibrahim, Jensen <i>et al.</i> 2018)	EWSR1::ATF1 fusion; mimics melanoma histologically	Extremities, tendons	Poor; high metastatic potential
Dedifferentiated liposarcoma (Thway 2019)	High-grade component alongside well-differentiated liposarcoma	Retroperitoneum, extremities	Poor if dedifferentiation is extensive
Extraskeletal myxoid chondrosarcoma (Stacchiotti, Baldi <i>et al.</i> 2020)	Low-grade malignancy; often presents with a myxoid stroma	Proximal extremities	Better prognosis compared to other subtypes
Malignant mixed tumour (Hornick and Fletcher 2003)	Rare; derived from salivary or glandular tissue; histologically diverse	Head, neck, salivary glands	Poor if high-grade
Malignant peripheral nerve sheath tumour (Yao, Zhou <i>et al.</i> 2023)	Associated with NF1 in many cases; spindle cell morphology	Trunk, proximal extremities	Poor if large or deep
Myxofibrosarcoma (Vanni, De Vita <i>et al.</i> 2022)	High recurrence rate; seen in elderly patients; myxoid stroma with pleomorphic spindle cells	Extremities, superficial tissues	Moderate, worse with deep tissue involvement
Myxoid liposarcoma (Abaricia and Hirbe 2018)	Round cell component predicts aggressive behaviour; sensitive to radiotherapy	Thigh, retroperitoneum	Good prognosis with localised disease

Pleomorphic leiomyosarcoma (Nicolas, Tamboli <i>et al.</i> 2010)	Derived from smooth muscle cells; high mitotic index and pleomorphism	Uterus, retroperitoneum	Poor prognosis, particularly in deep locations
Pleomorphic liposarcoma (Anderson and Jo 2019)	Aggressive, pleomorphic cells; lacks the typical myxoid component	Extremities	Poor; prone to metastasis
Spindle cell sarcoma NOS	Diagnosis of exclusion; similar to other spindle cell sarcomas	Extremities	Variable, depends on grade and size
Synovial sarcoma (Gazendam, Popovic <i>et al.</i> 2021)	Characterised by SS18::SSX fusion; monophasic or biphasic histology	Extremities, para-articular	Intermediate; 5-year survival ~50-60%
Undifferentiated pleomorphic sarcoma (Hames-Fathi, Nottley <i>et al.</i> 2022)	Formerly known as malignant fibrous histiocytoma; highly aggressive	Extremities, trunk	Poor; 5-year survival ~30-50%

1.2 The challenge to provide a timely and accurate diagnosis

Given their rarity and significant morphological heterogeneity, the accurate diagnosis of soft tissue sarcomas requires specialised expertise that is typically concentrated in high-volume tertiary referral centres. The challenges associated with diagnosing these tumours stem from their wide spectrum of histological subtypes, overlapping features with benign mimics, and the frequent need for advanced molecular and immunohistochemical testing. While the diversity of sarcomas reflects their complexity, this heterogeneity also poses significant barriers to achieving accurate and timely diagnoses, particularly in non-specialist settings (Hayes, Nixon *et al.* 2024).

For instance, spindle cell morphology can be observed in benign entities such as nodular fasciitis or in aggressive malignancies like leiomyosarcoma, emphasising the importance of a systematic diagnostic approach (Sbaraglia and Dei Tos 2019, Sbaraglia, Bellan *et al.* 2021). Diagnostic accuracy is significantly enhanced in centralised institutions, where multidisciplinary teams—including pathologists, radiologists, surgeons, and oncologists—collaborate to refine classifications and treatment plans. These teams integrate clinical data, multimodal imaging, morphology, immunohistochemistry, and molecular techniques like fluorescence in situ hybridisation (FISH) and next-generation sequencing to achieve precise diagnoses (Hayes, Nixon *et al.* 2024).

The identification of specific translocations, such as *SS18::SSX* in synovial sarcoma (Turc-Carel, Dal Cin *et al.* 1986), *EWSR1::ATF1* in clear cell sarcoma (Wang, Mayordomo *et al.* 2009), or *CIC::DUX4* in CIC-DUX4 sarcoma (Brahmi, Vanacker *et al.* 2022) has transformed diagnostic confidence and improved the classification of certain challenging cases. Similarly, immunohistochemical markers like MUC4 and NGS for *CTNNB1* mutations have been pivotal in distinguishing low-grade fibromyxoid sarcomas from mimics such as desmoid fibromatosis respectively (Doyle, Möller *et al.* 2011). Use of these molecular tests has facilitated the recognition of rare or atypically presenting sarcomas. Synovial sarcomas, traditionally associated with extremities, have increasingly been reported in visceral locations such as the

gastrointestinal tract (Requena, Longacre *et al.* 2024) and lungs (Roy, Das *et al.* 2012).

Despite advances in molecular diagnostics and multidisciplinary care, diagnostic inaccuracies remain a significant concern, particularly in non-specialist settings. In a study of 348 cases referred to a specialist sarcoma centre, major diagnostic discrepancies—those with the potential to significantly alter clinical management—were observed in 16.4% of cases (Thway, Wang *et al.* 2014). Additionally, 11.8% of cases had minor discrepancies, which, although not affecting treatment plans, highlight the challenges of accurate classification in this complex tumour group. Alarming, 23.5% of all discrepancies involved reclassification from benign to malignant or vice versa. Such discordances underline the inherent difficulties in interpreting soft tissue tumours in non-specialist environments, particularly given their rarity and diverse histological appearances.

The consequences of diagnostic inaccuracies are profound. For example, dedifferentiated liposarcomas, which necessitate aggressive surgical intervention, may be misdiagnosed as benign lipomas without molecular confirmation of *MDM2* amplification using immunohistochemistry or FISH (Gambella, Bertero *et al.* 2023). Similarly, misclassification of benign lesions as malignant can lead to unnecessary overtreatment, including unwarranted chemotherapy, radiotherapy, or radical surgery. The increasing complexity of soft tissue tumour diagnosis, driven by the integration of ancillary molecular techniques, requires both specialised expertise and access to timely testing facilities.

The study also highlighted the role of interpretational errors and the limited use of specific immunohistochemical markers in non-specialist settings as key contributors to diagnostic discrepancies. For instance, immunohistochemical markers such as h-caldesmon, CDK4, and beta-catenin, essential for leiomyosarcomas, well-differentiated/dedifferentiated liposarcomas, and desmoid fibromatosis, respectively, were either not utilised or misinterpreted in several cases (Thway, Wang *et al.* 2014). The findings emphasise the need for centralised review by specialist sarcoma pathologists, as recommended by the National Institute for Health and Care

Excellence (NICE) and the Royal College of Pathologists in the UK (Cyril Fisher 2022).

UK guidelines, including those from NICE, recommend that all suspected soft tissue sarcomas are managed within specialist centres by multidisciplinary teams (MDTs) to ensure timely and accurate diagnosis (Hayes, Nixon *et al.* 2024). Turnaround times for suspected sarcoma biopsies should ideally be within two weeks to facilitate prompt treatment planning, particularly for high-grade or rapidly progressing tumours. Specialist centres, such as the London Sarcoma Service, are equipped to streamline the diagnostic process through the integration of in-house testing, MDT discussions, and rapid turnaround times, particularly critical for high-grade sarcomas requiring urgent intervention.

However, differences between public and private healthcare systems can influence the diagnostic pathway. Private care pathways often outsource samples or rely on non-specialist testing, which can lead to delays in diagnosis. In contrast, specialist centres minimise such delays by prioritising in-house testing and expediting molecular diagnostics when needed, ensuring that patients receive timely treatment. Greater standardisation and coordination between public and private pathways could mitigate delays and ensure all patients benefit from the expertise available in specialist centres.

1.3 The genomics of soft tissue sarcomas

Soft tissue sarcomas exhibit a diverse range of genomic features, reflecting their heterogeneity and distinct biological behaviours. Broadly, they can be categorised into two groups based on analysis of their genomic profiles: translocation-driven sarcomas and those with complex genomic alterations. This dichotomy underscores the importance of genomic characterisation for understanding tumour biology, guiding treatment decisions, and refining prognostic predictions.

Translocation-associated sarcomas are characterised by specific chromosomal translocations that result in fusion oncoproteins. These fusions are critical drivers of oncogenesis, acting by dysregulating transcriptional networks and cellular pathways. In contrast, sarcomas with complex genomic profiles, such as undifferentiated pleomorphic sarcoma (UPS) and dedifferentiated liposarcoma (ddLPS), exhibit widespread genomic instability, including chromothripsis, copy number alterations, and aneuploidy. These patterns of genomic alteration have significant implications for clinical management, with translocation-driven sarcomas often benefiting from targeted therapeutic strategies, while complex genomic sarcomas pose greater challenges due to their heterogeneity and treatment resistance.

1.3.1 Translocation associated sarcomas

A defining feature of translocation-associated sarcomas is the presence of specific, recurrent chromosomal translocations that lead to the formation of fusion genes. For example, myxoid liposarcoma, which represents one of the subtypes included in this study, is characterised by a t(12;16)(q13;p11) translocation, resulting in the *FUS::DDIT3* fusion gene (Abaricia and Hirbe 2018). This fusion protein acts as an aberrant transcription factor, disrupting adipocytic differentiation and driving tumourigenesis.

Another key example is synovial sarcoma, which is defined by the t(X;18)(p11.2;q11.2) translocation, resulting in the *SS18::SSX* fusion gene (Gazendam, Popovic *et al.* 2021). This fusion disrupts chromatin remodelling and transcriptional regulation, driving oncogenesis. Synovial sarcoma predominantly

affects young adults and typically arises in the extremities. The identification of the *SS18::SSX* fusion gene is critical for diagnosis and has become essential in the pathological workup for these tumours. While FISH testing is routinely used, recent immunohistochemical markers have become available that can detect the novel fusion protein (Zaborowski, Vargas *et al.* 2020).

Beyond diagnosis, translocation-driven sarcomas hold promise for targeted therapies. While myxoid liposarcoma is highly sensitive to radiotherapy, ongoing research aims to identify molecular vulnerabilities linked to the *FUS::DDIT3* fusion that could lead to novel therapeutic interventions. The *FUS::DDIT3* fusion blocks adipocytic differentiation and leads to an increase in immature adipocytes. Trabectedin, a molecule originally extracted from sea squirts, has been shown to bind to this novel fusion and allows the tumour to differentiate thereby reducing its malignant potential (Craparotta, Mannarino *et al.* 2024).

Similarly, synovial sarcoma represents a promising candidate for immunotherapy approaches, including T-cell receptor-based therapies targeting the *SS18-SSX* fusion protein (Mavroeidis, Napolitano *et al.* 2024). The success of targeted therapies in other translocation-driven sarcomas, such as tyrosine kinase inhibitors in gastrointestinal stromal tumours (GISTs) (Serrano and Bauer 2022), shows the potential for similar approaches in myxoid liposarcoma, synovial sarcoma, amongst others.

Recent large-scale genomic sequencing studies performed at Memorial Sloan Kettering have refined the molecular landscape of translocation-associated sarcomas (Nacev, Sanchez-Vega *et al.* 2022). Nacev *et al.* identified additional recurrent alterations in synovial sarcoma, including *CDKN2A/B* deletions and *RB1* loss, suggesting that beyond the primary fusion event, secondary cell cycle dysregulation contributes to disease progression. In myxoid liposarcoma, *PIK3CA* mutations were present in 25% of cases, suggesting *PI3K* pathway activation as a potential therapeutic target (Gounder, Agaram *et al.* 2022). These findings reinforce the need for comprehensive molecular profiling to uncover additional oncogenic drivers as well as identifying therapeutic targets in translocation-driven sarcomas.

1.3.2 Complex genomic sarcomas

Sarcomas with complex genomic profiles, such as undifferentiated pleomorphic sarcoma (UPS) and dedifferentiated liposarcoma (DDLPS), are characterised by widespread genomic instability. Unlike translocation-driven sarcomas, these subtypes lack specific, recurrent chromosomal translocations and instead exhibit extensive chromothripsis, copy number alterations (CNAs), and whole-genome duplication (WGD) (Steele, Tarabichi *et al.* 2019, Steele, Abbasi *et al.* 2022).

1.3.2.1 Chromothripsis

Chromothripsis is particularly prevalent in sarcomas, affecting 54% of liposarcomas, 24% of fibrosarcomas, and 23% of sarcomas overall (Cortés-Ciriano, Lee *et al.* 2020). This catastrophic event results in hundreds of genomic rearrangements within single chromosomes, leading to oncogene amplification and tumour progression. In dedifferentiated liposarcoma, chromothripsis frequently targets MDM2 and CDK4, genes crucial for cell cycle regulation (Cortés-Ciriano, Lee *et al.* 2020).

Micronuclei formation and telomere crisis have been identified as primary mechanisms driving chromothripsis in sarcomas (Cortés-Ciriano, Lee *et al.* 2020). Micronuclei formation occurs when fragmented chromosomes become encapsulated outside the main nucleus, leading to defective DNA replication and chaotic rearrangements. Telomere crisis, caused by critically short telomeres, triggers chromosomal fusion events that further fuel genomic instability.

1.3.2.2 Whole genome doubling

Whole genome doubling (WGD) is a key feature of complex genomic sarcomas, occurring in a significant proportion of cases and contributing to extensive chromosomal imbalances and aneuploidy. This genomic event, in which the entire set of chromosomes is duplicated—sometimes multiple times—provides a selective advantage by increasing tolerance to additional structural alterations, thereby promoting tumour evolution and heterogeneity. Large-scale analyses have shown that WGD frequently precedes chromothripsis, reinforcing the hypothesis that

genome duplication acts as a catalyst for further genomic instability (Steele, Abbasi *et al.* 2022).

Sarcomas were among the tumour types with the highest levels of copy number alterations, with WGD-associated signatures strongly linked to poor prognosis. WGD has been identified as a recurrent event in undifferentiated pleomorphic sarcoma (UPS), leiomyosarcoma, and osteosarcoma, correlating it with higher mutational burdens, enhanced tumour aggressiveness, and poorer patient survival outcomes (Steele, Tarabichi *et al.* 2019, Nacev, Sanchez-Vega *et al.* 2022).

Additionally, a subset of sarcomas exhibited copy number patterns consistent with homologous recombination deficiency (HRD), suggesting that some WGD+ tumours may be vulnerable to PARP inhibitors or platinum-based chemotherapy (Steele, Abbasi *et al.* 2022). Given the impact of WGD on tumour evolution, understanding its role in driving treatment resistance and shaping the sarcoma genome remains a crucial area for further investigation.

1.3.2.3 Extrachromosomal DNA

Emerging evidence suggests that extrachromosomal DNA (ecDNA) contributes to sarcoma progression by amplifying oncogenes and driving tumour evolution. Unlike chromosomal amplifications, ecDNA consists of circular DNA fragments that promote rapid adaptation and therapy resistance (Kim, Nguyen *et al.* 2020, Bailey, Pich *et al.* 2024). A large-scale study identified ecDNA in 17.1% of tumours, with particularly high prevalence in liposarcomas (54.9%) (Kim, Nguyen *et al.* 2020).

Clinically, ecDNA-positive tumours exhibit increased metastasis rates, intratumoural heterogeneity, and resistance to cytotoxic chemotherapy (Bailey, Pich *et al.* 2024). In sarcomas, ecDNA frequently harbours oncogenes such as *MDM2*, *CDK4*, and *HMGA2*, particularly in dedifferentiated liposarcoma (Kim, Nguyen *et al.* 2020). There is also evidence that ecDNA may contribute to immune evasion, potentially impacting responses to immune checkpoint inhibitors (Bailey, Pich *et al.* 2024). Although ecDNA represents a promising therapeutic target, current research remains in early

stages. Potential approaches include inhibitors that disrupt ecDNA formation or reintegration, such as PARP inhibitors and chromatin-modulating drugs (Dong, He *et al.* 2023).

1.3.2.4 Hypermutation and Mismatch Repair Deficiency

While most soft tissue sarcomas exhibit a relatively low tumour mutational burden (TMB), a subset of undifferentiated pleomorphic sarcomas (UPS) and other complex genomic sarcomas exhibit hypermutation (defined as a TMB of > 10 mutations per megabase), often associated with mismatch repair deficiency (MMR-D) (TCGA 2017, Steele, Tarabichi *et al.* 2019). In a genomic study of undifferentiated sarcomas, approximately 15% of cases were found to harbour a hypermutator phenotype, with over 15,000 somatic mutations per tumour (Steele, Tarabichi *et al.* 2019). Notably, only 2.1% of 7,494 sarcomas sequenced in a study of 44 different sarcoma subtypes exhibited MMR deficiency. These tumours had a median TMB of 6.5 mutations/Mb, which is significantly higher than in MMR-proficient tumours of 2.4 mutations/Mb (Gounder, Agaram *et al.* 2022).

MMR-D in sarcomas can arise via several mechanisms, including germline or somatic pathogenic variants in *MLH1*, *MSH2*, *MSH6*, or *PMS2*; *MLH1* promoter hypermethylation; and structural alterations (e.g., deletions/rearrangements or loss of heterozygosity) that result in loss of MMR protein function (Steele, Tarabichi *et al.* 2019). Notably, these hypermutated sarcomas tend to have elevated immune infiltration and upregulation of immune-related gene expression signatures, suggesting potential sensitivity to immune checkpoint inhibitors (ICIs) (TCGA 2017, Steele, Tarabichi *et al.* 2019). Gounder *et al.* further highlighted that only 0.3% of sarcomas exhibit microsatellite instability (MSI-H), suggesting that while mismatch repair deficiency is present, MSI is rare, differentiating sarcomas from other MMR-D cancers like colorectal or endometrial carcinoma.

Hypermuted sarcomas with mismatch repair deficiency represent a distinct subgroup within complex genomic sarcomas that may have important therapeutic implications. Given their high tumour mutational burden (TMB) and increased

immune infiltration, these tumours show similarities to MMR-deficient colorectal and endometrial cancers, which have demonstrated strong responses to immune checkpoint blockade (e.g., anti-PD1/PD-L1 therapy) (Steele, Tarabichi *et al.* 2019, Shiravand, Khodadadi *et al.* 2022).

1.4 Risk factors for the development of soft tissue sarcomas

While most cases arise sporadically, several environmental, genetic, and viral factors have been implicated in sarcoma pathogenesis. These include inherited cancer predisposition syndromes, viral infections, and prior exposure to ionising radiation.

1.4.1 Inherited cancer predisposition syndromes

A subset of STS arise in individuals with inherited germline mutations in tumour suppressor genes, predisposing them to cancer development. The most well-characterised syndromes associated with soft tissue sarcomas include:

Li-Fraumeni syndrome: Caused by germline *TP53* mutations, predisposes individuals to a spectrum of malignancies, including rhabdomyosarcoma, leiomyosarcoma, and undifferentiated pleomorphic sarcoma (UPS). *TP53*-mutant sarcomas often exhibit early onset and high genomic instability (Correa 2016).

Neurofibromatosis type 1: This autosomal dominant disorder results from mutations in *NF1*, a tumour suppressor gene encoding neurofibromin. *NF1* patients have an increased risk of malignant peripheral nerve sheath tumours (MPNSTs), which arise from plexiform neurofibromas and exhibit complex genomic alterations (Gutmann, Ferner *et al.* 2017).

Retinoblastoma (*RB1*) gene mutations: Germline *RB1* mutations significantly increase the risk of osteosarcomas and soft tissue sarcomas (STS), particularly after radiotherapy. *RB1* loss disrupts cell cycle regulation, leading to uncontrolled proliferation and tumourigenesis. In hereditary retinoblastoma survivors, the risk of STS is substantially elevated, with incidence rising sharply after age 30. Compared to the general population, these individuals face a 500-fold higher risk of STS in irradiated regions (Kleinerman, Schonfeld *et al.* 2019).

1.4.2 Viral associations

Although viral oncogenesis is well-established in certain epithelial cancers for example HPV and cervical squamous cell carcinoma (Tjalma, Van Waes *et al.* 2005), its role in soft tissue sarcomas is less frequent but well-documented in specific subtypes.

Human herpesvirus 8 (HHV-8): Kaposi sarcoma (KS), caused by HHV-8/KSHV, is uniquely driven by viral oncogenes rather than clonal oncogenic transformation. KS tumours have an exceptionally low mutational burden and exhibit an angiogenic spindle cell proliferation, particularly in immunosuppressed individuals, including those with HIV/AIDS (Phipps, Bhinder *et al.* 2025).

Epstein-Barr virus (EBV) has been linked to leiomyosarcomas, particularly in immunosuppressed individuals, including post-transplant patients and those with HIV. EBV-associated smooth muscle tumours (EBV-SMTs) exhibit distinct molecular features, including lower genomic instability compared to conventional leiomyosarcomas but recurrent gains in oncogenes such as *RUNX1*, *CCND2*, and *ETS2*, suggesting a unique viral-driven oncogenesis pathway (Wah, Mok *et al.* 2023).

1.4.3 Radiation induced sarcomas

Exposure to ionising radiation is a well-recognised risk factor for secondary malignancies, particularly radiation induced sarcomas (RISs). While most cases occur following radiotherapy, RISs can also develop after exposure to other sources of ionising radiation. These tumours arise within previously irradiated tissues, typically 5 to 30 years post-treatment, though some cases have been reported after over five decades (Lesluyes, Baud *et al.* 2019). RISs account for <5% of all sarcomas, with undifferentiated pleomorphic sarcoma (UPS), angiosarcoma, and leiomyosarcoma being amongst the most common subtypes (Inchaustegui, Kon-Liao *et al.* 2023). Compared to sporadic soft tissue sarcomas (STS), RISs are more aggressive, often diagnosed late, and associated with poorer outcomes.

1.4.3.1 *The genomic and mutational landscape of radiation induced sarcomas*

Radiation-induced sarcomas (RISs) exhibit distinct genomic alterations that differentiate them from sporadic sarcomas. They are characterised by high levels of structural rearrangements and genomic instability, comparable to sarcomas with complex genetics (Lesluyes, Baud *et al.* 2019). A notable feature is the frequent deletion of *CDKN2A/CDKN2B* (9p21.3) (71% in RIS vs. 39% in sporadic sarcomas), leading to cell cycle deregulation via the RB1 and p53 pathways.

Unlike sporadic sarcomas, where deletions and mutations follow chromatin accessibility patterns, RISs exhibit a random distribution of deletions across the genome. In sporadic tumours, DNA damage and repair efficiency are influenced by chromatin structure, meaning mutations are more likely to occur in open (euchromatic) regions that are transcriptionally active and accessible to repair mechanisms. In contrast, radiation-induced DNA damage occurs in a stochastic manner, generating breaks indiscriminately across the genome, independent of chromatin accessibility (Behjati, Gundem *et al.* 2016). This suggests that RIS development is driven by direct radiation-induced DNA breaks, rather than selection for specific genomic vulnerabilities.

Following radiation therapy, cells attempt to repair double-strand breaks (DSBs), but RISs predominantly rely on non-homologous end-joining (NHEJ) and microhomology-mediated end-joining (MMEJ) - both of which are error-prone repair mechanisms. Unlike homologous recombination, which uses an intact template for accurate repair, NHEJ and MMEJ function without a template, leading to imprecise repair outcomes (Seol, Shim *et al.* 2018).

- **NHEJ** directly ligates broken DNA ends but frequently results in small insertions or deletions (indels) due to the loss or addition of nucleotides at the breakpoint.
- **MMEJ** is a subtype of alternative NHEJ (A-NHEJ) and it is even more error-prone requiring trimming of DNA ends before aligning short microhomology sequences (2–20 bp). This process always results in deletions, as the intervening sequence is lost during repair

RISs exhibit a high burden of MMEJ-associated deletions, often flanked by microhomology sequences, which can be detected through next-generation sequencing. MMEJ is strongly implicated in chromosomal instability, particularly in the formation of chromosomal translocations and complex structural variants. Unlike NHEJ, which predominantly generates small indels, MMEJ-mediated repair frequently results in large deletions and translocations, often flanked by short regions of microhomology. Studies in mammalian cells have demonstrated that MMEJ promotes chromosomal rearrangements when multiple DSBs occur simultaneously, leading to promiscuous end joining between incorrect chromosomes (Seol, Shim *et al.* 2018). The accumulation of these mutations over time drives tumour progression, contributing to the aggressive clinical behaviour of RISs (Behjati, Gundem *et al.* 2016).

RISs also harbour radiation-specific mutational signatures, including an excess of balanced inversions, a rare structural rearrangement type that is significantly enriched in radiation-associated malignancies compared to sporadic tumours (Behjati, Gundem *et al.* 2016). In post-radiotherapy angiosarcomas, *MYC* amplifications are present in 96% of cases, making them a defining molecular feature absent in the majority of sporadic angiosarcomas (Lesluyes, Baud *et al.* 2019).

1.4.3.2 *Post-radiotherapy angiosarcoma as a model for RISs*

Post-radiotherapy angiosarcomas (PRAs), particularly those arising after breast cancer treatment, are among the most well-characterised radiation-induced sarcomas (RISs) and provide key insights into the genomic consequences of radiotherapy (Lesluyes, Baud *et al.* 2019, Dermawan, Chi *et al.* 2023). Although angiosarcoma is not the primary focus of this thesis, PRAs serve as a valuable model for understanding radiotherapy-induced mutational landscapes, particularly in relation to mutation burden, copy number alterations, and DNA repair mechanisms. Comparative genomic analyses between 44 PRAs and 135 sporadic angiosarcomas revealed distinct molecular differences, reinforcing the radiation-driven oncogenesis of PRAs (Dermawan, Chi *et al.* 2023). A defining feature is the high frequency of *MYC* amplifications, detected in 75% of PRAs compared to just 13% of sporadic cases, making *MYC* a radiation-specific biomarker, particularly in breast/chest wall PRAs.

PRAs exhibit enrichment in *FLT4*, *CRKL*, *HRAS*, and *KMT2D* mutations, implicating MAP kinase and Hippo–Merlin pathway activation in their pathogenesis. Given these oncogenic drivers, PRAs develop more rapidly than other RISs, with a median latency of 8 years, significantly shorter than radiation-induced undifferentiated pleomorphic sarcomas (UPS) (18.5 years) and malignant peripheral nerve sheath tumours (MPNSTs) (12.5 years).

In addition to their shorter latency, PRAs harbour fewer *TP53* (9%) and *CDKN2A/B* deletions (2%) than other RIS subtypes, suggesting they follow a distinct molecular path to tumour development compared to radiation-induced UPS, which frequently harbours *TP53* mutations. Interestingly, despite their oncogenic changes, PRAs have a lower fraction of genome altered (FGA) than other RISs, indicating less extensive structural disruption compared to UPS and MPNSTs.

Mutational signature analysis of PRAs identified profiles associated with DNA repair deficiencies, including defective mismatch repair (MMR) and replication slippage, though specific COSMIC signatures were not reported in the study.

1.4.3.3 *Clinical outcomes and management of radiation induced sarcomas*

A systematic review of 1,371 RIS patients across 21 studies highlights worse survival, higher recurrence rates, and more limited treatment options compared to de novo soft tissue sarcomas (Inchaustegui, Kon-Liao *et al.* 2023).

Treatment approaches

Surgical resection remains the primary treatment, performed in 68% of cases, with limb-salvage surgery attempted in 74%. However, achieving negative margins (R0) is challenging, with only 58% of cases attaining clear margins, lower than in de novo STS. Chemotherapy is used in 29% of RIS patients, though its efficacy remains uncertain (Inchaustegui, Kon-Liao *et al.* 2023).

For recurrent or metastatic disease, chemotherapy remains the standard of care, with doxorubicin, ifosfamide, gemcitabine, and docetaxel commonly used. However, there is limited data on whether RISs respond differently to chemotherapy than sporadic STS, as most clinical trials do not stratify RIS patients separately (Dickson 2014).

Targeted therapies such as pazopanib (a tyrosine kinase inhibitor) and VEGF inhibitors (e.g., bevacizumab, sorafenib) have shown some efficacy in radiation-associated angiosarcomas, but their role in other RIS subtypes remains unclear.

Clinical outcomes: RIS vs. Sporadic STS

- **5-Year Overall Survival:** 45% in RIS compared to ~60% in sporadic STS, reflecting poorer prognosis.
- **Local Recurrence Rate:** 39% in RIS, significantly higher than the 6.5%–9% seen in sporadic STS, likely due to surgical challenges in irradiated tissue.
- **Metastasis Rate:** 27% in RIS, lower than the ~50% reported for high-grade sporadic STS, though still a significant concern.

RIS tumours are more aggressive, harder to resect, and have limited treatment options, contributing to high recurrence rates. While re-irradiation may improve local control, achieving negative margins remains a challenge.

1.5 The treatment of soft tissue sarcomas

Soft tissue sarcomas require a multimodal treatment approach, guided by tumour histology, size, location, and stage at presentation. The cornerstone of treatment is surgical excision, aiming for complete tumour removal with negative margins while preserving function. However, in many cases, radiotherapy plays a crucial role in improving local disease control, either as an adjunct to surgery or, in some cases, as a definitive treatment. The role of systemic therapy, including chemotherapy and targeted agents, is more selective and primarily applies to specific histological subtypes with higher chemosensitivity or in the setting of advanced disease.

In this section, I will first discuss surgical management, followed by the role of chemotherapy in both the neoadjuvant, adjuvant, and metastatic settings. I will then introduce the history of radiotherapy, providing context for its modern applications. This leads into discussions on preoperative vs. postoperative radiotherapy, histology-specific considerations, current UK radiotherapy guidelines, and concluding with a review of potential side effects.

1.5.1 Surgery

Surgical excision with negative margins is the primary treatment for localised STS. The UK guidelines emphasise that all sarcoma cases should be managed within a specialist multidisciplinary team (MDT) setting to ensure optimal outcomes (Hayes, Nixon *et al.* 2024). Key principles include:

- Wide local excision is the standard approach, aiming for microscopically negative (R0) margins.
- Planned marginal resection with RT may be an option where function-preserving surgery is required.
- Re-excision should be considered if positive margins (R1/R2) are found, unless adjuvant RT is deemed sufficient.
- Amputation is reserved for cases where limb-sparing surgery is not possible.

For tumours in functionally sensitive locations (e.g., retroperitoneum, head and neck), the balance between achieving clear margins and preserving organ function is crucial. Plastic surgical reconstruction is often required in limb-sparing approaches.

1.5.2 Chemotherapy

The role of chemotherapy in soft tissue sarcomas (STS) is largely histology subtype dependent. Unlike many other solid tumours, chemotherapy does not consistently improve overall survival in most adult-type STS. However, for certain subtypes with higher chemosensitivity, systemic therapy remains an integral component of treatment, particularly in the neoadjuvant, adjuvant, and metastatic settings.

Neoadjuvant and adjuvant chemotherapy

Neoadjuvant chemotherapy (administered before surgery) may be considered in selected high-risk cases, particularly for patients with large, deep, high-grade extremity or truncal STS, where it may help downstage the tumour and improve resectability. Certain histological subtypes, including myxoid liposarcoma, synovial sarcoma, and desmoplastic small round cell tumour (DSRCT), are known to be more responsive to chemotherapy and may benefit from this approach.

The role of adjuvant chemotherapy, given postoperatively, remains controversial. Current UK guidelines (Hayes, Nixon *et al.* 2024) do not recommend routine use of adjuvant chemotherapy but suggest that it can be considered for high-risk patients.

Risk stratification tools, such as the Sarcuator nomogram (Pasquali, Palmerini *et al.* 2022), provide an evidence-based means of identifying these patients. The Sarcuator integrates clinicopathological factors (discussed in more detail in section 1.6) such as tumour size, depth, grade, histological subtype, and patient age to predict 10-year overall survival and metastasis-free survival. Patients with a predicted 10-year overall survival of less than 50–60% are the most likely to benefit from systemic therapy. Thus, for these high-risk individuals, chemotherapy may be a reasonable option, though the decision must be weighed against potential toxicity and patient preferences.

Chemotherapy in metastatic STS

In patients with advanced or metastatic STS, chemotherapy is primarily palliative, aimed at controlling disease progression and alleviating symptoms rather than achieving cure. The choice of regimen depends on tumour histology, prior treatments, and patient performance status.

First-line therapy typically involves doxorubicin, either as a single agent or in combination with ifosfamide, particularly for patients requiring a higher response rate. Ifosfamide alone is often preferred in synovial sarcoma, which has shown particular chemosensitivity. For patients with leiomyosarcoma or undifferentiated pleomorphic sarcoma (UPS), the combination of gemcitabine and docetaxel is a commonly used alternative (Hayes, Nixon *et al.* 2024). Beyond conventional cytotoxic agents, targeted therapies such as trabectedin (Craparotta, Mannarino *et al.* 2024) have demonstrated efficacy in translocation-related sarcomas, particularly myxoid liposarcoma and leiomyosarcoma. Additionally, pazopanib, a tyrosine kinase inhibitor (TKI), is an option for patients with non-liposarcoma STS who have progressed on prior chemotherapy.

For select patients with oligometastatic pulmonary disease, metastasectomy may be considered, particularly in synovial sarcoma and leiomyosarcoma, where surgical removal of lung metastases has been associated with prolonged survival in carefully selected cases. However, the decision to proceed with surgical intervention requires a multidisciplinary approach, considering disease burden, response to systemic therapy, and patient fitness.

While chemotherapy remains a valuable tool in select high-risk patients, its use should be individualised, incorporating tumour biology, prognostic risk stratification (such as Sarculator), and patient-specific factors to guide clinical decision-making.

1.5.3 The history of radiotherapy

The history of radiotherapy is deeply intertwined with the discovery of X-rays and radioactivity, marking a transformative era in cancer treatment. The field began in 1895, when Wilhelm Röntgen discovered X-rays (Röntgen 1896), leading to their immediate application in medicine. Just a year later, Victor Despeignes in France attempted the first recorded use of X-rays to treat a patient with presumed gastric cancer (Despeignes 1896), though the patient ultimately succumbed to the disease. That same year, Emil Grubbe in Chicago claimed to be the first to use X-rays for treating a breast cancer patient, marking the beginning of radiotherapy as a clinical discipline. Almost simultaneously, Henri Becquerel discovered natural radioactivity in 1896 (Becquerel 1896), and in 1898, Marie and Pierre Curie isolated radium, a naturally radioactive element (Curie 1898). These pioneering breakthroughs established the foundation for external beam radiotherapy (EBRT) and brachytherapy (Figure 1.1).

Early clinical use and fractionation (1899–1930s)

By the early 1900s, X-ray therapy was being used for superficial tumours, particularly skin cancers. In Sweden (1899), Thor Stenbeck and Tage Sjögren successfully treated patients with skin cancer, proving that X-rays could eradicate tumours. However, deep-seated tumours remained challenging, as early X-ray machines had limited penetration and caused severe skin toxicity (Connell and Hellman 2009). A key case in New Haven (1902) involved Clarence Skinner, who may have cured one of the first deep-seated tumours with X-ray therapy, despite its limitations.

A major breakthrough in radiobiology came in 1911, when Claudius Regaud demonstrated that fractionating radiation doses—delivering them in small, repeated sessions instead of a single large dose—reduced normal tissue toxicity while maintaining tumour control (Foray 2012). Henri Coutard (1920s–1930s) later applied fractionated radiotherapy to head and neck cancers, establishing dose fractionation as a fundamental principle of modern radiotherapy (Prakash, Kumar Upadhyay *et al.* 2024).

Milestones in radiotherapy

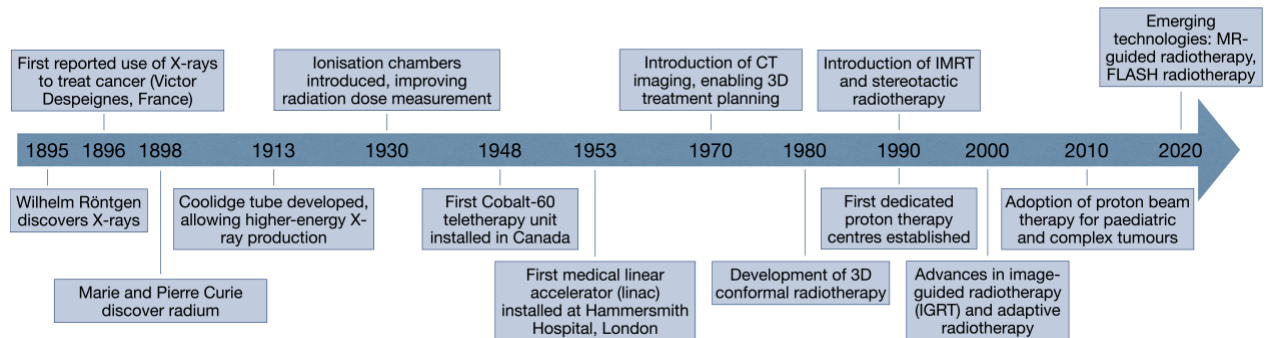


Figure 1.1. A timeline of the history of radiotherapy

The megavoltage era: from Kilovoltage to Linacs (1940s–1950s)

Early radiotherapy machines relied on Coolidge tubes (1913), which produced low-energy (kilovoltage) X-rays. These machines had limited penetration, causing severe skin damage while being ineffective for deep tumours. The development of megavoltage therapy revolutionised the field (Connell and Hellman 2009):

- **Cobalt-60 teletherapy (1948):** The first high-energy gamma ray therapy system was introduced, significantly improving deep tumour penetration while reducing skin toxicity.
- **First linear accelerator (1953, Hammersmith Hospital, London):** Linear accelerators (linacs) became the standard for high-energy external beam radiotherapy, delivering better dose control and sparing surrounding healthy tissues.

Advances in imaging and 3D treatment planning (1970s–1990s)

In the 1970s, the introduction of computed tomography (CT) imaging transformed radiotherapy by allowing precise tumour visualisation (Schulz, Stein *et al.* 2021). This led to the shift from 2D conventional radiotherapy to 3D conformal radiotherapy (3D-CRT), enabling tumour-specific dose sculpting and reducing radiation exposure to healthy tissues (Connell and Hellman 2009).

The 1990s saw a major breakthrough with the development of intensity-modulated radiotherapy (IMRT) (Cho 2018), which allowed radiation doses to be shaped with unprecedented precision around the tumour. This was followed by image-guided

radiotherapy (IGRT) (Grégoire, Guckenberger *et al.* 2020) and stereotactic body radiotherapy (SBRT) (Ricardi, Badellino *et al.* 2015) in the early 2000s, enabling high-dose, targeted treatment for small tumours, particularly in lung and brain cancers.

Proton therapy and the future of radiotherapy (1990s – Today)

One of the most significant modern advances is proton beam therapy (PBT). Unlike X-rays, protons deposit their energy at a precise depth (Bragg peak), minimising radiation exposure to surrounding normal tissues. While proton therapy was first proposed in 1946 by Robert Wilson, widespread clinical use only began in the 1990s, with the establishment of dedicated proton therapy centres. Today, PBT is widely used for paediatric cancers, base-of-skull tumours, and other radiation-sensitive malignancies, offering better tissue sparing than conventional photon therapy (Mohan 2022).

Radiotherapy continues to evolve with the development of cutting-edge technologies designed to improve treatment precision and minimise side effects. Adaptive radiotherapy (ART) utilises real-time imaging to continuously adjust radiation delivery throughout the course of treatment, allowing for modifications in response to changes in tumour size, shape, and position (Dona Lemus, Cao *et al.* 2024). MR-guided linear accelerators integrate magnetic resonance imaging (MRI) with radiotherapy, enabling real-time tumour tracking and improving accuracy, particularly for tumours in anatomically complex or mobile regions (Bryant, Weygand *et al.* 2023). Another emerging innovation is Flash radiotherapy, which delivers ultra-high dose rates in a single fraction, potentially reducing toxicity to surrounding healthy tissues while maintaining effective tumour control (Lin, Gao *et al.* 2021). These advancements represent the future of radiotherapy, aiming to further enhance patient outcomes through increased personalisation and precision.

Global impact of radiotherapy

Over more than a century, radiotherapy has evolved from the crude application of X-rays to a highly sophisticated and indispensable pillar of cancer treatment. What began as an experimental approach in the late 19th century is now a precisely targeted, technologically advanced therapy, capable of eradicating tumours,

preserving organ function, and providing symptom relief for millions of patients worldwide.

Radiotherapy plays a critical role in both curative and palliative care, often used alongside surgery and systemic therapies to maximise treatment effectiveness. With over 50% of all cancer patients requiring radiotherapy at some stage of their treatment, it remains one of the most effective and widely utilised cancer treatments, contributing to approximately 40% of all cancer cures (Baskar, Lee *et al.* 2012).

1.5.4 Preoperative vs. postoperative radiotherapy

Radiotherapy (RT) plays a key role in the management of soft tissue sarcomas (STS), particularly in improving local disease control and enabling limb-sparing surgery. According to the UK guidelines for the management of STS (Hayes, Nixon *et al.* 2024), RT should be considered for tumours with a high risk of local recurrence, either in the preoperative or postoperative setting. Historically, postoperative RT was the standard approach following surgery, based on early trials demonstrating a significant reduction in local recurrence rates from above 30% to below 10%, although no survival benefit was observed (Pisters, Harrison *et al.* 1996, Yang, Chang *et al.* 1998, Gronchi 2015).

Preoperative RT has become increasingly favoured due to its reduced long-term toxicity compared to postoperative RT (O'Sullivan, Davis *et al.* 2002). While both approaches achieve similar local control rates, preoperative RT allows for smaller treatment fields and lower doses (50 Gy vs. 60–66 Gy in the postoperative setting), reducing the risk of late effects such as fibrosis and joint stiffness (Hayes, Nixon *et al.* 2024, Noeuvéglise, Tessier *et al.* 2024). However, acute wound healing complications are more common with preoperative RT, necessitating careful patient selection.

Gronchi (2015) also highlighted the importance of individualising RT use, noting that while most STS patients historically received RT routinely, many derived no clear benefit. 70% of patients would not have had a recurrence even without RT, and 10%

recur despite it. This stresses the importance of selecting patients for RT based on their individual recurrence risk rather than applying it universally.

1.5.5 Histology-specific considerations in radiotherapy

STS is not a single disease but comprises over 80 histological subtypes, each with varying sensitivity to RT. Myxoid liposarcoma (MLS) is particularly radiosensitive, with over 50% of tumours demonstrating a major response to preoperative RT, often leading to significant tumour shrinkage before surgery (Chung, Deheshi *et al.* 2009). The UK guidelines (Hayes, Nixon *et al.* 2024) also emphasise this radiosensitivity and support the use of preoperative RT for MLS, particularly for borderline-resectable tumours, where tumour downsizing may facilitate surgical excision.

Conversely, certain other STS subtypes, such as undifferentiated pleomorphic sarcoma (UPS) and pleomorphic liposarcoma, have a higher baseline risk of local recurrence, making RT a critical component of their management. For low-risk subtypes, such as well-differentiated liposarcoma, RT may not be necessary if the tumour is completely resected with adequate margins.

1.5.6 Current UK radiotherapy guidelines

Radiotherapy, either preoperative or postoperative, combined with surgery provides similar local control and survival rates to radical resection. The addition of RT therefore allows for limb-sparing surgery. In many UK centres, preoperative RT is routine practice, with the standard regimen being 50 Gy over five weeks, followed by surgery four to six weeks after RT completion (Hayes, Nixon *et al.* 2024).

Despite this approach, clinical, radiological and pathological responses to RT vary both between and within STS histological subtypes (Messiou, Bonvalot *et al.* 2016, Kocakavuk, Anderson *et al.* 2021). While good responses are typically observed in myxoid liposarcoma, predicting clinical or pathological outcomes for RT across all sarcoma subtypes remains challenging, especially for high-grade tumours.

The UK guidelines (Hayes, Nixon *et al.* 2024) recommend a multidisciplinary approach, where RT decisions should be tailored based on tumour histology, grade, location, and resectability. Key recommendations include:

- **Preoperative RT** (50 Gy in 25 fractions) is preferred when feasible, particularly in histologies with high radiosensitivity (e.g., MLS) or where surgical margins are expected to be close.
- **Postoperative RT** (60–66 Gy in 30–33 fractions) is used when margins are positive or close after surgery.
- **Surgery alone** may be sufficient for low-grade tumours or cases where wide-margin resection is achievable without functional compromise.

Despite these advances, the optimal role of RT in some STS subtypes remains uncertain, as large randomised controlled trials are difficult due to the rarity of the disease. As Gronchi (2014) noted, evidence generation in rare cancers requires collaborative efforts, and retrospective data often guide difficult decision-making.

1.5.7 Side effects of radiotherapy

Radiotherapy remains central to modern cancer treatment, with advancements in precision techniques significantly improving tumour targeting while minimising damage to surrounding normal tissues. However, despite these improvements, radiation-induced side effects remain a major challenge, affecting both short- and long-term patient outcomes.

Radiation side effects can be broadly classified into acute and late effects. Acute effects occur during or shortly after treatment and are typically caused by direct damage to rapidly dividing cells. These include mucositis, skin reactions (radiodermatitis), nausea, fatigue, and inflammation of irradiated organs, such as pneumonitis following lung irradiation. While many acute effects resolve after treatment, they can be debilitating and impact a patient's quality of life (Barazzuol, Coppes *et al.* 2020).

Late side effects of radiotherapy can develop months to years after treatment, often leading to chronic and sometimes irreversible complications. These effects are primarily driven by persistent inflammation, fibrosis, vascular damage, and stem cell depletion in irradiated tissues. In soft tissue sarcoma patients, particularly those treated with limb-sparing surgery and radiotherapy, these late toxicities can significantly impact mobility, function, and quality of life. The severity of complications depends on radiation dose, treatment technique, and the volume of normal tissue exposed (Barazzuol, Coppes *et al.* 2020).

One of the most common and debilitating late effects is radiation-induced fibrosis, which leads to stiffness, contractures, and reduced limb mobility. When large portions of a joint are included in the radiation field, fibrosis can result in permanent contractures, limiting range of motion and functional independence. Studies have shown that 20% of STS patients develop joint contractures following radiotherapy, which can impair daily activities and require long-term physiotherapy or orthotic support (Stinson, Delaney *et al.* 1991).

Muscle atrophy and weakness are also frequent complications, with up to 20% of patients experiencing significant reductions in strength. This is often accompanied by radiation-induced damage to nerves and blood vessels, leading to neuropathy, chronic pain, and sensory deficits. In some cases, pain is severe enough to require long-term analgesic management. Lymphoedema, a result of lymphatic damage, affects nearly one in five patients, causing persistent limb swelling, heaviness, and an increased risk of recurrent infections (Stinson, Delaney *et al.* 1991).

Bone complications are another major concern, with 6% of patients developing pathological fractures due to radiation-induced bone fragility. This is particularly problematic in weight-bearing bones, such as the femur or tibia, where fractures can lead to prolonged immobility, surgical interventions, or even limb amputation in severe cases. Additionally, vascular damage within irradiated tissues can impair healing, increasing the risk of chronic ulcers and infection, particularly in the lower extremities (Stinson, Delaney *et al.* 1991).

The cumulative burden of these late toxicities can severely impact mobility, independence, and overall quality of life. Many patients require assistive devices such as canes, crutches, or orthotic braces, while some may face permanent functional disability. The use of modern radiotherapy techniques, such as intensity-modulated radiotherapy (IMRT) and proton therapy has helped to reduce late toxicity, but radiation-related functional impairments remain a significant long-term challenge in patients. Careful patient selection, treatment planning, and rehabilitation strategies are essential to minimise these effects while maintaining optimal oncological control.

1.6 Prognostic factors in soft tissue sarcoma

The prognosis of soft tissue sarcomas is influenced by a range of factors, including histological grade, tumour size, anatomical location, stage at diagnosis, resection margins, and patient-related characteristics such as age and performance status (Lebas, Le Fèvre *et al.* 2023, Díaz Casas, Villacrés *et al.* 2024). Understanding these factors is critical for risk stratification, guiding treatment decisions, and providing accurate survival estimates. These factors will be briefly discussed below.

1.6.1 Histological grade

Histological grade is one of the strongest independent prognostic factors in STS and is incorporated into widely used staging systems such as the AJCC (American Joint Committee on Cancer) staging system. The grading system recommended by the European Organisation for Research and Treatment of Cancer (EORTC) is the French Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC) grading system (Cyril Fisher 2022).

The FNCLCC system scores tumours based on three separate categories (Table 2) (Trojani, Contesso *et al.* 1984, Guillou, Coindre *et al.* 1997):

The individual scores for tumour differentiation, mitotic count, and necrosis are summed to determine the histological grade of the tumour. Tumours are classified as follows:

Grade 1 (low grade): Total score of 2 or 3

Grade 2 (intermediate grade): Total score of 4 or 5

Grade 3 (high grade): Total score of 6, 7, or 8

High-grade sarcomas are strongly associated with worse clinical outcomes, including increased rates of local recurrence, distant metastasis, and poorer overall survival (Lee, Kim *et al.* 2021).

Table 2. FNCLCC grading system

Soft tissue sarcoma grading criteria set by the French Fédération Nationale des Centres de Lutte Contre le Cancer. Adapted from the RCPATH soft tissue sarcoma dataset (Cyril Fisher 2022).

Category	Score	Criteria
Tumour Differentiation	1	Sarcoma histologically very similar to normal adult mesenchymal tissue
	2	Sarcoma of defined histological subtype (e.g. myxofibrosarcoma)
	3	Sarcoma of uncertain type, embryonal, and undifferentiated sarcomas
Mitosis Count	1	0–9 mitoses per 10 high-power fields (HPF) (2 sq mm)
	2	10–19 mitoses per 10 HPF (2 sq mm)
	3	>20 mitoses per 10 HPF (2 sq mm)
Microscopic Tumour Necrosis	0	No necrosis
	1	<50% tumour necrosis
	2	>50% tumour necrosis

A recent study highlighted the significant impact of histological grade on prognosis. For instance, grade 3 sarcomas were found to have a significantly higher risk of local recurrence compared to lower-grade sarcomas and a markedly increased risk of distant metastasis (Díaz Casas, Villacrés *et al.* 2024). Poorly differentiated tumours, as reflected in higher FNCLCC grading scores, are predictive of aggressive behaviour, systemic progression, and increased mortality. The study further emphasised that tumour size (>5 cm), deep fascial involvement, and inadequate surgical margins exacerbate these risks.

1.6.2 Tumour size

Tumour size is a key component of AJCC staging and directly correlates with metastatic risk. The most recent AJCC 8th edition subdivides size categories depending on the site. For tumours arising in the extremities, the superficial trunk, and retroperitoneum, the cut-offs are as follows (Cyril Fisher 2022):

T1: ≤ 5 cm

T2: >5 cm and ≤ 10 cm

T3: >10 cm and ≤ 15 cm

T4: >15 cm

Tumours arising in the head and neck or in the thoracic or abdominal viscera have their own specific staging criteria.

Larger tumours (>5 cm) are associated with:

Worse disease-free survival (DFS): Larger tumours are more likely to recur locally or metastasise, reducing the duration of DFS. For example, tumours >10 cm exhibit a threefold increase in metastatic risk compared to tumours ≤ 5 cm (Lee, Kim *et al.* 2021, Lebas, Le Fèvre *et al.* 2023).

Higher risk of distant metastases: The probability of metastasis to the lungs or other distant sites increases significantly with tumour size. In sarcomas larger than 10 cm, lung metastases are particularly common, further contributing to poor outcomes (Lebas, Le Fèvre *et al.* 2023, Díaz Casas, Villacrés *et al.* 2024).

In addition to size, the anatomical location of the tumour also significantly impacts prognosis. Retroperitoneal sarcomas are particularly challenging due to their location and often reach a substantial size before detection, with many exceeding 15 cm at diagnosis. This delayed presentation arises from their asymptomatic progression until they exert mass effects on adjacent organs. Consequently, surgical excision is more complex, often requiring multi-organ en bloc resections to achieve negative margins. Despite such efforts, retroperitoneal sarcomas are associated with worse survival outcomes compared to sarcomas in the extremities, with five-year overall

survival (OS) rates ranging from 39% to 70% depending on the subtype and resection margins (Guo, Zhao *et al.* 2022). Moreover, local recurrence remains the primary cause of mortality, with rates as high as 75% in patients without distant metastases.

The prognostic significance of tumour size highlights the importance of early diagnosis and intervention. Smaller tumours identified and treated at an earlier stage are associated with improved outcomes, stressing the need for vigilance in diagnosing and staging soft tissue sarcomas. This has led to one of the key recommendations from the UK guidelines on the management of soft tissue sarcomas which states that “any patient with an unexplained lump that is increasing in size, should be considered for a direct access ultrasound scan to be performed within 2 weeks” (Hayes, Nixon *et al.* 2024).

1.6.3 Anatomical site

The anatomical location of STS influences prognosis due to differences in surgical resectability, metastatic potential, and response to treatment.

Extremity sarcomas (limbs) generally have a better prognosis than those in retroperitoneal or visceral locations (Lebas, Le Fèvre *et al.* 2023, Díaz Casas, Villacrés *et al.* 2024). Retroperitoneal sarcomas as previously discussed tend to be diagnosed at larger sizes, often >15 cm, making complete resection difficult, with worse five-year survival rates (~40-50%) (Guo, Zhao *et al.* 2022).

1.6.4 Stage at diagnosis

In the UK, STS is staged using the AJCC TNM system, which incorporates tumour size (T), lymph node involvement (N), and distant metastases (M) (Cyril Fisher 2022, Hayes, Nixon *et al.* 2024).

Unlike carcinomas, which typically metastasise through lymphatic channels, soft tissue sarcomas (STS) predominantly spread haematogenously, with the lungs, liver,

and bones being the most common metastatic sites (Pennacchioli, Tosti *et al.* 2012). While lymphatic involvement is rare in STS compared to carcinomas, certain subtypes exhibit a higher propensity for lymph node metastasis (LNM). For example, one study found clear cell sarcoma, epithelioid sarcoma, angiosarcoma, and small cell sarcoma had LNM rates of 15.9%, 13.1%, 6.1%, and 19.1%, respectively. These rates are markedly higher than the overall LNM incidence in STS, which remains low at approximately 3.5% (Keung, Chiang *et al.* 2018).

The likelihood of LNM varies not only by histological subtype but also by tumour location. For example, head, neck, and visceral sarcomas demonstrate slightly higher LNM rates compared to those arising in the extremities with 5.8% of head and neck sarcomas and 5.1% of intra-abdominal sarcomas having nodal involvement, compared to just 2% in the extremities.

The clinical significance of LNM in STS is profound. In the absence of distant metastases (M0), LNM is associated with a worse overall survival (OS). For instance, patients with isolated LNM (pN1M0) experience a median OS of 2.4 years, significantly shorter than the 8.5 years seen in patients without nodal or distant metastases (N0M0). Moreover, histological subtypes like angiosarcoma and clear cell sarcoma tend to have particularly poor prognoses when associated with LNM, with median OS as low as 19.4 months for angiosarcoma.

These findings show the importance of accurately staging lymph node involvement, particularly in histologies at higher risk of nodal spread. However, current practice often lacks consistency in pathologically evaluating lymph nodes in STS. A significant proportion of cases rely on clinical rather than pathological confirmation of nodal disease, which may affect staging accuracy and subsequent management.

The M stage in the AJCC TNM system reflects the presence of distant metastases, a critical determinant of prognosis in soft tissue sarcomas. As mentioned earlier, haematogenous spread, rather than lymphatic spread, is the predominant route of metastatic dissemination in STS. The lungs are the most common site of metastases, and account for over 90% of metastases, followed by the liver and bones.

Patients presenting with distant metastases (M1 stage) have significantly poorer outcomes compared to those with localised disease. Median overall survival (OS) for patients with metastatic STS is approximately 12 months, although recent advancements in treatment, including systemic therapies, have extended survival in some cases to around 18 months (Hayes, Nixon *et al.* 2024). This remains markedly worse than the survival rates observed in early-stage disease.

The presence of distant metastases often limits curative treatment options, with management typically focusing on systemic therapies such as chemotherapy or targeted agents. In some cases, metastasectomy (e.g., resection of pulmonary metastases) may be considered, especially in patients with limited disease burden, though this approach is not suitable for all subtypes (Sardenberg, Figueiredo *et al.* 2010).

Given the poor prognosis associated with metastatic STS, accurate staging with imaging modalities such as CT scans and PET scans is essential. Early detection of metastases informs treatment planning, helping clinicians determine whether the goal of care should be curative or palliative.

A study from the UK, analysing data on soft tissue sarcoma cases between 2013 and 2017, provides recent survival rates based on disease stage (Bacon, Wong *et al.* 2023):

Stage I: Low-grade, small tumours with no metastases have the most favourable prognosis, with a five-year survival rate of approximately 85-90%.

Stage II/III: Patients with high-grade or larger tumours, who often require multimodal treatment including surgery, radiotherapy, and systemic therapy, have a five-year survival rate ranging from 50-70%.

Stage IV: The presence of distant metastases confers a significantly worse prognosis, with a five-year survival rate of less than 20%, reflecting the aggressive nature of advanced STS and the limited efficacy of systemic therapies in metastatic disease.

1.6.5 Resection margin status

Surgical resection is the mainstay of curative treatment for soft tissue sarcomas (STS), with margin status being a critical determinant of prognosis and recurrence risk. The Royal College of Pathologists' dataset highlights the importance of accurate margin reporting, including measurements in millimetres and tissue type at the margin (e.g., fascia, muscle, fat), with margins classified as follows (Cyril Fisher 2022):

- **R0 Resection (Negative Margins):** Clear surgical margins with no tumour cells at the inked edge are associated with the lowest recurrence rates and best survival outcomes, making this the gold standard for STS surgery.
- **R1 Resection (Microscopic Residual Disease):** Margins with microscopic tumour presence increase recurrence risk. Adjuvant radiotherapy is often employed to improve local control and reduce recurrence.
- **R2 Resection (Macroscopic Residual Disease):** Macroscopic tumour left behind leads to a significantly poorer prognosis, with five-year survival rates below 30%. R2 resections are generally considered palliative.

The European Society for Medical Oncology – European Reference Network for rare adult solid cancers (ESMO-EURACAN) guidelines (Casali, Abecassis *et al.* 2018) recommend R0 resection as the primary objective, particularly for extremity tumours where wide margins are achievable. For R1 resections, reoperation in a reference centre is advised if adequate margins can be achieved without major morbidity. In cases where re-excision is not feasible, adjuvant radiotherapy is strongly recommended to improve local control.

For retroperitoneal sarcomas or tumours near critical structures, achieving R0 margins may not be feasible. ESMO guidelines state that planned R1 resections can yield acceptable outcomes when combined with radiotherapy. The impact of R1 margins varies by tumour subtype, location, and margin characteristics.

Additionally, marginal excision may be appropriate for certain low-grade tumours, such as atypical lipomatous tumours, where local recurrence rates are low, and radiotherapy may not be necessary. These situations highlight the need for individualised surgical strategies that balance safety with patient quality of life. Indication of the use of a marginal excision should be put on the pathology request form when a specimen is submitted for pathological assessment.

1.6.6 Patient age, performance status, and socioeconomic background

Age, performance status, and socioeconomic factors significantly influence survival outcomes in patients with soft tissue sarcomas (STS). Older patients, particularly those aged >65 years, tend to have poorer prognoses. A population-based study from England reported that while the overall five-year net survival for STS was 65%, survival rates were significantly lower in older age groups, attributed to comorbidities and reduced tolerance for aggressive treatments (Bacon, Wong *et al.* 2023).

Younger patients with good performance status (ECOG 0-1) demonstrate better long-term survival, even in cases of high-grade disease. The European Society for Medical Oncology (ESMO) guidelines emphasise the need for individualising treatment plans to account for age, tumour biology, and performance status (Casali, Abecassis *et al.* 2018)

Socioeconomic background also impacts outcomes in STS. The Bacon *et al.* study highlights disparities in survival linked to socioeconomic status. Patients in the most deprived quintile had a five-year net survival rate of approximately 55%, compared to 70% in the least deprived quintile. These disparities may reflect inequalities in access to care, delayed diagnoses, or differences in treatment availability (Bacon, Wong *et al.* 2023).

1.7 The biological effects of radiotherapy

Radiotherapy primarily exerts its therapeutic effects by inducing extensive DNA damage in cancer cells, ultimately leading to cell death. The energy from ionising radiation generates DNA double-strand breaks (DSBs), which, if left unrepaired or misrepaired, trigger apoptosis or mitotic catastrophe. Ionising radiation also generates reactive oxygen species (ROS), which cause single-strand breaks (SSBs) and oxidative base modifications, further compromising genomic integrity (Horsman MR 2009). In addition to direct DNA cleavage, radiation can induce chemical modifications such as 5' hydroxyls, 3' phosphates, and covalent DNA-protein crosslinks, which require resolution before DSB repair can occur (Borrego-Soto, Ortiz-Lopez *et al.* 2015).

The ability of a tumour to repair radiation-induced DNA damage influences its sensitivity to treatment. Tumours deficient in homologous recombination repair (HRR), such as those with *BRCA1/2* mutations, exhibit heightened radiosensitivity due to their inability to accurately repair DSBs (Ernestos, Nikolaos *et al.* 2010). In contrast, tumours with an overactive non-homologous end-joining (NHEJ) pathway (discussed previously in section 1.4.3.1) may demonstrate radiation resistance, as NHEJ rapidly, albeit error-prone, ligates broken DNA ends (Morgan and Lawrence 2015). The relative reliance on different DNA repair pathways contributes to tumour-specific responses to radiotherapy. This underlies the rationale for radiosensitising agents such as PARP inhibitors, which exploit DNA repair deficiencies to enhance radiation efficacy (Angel, Zarba *et al.* 2021).

1.7.1 Mutational signatures associated with radiotherapy

Mutational signatures are distinct patterns of somatic mutations found in cancer genomes that reflect the biological processes causing genetic alterations. These signatures arise due to endogenous processes, such as spontaneous deamination or replication errors, or exogenous exposures, such as ionising radiation, ultraviolet light, or chemotherapy. By analysing large sequencing datasets, computational methods allow us to identify and categorise these signatures, helping to reveal the underlying mechanisms of mutagenesis.

One key approach for identifying mutational signatures is trinucleotide context analysis. Mutations are classified based on the flanking bases surrounding a mutated site, leading to 96 possible mutation types. This level of granularity is crucial because many mutagens exhibit sequence specificity, and analysing mutations in their sequence context improves the accuracy of signature extraction.

To systematically identify mutational signatures, non-negative matrix factorization (NMF) is applied to mutation frequency data across multiple cancer genomes. NMF is a mathematical technique that deconstructs complex mutational spectra into a set of underlying signatures. Each extracted signature is then compared against known reference signatures, such as those curated in COSMIC, allowing for the attribution of mutations to specific mutational processes (Alexandrov, Nik-Zainal *et al.* 2013).

Mutational signatures are broadly classified into several categories, including single base substitution (SBS) signatures, doublet base substitution (DBS) signatures, insertion-deletion (ID) signatures, and, more recently, copy number alteration (CNA) signatures (Alexandrov, Kim *et al.* 2020, Steele, Abbasi *et al.* 2022).

1.7.2 Radiation induced mutational signatures

Several studies have identified mutational signatures enriched in post-radiotherapy malignancies. Behjati *et al.* (2016) performed whole-genome sequencing on radiation-associated secondary malignancies and identified two characteristic mutational features:

- An excess of balanced inversions, a rare form of structural rearrangement.
- An increased burden of small deletions, which were validated in a separate cohort of prostate cancer patients who had received radiotherapy.

Research on thyroid cancers arising after the Chernobyl nuclear disaster found a radiation dose-dependent increase in small deletions and simple/balanced structural variants, particularly an increased deletion-to-SNV ratio (Morton, Karyadi *et al.* 2021). These findings suggest that radiation-induced DSBs are frequently repaired

through error-prone mechanisms such as NHEJ, leading to characteristic genomic alterations, including deletions.

1.7.3 Radiation-induced APOBEC mutagenesis

Beyond direct DSB induction, ionising radiation has been implicated in the activation of endogenous mutagenic processes, particularly the APOBEC cytosine deaminase family. APOBEC enzymes preferentially induce C>T or C>G substitutions at TpC dinucleotides, corresponding to mutational signatures SBS2 and SBS13 (Alexandrov, Kim *et al.* 2020).

Kocakavuk *et al.* (2021) identified enrichment of SBS2 and SBS13 in gliomas and metastatic tumours following radiotherapy, suggesting that APOBEC mutagenesis might occur in post-radiotherapy tumours. However, rather than direct induction by radiation, the study proposed that APOBEC-driven mutagenesis may be a secondary effect of DNA damage repair. This aligns with previous findings that APOBEC enzymes act on single-stranded DNA, which can be transiently generated during the repair of radiation-induced DSBs (Schlegel, Jodelka *et al.* 2006).

Crucially, Morton *et al.* (2021) found no significant association between radiation dose and APOBEC mutational signatures (SBS2/SBS13) in thyroid cancers from the Chernobyl cohort. Although SBS2 and SBS13 made up 6.2% and 6.4% of attributed mutations, their presence was not linked to radiation exposure. This suggests that APOBEC activity may be involved in tumour progression but is not a direct consequence of radiation exposure.

1.7.4 ID8 and DNA Repair pathways

Insertion-deletion signature ID8, which is associated with microhomology-mediated end joining (MMEJ), has been identified in multiple radiation-exposed tumour cohorts. A significant increase in ID8 was noted in gliomas following radiotherapy in the GLASS cohort (Kocakavuk, Anderson *et al.* 2021) and in post-radiotherapy papillary thyroid carcinomas (Morton, Karyadi *et al.* 2021). However, further analysis

suggested that classical non-homologous end joining (c-NHEJ), rather than MMEJ, is the primary repair mechanism for radiation-induced DSBs.

The identification of clonal small deletions and the enrichment of radiation-associated ID8 signatures suggest that radiotherapy induces genomic instability through error-prone DSB repair pathways, predominantly classical NHEJ. These findings highlight the role of radiotherapy in shaping the mutational landscape of soft tissue sarcomas and the importance of understanding these alterations in the context of tumour biology. Further characterisation of these mutational processes will help refine our understanding of how radiotherapy impacts tumour evolution and may provide insights into potential therapeutic vulnerabilities.

1.8 The pathological assessment of the response to radiotherapy

Assessing the pathological response to radiotherapy in soft tissue sarcomas presents significant challenges for histopathologists. This process involves both macroscopic and microscopic evaluations, each of which is subject to limitations in sampling and interpretation.

At a macroscopic level, the primary issue is accurate recording of the proportion of necrosis and of the residual tumour. In the UK, Royal College of Pathologists (RCPATH) guidelines recommend documenting tumour size, colour, consistency, and necrosis as a percentage of the total tumour mass (Cyril Fisher 2022). However, quantifying necrosis macroscopically is prone to error, as it relies on gross appearance rather than cellular assessment.

At a microscopic level, the challenge lies in distinguishing treatment-related changes from tumour-related necrosis and ensuring representative sampling of viable tumour. The RCPATH guidelines recommend sampling one block per 10 mm of the tumour's longest dimension, with a maximum of 12 blocks, though high-grade tumours may require fewer (Cyril Fisher 2022). However, this approach inherently biases against necrotic areas, as pathologists typically prioritise viable tumour for microscopic examination.

To improve standardisation, the European Organization for Research and Treatment of Cancer - Soft Tissue and Bone Sarcoma Group (EORTC-STBSG) has proposed an alternative approach: sampling and blocking an entire representative central slice of the tumour (Wardelmann, Haas *et al.* 2016). However, selecting a truly "representative" slice remains subjective and introduces its own sampling bias.

1.8.1 Histopathological features of radiotherapy response

Histological features suggestive of radiotherapy-induced changes include:

- Necrosis
- Ghost cells (cells with loss of nuclear and cytoplasmic detail)
- Granulation tissue and fibrosis
- Hemosiderin deposition
- Foamy macrophages
- Calcifications and inflammatory changes.

An important limitation of assessing tumour necrosis alone is that pre-treatment necrosis cannot be reliably distinguished from post-radiotherapy necrosis.

1.8.2 Correlation with radiological response criteria

To improve response assessment, correlation with radiological imaging has been proposed. The Response Evaluation Criteria in Solid Tumours (RECIST), which relies on tumour shrinkage, is often not useful in soft tissue sarcomas, as most do not significantly decrease in size following radiotherapy (Betgen, Haas *et al.* 2013, Wardelmann, Haas *et al.* 2016). In some cases, tumours may even increase in size due to cystic transformation or haemorrhage, a phenomenon known as pseudoprogression.

An alternative is the Choi criteria, originally developed for gastrointestinal stromal tumours (GISTs). These criteria assess response based on both size reduction and changes in tumour density on imaging. Studies have shown that the Choi criteria outperform RECIST in predicting soft tissue sarcoma response to chemotherapy and radiotherapy outcome (Stacchiotti, Verderio *et al.* 2012).

Additionally, MRI techniques such as diffusion-weighted imaging (DWI) and contrast-enhanced MRI have been explored as tools for estimating tumour necrosis post-radiotherapy. One study demonstrated that MRI-derived necrosis percentages correlated well with histopathological assessment, though distinguishing true

necrosis from oedema or tumour recurrence remains challenging (Monsky, Jin *et al.* 2012, Nichelli and Casagrande 2021)

1.8.3 Lack of consensus in pathological reporting

Despite multiple proposed approaches, no consensus exists on the optimal method for reporting pathological response to radiotherapy in sarcomas. Some studies report percentage necrosis, while others quantify viable tumour percentage. The EORTC-STBSG scoring system, which categorises response into five tiers based on viable tumour percentage, was tested in an independent cohort of 100 sarcoma patients' post-radiotherapy. However, neither the EORTC response score nor percentage viable tumour were prognostic (Wardelmann, Haas *et al.* 2016).

1.8.4 Implications for tissue processing and molecular analysis

From a molecular research perspective, variability in sampling and tissue processing poses challenges for genomic and transcriptomic studies. DNA and RNA extraction from formalin-fixed, paraffin-embedded (FFPE) samples requires sufficient viable tumour cells, as excessive necrosis can interfere with sequencing protocols. This highlights the importance of optimising tumour sampling methods, particularly in studies seeking to define molecular biomarkers of radiotherapy response.

1.8.5 Conclusion

The lack of standardisation in pathological response assessment presents a major challenge in understanding the effects of radiotherapy in sarcomas. While percentage necrosis remains the most commonly reported metric, it is inherently limited by sampling bias and the inability to distinguish pre-existing from treatment-induced necrosis. Alternative methods, such as MRI-derived necrosis measurements and Choi criteria, may improve response evaluation but require further validation.

Future efforts should focus on refining the pathological assessment of radiotherapy response by incorporating objective and reproducible metrics that better reflect tumour biology. While radiological and histopathological criteria remain the current standard, they have limitations in accurately capturing the molecular impact of radiotherapy. By characterising the genomic alterations and mutational processes induced by radiotherapy, this thesis aims to understand the biological consequences of treatment, which could ultimately inform the development of molecular biomarkers for response assessment in soft tissue sarcomas.

1.9 Rationale and objectives of this thesis

Radiotherapy is a key component in the management of soft tissue sarcomas, enabling limb-sparing surgery and improving local control. However, its clinical efficacy varies significantly across histological subtypes, and the absence of robust molecular biomarkers limits the ability to personalise treatment. There are no published studies using clinical material describing the short-term molecular changes that occur during and immediately after therapy. In the setting of local recurrence or metastasis, where patients have received radiotherapy, prior studies have identified broad genomic changes including mutational signatures and activation of DNA repair pathways, however the precise molecular mechanisms underpinning radiotherapy response remain poorly understood. This is largely due to the lack of comprehensive multi-omics analyses directly comparing pre- and post-radiotherapy tumour samples, as well as the limited ability to distinguish treatment-induced changes from tumour-intrinsic alterations.

Although previous work has characterised some mutational consequences of radiation exposure in secondary malignancies and radiation-induced sarcomas, studies focusing on primary soft tissue sarcomas treated with neoadjuvant radiotherapy remain limited. Moreover, existing analyses often rely on bulk sequencing approaches with low sensitivity, making it challenging to resolve low-frequency mutations, copy number alterations, or changes in the tumour microenvironment. Additionally, the histology-specific variability in genomic and transcriptomic responses to radiotherapy remains unclear, despite growing evidence that certain subtypes, such as myxoid liposarcoma, demonstrate a higher sensitivity to treatment.

Furthermore, while histopathological and radiological criteria are used to assess radiotherapy response, their accuracy in predicting treatment outcomes is limited. The lack of standardised pathological criteria complicates efforts to integrate molecular findings into clinical practice, and there remains an urgent need to define reproducible and biologically meaningful markers of radiotherapy response.

To address these gaps, this thesis applies whole-exome sequencing (WES), whole-genome sequencing (WGS), RNA sequencing (RNAseq), and high-sensitivity sequencing methods (NanoSeq) to characterise the genomic and transcriptomic alterations induced by radiotherapy in soft tissue sarcomas. By leveraging these high-resolution sequencing approaches, this study aims to provide a comprehensive understanding of how the genomic and transcriptomic landscape of soft tissue sarcomas is altered following radiotherapy, while also exploring potential predictive biomarkers of treatment response. These findings will enhance our understanding of radiotherapy-induced mutagenesis, define molecular changes associated with treatment response, and lay the groundwork for future biomarker development to guide personalised therapeutic strategies.

The specific objectives of this thesis are:

1. To characterise radiotherapy-induced genomic alterations.

- Quantify and compare somatic mutations, including SNVs and indels, in pre- and post-radiotherapy tumour samples.
- Investigate copy number alterations (CNAs) and mutational signatures to identify radiotherapy-induced genomic changes.

2. To analyse the transcriptional landscape in response to radiotherapy.

- Identify differentially expressed genes and enriched biological pathways in pre- and post-radiotherapy tumour samples.
- Explore how transcriptomic responses differ between sarcoma histological subtypes.

3. To identify histology-specific variability.

- Determine whether different sarcoma subtypes exhibit distinct genomic and transcriptomic responses to radiotherapy.

4. To establish a molecular framework for understanding radiotherapy response.

- Use multi-omics data to refine our understanding of radiotherapy-induced mutagenesis.
- Provide a foundation for future biomarker discovery by characterising molecular patterns associated with treatment response.

Chapter 2. Materials & Methods

2.1 Ethics approval and patient sample selection

This study was approved by the NHS Health Research Authority (REC reference 16/NW/0769). Patient tissue and clinical data were obtained from the Royal National Orthopaedic Hospital biobank in collaboration with the London Sarcoma Service. All patients had provided informed consent for research use.

A total of 122 patients were assessed for inclusion in the study. These patients were identified by Dr Nischalan Pillay and Dr Mahbub Ahmed. Eligibility criteria were:

- Histologically confirmed soft tissue sarcoma. All included cases underwent histological review to confirm diagnosis and eligibility (see section 2.2).
- Treatment at the London Sarcoma Service with standard of care pre-operative (neoadjuvant) radiotherapy (see section 2.1.1).
- Availability of diagnostic (pre-radiotherapy) tumour tissue and, where applicable, matched post-radiotherapy resection tissue.
- Adequate tissue quality and quantity for downstream molecular assays (RNA-seq, NanoSeq, WES) and histological review.
- Sufficient clinical and treatment data to confirm radiotherapy dose, fractionation, timing, and follow up data.

Sixty-one patients were excluded for the following reasons:

Post-operative radiotherapy only (n = 24): Patients received radiotherapy only after surgery, with no pre-radiotherapy (diagnostic) tumour tissue available.

Radiotherapy not completed (n = 3): Radiotherapy was not given, stopped early, declined, or administered with palliative intent only.

No viable tumour in available tissue (n = 6): Resection specimens showed complete pathological response or extensive necrosis, leaving no viable tumour for analysis.

Insufficient or unavailable tissue (n = 12): Diagnostic or resection blocks were missing, untraceable, or contained insufficient tumour material for sequencing; in some cases, paired normal tissue was not available.

Eligibility not confirmed (n = 16): Treated outside the specialist London Sarcoma Service, preventing confirmation of radiotherapy dose, timing, and delivery.

The final study cohort comprised 61 patients. Of these:

- NanoSeq was performed on 10 samples, 9 of which passed quality control (QC).
- DNA sequencing (WES/WGS) was performed on 70 samples, 67 passed QC.
- RNA-seq was performed on 122 samples, 117 passed QC.

A Consort-style schematic summarising the inclusion and exclusion process is shown in Figure 2.1.

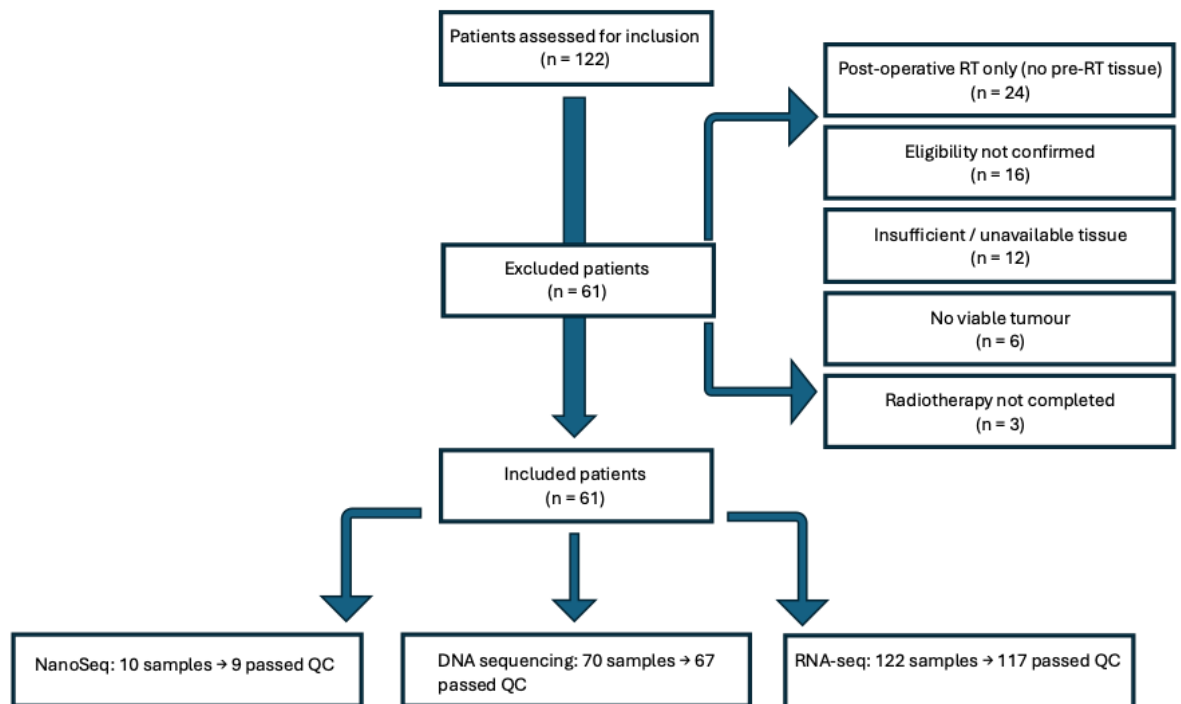


Figure 2.1. Consort diagram of patient inclusion and sample quality control flowchart

The diagram summarises patient selection, reasons for exclusion, and the number of samples processed for sequencing assays.

2.1.1 Radiotherapy treatment

All patients received standard of care treatment through the London sarcoma service. All patients received 50 Gy of intensity modulated radiotherapy (IMRT) delivered in 25 fractions (see section 1.5.6).

The median interval between completion of neoadjuvant radiotherapy and surgical resection was 46 days (range: 13–84 days), with a mean of approximately 46.1 days. This was broadly consistent across the cohort, reflecting the standard practice in our centre of scheduling surgery around six to seven weeks after completion of treatment.

2.1.2 Defining disease progression

In the introduction, I noted that there are no universally established histological or radiological standards for defining response or progression in this context. Therefore,

in the later chapters, when distinguishing between progressors and responders, I rely on clinical follow-up data.

Patients were classified as **progressors** if they experienced either local recurrence or death from disease during the follow-up period. **Responders** were defined as those with no evidence of recurrence or disease-related mortality within the follow-up timeframe.

Of the 61 patients in the study, 25 were classified as progressors. This group included 4 patients who developed local recurrence (median follow-up: 2,448 days; range: 2,124–2,602 days) and 21 patients who died of disease (median follow-up: 892 days; range: 223–2,637 days). The combined progressor group had a median follow-up time of 955 days (range: 223–2,637 days).

The remaining 36 patients were classified as responders, with a median follow-up time of 2,584 days (range: 1,391–3,744 days).

A final list of included patients, their clinical characteristics, and sequencing data availability is presented in Figure 2.2.

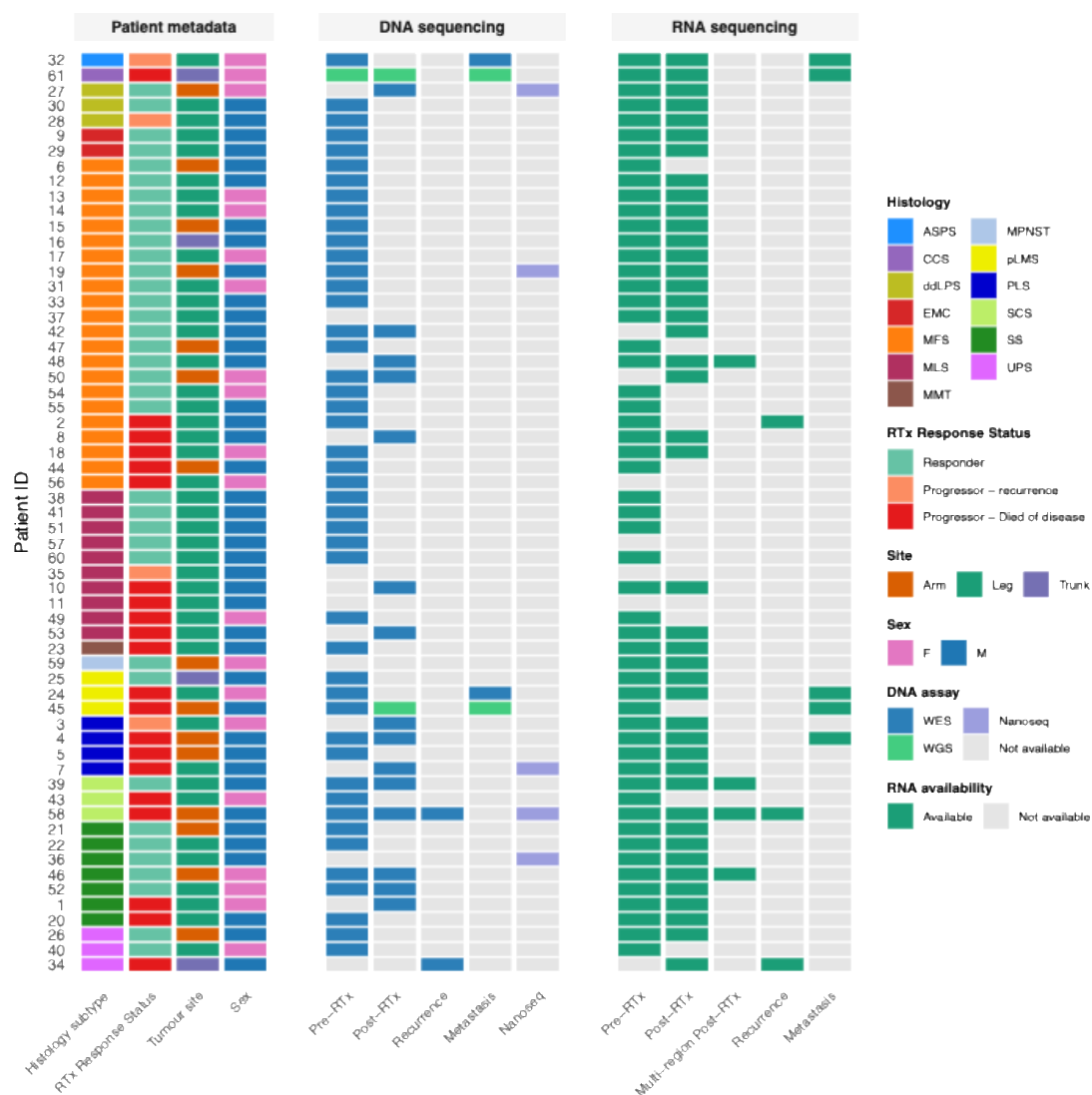


Figure 2.2. Overview of the London Sarcoma Service patient cohort, clinical features, and sequencing data availability

The heatmap summarises clinical metadata and sequencing data availability for 61 patients in the study. The **Patient metadata** panel shows histological subtype, radiotherapy (RTx) response status, tumour site, and sex. Histological subtypes include: MFS (myxofibrosarcoma), mLPS (myxoid liposarcoma), SS (synovial sarcoma), pLPS (pleomorphic liposarcoma), SpCS (spindle cell sarcoma), UPS (undifferentiated pleomorphic sarcoma), ddLPS (dedifferentiated liposarcoma), pLMS (pleomorphic leiomyosarcoma), EMC (extraskelletal myxoid chondrosarcoma), ASPS (alveolar soft part sarcoma), CCS (clear cell sarcoma), MLS (myxoid liposarcoma), MMT (malignant myoepithelial tumour), and MPNST (malignant peripheral nerve sheath tumour). RTx response status is categorised as: Responder, Progressor – recurrence, and Progressor – died of disease. Tumour site is recorded as arm, leg, or trunk; sex as female (F) or male (M). The **DNA sequencing** panel shows availability of whole exome sequencing (WES), whole genome sequencing (WGS), and NanoSeq data across the following time points: pre-RTx, post-RTx, recurrence, and metastasis. The **RNA sequencing**

panel shows RNA-seq data availability at pre-RTx, post-RTx, multi-region post-RTx, recurrence, and metastasis. Green shading indicates available data, grey shading indicates missing data.

2.2 Histological assessment of samples

Histological assessment of all available biopsy, resection, and, where applicable, recurrence and metastatic samples was performed by two pathologists, Steven Nottley (SN) and Dr Nischalan Pillay (NP). Samples were excluded from nucleic acid extraction if they exhibited insufficient tumour cellularity (<20%) or excessive necrosis, as these factors could compromise sequencing quality. Each sample was systematically evaluated for tumour cellularity, percentage of viable tumour, and proportion of necrosis (Figure 2.3).

2.3 Nuclei acid extraction

For formalin-fixed paraffin-embedded (FFPE) tissue, 4 × 10 µm sections were cut from each tumour block and nucleic acids extracted using the *truXTRAC FFPE total NA Ultra Kit – Column* (Covaris) according to the manufacturer's protocol. For each patient, both DNA and RNA were obtained from the same FFPE block, ensuring that whole-exome sequencing (WES) and RNA-seq data originated from the same physical tumour sample. Whole-genome sequencing (WGS) and NanoSeq analyses were performed on DNA extracted from fresh frozen tissue taken from the designated pre-, post-, recurrence-, or metastasis-radiotherapy sample as recorded. Matched germline DNA was isolated from the patient's peripheral blood.

2.4 Sequencing protocols

DNA Sequencing:

DNA samples were sent to MacroGen (Amsterdam) for library preparation using the Twist Human Core Exome library kit (Twist Bioscience). Quality control (QC) was performed, and samples meeting QC criteria were sequenced on an Illumina NovaSeq platform. Whole exome sequencing (WES) was conducted using 150 bp paired-end (PE) libraries to achieve a targeted mapped coverage of 250X.

For germline whole-exome sequencing, normal DNA from matched blood samples underwent library preparation using the Twist Human Core Exome kit and were

sequenced on the Illumina NovaSeq platform with 150 bp paired-end reads at 50X mapped coverage.

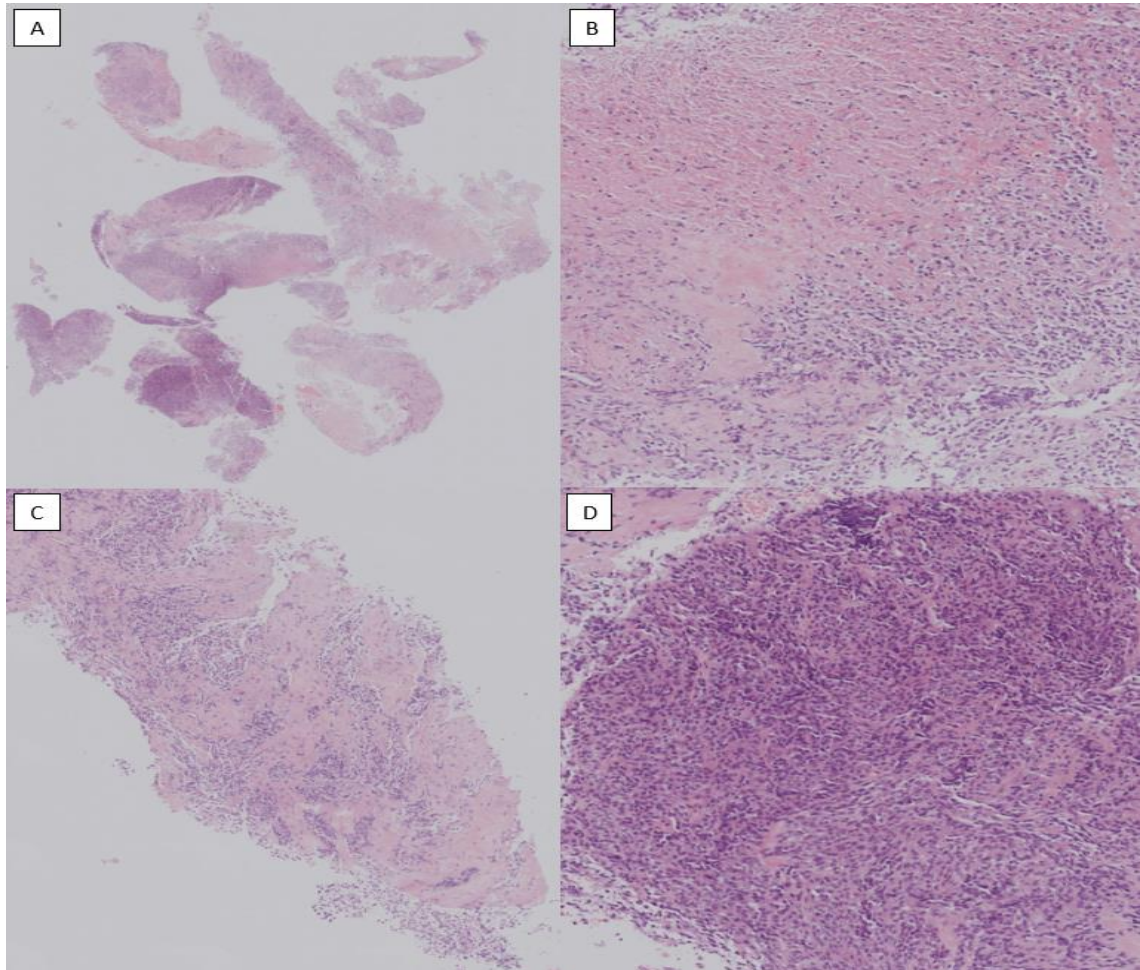


Figure 2.3. Histological assessment of tumour samples

Representative histological sections demonstrating tumour cellularity and necrosis evaluation. Tumour samples were assessed for cellularity, percentage of viable tumour, and extent of necrosis, key factors influencing sequencing quality. The case shown was assessed as high tumour cellularity (~50%) with ~20% necrosis. **(A)** Low-power view of core biopsies demonstrating tumour architecture. **(B)** Medium-power view highlighting areas of tumour necrosis (upper left). **(C)** Medium-power view showing tumour-stroma interaction and variability in tumour cellularity. **(D)** High-power view of a region with high tumour cellularity, with scattered admixed inflammatory cells.

RNA Sequencing:

RNA was extracted from FFPE tissue using the truXTRAC FFPE total NA Ultra Kit – Column (Covaris). As expected for FFPE-derived material, RNA integrity numbers (RIN) were low, ranging from 1.1 to 2.3. Samples were sent to MacroGen (Amsterdam) for further quality control, where RNA quality was assessed using DV200 metrics (percentage of RNA fragments >200 nucleotides). Samples with DV200 >40% proceeded to library preparation using the TruSeq RNA Exome kit (Illumina) using 100ng of input RNA. Sequencing was performed on an Illumina NovaSeq 6000 S4 platform with 100 bp paired-end reads, generating ~50 million read pairs (~100 million total reads) per sample, providing sufficient depth for downstream transcriptomic analyses. No correlation was observed between RIN values and downstream expression data quality, but samples failing the DV200 threshold were not sequenced.

2.5 Alignment, variant calling, and annotation**2.5.1 Computational Resources**

The alignment, variant calling, and generation of ensemble VCFs were performed on the UCL Myriad high-performance computing (HPC) cluster. The alignment stage used 32 CPU cores per job, while variant calling was performed with 8-core jobs. Java (v1.8.0_92) and Samtools (v1.9) were used across multiple stages, including alignment, variant calling, and panel of normals generation.

2.5.2 Alignment and pre-processing

Whole exome sequencing (WES) and whole genome sequencing (WGS) raw data in the form of fastq files were obtained from MacroGen. They were processed using the bcbio pipeline (Chapman, Kirchner *et al.* 2021), a community-developed bioinformatics framework for variant calling and analysis. The sequencing reads were aligned to the hg38 reference genome (1000g-20150219) using BWA-MEM (v0.7.17) (Li 2013), which is derived from the 1000 Genomes Project's release of hg38.

The alignment pipeline included duplicate marking (Picard v2.27.4), base quality score recalibration, and local realignment around indels using the Genome Atlas Toolkit - GATK (v4.2.6.1). Alignment parameters were optimised for paired-end sequencing, with multi-threading enabled (-t 8) for efficient processing. Quality control checks were further enhanced using Qualimap (v2.2.2d) to assess alignment metrics and sequencing depth.

2.5.3 Variant calling

Variant calling was conducted using a tumour-normal paired analysis approach to identify somatic mutations. The samples were processed with a configuration file specifying the use of three different variant callers: Mutect2 (GATK v4.2.6.1) (Benjamin, Sato *et al.* 2019), Strelka2 (v2.9.10) (Kim, Scheffler *et al.* 2018), and VardDict (v1.8.2) (Lai, Markovets *et al.* 2016). These tools were run in ensemble mode, requiring at least two variant callers to support a given mutation for inclusion in the final variant call set.

Variant calling was restricted to the exonic regions defined by the Twist Bioscience Exome Capture (hg38) target regions (Twist_Exome_Target_hg38.bed) available from (<https://www.twistbioscience.com/resources/data-files/ngs-human-core-exome-panel-bed-file>).

2.5.4 Panel of Normals (PON) generation

To improve the accuracy of somatic variant calling and reduce false positives, a Panel of Normals (PON) was generated using GATK Mutect2 (v4.2.5.0). Since the samples were derived from formalin-fixed paraffin-embedded (FFPE) tissue, they are prone to sequencing artefacts and technical noise. The PON helps to filter out recurrent sequencing artefacts, FFPE-induced damage, and germline variants that might otherwise be misclassified as somatic mutations in tumour samples.

1. Mutect2 Variant Calling:

- Each of the 54 normal samples was processed individually to generate normal VCFs.

- Variant calling used the hg38 reference genome and gnomAD (v2.1.1, af-only-gnomad.hg38.vcf.gz).

2. GenomicsDBImport:

- Normal VCFs were combined into a GenomicsDB database using GATK GenomicsDBImport (v4.2.5.0), with intervals defined by the Twist Bioscience Exome Capture (hg38) target regions.

3. Final PON Creation:

- GATK CreateSomaticPanelOfNormals (v4.2.5.0) was used to generate a single PON VCF file for downstream somatic mutation filtering.

2.5.5 Functional annotation

Post-variant calling annotation was performed using Variant Effect Predictor (Ensembl-VEP, v104.3) (McLaren, Gil *et al.* 2016), which provides functional consequences for each variant. Additional annotation was carried out against ClinVar (2021-01-10) (Landrum, Lee *et al.* 2016) and dbSNP (v154-20210112) (Sherry, Ward *et al.* 2001).

2.5.6 Quality control and filtering

To ensure the reliability and accuracy of mutation analysis in WES and WGS data, stringent filtering criteria were applied to distinguish high-quality somatic mutations from sequencing artefacts and low-confidence calls. Given that the samples were derived from formalin-fixed paraffin-embedded (FFPE) tissue, filtering steps were necessary to mitigate common FFPE-induced artefacts.

The first stage of filtering involved applying the PASS filters from each of the three variant callers, ensuring that only high-confidence variants were retained. To further eliminate potential sequencing artefacts, the DKFZ Bias Filter (<https://github.com/DKFZ-ODCF/DKFZBiasFilter>) was employed to exclude variants affected by strand bias or sequencing damage. Additionally, variants with a

frequency greater than 0.0004 in any gnomAD (v2.1.1) subpopulation were removed to focus on somatic mutations and exclude common polymorphisms.

Post-filtering quality control was conducted using MultiQC (v1.13a) (Ewels, Magnusson *et al.* 2016) to aggregate reports from all processing steps, ensuring uniformity in sequencing depth, mapping quality, and variant calling accuracy. Three biopsy-derived WES samples failed this QC step and were removed from downstream analysis.

2.5.7 Consensus and manual review

For WES data, a consensus approach was implemented: a variant was retained if identified by at least two out of the three variant callers. Single nucleotide variants (SNVs) not included in the ensemble VCF underwent manual review using the Integrative Genomics Viewer (IGV) (Robinson, Thorvaldsdóttir *et al.* 2011), where visually validated variants were "rescued" and included in the final dataset. The column `ensemble_or_rescue` in the mutation data indicates whether a variant was included in the ensemble VCF ('e') or manually rescued ('r').

Given the lower frequency and complexity of insertions and deletions (indels), all indels identified by the mutation callers were manually reviewed using IGV to ensure accuracy.

For WGS data, the same ensemble approach was used; however, due to the significantly higher number of variants, manual review was not conducted. Instead, only variants present in the ensemble VCF (i.e., identified by at least two of the three variant callers) were retained in the final dataset.

2.5.8 Final filtering criteria

After compiling the final list of ensemble and rescued mutations, additional criteria were applied to refine the dataset further:

- **Coverage:** A minimum depth of 20x was required for SNVs in both tumour and normal samples, while INDELs required a minimum depth of 30x.
- **Variant Allele Frequency (VAF):** SNVs were required to have a tumour VAF of at least 5% and a normal VAF of no more than 1%. INDELs required a tumour VAF of at least 5% and a normal VAF of no more than 1%.
- **Read Counts:** SNVs were retained if they had at least 4 alternate reads in the tumour sample and no more than 1 alternate read in the normal sample. INDELs required at least 5 alternate reads in the tumour sample and no more than 1 alternate read in the normal sample.
- **Homopolymer Regions:** Mutations in homopolymer regions were filtered out due to the increased likelihood of sequencing-induced artefacts in these repetitive sequences.
- **Mapping Quality:** A minimum mapping quality (MAPQ) score of 60 (Phred-scaled) was required for inclusion.
- **Directional Presence:** Mutations had to be detected in both forward and reverse strands to rule out strand-specific artefacts.

By implementing these stringent filtering steps, a high-confidence dataset of somatic mutations was generated, balancing sensitivity and specificity to ensure reliable and reproducible results.

2.5.9 REVEL score annotation

The final list of high-quality mutations was annotated with REVEL (Rare Exome Variant Ensemble Learner) scores, which predict the pathogenicity of missense mutations. REVEL is an ensemble-based tool that integrates multiple pathogenicity predictors, including SIFT, PolyPhen-2, MutationTaster, and CADD, to generate a consensus score (Ioannidis, Rothstein *et al.* 2016).

Precomputed REVEL scores were downloaded from Zenodo (<https://zenodo.org/records/7072866>). The REVEL annotation file (revel_with_transcript_ids.txt, 6.5GB) was processed in R Studio in chunks to

efficiently match variants based on chromosome, genomic position, reference allele, and alternate allele.

2.6 COSMIC cancer genes

The COSMIC Cancer Gene Census (v100) was used to identify and annotate cancer-related genes in this study. The gene list was obtained from the Catalogue of Somatic Mutations in Cancer (COSMIC) database, accessible at: <https://cancer.sanger.ac.uk/cosmic/download/cosmic/v100/cancergenecensus>

2.7 Copy number analysis

2.7.1 ASCAT

Copy number alterations were analysed using ASCAT (v3.1.0) (Van Loo, Nordgard *et al.* 2010), a method designed to estimate tumour purity, ploidy, and allele-specific copy number from sequencing data. Since formalin-fixed paraffin-embedded (FFPE) samples can introduce technical biases, ASCAT was used to correct for such artefacts and infer accurate copy number states. This was performed on the UCL Myriad HPC cluster using R (v4.2.0).

Preprocessing steps

- Tumour and matched normal BAM files were processed using ASCAT's `ascat.prepareHTS` function.
- Allele-specific log R ratios (LogR) and B allele frequencies (BAF) were extracted for each sample.

Reference files included:

- hg38 reference genome
- Twist Exome Capture target regions
- Battenberg allele and loci indices
- Problematic loci exclusion file

ASCAT segmentation and Copy Number estimation

- LogR and BAF values were corrected for GC content and replication timing biases.
- ASCAT segmentation (ascat.aspcf) was performed to identify breakpoints in copy number alterations, applying a penalty of 70 to reduce over-segmentation.
- Final copy number profiles were generated using ascat.runAscat, estimating tumour ploidy and purity.

2.7.2 Copy Number analysis methodology

ASCAT segmentation files were processed alongside sample metadata, with analysis restricted to exonic regions based on Twist Bioscience Exome Capture (hg38). Log2 copy number values from the segmentation files were used to quantify deviations from the diploid state. The primary objectives were to quantify the fraction of the genome altered (FGA), the number of breakpoints, and the total number of copy number alterations (CNAs).

2.7.2.1 Fraction of Genome Altered (FGA) Calculation

To assess genomic instability, FGA was calculated using two complementary methods:

Method 1: FGA Based on Log2 Copy number alteration (log2 CNV)

This method identifies genomic regions where the total copy number deviates significantly from the diploid state. The $\log_2$ CNV was calculated as:

$$\log_2 \left(\frac{n_{\text{Major}} + n_{\text{Minor}}}{2} \right)$$

Segments were considered altered if $\log_2$ CNV exceeded ± 0.2 , indicating a substantial deviation from the expected diploid copy number (2 copies per region). FGA was then computed as:

$$FGA(\text{Method 1}) = \frac{\text{Total length of altered segments}}{\text{Total targeted genomic length}}$$

This method primarily captures copy number gains and losses based on deviations from the expected diploid state.

Method 2: FGA based on Copy Number deviations and allele-specific imbalances

This method I developed extends Method 1 by incorporating allele-specific imbalances, which can indicate loss of heterozygosity (LOH) and regions of allelic imbalance. Segments were considered altered if either:

1. **Total copy number $\neq 2$** , indicating a gain or loss relative to the diploid state.
2. **Major and minor allele counts differed**, signifying allelic imbalance.

FGA (Method 2) was calculated using a similar formula to Method 1, but with broader criteria for defining altered segments, incorporating both copy number changes and allele-specific imbalances.

$$FGA(\text{Method 2}) = \frac{\text{Total length of altered segments (based on copy number and allele-specific imbalances)}}{\text{Total targeted genomic length}}$$

2.7.2.2 Additional Copy Number metrics – breakpoints and CNAs

Breakpoints: The number of transitions between adjacent copy number segments within each sample. A breakpoint is defined as a change in copy number state within a chromosome, representing genomic rearrangements or instability. The total number of breakpoints was computed per sample by summing the transitions across all chromosomes.

Total CNAs: The total number of distinct copy number alteration (CNA) events identified in each sample. A CNA was considered unique based on the chromosome, start position, and end position of each segment.

2.7.2.3 Statistical analysis

Pairwise comparisons of pre- and post-radiotherapy samples were conducted using two-tailed t-tests.

2.8 Nanoseq

To achieve higher sensitivity in detecting low-frequency mutations, I used NanoSeq, a duplex sequencing approach designed to minimise sequencing errors and improve the resolution of somatic mutation detection (Abascal, Harvey *et al.* 2021). Unlike conventional whole-genome sequencing (WGS) or whole-exome sequencing (WES), which are often limited by sequencing errors and low sensitivity for detecting rare variants, NanoSeq employs duplex sequencing which is an error-correction technique involving sequencing both strands of a DNA molecule independently. This method significantly reduces background noise, achieving an error rate of fewer than five errors per billion base pairs which is two orders of magnitude lower than typical somatic mutation loads. This allows for accurate detection of low-frequency single nucleotide variants (SNVs) and small insertions and deletions (indels), particularly in samples with low tumour purity or minimal mutational burden.

DNA from fresh frozen tumour tissue samples from five separate patients, each with paired pre- and post-radiotherapy specimens were sent to the Martincorena group at the Wellcome Trust Sanger Institute, where they underwent sequencing following the NanoSeq protocol. The resulting variant call format (VCF) files and somatic mutation counts were returned for downstream analysis. I subsequently performed mutational signature analysis on these data to characterise the mutation patterns and assess genomic changes induced by radiotherapy.

2.9 Mutational signature analysis

Mutational signatures were identified using SigProfilerExtractor (v1.1.24), applying non-negative matrix factorization (NMF) to decompose mutation profiles into their underlying processes based on trinucleotide context. Analyses were performed

using Python (v3.10.12), with SBS96 and ID83 mutational contexts extracted relative to the GRCh38 reference genome (Alexandrov, Nik-Zainal *et al.* 2013).

2.10 Processing of RNAseq data

Raw RNAseq data were processed using bcbio-nextgen, aligning reads to hg38 with STAR and quantifying transcript expression and generating counts with Salmon (v1.9.0) (Patro, Duggal *et al.* 2017). Fusion transcripts were identified using EricScript (Benelli, Pescucci *et al.* 2012).

2.11 Differential Gene Expression analysis

Differential gene expression analyses were performed using DESeq2 (v1.42.1) (Love, Huber *et al.* 2014) in R (v3.2.1.). The analysis in the pre vs post-radiotherapy comparison controlled for inter-patient variability by including Patient ID and Histology as covariates in the model. Raw transcript counts were obtained from Salmon (v1.9.0) quantification and processed using tximport (v1.30.0) to aggregate transcript-level estimates to gene-level counts.

Data pre-processing steps included:

- Filtering out genes with fewer than 10 counts in at least three samples to remove low-expression genes.
- Variance stabilising transformation (VST) for quality control.
- Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) to assess sample clustering using the inbuilt plotPCA function from DESeq2 and umap (v0.2.10.0) R packages respectively.

Differential expression analysis was conducted using a paired Wald test with a design formula of $\sim$ Histology + Patient_ID + Sample_type, comparing post-radiotherapy to pre-radiotherapy samples.

Log₂ fold change shrinkage was applied using the apegglm method (Zhu, Ibrahim *et al.* 2018) to reduce noise in low-expressed genes. Adjusted p-values were computed

using the Benjamini-Hochberg correction, and significant genes were defined as those with:

- **adjusted p-value (padj) < 0.05**
- **Log₂ fold change > ± 1**

2.12 Gene Set Enrichment Analysis

To identify biological pathways enriched in differentially expressed genes, GSEA was performed using fgsea (v1.28.0) in R (Korotkevich, Sukhov *et al.* 2016). The ranked gene list was generated based on signed $-\log_{10}(\text{p-value})$ weighted by $\log_2$ fold change, prioritising both significance and effect size. The analysis used Hallmark gene sets from the Molecular Signatures Database (Liberzon, Birger *et al.* 2015). Pathways were considered significantly enriched if:

- Adjusted p-value (padj) < 0.05.
- Normalised Enrichment Score (NES) > 0 (upregulated) or NES < 0 (downregulated).

2.13 Gene Ontology analysis

Gene Ontology (GO) enrichment analysis for Biological Processes (BP) was performed in R (v4.3.1) using clusterProfiler (v4.10.1) (Yu, Wang *et al.* 2012), identifying overrepresented pathways among differentially expressed genes. Genes with adjusted p-value (padj) < 0.05 were analysed, with separate tests for upregulated and downregulated genes. Benjamini-Hochberg correction was applied.

2.14 PROGENy pathway activity analysis

To infer pathway activity from gene expression data, PROGENy (v1.24.0) (Schubert, Klinger *et al.* 2018) was used to estimate pathway scores based on a set of predefined pathway-responsive genes. Transcript abundance was quantified using Salmon (v1.9.0) to generate TPM values. These values were then transformed using

variance stabilising transformation (VST) from DESeq2 (v1.42.1) to normalise expression data.

PROGENy scores were computed for 14 canonical signalling pathways. To account for multi-region sampling, expression values from multiple regions of the same tumour were averaged (mean) before pathway inference. Statistical comparisons were performed between progressors and responders, using Wilcoxon rank-sum tests to assess pathway activity differences with significance thresholds set at $p < 0.05$.

2.15 Immune cell inference with xCell

Immune cell proportions were estimated using xCell (v1.1.0) (Aran, Hu *et al.* 2017), a gene signature-based deconvolution method. Transcript abundance was quantified using Salmon (v1.9.0), and TPM values were normalised using variance stabilising transformation (VST) from DESeq2 (v1.42.1). To account for multi-region sampling, expression values from multiple tumour regions were averaged (mean) before deconvolution.

Immune cell compositions were compared between paired pre- and post-radiotherapy samples, as well as between progressors and responders. Wilcoxon signed-rank tests were used to assess differences between conditions, with statistical significance defined as adjusted $p < 0.05$ (Benjamini-Hochberg correction).

2.16 Gene expression-based modelling of disease progression

To develop a gene expression-based model for predicting disease progression following radiotherapy, transcript abundance was quantified using Salmon (v1.9.0), and variance stabilising transformation (VST) from DESeq2 (v1.42.1) was applied. Only post-radiotherapy samples that passed quality control were included.

Feature selection was performed using LASSO (Least Absolute Shrinkage and Selection Operator) regression (Tibshirani 2018), implemented via the glmnet package (v4.1.8) in R (Hastie and Qian 2014). The optimal regularisation parameter (λ) was determined through 10-fold cross-validation, selecting genes most predictive of progression. A Random Forest (RF) classifier was then trained on the selected genes using caret (v6.0.94) (Kuhn, Wing *et al.* 2020) and randomForest (v4.7.1.2) (Liaw and Wiener 2002), with stratified 10-fold cross-validation to evaluate model performance.

2.17 Data management and storage

All data generated in this study is securely stored on UCL servers, with access restricted to members of Dr Nischalan Pillay's research group. Data management and sharing were conducted in accordance with the UCL Research Data Policy, ensuring compliance with institutional and ethical guidelines. The full policy can be accessed at:
https://rdr.ucl.ac.uk/articles/presentation/UCL_Research_Data_Policy_2024/25579800/1?file=45790287

Chapter 3. Genomic responses to neoadjuvant radiotherapy in soft tissue sarcomas.

3.1 Introduction

Radiotherapy (RT) is a key component of multimodal therapy for soft tissue sarcomas (STS), improving local disease control and enabling limb-sparing surgery. However, the molecular consequences of RT on tumour genomes remain poorly defined. While previous research has described general mutational processes associated with radiation exposure, there is limited understanding of how neoadjuvant RT specifically impacts primary STS. This chapter aims to address this knowledge gap by characterising genomic alterations in pre- and post-radiotherapy STS samples, focusing on mutational burden, copy number changes, and mutational signatures.

Cohort overview

This genomic study includes 56 patients treated at the London Sarcoma Service, encompassing 12 different histological subtypes of STS. Tumour samples were collected pre- and/or post-RT, providing a unique opportunity to assess direct genomic changes induced by therapy. Whole exome sequencing (WES) and Whole genome sequencing (WGS) was used to analyse broad genomic alterations, while NanoSeq, a high-sensitivity sequencing approach, was applied to detect low-frequency mutations, enhancing resolution beyond traditional WES.

Objectives of this chapter

This chapter specifically addresses the following questions:

- What is the mutational burden of pre- and post-radiotherapy STS samples?
- Are there specific mutational signatures enriched in post-radiotherapy samples, indicative of RT-induced DNA damage?
- Does RT induce copy number alterations (CNAs), and do these changes vary by histological subtype?
- How do sequencing approaches (WES vs. NanoSeq) compare in capturing these RT-induced genomic changes?

By integrating both bulk sequencing data and high-sensitivity targeted sequencing, this chapter aims to distinguish therapy-driven mutations from pre-existing tumour-intrinsic alterations.

Structure of this chapter

The chapter begins with an analysis of the baseline mutational landscape of pre-treatment STS samples, contextualising them against prior large-scale sarcoma genomic studies (e.g., TCGA, MSKCC cohorts). This is followed by comparative analyses of somatic mutations in pre- and post-RT samples, including single nucleotide variants (SNVs), small insertions/deletions (indels), and their ratios. Copy number alterations (CNAs) are then assessed across the cohort, with a particular focus on subtype-specific responses to RT. Finally, mutational signature analysis is performed, leveraging both WES and NanoSeq data to identify potential RT-specific mutagenic processes.

3.2 The mutational landscape of the London Sarcoma Service cohort.

In this section I characterise the single nucleotide variants (SNVs) and insertion/deletion (indel) mutations within this London Sarcoma Service (LSS) cohort of soft tissue sarcomas. I begin with quantifying the tumour mutational burden across the cohort and then move on to describing both the frequency and distribution of somatic mutations across the subtypes examined with particular attention paid to known cancer genes and relating the findings in this study to what has previously been shown in the current literature.

3.2.1 Tumour mutational burden varies across and within different subtypes

A total of 70 samples, representing 56 unique patients, underwent somatic mutation analysis. This comprised 65 WES and 5 WGS samples. Following the stringent criteria outlined in the methods, a total of 21,510 mutations were identified across 70 samples. The median number of SNVs and INDELs called for the 65 WES samples was 33 and 2, and for the 5 WGS samples was 3873 and 145 respectively.

The tumour mutational burden (TMB) was calculated for the samples by normalising the number of mutations detected to the size of the genomic region sequenced. For the WES samples, the TMB was determined by dividing the total number of mutations by the total exome size. Specifically, this was the amount that was sequenced and was calculated using the sum of the targeted regions listed in the Twist Exome Target hg38 bed file, which is approximately 33 Mb. For WGS samples, the TMB was calculated by dividing the total number of mutations by the entire genome size, which is approximately 3200 Mb. The results from both WES and WGS samples were then combined. The TMB ranged between 0.5 and 1.7 mutations/Mb with a median TMB across the samples of 1.07 mutations/Mb (Figure 3.1). A TCGA analysis of 206 soft tissue sarcomas representing 6 different subtypes found an average TMB of 1.06 per Mb (2017), therefore my results are consistent with this study.

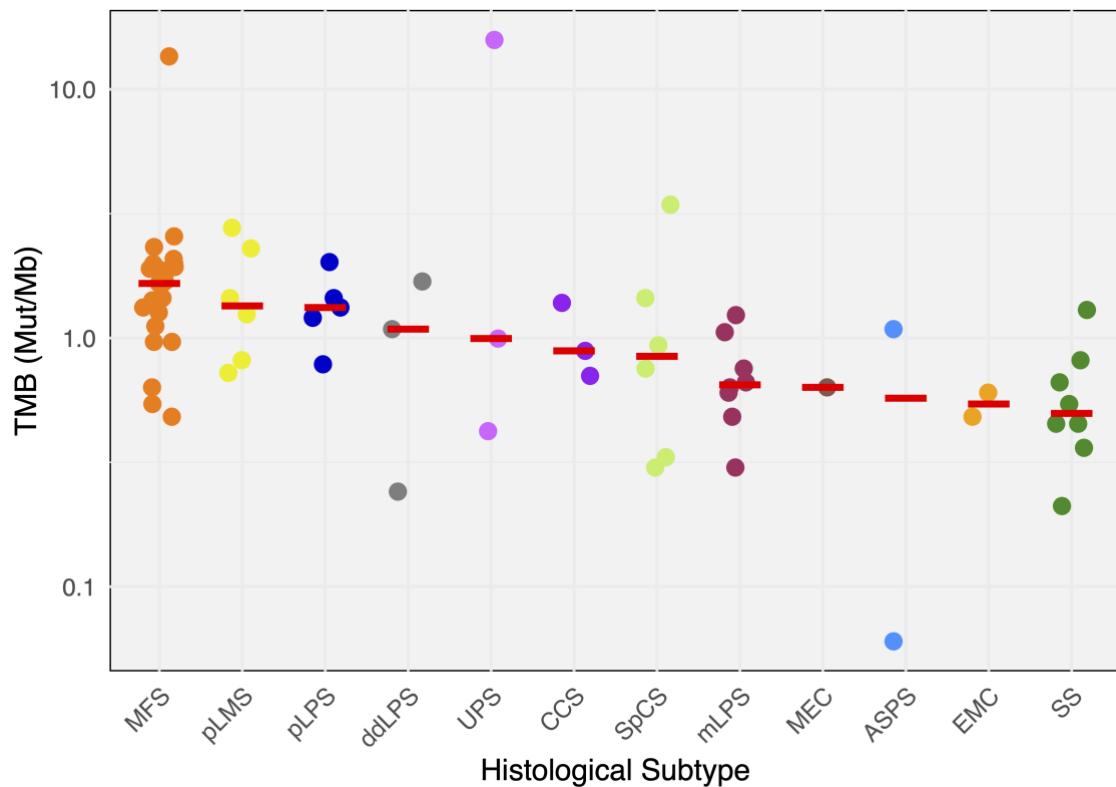


Figure 3.1. Distribution of Tumour Mutational Burden by sarcoma subtype.

This scatter plot shows the distribution of tumour mutational burden (TMB) across various sarcoma subtypes, measured in mutations per megabase (Mut/Mb) and displayed on a logarithmic scale. Each dot represents an individual sample, with colours indicating different sarcoma subtypes. Red horizontal lines indicate the median TMB for each subtype. TMB values were normalised to the sequenced genomic region size, with whole-exome sequencing (WES) samples normalised to the exome and whole-genome sequencing (WGS) samples to the entire genome. Subtypes are ordered by decreasing median TMB, from left to right: MFS (Myxofibrosarcoma, 1.7), pLMS (Pleomorphic Leiomyosarcoma, 1.4), pLPS (Pleomorphic Liposarcoma, 1.3), ddLPS (Dedifferentiated Liposarcoma, 1.1), UPS (Undifferentiated Pleomorphic Sarcoma, 1.0), CCS (Clear Cell Sarcoma, 0.9), SpCS (Spindle Cell Sarcoma, 0.8), mLPS (Myxoid Liposarcoma, 0.6), MEC (Myoepithelial Carcinoma, 0.6), ASPS (Alveolar Soft Part Sarcoma, 0.6), EMC (Extraskelatal Myxoid Chondrosarcoma, 0.5), and SS (Synovial Sarcoma, 0.5).

Most of the tumour samples exhibit TMB values that are closely clustered around the median value, indicating a relatively consistent mutational burden within most histological subtypes. However, there are notable exceptions where outliers significantly deviate from the median. Specifically, myxofibrosarcoma (MFS) and undifferentiated pleomorphic sarcoma (UPS) exhibit the most significant deviations from the median TMB, with maximum values reaching 13.6 and 15.8 mutations per megabase (Mut/Mb) respectively. These differences of 11.9 Mut/Mb from the median for MFS and 14.8 Mut/Mb for UPS indicate substantial variability within these subtypes, driven by outlier tumour samples with exceptionally high mutational burdens. Spindle cell sarcoma (NOS) (which is not a specific subtype) also shows notable variability, with a maximum TMB of 3.44 Mut/Mb, resulting in a difference of 2.59 Mut/Mb from the median, though this is less pronounced compared to the variability observed in MFS and UPS. In contrast, other subtypes display minimal deviations from the median, indicating a more uniform TMB across samples.

3.2.2 Identification of somatic mutations within known cancer genes

I next aimed to investigate the specific mutations occurring in known cancer-related genes within these sarcoma samples. Figure 3.2 shows an oncoplot, which represents the mutation profiles across different samples and sarcoma subtypes.

The oncoplot was constructed by filtering the high-quality mutation dataset comprising the identified 21,510 mutations from 56 patients to include only those mutations found in genes known to be associated with cancer. Using the COSMIC cancer gene list (version 100) (Sondka, Bamford *et al.* 2018) which comprises 581 genes known to be associated with cancer to filter this dataset, I identified 530 mutations across 45 samples. Variants with less impactful on protein function such as silent mutations and mutations in non-coding regions (those occurring in regions labelled as Intron, 3'UTR, 5'UTR, 3'Flank, and 5'Flank) were excluded to focus on mutations that are more likely to have functional consequences. This resulted in a final set of 130 mutations seen in 51 samples representing 41 unique patients.

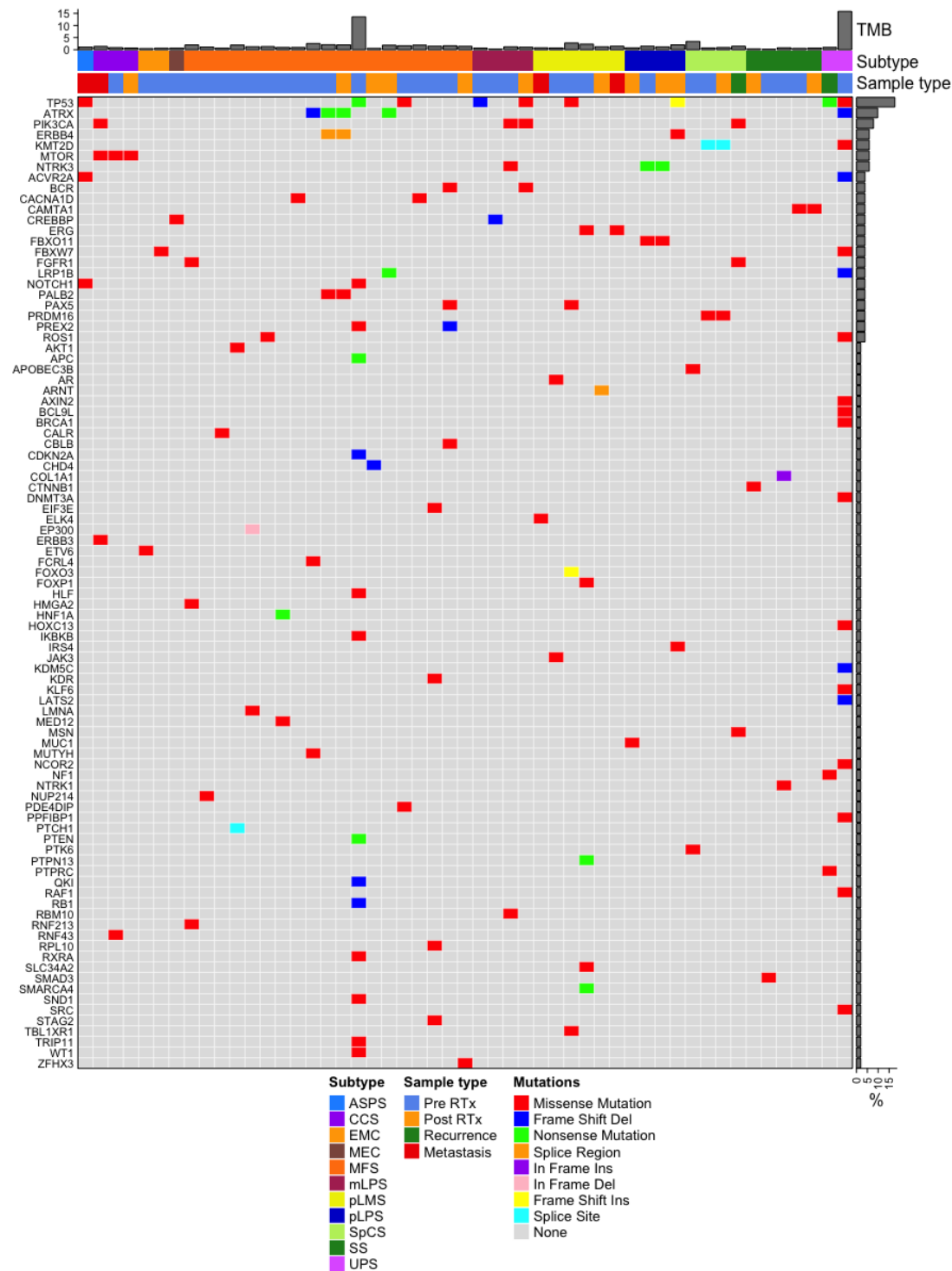


Figure 3.2 Oncoplot of somatic mutations in known cancer genes.

This heatmap visualises the different types of somatic mutations identified within the COSMIC cancer genes. Each column represents an individual sample, while each row corresponds to a specific gene. The samples are annotated at the top with their Tumour Mutational Burden (TMB), histological subtype, and sample type, as indicated in the legend. The genes are ordered by in decreasing frequency of mutations across all samples.

3.2.2.1 Analysis of missense mutations

Most of the mutations were missense mutations, with 94 instances identified, making it the most common type of mutation observed. Missense mutations involve a single nucleotide change that result in the substitution of one amino acid for another in the protein product, potentially altering the function of the protein. This predominance of missense mutations is consistent with expectations, as these mutations are the most common type of somatic mutations found in cancer (Vogelstein, Papadopoulos *et al.* 2013). The most common genes with missense mutations in this cohort were *TP53*, *PIK3CA*, *MTOR* with 5, 4, and 3 instances identified respectively.

To evaluate the functional relevance of the missense mutations observed in this cohort, the Rare Exome Variant Ensemble Learner (REVEL) (Ioannidis, Rothstein *et al.* 2016) was employed. REVEL is a machine learning-based tool that integrates scores from multiple pathogenicity prediction methods, including SIFT (Kumar, Henikoff *et al.* 2009), PolyPhen-2 (Adzhubei, Schmidt *et al.* 2010), MutationAssessor (Reva, Antipin *et al.* 2011), and others. It is specifically designed to predict the pathogenicity of missense mutations based on their likelihood to affect protein function. The REVEL score ranges from 0 to 1, with higher scores indicating a greater likelihood of pathogenicity. Scores ≥ 0.75 were classified as "High," 0.5–0.74 as "Moderate," and < 0.5 as "Likely Benign," according to thresholds previously established in the literature (Garcia, de Andrade *et al.* 2022, Hopkins, Wakeling *et al.* 2023).

Using REVEL to analyse and categorise the missense mutations identified within the COSMIC cancer genes in this cohort, 14 mutations were categorised as "High" pathogenicity (Table 3.1) and 16 as "Moderate" pathogenicity (Table 3.2). The remaining mutations were classified as "Likely Benign".

To determine whether these mutations had been previously reported in published studies, I queried the cBioPortal for Cancer Genomics. cBioPortal (Cerami, Gao *et al.* 2012) is a publicly accessible platform offering integrative visualization and analysis tools for large-scale cancer genomics datasets. Using this platform, I cross-referenced the mutated genes identified in this cohort against all sarcoma datasets,

including TCGA and MSKCC studies (Barretina, Taylor *et al.* 2010, 2017, Gounder, Agaram *et al.* 2022, Nacev, Sanchez-Vega *et al.* 2022), to evaluate whether the mutations were previously reported or represented novel findings within their respective histological subtypes.

Table 3.1. High pathogenicity missense mutations

Missense mutations classified as "High" pathogenicity by REVEL (scores ≥ 0.75). In the Gene column, bold entries represent novel mutations. Numbers in parentheses indicate the frequency of mutations in that gene within the corresponding tumour type, based on the cBioPortal database. Sample types include Pre RT (pre-radiotherapy), Post RT (post-radiotherapy), and Recurrence/Metastasis.

Patient ID	Histology	Sample type	Gene
32	Alveolar soft part sarcoma	Metastasis	GNAQ , TP53 (4/66)
6	Myxofibrosarcoma	Pre RT	CBLB , PAX5 (1/194)
12	Myxofibrosarcoma	Pre RT	FGFR1
44	Myxofibrosarcoma	Pre RT	NOTCH1 (1/194)
54	Myxofibrosarcoma	Pre RT	TP53 (66/194)
55	Myxofibrosarcoma	Pre RT	CACNA1D (1/194)
53	Myxoid liposarcoma	Post RT	TP53 (25/183)
25	Pleomorphic leiomyosarcoma	Pre RT	TP53 (677/1228)
58	Spindle cell sarcoma (NOS)	Recurrence	MSN
40	Undifferentiated pleomorphic sarcoma	Pre RT	HOXC13 , RAF1 (1/590), TP53 (267/590)

Table 3.2. Moderate pathogenicity missense mutations

Missense mutations classified as "Moderate" pathogenicity by REVEL (scores 0.5–0.74). In the Gene column, bold entries represent novel mutations. Numbers in parentheses indicate the frequency of mutations in that gene within the corresponding tumour type, based on the cBioPortal database. Sample types include Pre RT (pre-radiotherapy), Post RT (post-radiotherapy), and Recurrence/Metastasis.

Patient ID	Histology	Sample type	Gene
32	Alveolar soft part sarcoma	Metastasis	ACVR2A
61	Clear cell sarcoma	Metastasis	ERBB3
8	Myxofibrosarcoma	Post RT	<i>ZFHX3</i> (4/194)
16	Myxofibrosarcoma	Pre RT	AKT1
17	Myxofibrosarcoma	Pre RT	LMNA
48	Myxofibrosarcoma	Post RT	<i>ATRX</i> (22/194)
50	Myxofibrosarcoma	Post RT	<i>KMT2C</i> (4/194)
24	Pleomorphic leiomyosarcoma	Pre RT	<i>JAK3</i> (1/1228)
45	Pleomorphic leiomyosarcoma	Pre RT	<i>FOXP1</i> (4/1228), SLC34A2
58	Spindle cell sarcoma	Recurrence	<i>FGFR1</i> (2/590), <i>PIK3CA</i> (17/590)
1	Synovial sarcoma	Post RT	<i>CTNNB1</i> (10/353)
20	Synovial sarcoma	Pre RT	SMAD3
40	Undifferentiated pleomorphic sarcoma	Pre RT	<i>BCL9L</i> (1/590), FBXW7

High pathogenicity mutations, including those in *CACNA1D*, *NOTCH1*, *PAX5*, *RAF1*, and *TP53* were found to have been previously reported in their respective subtypes in cBioPortal, reinforcing their clinical and biological relevance (Table 3.1). For example, *TP53* mutations were frequently observed across multiple subtypes, consistent with its role as a key tumour suppressor gene. However, novel mutations were also identified in this analysis, including *GNAQ* in alveolar soft part sarcoma, *CBLB* and *FGFR1* in myxofibrosarcoma, *MSN* in spindle cell sarcoma, and *HOXC13* in undifferentiated pleomorphic sarcoma, highlighting potentially unexplored pathogenic mechanisms in these tumours.

Moderate pathogenicity mutations similarly included both previously reported and novel findings. Reported mutations included *ATRX*, *BCL9L*, *CTNNB1*, *FGFR1*, *FOXP1*, *JAK3*, *KMT2C*, *PIK3CA*, and *ZFH3* which were detected in their respective tumour subtypes (Table 3.2). Novel mutations identified in this category included *ACVR2A* in alveolar soft part sarcoma, *ERBB3* in clear cell sarcoma, *AKT1* and *LMNA* in myxofibrosarcoma, *SLC34A2* in pleomorphic leiomyosarcoma, *SMAD3* in synovial sarcoma, and *FBXW7* in undifferentiated pleomorphic sarcoma.

These findings highlight the well-documented heterogeneity of soft tissue sarcomas, as evidenced by the identification of both well-characterised mutations, such as *TP53*, and novel mutations, including *HOXC13* and *GNAQ*. By integrating REVEL scores with publicly available cancer genomics datasets, such as those in cBioPortal, this analysis provides a practical framework for prioritising mutations based on predicted pathogenicity. High pathogenicity mutations, such as those in *TP53* and *NOTCH1*, are well-known to drive cancer progression and likely have direct functional consequences. In contrast, moderate pathogenicity mutations, such as those in *ATRX* and *KMT2C*, may play a contributory role, potentially interacting with other genomic alterations to influence tumour behaviour. Experimental validation would be required to elucidate the functional significance of these missense mutations and better understand their roles in sarcoma pathogenesis.

3.2.2.2 Analysis of frameshift and nonsense mutations

In addition to missense mutations, there were 13 instances each of frameshift deletions and nonsense mutations. Frameshift deletions, which occur when nucleotides are deleted from the DNA sequence in numbers that are not multiples of three, result in a shift of the amino acid codon reading frame, often leading to a truncated, non-functional protein. Nonsense mutations introduce a premature stop codon into the DNA sequence, also resulting in a truncated protein. Both types of mutations are typically associated with loss of function and are frequently observed in tumour suppressor genes. As in this cohort, frameshift deletions were identified in the well-known tumour suppressor genes *ATRX*, *CDKN2A*, *RB1*, and *TP53*. Nonsense mutations were identified in the tumour suppressor genes *APC*, *ATRX*, *HNF1A*, *PTEN*, *SMARCA4*, and *TP53*.

There were also smaller numbers of other mutation types: frameshift insertions (2 instances - *FOXO3*, and *TP53*), in-frame deletions (1 instance - *EP300*), in-frame insertions (1 instance - *COL1A1*), splice region mutations (3 instances – *ARNT*, and *ERBB4*), and splice site mutations (3 instances - *KMT2D*, and *PTCH1*). Frameshift insertions, like deletions, can cause significant disruption to the resulting protein, while in-frame deletions and insertions result in the addition or removal of amino acids without altering the overall reading frame, which may or may not affect protein function depending on the location and context. Splice region and splice site mutations affect the process by which introns are removed from pre-mRNA, potentially leading to the inclusion of intronic sequences in the mRNA or the exclusion of exonic sequences, which can have an adverse effect on the resulting protein.

3.2.2.3 Recurrently mutated genes in the LSS cohort

Analysis of the LSS cohort identified 17 genes that were recurrently mutated in more than one patient. This includes missense mutations with any REVEL score. Table 3.3 provides a summary of these genes, including the specific sarcoma subtypes affected and the number of patients in which each gene was mutated. The most frequently mutated genes were *TP53*, *ATRX*, and *PIK3CA*, which were found to be mutated in 9, 4, and 4 patients, respectively.

TP53 was found to be mutated across several sarcoma subtypes, including pleomorphic liposarcoma, myxoid liposarcoma, pleomorphic leiomyosarcoma, undifferentiated pleomorphic sarcoma, and myxofibrosarcoma. These findings are consistent with those from a large study performed using the MSK IMPACT targeted sequencing (Zehir, Benayed *et al.* 2017) panel on 2,138 sarcomas. They reported *TP53* mutations in 68% of pleomorphic liposarcoma cases, 3% of myxoid liposarcoma cases, 45% of pleomorphic leiomyosarcoma cases, 43% of undifferentiated pleomorphic sarcoma cases, and 26% of myxofibrosarcoma cases (Nacev, Sanchez-Vega *et al.* 2022). In their cohort of 13 alveolar soft part sarcomas, none had a *TP53* mutation. This tumour is typically caused by a *ASPSCR1::TFE3* fusion (Sicinska, Kola *et al.* 2024). In the LSS cohort the *TP53* mutation was identified only in the metastasis sample but not in the primary tumour.

Similarly, *ATRX* mutations in the LSS cohort were identified in myxofibrosarcoma and undifferentiated pleomorphic sarcoma, which is in line with the same study that found *ATRX* mutations in 10% of myxofibrosarcoma cases and 18% of undifferentiated pleomorphic sarcoma cases (Nacev, Sanchez-Vega *et al.* 2022). Both *TP53* and *ATRX* have also been found to be recurrently mutated in 27 and 38% respectively of a cohort of 76 undifferentiated pleomorphic sarcomas (Steele, Tarabichi *et al.* 2019) from the London Sarcoma Service.

Table 3.3. COSMIC cancer genes mutated in multiple patients across cohort.

This table lists the COSMIC cancer-related genes that were found to be mutated in more than one patient across the different sarcoma subtypes. Alveolar Soft Part Sarcoma (ASPS), Clear Cell Sarcoma (CCS), Dedifferentiated Liposarcoma (ddLPS), Extraskeletal Myxoid Chondrosarcoma (EMC), Myoepithelial Carcinoma (MEC), Myxofibrosarcoma (MFS), Myxoid Liposarcoma (mLPS), Pleomorphic Leiomyosarcoma (pLMS), Pleomorphic Liposarcoma (pLPS), Spindle Cell Sarcoma (SpCS), Synovial Sarcoma (SS), Undifferentiated Pleomorphic Sarcoma (UPS).

Gene	Histological subtype(s)	No. of patients
<i>TP53</i>	pLPS, mLPS, pLMS, ASPS, UPS, MFS	9
<i>ATRX</i>	MFS, UPS	4
<i>PIK3CA</i>	mLPS, SpCS, CCS	4
<i>ACVR2A</i>	ASPS, UPS	2
<i>BCR</i>	MFS, mLPS	2
<i>CACNA1D</i>	MFS	2
<i>CREBBP</i>	MEC, mLPS	2
<i>ERBB4</i>	pLPS, MFS	2
<i>FBXW7</i>	EMC, UPS	2
<i>FGFR1</i>	MFS, SpCS	2
<i>KMT2D</i>	UPS, SpCS	2
<i>LRP1B</i>	UPS, MFS	2
<i>NOTCH1</i>	ASPS, MFS	2
<i>NTRK3</i>	pLPS, mLPS	2
<i>PAX5</i>	MFS, pLMS	2
<i>PREX2</i>	MFS	2
<i>ROS1</i>	MFS, UPS	2

PIK3CA mutations were observed in myxoid liposarcoma, spindle cell sarcoma, and clear cell sarcoma. Likewise *PIK3CA* mutations were present in 27% of myxoid liposarcoma cases and 5% of spindle cell sarcoma cases of the MSK IMPACT study (Nacev, Sanchez-Vega *et al.* 2022). Unlike in this cohort, they did not report a *PIK3CA* mutation in their cohort of 16 clear cell sarcomas.

3.2.2.4 Somatic mutations identified in each histological subtype

Table 3.4 summarises the COSMIC cancer-related genes mutated within each histological subtype, along with the number of patients affected for each gene. The data highlight both shared mutations within subtypes and unique mutations in individual cases. It is evident that some tumour subtypes harbour a greater number of somatic mutations within cancer genes compared to others, even after accounting for differences in sample sizes across subtypes.

For instance, myxofibrosarcoma (MFS) exhibited a diverse range of mutations across 43 different cancer-related genes, with *ATRX*, *CACNA1D*, *PREX2*, and *TP53* being mutated in 3, 2, 2, and 2 patients, respectively. Similarly, undifferentiated pleomorphic sarcoma (UPS) showed mutations in 21 different cancer genes, with *TP53* mutations found in 2 patients. These findings suggest that MFS and UPS may have a more heterogeneous mutational landscape, potentially contributing to their variable clinical behaviour and treatment responses.

Table 3.4. COSMIC cancer genes mutated in each sarcoma subtype

This table presents a detailed overview of the cancer-related genes that were mutated in each sarcoma subtype, along with the number of different patients in which each gene was mutated shown in parenthesis. Genes highlighted in bold are mutated in multiple patients.

Subtype	No. of Patients	Genes (No. of Patients)
Myxofibrosarcoma	18	ATRX (3), CACNA1D (2), PREX2 (2), TP53 (2) , AKT1 (1), APC (1), BCR (1), CALR (1), CBLB (1), CDKN2A (1), CHD4 (1), EIF3E (1), EP300 (1), ERBB4 (1), FCRL4 (1), FGFR1 (1), HLF (1), HMGA2 (1), HNF1A (1), IKBKB (1), KDR (1), LMNA (1), LRP1B (1), MED12 (1), MUTYH (1), NOTCH1 (1), NUP214 (1), PALB2 (1), PAX5 (1), PDE4DIP (1), PTCH1 (1), PTEN (1), QKI (1), RB1 (1), RNF213 (1), ROS1 (1), RPL10 (1), RXRA (1), SND1 (1), STAG2 (1), TRIP11 (1), WT1 (1), ZFH3 (1)
Synovial sarcoma	4	CAMTA1 (1), COL1A1 (1), CTNNB1 (1), NTRK1 (1), SMAD3 (1)
Myxoid Liposarcoma	4	PIK3CA (2), TP53 (2) , BCR (1), CREBBP (1), NTRK3 (1), RBM10 (1),
Pleomorphic Leiomyosarcoma	3	AR (1), ARNT (1), ELK4 (1), ERG (1), FOXO3 (1), FOXP1 (1), JAK3 (1), PAX5 (1), PTPN13 (1), SLC34A2 (1), SMARCA4 (1), TBL1XR1 (1), TP53 (1)
Pleomorphic Liposarcoma	3	ERBB4 (1), FBXO11 (1), IRS4 (1), MUC1 (1), NTRK3 (1), TP53 (1)
Extraskeletal Myxoid Chondrosarcoma	2	ETV6 (1), FBXW7 (1)
Spindle cell sarcoma	2	APOBEC3B (1), FGFR1 (1), KMT2D (1), MSN (1), PIK3CA (1), PRDM16 (1), PTK6 (1)
Undifferentiated Pleomorphic Sarcoma	2	TP53 (2) , ACVR2A (1), ATRX (1), AXIN2 (1), BCL9L (1), BRCA1 (1), DNMT3A (1), FBXW7 (1), HOXC13 (1), KDM5C (1), KLF6 (1), KMT2D (1), LATS2 (1), LRP1B (1), NCOR2 (1), NF1 (1), PPFIBP1 (1), PTPRC (1), RAF1 (1), ROS1 (1), SRC (1)
Alveolar Soft Part Sarcoma	1	ACVR2A (1), NOTCH1 (1), TP53 (1)
Clear Cell Sarcoma	1	ERBB3 (1), MTOR (1), PIK3CA (1), RNF43 (1)
Myoepithelial Carcinoma	1	CREBBP (1)

In contrast, subtypes such as myoepithelial carcinoma (MEC), extraskeletal myxoid chondrosarcoma (EMC), and alveolar soft part sarcoma (ASPS) had fewer mutated cancer-related genes (1, 2, and 3 genes, respectively). This observation suggests a more limited role for single nucleotide variants (SNVs) and small insertions or deletions (Indels) as potential driver mutations in these tumour types. Notably, subtypes such as synovial sarcoma, extraskeletal myxoid chondrosarcoma, myxoid liposarcoma, myoepithelial carcinoma, alveolar soft part sarcoma, and clear cell sarcoma are known to harbour specific gene fusions that drive tumourigenesis. In contrast, subtypes like myxofibrosarcoma, pleomorphic liposarcoma, pleomorphic leiomyosarcoma, undifferentiated pleomorphic sarcoma, spindle cell sarcoma, and dedifferentiated liposarcoma typically lack recurrent fusion genes, reflecting a different molecular pathogenesis.

Integration of RNA sequencing data obtained for many of these patients (detailed in Chapter 4) corroborated the presence of fusion mRNA transcripts in many cases. For example:

- Patient 9 and Patient 29, both diagnosed with EMC, were found to have canonical *EWSR1::NR4A3* and *TCF12::NR4A3* fusions, respectively.
- Patient 32, diagnosed with ASPS, exhibited the characteristic *ASPSCR1::TFE3* fusion.

These results underscore the complexity of sarcoma genomics, where some subtypes are driven predominantly by gene fusions, while others display a broader spectrum of somatic mutations.

3.2.3 Summary

The findings presented in this section confirm that the mutation profiles observed within this cohort are consistent with previously reported data for soft tissue sarcomas, reinforcing the reliability of both the sequencing and somatic mutation calling methods used. This alignment with existing literature provides a reliable foundation for the comparative analyses of pre- and post-radiotherapy samples presented in the following sections.

3.3 Comparative analysis of somatic mutation burden in pre- and post-radiotherapy sarcoma samples.

While radiotherapy is known to cause DNA damage, its direct effect on the mutation burden of soft tissue sarcomas remains uncertain. Previous studies suggest that radiotherapy does not significantly increase the overall tumour mutational burden (TMB) in STS, which typically exhibits a low baseline mutation rate of approximately 1 mutation per megabase (Mb) (see Section 1.3). However, this does not exclude the possibility of subtle mutational shifts following treatment, particularly in the distribution of single nucleotide variants (SNVs) and insertion/deletion mutations (indels). In other cancers, radiation exposure has been associated with an increase in indel mutations, particularly deletions, due to the involvement of error-prone DNA repair mechanisms such as non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ) (see Section 1.4.3.1). Given that radiotherapy induces double-strand breaks (DSBs), it is reasonable to hypothesise that the Indel-to-SNV ratio in STS may be altered post-treatment, with a shift toward a higher frequency of deletions relative to insertions.

The extent of these changes is likely to vary across different histological subtypes of STS. Although no prior studies have systematically analysed pre- vs. post-radiotherapy mutation burden in sarcomas, some subtypes may be more prone to radiation-induced genomic changes based on their baseline levels of genomic instability. For example, dedifferentiated liposarcoma (ddLPS) and myxofibrosarcoma (MFS) are known to exhibit extensive chromosomal alterations and ongoing genome evolution, making them potential candidates for higher mutation burden post-radiotherapy. In contrast, synovial sarcoma (SS) and clear cell sarcoma (SpCS) typically have comparatively more stable genomes with characteristic fusion drivers, suggesting they may accumulate fewer additional mutations following radiotherapy. If significant differences emerge between histologies, this could indicate that intrinsic genomic stability influences the extent of radiotherapy-induced mutagenesis, an important consideration for future research into subtype-specific responses to treatment.

To investigate these potential effects, this section systematically compares the somatic mutation burden in pre- and post-radiotherapy sarcoma samples using whole exome sequencing (WES). The analysis begins with an unpaired comparison of all pre- and post-radiotherapy samples to determine whether radiotherapy leads to a global increase in SNVs or indels across the cohort. This is followed by a histology-specific analysis, where mutation counts are examined separately for subtypes with sufficient sample numbers to detect statistical differences. Additionally, a paired comparison of patients with available matched pre- and post-radiotherapy samples is conducted to assess individual-level mutational changes, which helps account for inter-patient variability. Lastly, the Indel-to-SNV ratio and Deletion-to-Insertion ratio are examined to determine whether radiotherapy induces a preferential shift toward indel formation, particularly deletions, which would be consistent with known mechanisms of radiation-induced mutagenesis.

Although a significant increase in TMB is not expected, the findings from this analysis will clarify whether radiotherapy results in more subtle but biologically relevant mutational changes in STS. If an increase in the Indel-to-SNV ratio or a shift toward deletions is observed, this would suggest that radiotherapy is driving specific mutational processes in these tumours, likely through its impact on DNA repair pathways. Furthermore, if distinct histology-specific differences emerge, this could indicate that some STS subtypes are more prone to radiation-induced genomic alterations than others, potentially due to their baseline genomic stability or inherent DNA repair mechanisms. These results will provide the groundwork for later sections, which will explore copy number alterations and mutational signatures to further characterise the genomic consequences of radiotherapy in STS.

3.3.1 Unpaired comparison of all pre- and post-radiotherapy samples.

The initial analysis aimed to determine if there was a difference in the number of SNVs or indels following radiotherapy. There was no significant difference in the number of SNVs ($p = 0.88$) or indels ($p = 0.17$) when comparing all pre- to all post-radiotherapy whole exome sequencing (WES) samples (Figure 3.3). For SNVs, 46 pre-radiotherapy samples were compared to 15 post-radiotherapy samples, with median SNV counts of 32.5 and 33, respectively. For indels, 37 pre-radiotherapy samples were compared to 12 post-radiotherapy samples, with median indel counts of 3 in the pre-radiotherapy group and 2 in the post-radiotherapy group.

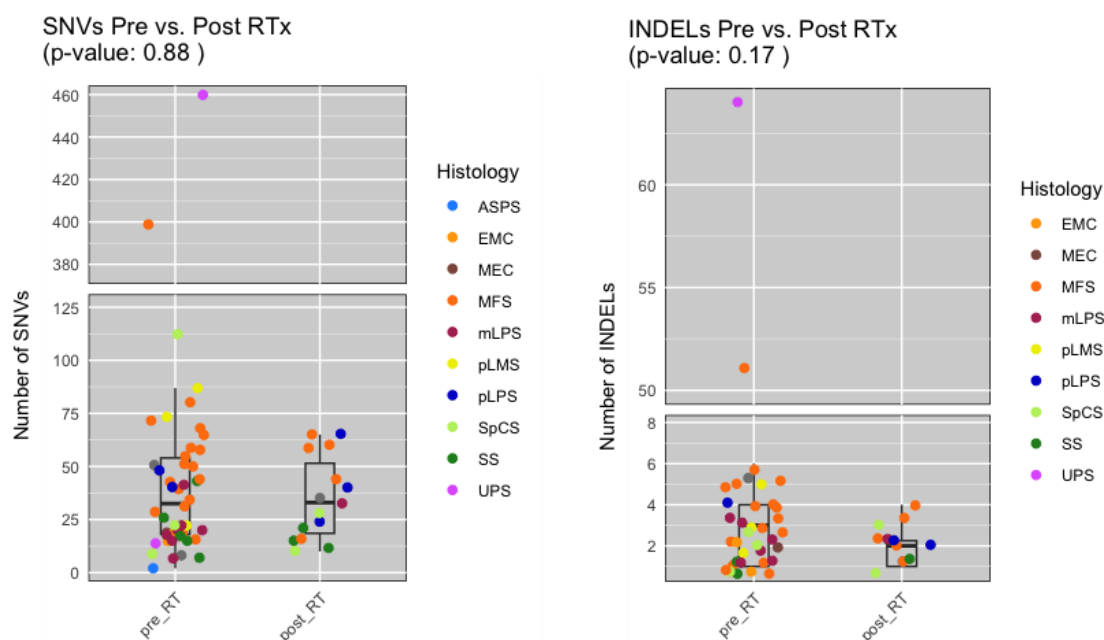


Figure 3.3. Comparison of the frequency of SNVs and indels in pre- and post-radiotherapy sarcoma samples.

This figure shows the distribution of the number of SNVs (left panel) and Indels (right panel) in tumour samples collected before (pre-RT) and after radiotherapy (post-RT). Each point represents a tumour sample, coloured by its histological subtype. The p-values for the comparisons between pre-RT and post-RT were calculated using the Wilcoxon rank-sum test (unpaired) and are shown in the title of each panel. The boxes display the interquartile range with the median indicated by the horizontal line, and the whiskers extend to 1.5 times the interquartile range. The y-axis is broken to better visualise samples with lower mutation counts, while still displaying the two outliers. Alveolar Soft Part Sarcoma (ASPS), Dedifferentiated Liposarcoma (ddLPS), Extraskeletal Myxoid Chondrosarcoma (EMC), Myoepithelial Carcinoma (MEC), Myxofibrosarcoma (MFS), Myxoid Liposarcoma (mLPS), Pleomorphic Leiomyosarcoma (pLMS), Pleomorphic Liposarcoma (pLPS), Spindle Cell Sarcoma (SpCS), Synovial Sarcoma (SS), Undifferentiated Pleomorphic Sarcoma (UPS).

3.3.2 Unpaired comparison of pre- and post-radiotherapy samples within sarcoma subtypes

To assess whether radiotherapy impacted the number of SNVs and indels within individual subtypes, I focused on subtypes with sufficient pre- and post-radiotherapy samples for statistical analysis. Six subtypes met this criterion: dedifferentiated liposarcoma, myxofibrosarcoma, myxoid liposarcoma, pleomorphic liposarcoma, spindle cell sarcoma, and synovial sarcoma. No statistically significant differences were observed in the number of SNVs or indels post-radiotherapy across any of these subtypes (Figure 3.4).

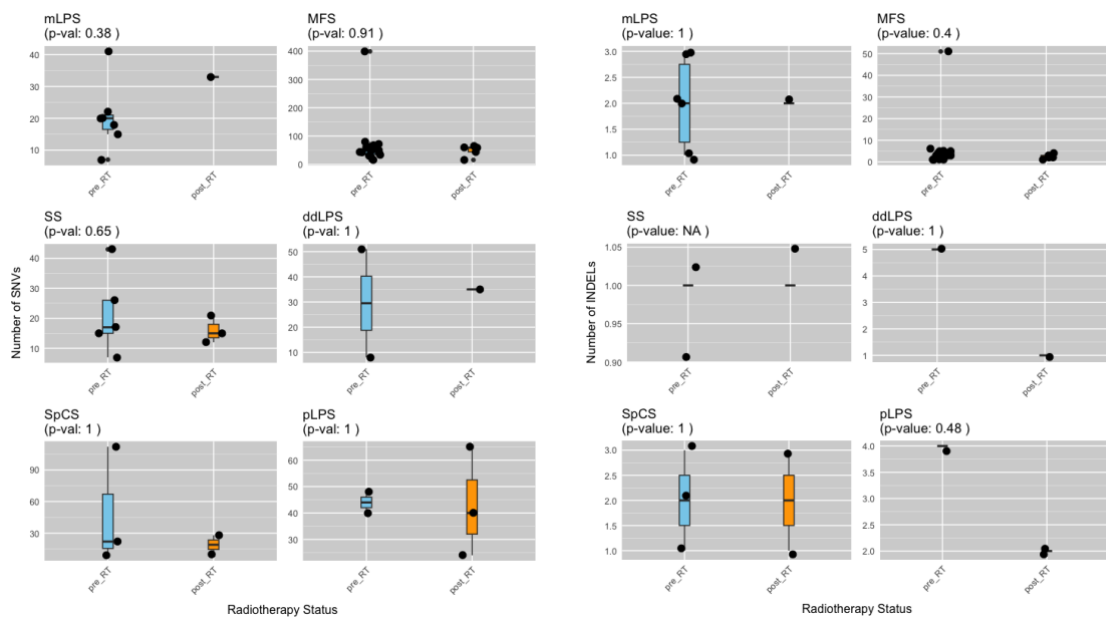


Figure 3.4. Comparison of the frequency of SNVs and indels pre- and post-radiotherapy across different histological subtypes.

This figure illustrates the number of SNVs (left panel) and indels (right panel) in tumour samples before and after radiotherapy separated by histological subtype. The p-values for the comparisons between pre-RT and post-RT were calculated using the Wilcoxon rank-sum test (unpaired) and are shown in the title of each panel. The boxes display the interquartile range with the median indicated by the horizontal line, and the whiskers extend to 1.5 times the interquartile range. Dedifferentiated Liposarcoma (ddLPS), Myxofibrosarcoma (MFS), Myxoid Liposarcoma (mLPS), Pleomorphic Liposarcoma (pLPS), Spindle Cell Sarcoma (SpCS), Synovial Sarcoma (SS).

3.3.3 Paired analysis of patients with matching pre- and post-radiotherapy samples

I conducted a paired analysis on patients for whom both pre- and post-radiotherapy samples were available. There was no significant difference in the number of SNVs ($p = 0.49$) or indels ($p = 0.36$) following radiotherapy (Figure 3.5).

In 6 out of the 7 patients, the number of indels remained stable post-radiotherapy. Notably, in patient 50 (diagnosed with myxofibrosarcoma), there was a loss of one indel after radiotherapy. Specifically, this was a TG frameshift deletion mutation in the *EPHA1* gene, detected in the pre-radiotherapy biopsy but absent in the post-radiotherapy resection specimen. This absence in the post-treatment sample could reflect a potential treatment effect, wherein the mutation bearing cells were selectively killed by radiotherapy, or it may be attributed to sampling variation between the biopsy and resection specimen. Specifically, the mutation may have been present only in a subclone of the tumour that was sampled in the biopsy but not in the resected specimen, leading to its absence in the post-radiotherapy sample. This kind of variation is not uncommon, as tumour heterogeneity can result in certain mutations being detected in one sample but not in another (Gerlinger, Rowan *et al.* 2012, Jamal-Hanjani, Wilson *et al.* 2017).

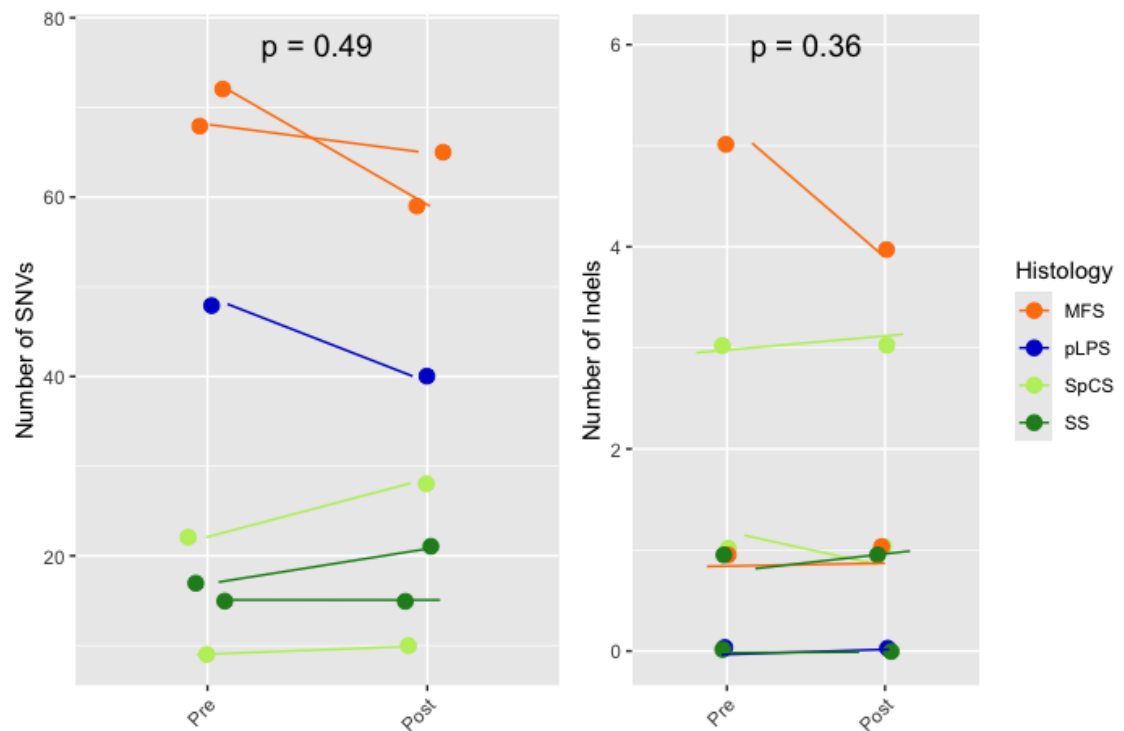


Figure 3.5. Comparison of the frequency of SNVs and indels in paired pre- and post-radiotherapy sarcoma samples.

The results of 7 patients with paired pre- and post-radiotherapy whole exome sequencing SNV and Indel mutation data are presented here. No significant difference is seen in the number of SNVs or Indels. Paired t-test. Myxofibrosarcoma (MFS), Pleomorphic Liposarcoma (pLPS), Spindle Cell Sarcoma (SpCS), Synovial Sarcoma (SS).

3.3.4 Assessment of the Indel-to-SNV and the Deletion-to-Insertion ratio

In addition to comparing the frequency of SNVs and indels, I further investigated specific metrics to assess the genomic impact of radiotherapy. The two key metrics analysed were the Indel-to-SNV ratio and the Deletion-to-Insertion ratio across pre-radiotherapy, post-radiotherapy, metastasis, and recurrence samples.

These ratios specifically quantify the balance between different types of mutations, such as the ratio of indels relative to SNVs and the tendency for deletions over insertions. An increased Indel-to-SNV ratio may indicate that radiotherapy has heightened the frequency of double-strand breaks, leading to more error-prone repair processes like non-homologous end joining (NHEJ), which can result in a higher occurrence of insertions and deletions compared to point mutations. This would align with previous studies, such as an analysis of radiation-associated gliomas, where an increased burden of indels relative to the overall mutation profile was observed following radiotherapy exposure (Kocakavuk, Anderson *et al.* 2021). Additionally, findings from radiation-exposed populations, such as those studied in the aftermath of the Chernobyl disaster, have also highlighted similar increases in DNA repair-associated mutational patterns, emphasising the role of radiation in driving genomic instability (Morton, Karyadi *et al.* 2021).

Meanwhile, a higher Deletion-to-Insertion ratio could suggest that radiotherapy-driven DNA damage is more likely to be resolved through deletion events, potentially due to specific vulnerabilities in the genome to radiation-induced breaks. This is consistent with the findings from prior research that identified a propensity for deletions in radiation-exposed tissues (Behjati, Gundem *et al.* 2016).

The deletion-to-insertion ratio was assessed across pre-radiotherapy, post-radiotherapy, metastasis, and recurrence samples (Figure 3.6A). The median ratio was 1.0 in both pre- and post-radiotherapy samples, indicating no significant change following treatment. Metastasis samples exhibited a higher median ratio of 1.89, while recurrence samples showed the highest median ratio of 4.0, suggesting a greater prevalence of deletions compared to insertions in these groups. A statistically significant difference was observed between pre-radiotherapy and recurrence

samples ($p = 2e-05$), while differences between pre-radiotherapy and post-radiotherapy or metastasis samples were not significant.

The indel-to-SNV ratio was also analysed (Figure 3.6B), revealing a median value of 0.067 in pre-radiotherapy samples, which slightly decreased to 0.048 in post-radiotherapy samples. Metastasis samples showed a similar median ratio of 0.056, whereas recurrence samples had a notably higher median of 0.221, indicating an increased indel burden in cases of recurrent disease. However, the differences in the indel-to-SNV ratio between pre-radiotherapy and the other sample types did not reach statistical significance.

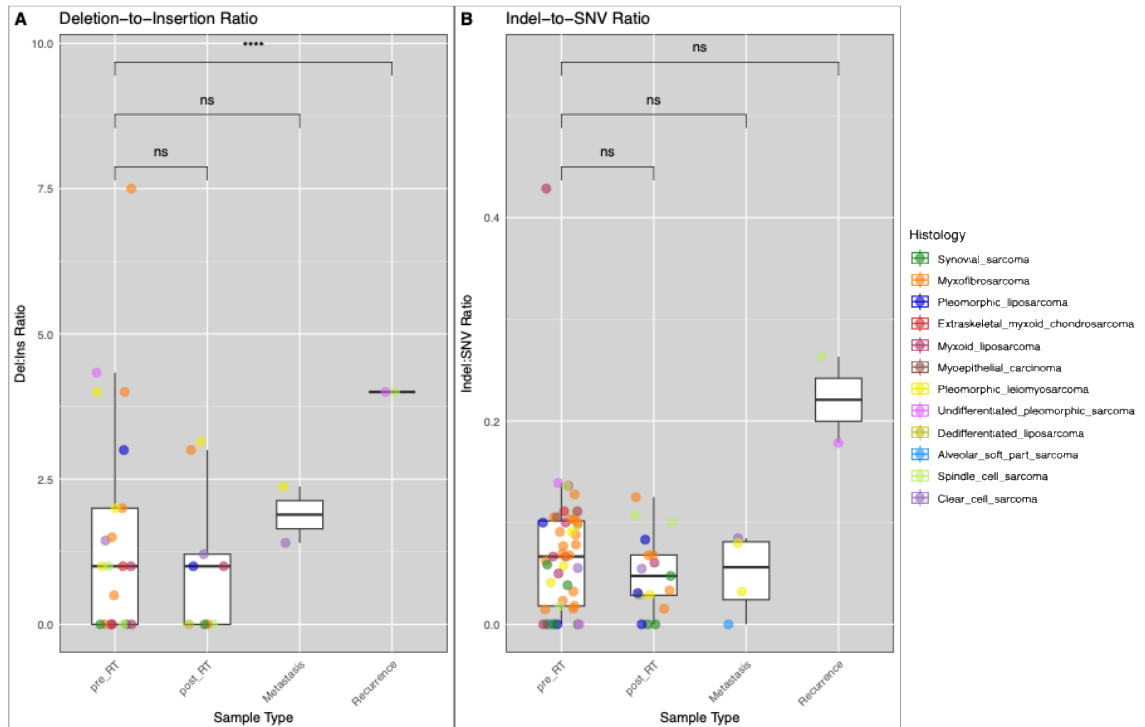


Figure 3.6. Comparison of Deletion-to-Insertion and Indel-to-SNV Ratios

Boxplots of **(A)** Deletion-to-Insertion ratio and **(B)** Indel-to-SNV ratio across pre-radiotherapy (pre_RT), post-radiotherapy (post_RT), metastasis, and recurrence samples. Each point represents an individual tumour sample, coloured by histological subtype. The statistical comparisons between pre-radiotherapy and other sample types (post-radiotherapy, metastasis, and recurrence) are shown above each boxplot, with p-values obtained using t-tests. Asterisks indicate levels of statistical significance (**** $p < 0.0001$), while 'ns' denotes non-significant differences. The centre line of the box plot represents the median value, with the edges of the box indicating the interquartile range, and the whiskers extending to 1.5 times the interquartile range.

3.3.5 Case study of somatic mutation dynamics in patient 58: a longitudinal analysis across pre-, post-radiotherapy, and recurrence stages.

Longitudinal analyses of somatic mutations provide valuable insights into the evolutionary dynamics of tumours in response to treatment and disease progression. This case study examines the mutational landscape of patient 58, a 53-year-old male diagnosed with spindle cell sarcoma of the triceps, by sequencing tumour specimens collected at three key time points: pre-radiotherapy, post-radiotherapy, and at recurrence. The analysis captures the relative stability of somatic mutations immediately following radiotherapy and highlights the emergence of additional mutations at recurrence, reflecting possible clonal evolution over time.

The patient commenced radiotherapy 25 days after diagnosis, receiving a total of 50 Gy in 25 fractions. The tumour was resected 32 days post-radiotherapy (89 days after diagnosis), with final staging recorded as ypT2b Nx Mx (TNM8). Recurrence was detected 71 days after the initial resection and excised 154 days later (186 days post-radiotherapy). Sadly, the patient succumbed to the disease 104 days after recurrence resection, 347 days post-diagnosis (Figure 3.7).

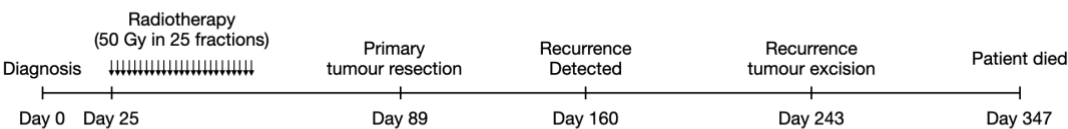


Figure 3.7 Timeline of treatment and disease progression for patient 58

This timeline summarises the patient's treatment course, from diagnosis through radiotherapy, surgical interventions, recurrence, and eventual disease progression.

Sequencing analysis revealed comparable numbers of SNVs in the pre- and post-radiotherapy samples, with 22 SNVs identified pre-radiotherapy and 28 post-radiotherapy (Figure 3.8). Of these, 19 SNVs (86%) persisted across both time points, indicating minimal genomic alterations immediately following treatment. Three SNVs were unique to the pre-radiotherapy sample, while 26 additional SNVs were identified in the recurrence specimen that were absent in earlier samples. Additionally, 12 SNVs were shared in the recurrence specimen: 11 common to both the pre- and post-radiotherapy samples, and one exclusive to the pre-radiotherapy sample.

Analysis of indels showed no new mutations immediately post-radiotherapy. The same indels in genes *C4A*, *PLSCR4*, and *ZG16B* were present in both the pre- and post-radiotherapy samples. In contrast, the recurrence specimen displayed seven additional indels in *ARGLU1*, *C3AR1*, *CHD8*, *FAM171B*, *SLC5A2*, *SPATS2L*, and *PRKCZ*, which were absent from the earlier samples. This suggests ongoing clonal evolution and selection in the intervening period post-radiotherapy.

The timing of sample collection provides additional context. The post-radiotherapy sample was collected 32 days after treatment, while the recurrence specimen was obtained 186 days later. The additional SNVs and indels in the recurrence sample may reflect clonal expansion of mutations emerging after radiotherapy. Mutations induced by radiotherapy might not have reached detectable levels within the initial 32-day window, particularly without sufficient clonal growth to exceed the variant allele frequency (VAF) thresholds for this bulk WES analysis. Alternatively, some of the mutations observed in the recurrence specimen could have been subclonal in the post-radiotherapy sample but fell below detection limits.

To further investigate potential subclonal dynamics, I attempted to reconstruct the clonal architecture using Conipher (Grigoriadis, Huebner *et al.* 2024), a tool for subclonal phylogenetic analysis. Unfortunately, the low number of detectable mutations in this case precluded the identification of definitive subclonal relationships.

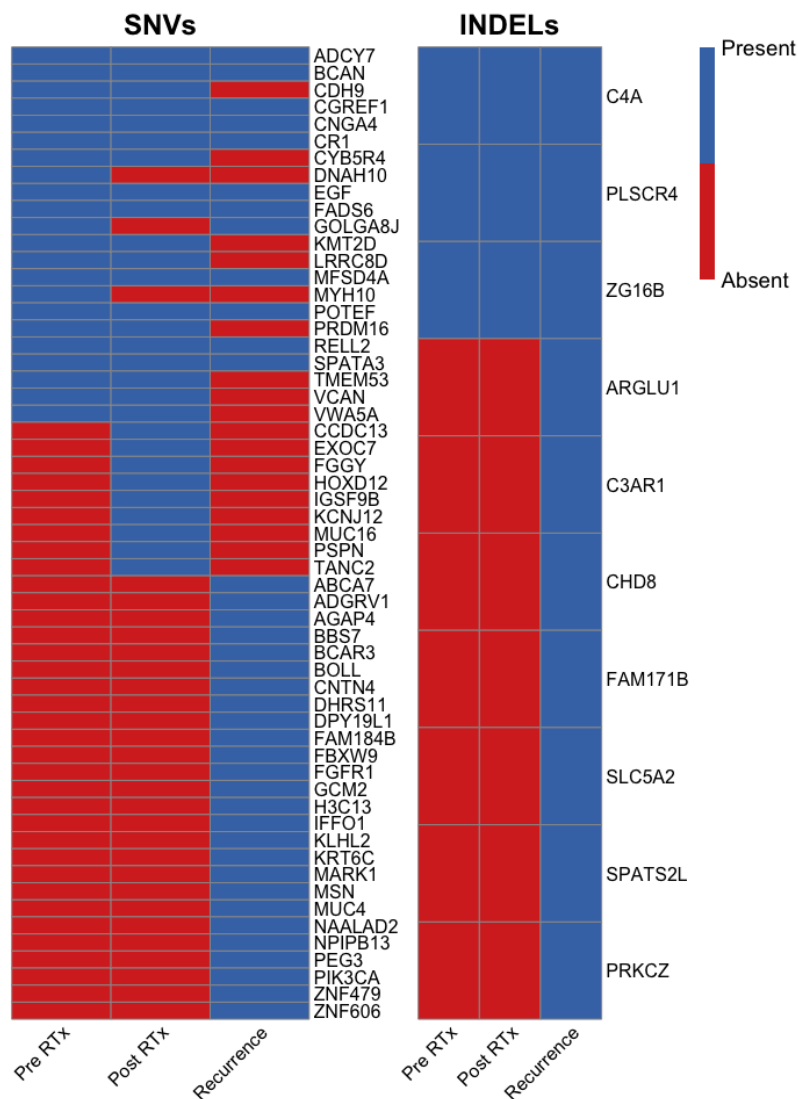


Figure 3.8. Temporal evolution of somatic mutations in patient 58's sarcoma specimens across treatment stages.

Heatmaps show the presence or absence of SNVs and Indel mutations across three stages Pre-, Post-radiotherapy, and recurrence for patient 58 who was diagnosed with a spindle cell sarcoma.

3.3.6 Summary

The analysis presented in this section indicates that radiotherapy does not induce significant changes in the frequency of somatic SNVs or indels in soft tissue sarcoma samples, as shown by both unpaired and paired comparisons across multiple sarcoma subtypes. Across different subtypes and individual cases, mutational profiles remained largely consistent before and after radiotherapy, suggesting that any radiotherapy-related mutational shifts may be subtle or influenced by time-dependent factors, rather than being immediately apparent. The case study of patient 58 suggests that time may be necessary for clonal outgrowth to reveal radiotherapy-related mutations, as indicated by the emergence of unique mutations in the recurrence sample. Alternatively, whole-exome sequencing may have limitations in detecting subtle mutational differences in sarcomas with inherently low tumour mutational burden (TMB), highlighting the need for more sensitive approaches in future studies.

3.4 Comparison of copy number alterations pre- and post-radiotherapy.

Copy number alterations (CNAs) play a critical role in cancer progression, influencing tumour growth, metastasis, and treatment response (Zack, Schumacher *et al.* 2013, Steele, Abbasi *et al.* 2022). Several studies have demonstrated that CNAs can also affect radiotherapy outcomes, though research has predominantly focused on other cancer types. For instance, in prostate cancer, CNAs in *PTEN* (loss) and *c-MYC* (gain) have been linked to an increased risk of biochemical relapse after radiotherapy, suggesting that genomic instability can influence treatment failure (Zafarana, Ishkanian *et al.* 2012). Similarly, in breast cancer, CNAs at chromosome 8p11-12 have been shown to predict poor survival and resistance to both chemotherapy and radiotherapy, reinforcing the role of CNA-driven tumour evolution (Moelans, van Maldegem *et al.* 2018). In lung cancer, somatic CNAs have been associated with progression-free survival following radiotherapy, with high SCNA levels correlating with poorer outcomes, particularly in lung adenocarcinoma (Kou, Wu *et al.* 2021). Additionally, CNAs in genes involved in DNA repair (*PRMT5* and *APE1*) have been implicated in radiation resistance in oral squamous cell carcinoma, suggesting that structural genomic alterations can impact sensitivity to treatment (Izumi, Rychahou *et al.* 2023).

Despite these findings in other malignancies, the impact of radiotherapy on CNA dynamics in soft tissue sarcomas remains poorly understood. Some evidence suggests that radiation exposure can induce large deletions and increase genome-wide instability, as observed in gliomas, where post-radiotherapy samples exhibited an enrichment of large deletions spanning chromosome-arm lengths (Kocakavuk, Anderson *et al.* 2021). However, studies specifically analysing pre- vs. post-radiotherapy CNAs in sarcomas are limited, with most research focusing on radiation-induced secondary sarcomas rather than the genomic consequences of radiotherapy in primary tumours (Lesluyes, Baud *et al.* 2019).

In this section, I present a systematic comparison of CNAs detected in pre- and post-radiotherapy soft tissue sarcoma samples, focusing on the extent of genome-wide copy number changes. Given the well-documented CNA-driven effects in other

cancers, understanding how radiotherapy shapes the genomic architecture of sarcomas helps us better understand tumour evolution, treatment resistance, and potential therapeutic vulnerabilities.

3.4.1 Assessing Copy Number Alterations: methods and metrics.

Copy number alterations were identified using the ASCAT (Allele-Specific Copy number Analysis of Tumours) algorithm (Van Loo, Nordgard *et al.* 2010), which accounts for tumour purity and ploidy, allowing accurate detection of both total copy number changes and allele-specific imbalances. Using ASCAT I compared the fraction of the genome altered, the total number of CNAs, and the total number of breakpoints in pre- and post-radiotherapy samples.

The fraction of genome altered (FGA) was calculated as the total length of altered genomic segments divided by the total length of the sequenced genomic regions. To quantify the FGA, two distinct methods were employed to determine whether a genomic segment was altered.

Method 1 focuses on regions where the total copy number deviates from the expected diploid state, identifying significant gains or losses of chromosomal material. This method uses the formula $\log_2 \text{CNV (copy number alteration)} > \pm 0.2$ to determine whether a genomic segment is altered. This approach has been widely used in previous studies (Xi, Lee *et al.* 2016, Rizvi, Sanchez-Vega *et al.* 2018, Caso, Sanchez-Vega *et al.* 2020, Pariyar, Johns *et al.* 2021), but it does not consider allele-specific information.

Therefore, I developed Method 2, which expands on the rationale of Method 1 by incorporating allele-specific imbalances. In Method 1, only total copy number changes are considered, meaning that while it captures significant deviations in the overall number of copies for a given region it does not differentiate between the major and minor alleles within that region. This limitation can overlook more subtle forms of genomic instability, such as loss of heterozygosity (LOH) or regions where the

major and minor alleles are imbalanced, such as one allele being gained while the other remains unchanged or is lost. Method 2 improves upon this by detecting both total copy number changes and imbalances between the two alleles. By capturing regions where the major and minor alleles differ, Method 2 provides a more nuanced view of genomic instability, identifying additional forms of chromosomal alteration that Method 1 might miss.

Additionally, two other metrics were analysed, the total number of CNAs, and the total number of breakpoints. These metrics provide complementary insights into the possible role of radiotherapy in affecting genomic instability. The total number of CNAs represents distinct regions of the genome where there has been a gain or loss of chromosomal material. Each CNA corresponds to an independent event, such as a loss or gain, where the copy number has deviated from the normal diploid state. To calculate the total number of CNAs, I counted the unique chromosomal regions with altered copy numbers. This metric reflects the extent of genomic regions affected by copy number alterations. By comparing the number of CNAs pre- and post-radiotherapy, I aimed to assess whether radiotherapy increases overall genomic instability by introducing new copy number changes.

Conversely, the total number of breakpoints represents transitions between different copy number states within the genome. A single CNA can contain multiple breakpoints if there are shifts between copy number states, for example, following genomic episodes of chromothripsis (Stephens, Greenman *et al.* 2011). Breakpoints provide a measure of the structural complexity of the genome, indicating the frequency of genomic rearrangements. Even if the number of CNAs remains constant, an increase in breakpoints suggests more intricate structural changes. To calculate the total number of breakpoints, I counted the transitions between altered segments identified by ASCAT. Comparing breakpoints pre- and post-radiotherapy reveals whether radiotherapy induces additional rearrangements or increases genomic complexity.

By analysing the fraction of the genome altered, along with both the number of CNAs and breakpoints, it is possible to obtain a more comprehensive understanding of whether radiotherapy influences copy number.

3.4.2 Comparison of copy number alteration metrics across all pre- and post-radiotherapy samples

The analysis demonstrated that radiotherapy induces chromosomal alterations detectable at the genomic level. Using Method 1, which evaluates total copy number changes, a significant increase in the fraction of genome altered (FGA) was observed post-radiotherapy, with the median FGA increasing from 31% to 64% ($p = 0.04$) (Figure 3.9). This finding highlights the impact of radiotherapy on large-scale genomic stability. In contrast, incorporating allele-specific imbalances with Method 2 resulted in an increase in FGA from 38% to 81%; however, this change did not reach statistical significance ($p = 0.1$). This suggests that the primary genomic changes induced by radiotherapy may involve broader chromosomal alterations rather than allele-specific imbalances.

The total number of breakpoints, which reflect the frequency of transitions between different copy number states, did not significantly change between pre- and post-radiotherapy samples, with median values of 34 and 38, respectively ($p = 0.18$). This suggests that, although radiotherapy introduces chromosomal alterations, it does not significantly increase the frequency of genomic rearrangements or structural complexity. Similarly, the total number of CNAs showed no significant difference between pre- and post-radiotherapy samples, with median values of 57 and 61, respectively ($p = 0.18$), indicating that the overall number of copy number events remains stable.

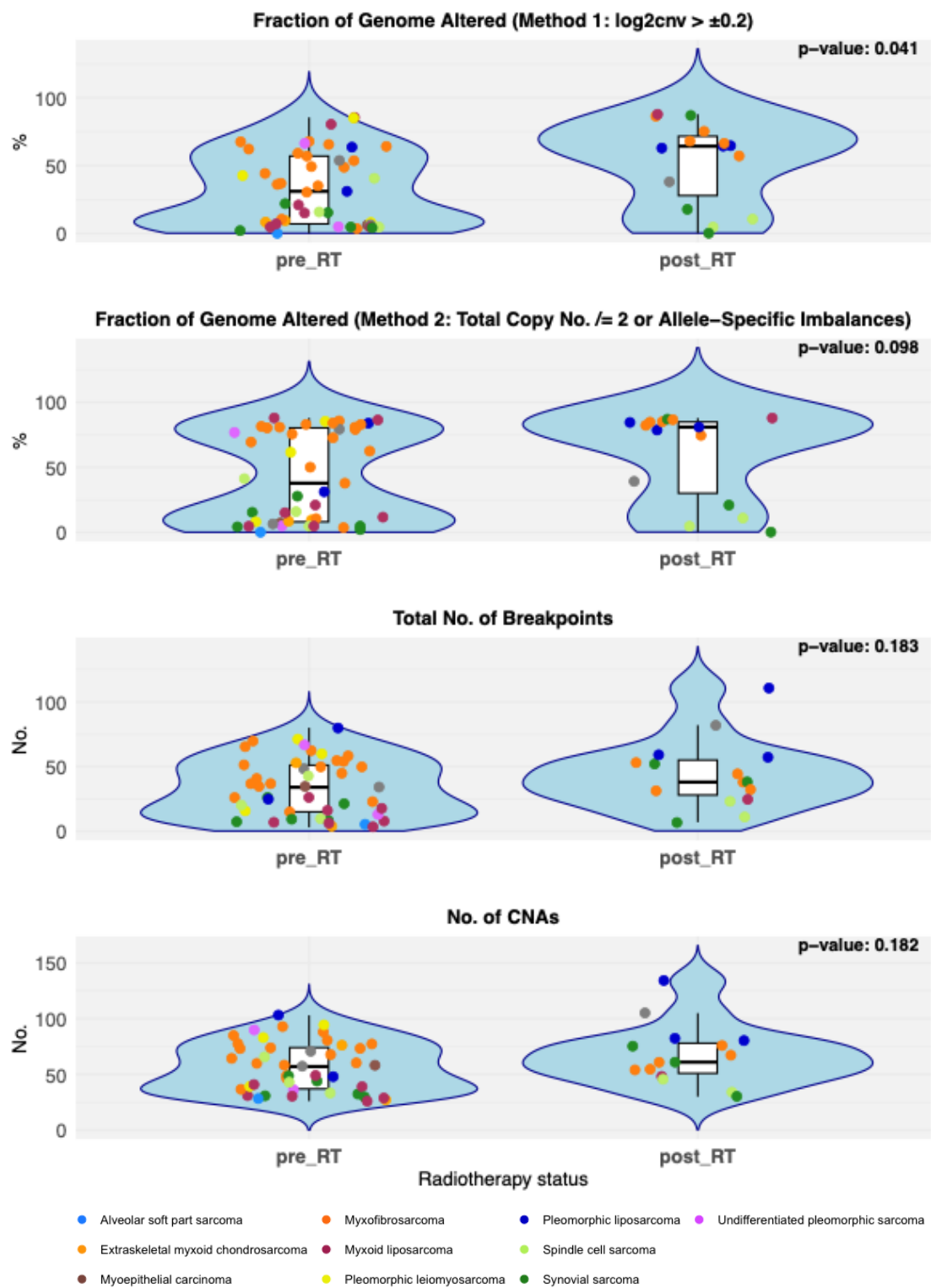


Figure 3.9 Comparison of copy number alteration metrics in pre- and post-radiotherapy samples

This figure compares the fraction of genome altered (FGA), total number of breakpoints, and total number of copy number alterations (CNAs) between pre- and

post-radiotherapy samples across multiple histological soft tissue sarcoma subtypes. The **fraction of genome altered (Method 1)** panel shows regions where total copy number deviates from the expected diploid state ($\log_2 \text{CNV} > \pm 0.2$). The **fraction of genome altered (Method 2)** panel includes both total copy number changes and allele-specific imbalances (total copy number $\neq 2$ or allele-specific imbalances). The **total number of breakpoints** panel represents transitions between distinct copy number states. The **total number of CNAs** panel shows the total count of unique CNAs per sample. Each data point is coloured based on histological subtype, and the radiotherapy status (pre-RT or post-RT) is shown along the x-axis. P-values were calculated using an unpaired t-test. The violin plot shows the distribution of the data, with the width representing the density of values at different levels. The centre line of the embedded box plot represents the median value, with the edges of the box indicating the interquartile range, and the whiskers extending to 1.5 times the interquartile range.

3.4.3 Subtype specific analysis of Copy Number Alteration metrics in pre- and post-radiotherapy samples.

To explore whether the changes in copy number might be specific to certain subtypes, I stratified the analysis by histological subtypes with sufficient pre- and post-radiotherapy samples. These subtypes included synovial sarcoma, myxofibrosarcoma, pleomorphic liposarcoma, and spindle cell sarcoma. Of these four subtypes, myxofibrosarcoma (Figure 3.10) was the only one to show a significant difference between pre- and post-radiotherapy samples. In this subgroup, there was a significant increase in the fraction of genome altered post-radiotherapy using both Method 1 (49% to 68%, $p = 0.004$) and Method 2 (79% to 85%, $p = 0.02$), suggesting a subtype specificity to radiotherapy-induced copy number changes in this subtype.

Although there was a decrease in both the median total number of breakpoints (from 50 to 38) and CNAs (from 73 to 61) in the post-radiotherapy samples, these changes were not statistically significant (both $p = 0.3$). The other subtypes, pleomorphic liposarcoma, synovial sarcoma, and spindle cell sarcoma, showed no significant differences across the metrics tested, indicating that the genomic response to radiotherapy may vary among different sarcoma types.

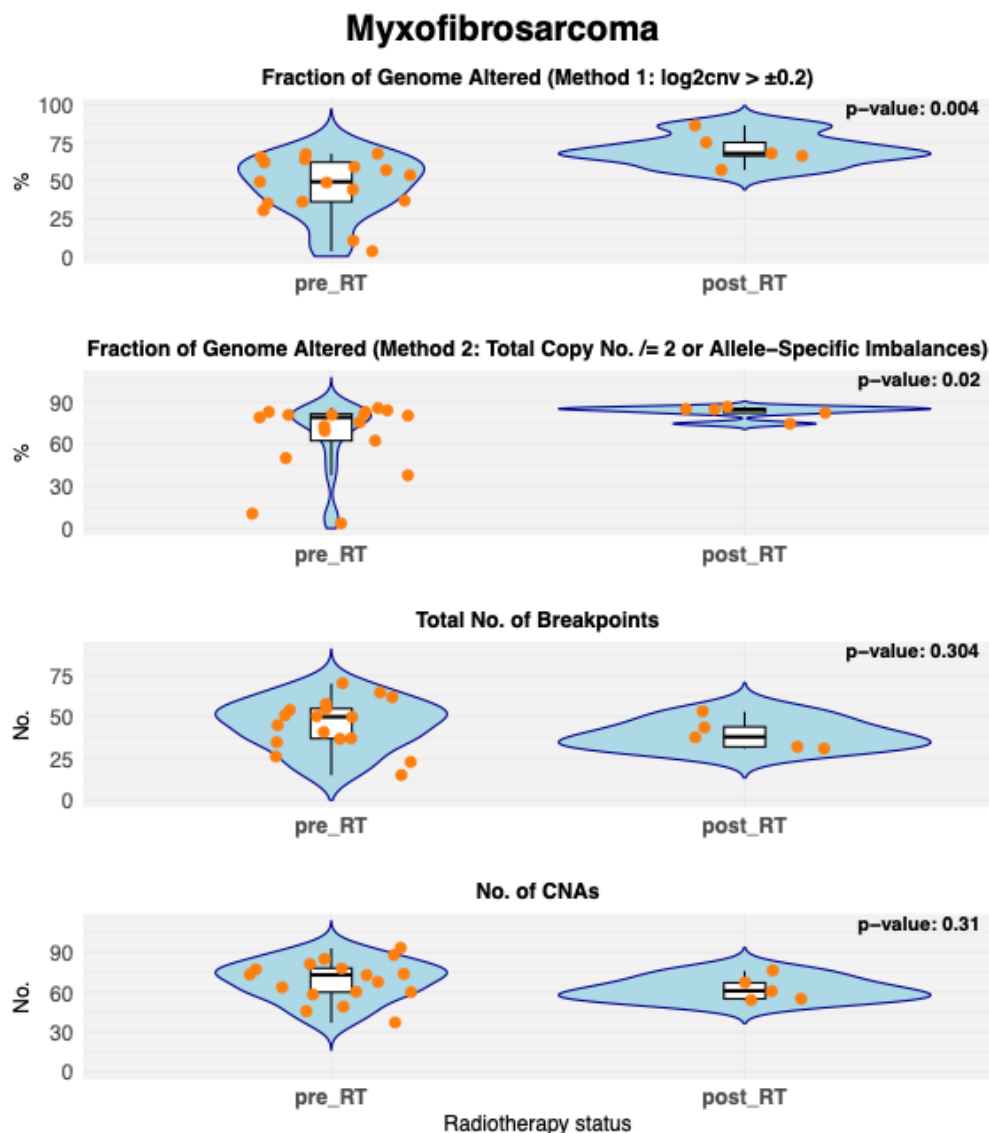


Figure 3.10. Comparison of copy number alteration metrics in pre- and post-radiotherapy myxofibrosarcoma samples

This figure compares the fraction of genome altered (FGA), total number of breakpoints, and total number of copy number alterations (CNAs) between pre- and post-radiotherapy myxofibrosarcoma samples. The **fraction of genome altered (Method 1)** panel shows regions where total copy number deviates from the expected diploid state ($\log_2 \text{CNV} > \pm 0.2$). The **fraction of genome altered (Method 2)** panel includes both total copy number changes and allele-specific imbalances (total copy number $\neq 2$ or allele-specific imbalances). The **total number of breakpoints** panel represents transitions between distinct copy number states. The **total number of CNAs** panel shows the total count of unique CNAs per sample. P-values were calculated using an unpaired t-test. The violin plot shows the distribution of the data, with the width representing the density of values at different levels. The centre line of the embedded box plot represents the median value, with the edges of the box indicating the interquartile range, and the whiskers extending to 1.5 times the interquartile range.

3.4.4 Comparative analysis of Copy Number Alteration metrics in paired pre- and post-radiotherapy samples.

I performed a paired analysis with the six patients for whom I had matching pre- and post-radiotherapy samples (Figure 3.11). There was no significant difference in the fraction of genome altered, total number of breakpoints, or total number of copy number alterations.

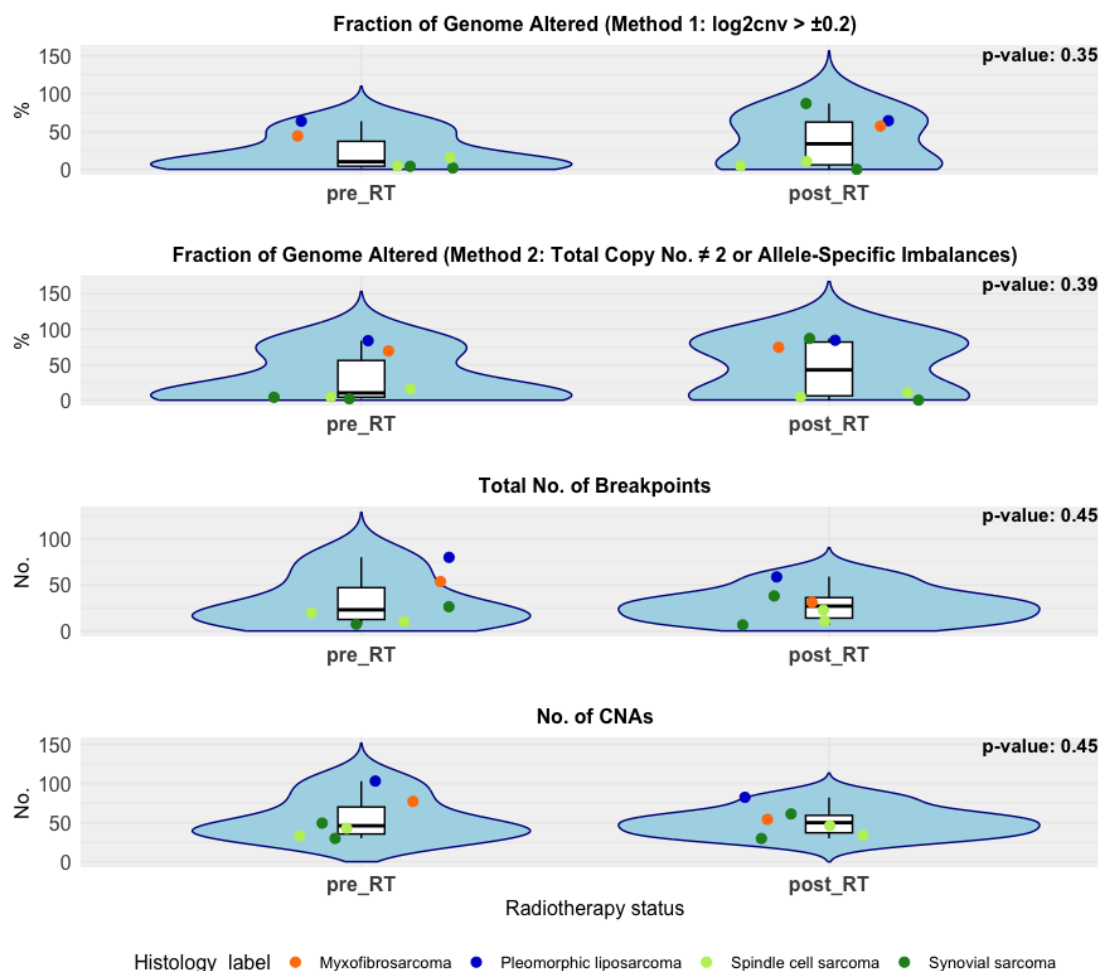


Figure 3.11. Comparison of copy number alteration metrics in paired pre- and post-radiotherapy samples.

This figure compares the fraction of genome altered (FGA), total number of breakpoints, and total number of copy number alterations (CNAs) between the 6 patients with paired pre- and post-radiotherapy samples. The **fraction of genome altered (Method 1)** panel shows regions where total copy number deviates from the expected diploid state ($\log_2 \text{CNV} > \pm 0.2$). The **fraction of genome altered (Method 2)** panel includes both total copy number changes and allele-specific imbalances (total copy number $\neq 2$ or allele-specific imbalances). The **total number of breakpoints** panel represents transitions between distinct copy number states. The **total number of CNAs** panel shows the total count of unique CNAs per sample. P-values were calculated using a paired t-test. The violin plot shows the distribution of the data, with the width representing the density of values at different levels. The centre line of the embedded box plot represents the median value, with the edges of the box indicating the interquartile range, and the whiskers extending to 1.5 times the interquartile range.

3.4.5 Summary

In this section, the comparison of copy number alterations (CNAs) between pre- and post-radiotherapy samples highlighted both global and subtype-specific changes. Post-radiotherapy samples demonstrated no significant increase in the overall number of CNAs; however, an increased Fraction of Genome Altered (FGA) was seen. This suggests an expansion of regions with chromosomal gains or losses following radiotherapy. Specifically, myxofibrosarcoma showed notable differences in the FGA post-treatment, highlighting potential heterogeneity in radiotherapy response among different sarcoma types. Importantly, the lack of a marked increase in CNA counts but an expansion in FGA may imply that radiotherapy could lead to an enlargement of pre-existing altered regions rather than the formation of entirely new CNAs.

3.5 Mutational signature analysis

In this section, I aimed to assess whether there were differences in the mutational signatures between pre- and post-radiotherapy tumour samples. Mutational signatures are characteristic patterns of mutations that reflect the underlying mechanisms of acquired DNA damage and repair (Alexandrov, Nik-Zainal *et al.* 2013, Alexandrov, Kim *et al.* 2020). Analysing these signatures can give clues as to both the mechanisms and aetiology of tumourigenesis as well as how radiotherapy can influence the mutational landscape of tumours.

Previous studies have identified specific mutational signatures associated with radiation-induced damage, such as SBS18, which is linked to oxidative stress caused by radiotherapy, and indel signatures like ID8, which reflect error-prone DNA repair mechanisms such as non-homologous end joining (NHEJ) (Kocakavuk, Anderson *et al.* 2021).

I used SigProfiler (Alexandrov, Nik-Zainal *et al.* 2013) to decompose the single base substitution (SBS) and Indel (ID) mutations identified in the whole exome (WES) and whole genome (WGS) tumour samples into COSMIC mutational signatures.

3.5.1 Analysis of Single base substitution (SBS) signatures

Five SBS signatures were identified across both the WES and WGS datasets (Figure 3.12 and Figure 3.13 respectively). These include SBS1, SBS2, SBS3, SBS5, and SBS13, which were detected in both WES and WGS tumour samples. Additionally, a sixth signature, SBS15, was observed exclusively in the WES samples.

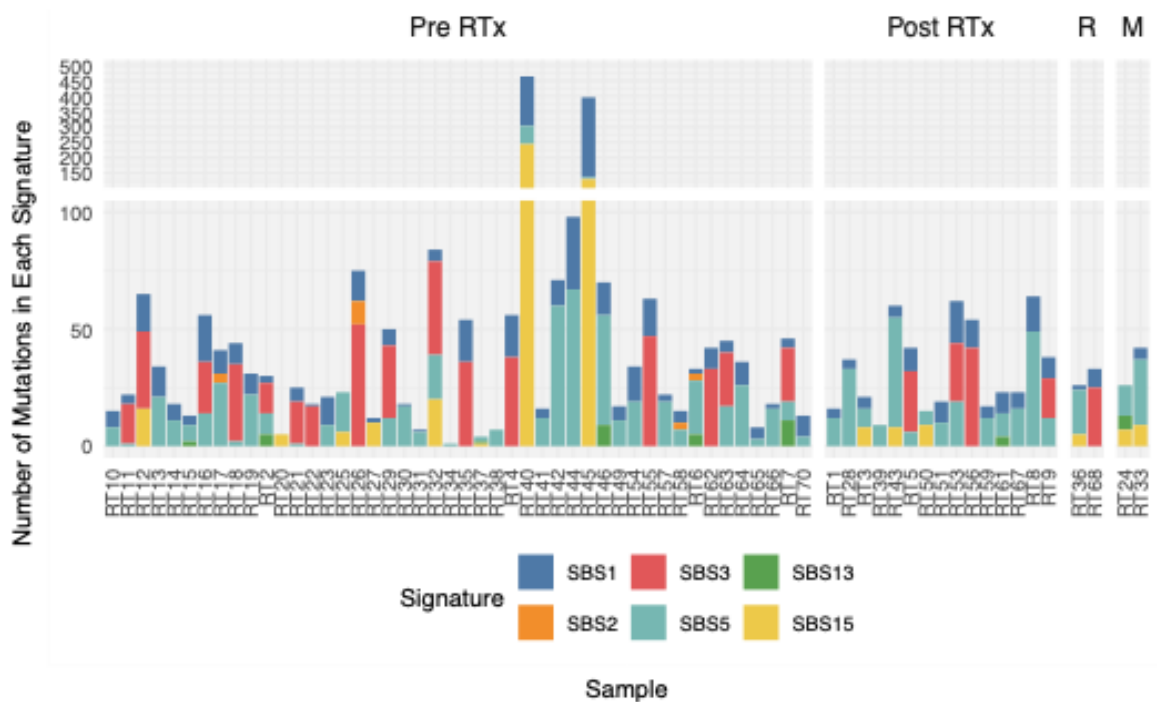


Figure 3.12. Distribution of SBS mutational signatures across pre-radiotherapy, post-radiotherapy, recurrence, and metastasis samples in WES data.

The bar plot illustrates the number of somatic single base substitution (SBS) mutations attributed to various mutational signatures for each sample type. The samples are grouped based on their treatment status, including pre-radiotherapy (Pre RT), post-radiotherapy (Post RT), recurrence (R), and metastasis (M). Each bar represents the contribution of specific SBS signatures within individual samples, with different colours corresponding to each signature.

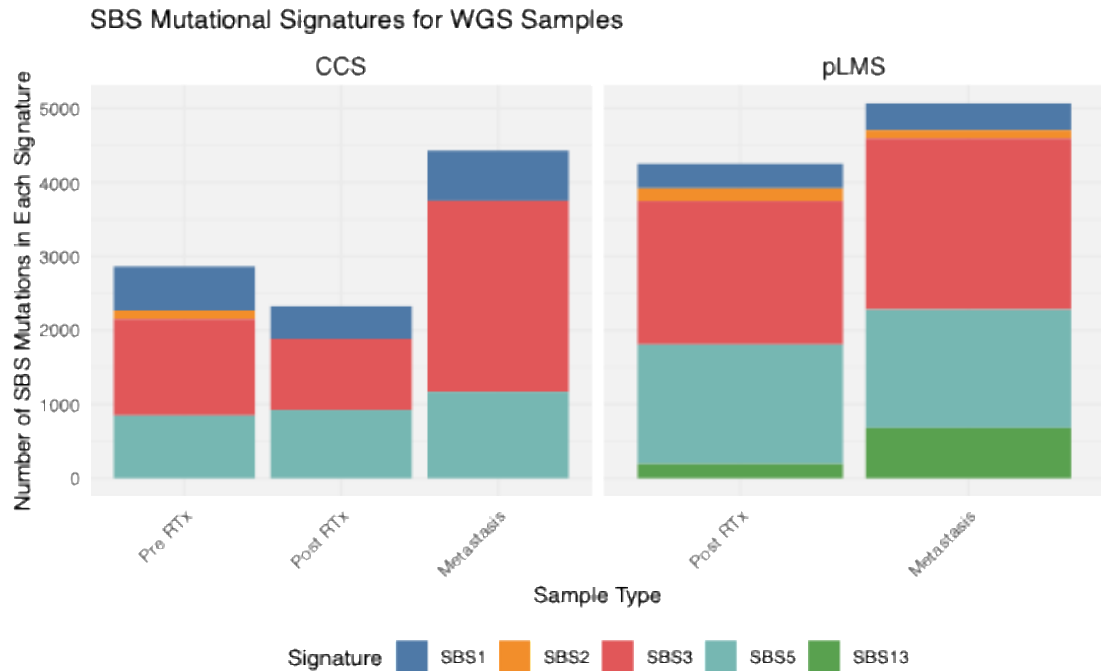


Figure 3.13. Distribution of SBS mutational signatures across pre-radiotherapy, post-radiotherapy, and metastasis samples in WGS data.

The bar plot illustrates the number of somatic single base substitution (SBS) mutations attributed to various mutational signatures for two separate patients. Each bar represents the contribution of specific SBS signatures within individual samples, with different colours corresponding to each signature. Pre-radiotherapy (Pre RT) and post-radiotherapy (Post RT). Clear cell sarcoma (CCS), pleomorphic leiomyosarcoma (pLMS).

The SBS signatures identified in these samples reflect a range of biological processes associated with cancer development. SBS1, often called “clock-like”, is linked to aging and arises from the spontaneous or enzymatic deamination of methylated cytosines, resulting in C>T mutations that accumulate over time; this signature is common across various cancer types (Alexandrov, Nik-Zainal *et al.* 2013, Alexandrov, Kim *et al.* 2020). Both SBS2 and SBS13 are associated with the activity of APOBEC enzymes, which cause C>T and C>G mutations specifically at TpC dinucleotides. The presence of these APOBEC-associated mutations indicates increased genomic instability, which is a hallmark of many cancers, and these signatures are seen in a wide range of cancer types including sarcomas (Alexandrov, Nik-Zainal *et al.* 2013, Alexandrov, Kim *et al.* 2020). SBS3 while not specific is suggestive of homologous recombination deficiency (HRD), and is commonly linked to mutations in *BRCA1* or *BRCA2* genes and frequently observed in cancers such

as breast and ovarian cancer, but has also been seen in sarcomas (Alexandrov, Nik-Zainal *et al.* 2013, Alexandrov, Kim *et al.* 2020). SBS5, is another "clock-like" signature with a currently undetermined cause. It is found across a wide range of tissues and correlates with age, accumulating mutations in both normal and cancerous cells.

Lastly, SBS15 is associated with defective DNA mismatch repair (MMR) and is often found in tumours with microsatellite instability (MSI), making it particularly prevalent in cancers like colorectal and endometrial cancer that exhibit MMR deficiency (Alexandrov, Nik-Zainal *et al.* 2013, Alexandrov, Kim *et al.* 2020).

I have split the tumour samples in to show whether the samples are pre- or post-radiotherapy, metastasis or recurrence samples. All 6 SBS signatures (SBS 1, 2, 3, 5, 13, and 15) were identified in the pre-radiotherapy samples (Figure 3.12). In the post-radiotherapy samples the same signatures except for SBS2 were identified. In the 2 recurrence samples SBS signatures 1, 3, 5, 15 were seen. Lastly in the 2 metastasis samples the SBS signatures 1, 5, 13, 15 were identified. In Figure 3.13, the WGS samples showed SBS1, SBS3, and SBS5 in all samples. SBS2 was identified in the pre-radiotherapy clear cell sarcoma sample but this was not present in the post-radiotherapy or metastasis sample. SBS2 and SBS13 was identified in post-radiotherapy and metastasis samples of pleomorphic leiomyosarcoma.

There are differences in the SBS signatures identified in the WES samples depending on the histological subtype (Figure 3.14). The clock-like signatures SBS1 and SBS5 are seen in all subtypes sequenced.

SBS2 (APOBEC related) was seen in single case in each of pleomorphic leiomyosarcoma (1 of 4 samples), pleomorphic liposarcoma (1 of 5), synovial sarcoma (1 of 8) and myxofibrosarcoma (1 of 23). SBS13 (also APOBEC related) was seen in myxoid liposarcoma (1 of 8 samples), pleomorphic leiomyosarcoma (2 of 4 samples), pleomorphic liposarcoma (1 of 5) and myxofibrosarcoma (3 of 23).

SBS3 (associated with homologous recombination deficiency) was seen in dedifferentiated liposarcoma (1 of 3 samples), myxoid liposarcoma (1 of 8),

pleomorphic leiomyosarcoma (1 of 4), pleomorphic liposarcoma (2 of 5), spindle cell sarcoma (1 of 6), synovial sarcoma (2 of 8), and myxofibrosarcoma (13 of 23).

Lastly SBS15 (associated with MMR deficiency) was seen in alveolar soft part sarcoma (1 of 2 samples), myxoid liposarcoma (1 of 8), pleomorphic leiomyosarcoma (2 of 4), pleomorphic liposarcoma (1 of 5), synovial sarcoma (2 of 8), myxofibrosarcoma (4 of 23), and undifferentiated pleomorphic sarcoma (3 of 3).



Figure 3.14. Distribution of SBS mutational signatures across sarcoma subtypes.

The bar plot illustrates the number of somatic single base substitution (SBS) mutations attributed to various mutational signatures for each sample broken up by histological subtype. Alveolar Soft Part Sarcoma (ASPS), Dedifferentiated Liposarcoma (ddLPS), Extraskeletal Myxoid Chondrosarcoma (EMC), Myoepithelial Carcinoma (MEC), Myxoid Liposarcoma (mLPS), Pleomorphic Leiomyosarcoma (pLMS), Pleomorphic Liposarcoma (pLPS), Spindle Cell Sarcoma (SpCS), Synovial Sarcoma (SS), Myxofibrosarcoma (MFS), and Undifferentiated Pleomorphic Sarcoma (UPS).

3.5.2 Analysis of Indel (ID) signatures

There were different ID signatures identified in the WES and WGS data (Figure 3.15 and Figure 3.16 respectively). ID1, ID2, ID9, and ID23 were identified in both WES and WGS tumour samples. Additionally, ID7, ID8, and ID10 signatures were seen in WES tumour samples.

The ID (Indel) signatures identified in these samples reflect and are attributed to various mutational processes. ID1 and ID2 are commonly associated with aging and reflects the accumulation of small insertions (ID1) and deletions (ID2) over time, often linked to cell division and DNA replication errors. This signature is observed across numerous cancer types. ID1 and ID2 are also frequently associated with DNA mismatch repair (MMR) deficiency, leading to replication slippage and indel mutations, particularly in repetitive DNA regions (Alexandrov, Nik-Zainal *et al.* 2013, Alexandrov, Kim *et al.* 2020).

ID7 is linked to MMR deficiency, similar to ID2, and is often observed in tumours with MSI and has been identified previously in gastric adenocarcinoma (Alexandrov, Kim *et al.* 2020). ID8 is associated double stranded break repair by non-homologous end joining and has been seen in tissues post-radiotherapy (Alexandrov, Kim *et al.* 2020, Kocakavuk, Anderson *et al.* 2021). ID23 is associated with aristolochic acid exposure (Senkin, Moody *et al.* 2024). ID9 and ID10 have an unknown aetiology.

Different signatures were seen depending on where the samples are pre- or post-radiotherapy, metastasis or recurrence samples (Figure 3.15). ID2 ID7, ID8, ID9, ID10, ID23 were identified in the pre-radiotherapy samples. ID7, ID8, ID10, and ID23 were the only signatures identified in the post-radiotherapy samples. ID1, ID8, ID23 were seen in the recurrence samples, and ID23 was seen in the metastasis sample.

There were no changes in the ID signatures called for the two patients with WGS performed on the tumour samples in their pre- or post-radiotherapy samples (Figure 3.16). The patient with clear cell sarcoma was found to have signatures ID1, ID2, and ID9 in the each of the pre-radiotherapy, post-radiotherapy, and metastasis

samples. The patient with pleomorphic leiomyosarcoma likewise had ID1, ID2, ID9, and ID23 in both the post-radiotherapy and metastasis samples.

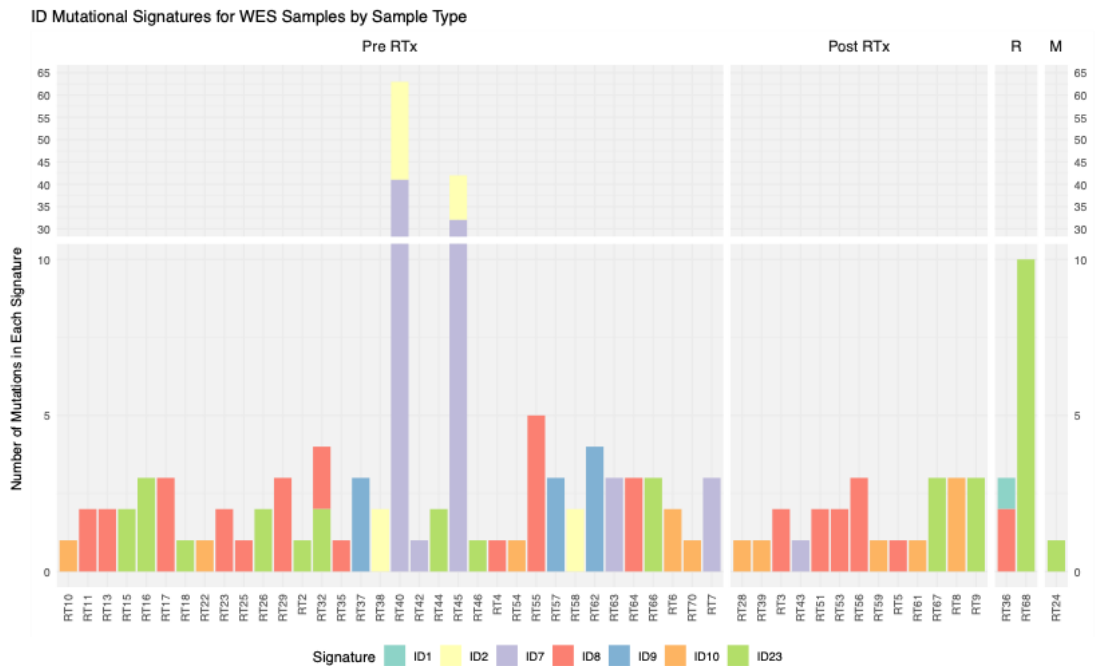


Figure 3.15. Distribution of indel mutational signatures across pre-radiotherapy, post-radiotherapy, recurrence, and metastasis samples in WES data.

The bar plot illustrates the number of Indel (ID) mutations attributed to various mutational signatures for each sample type. The samples are grouped based on their treatment status, including pre-radiotherapy (Pre RT), post-radiotherapy (Post RT), recurrence (R), and metastasis (M). Each bar represents the contribution of specific ID signatures within individual samples, with different colours corresponding to each signature.



Figure 3.16. Distribution of indel mutational signatures across pre-radiotherapy, post-radiotherapy, and metastasis samples in WGS data.

The bar plot illustrates the number of Indel (ID) mutations attributed to various mutational signatures for two separate patients. Each bar represents the contribution of specific ID signatures within individual samples, with different colours corresponding to each signature. Pre-radiotherapy (Pre RT) and post-radiotherapy (Post RT). Clear cell sarcoma (CCS), pleomorphic leiomyosarcoma (pLMS).

ID signatures by histological subtype

No ID signatures were seen in all subtypes examined (Figure 3.17).

ID1 was identified in undifferentiated pleomorphic sarcoma (1 of 2 samples).

ID2 was seen in spindle cell sarcoma (1 of 6), synovial sarcomas (1 of 3), myxofibrosarcoma (1 of 20), and undifferentiated pleomorphic sarcoma (1 of 2).

ID7 was seen in myxofibrosarcoma (5 of 20 samples), and undifferentiated pleomorphic sarcoma (1 of 2).

ID8 was seen in dedifferentiated liposarcoma (1 of 2 samples), myoepithelial carcinoma (1 of 1), myxoid liposarcoma (1 of 6), pleomorphic leiomyosarcoma (1 of 4), pleomorphic liposarcoma (3 of 5), myxofibrosarcoma (9 of 20 samples), and undifferentiated pleomorphic sarcoma (1 of 2).

ID9 was seen in myxoid liposarcoma (2 of 5 samples), and myxofibrosarcoma (1 of 20).

ID10 was seen in dedifferentiated liposarcoma (1 of 2 samples), Extraskeletal myxoid chondrosarcoma (1 of 1), myxoid liposarcoma (3 of 6), pleomorphic liposarcoma (2 of 5), spindle cell sarcoma (1 of 6), and synovial sarcoma (2 of 3).

Lastly ID23 was seen in pleomorphic leiomyosarcoma (3 of 4 samples), spindle cell sarcoma (4 of 6), and myxofibrosarcoma (6 of 20). Given that ID23 is associated with aristolochic acid exposure, it is possible that this represents a spurious signature. Alternatively, this finding could suggest a previously unrecognised role of Aristolochic acid exposure as a risk factor in the development of these sarcomas.



Figure 3.17. Distribution of Indel Mutational Signatures across sarcoma subtypes.

The bar plot illustrates the number of Indel (ID) mutations attributed to various mutational signatures for each sample broken up by histological subtype. Alveolar Soft Part Sarcoma (ASPS), Dedifferentiated Liposarcoma (ddLPS), Extraskelletal Myxoid Chondrosarcoma (EMC), Myoepithelial Carcinoma (MEC), Myxoid Liposarcoma (mLPS), Pleomorphic Leiomyosarcoma (pLMS), Pleomorphic Liposarcoma (pLPS), Spindle Cell Sarcoma (SpCS), Synovial Sarcoma (SS), Myxofibrosarcoma (MFS), and Undifferentiated Pleomorphic Sarcoma (UPS).

3.5.3 Identification of mismatch repair deficiency

Interestingly, two patient samples - RT40 (undifferentiated pleomorphic sarcoma) and RT45 (myxofibrosarcoma), both had a very high tumour mutational burden relative to the other cases (previously discussed in section 3.2) have high numbers and proportions of SBS15 and ID7 mutations which are associated with MMR.

Examining the mutations identified in the WES data, no mutations were identified in the genes related to MMR (*MLH1*, *MSH2*, *MSH6*, and *PMS2*) (Pećina-Šlaus, Kafka *et al.* 2020). Both samples did show mutations within *TP53*, a gene associated with genomic instability, which could contribute to an overall increase in the tumour mutation burden. To further explore potential MMR deficiencies, I examined the associated bulk RNAseq data to review the expression levels of the MMR genes.

RNA expression analysis revealed relatively low expression of *MLH1* in sample RT40 and *MSH2* in sample RT45 compared to the other 115 sequenced samples (Figure 3.18). Low expression of *MLH1* could potentially result from promoter hypermethylation, a known mechanism for *MLH1* silencing in cancers with microsatellite instability (MSI) (Kane, Loda *et al.* 1997). Although *MSH2* is not typically silenced by promoter hypermethylation, alternative regulatory mechanisms—such as loss of heterozygosity (LOH) or post-translational instability due to reduced *MSH6* levels—could explain the observed reduction in *MSH2* expression in RT45.

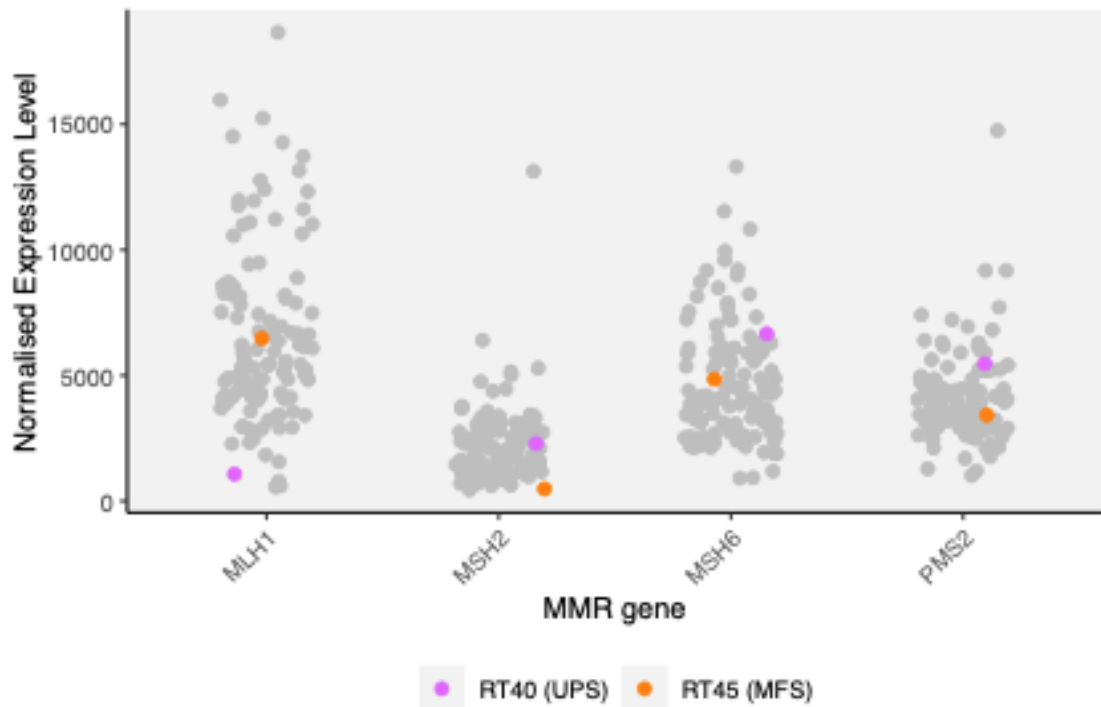


Figure 3.18. Expression of mismatch repair genes

Normalised RNA gene expression levels of mismatch repair (MMR) genes in 117 soft tissue sarcoma, highlighting *MLH1*, *MSH2*, *MSH6*, and *PMS2*. Each dot represents the Normalised expression level of a specific MMR gene in an individual tumour sample. Grey dots correspond to the expression levels across all other samples, providing a reference distribution for each gene. The samples of interest, RT40 (Undifferentiated Pleomorphic Sarcoma, UPS) and RT45 (Myxofibrosarcoma, MFS), are highlighted in purple and orange, respectively.

To investigate further, I conducted a copy number analysis using the ASCAT output and identified several regions with LOH involving MMR genes in both samples. In RT40 (Undifferentiated Pleomorphic Sarcoma), a homozygous deletion of the *MLH1* gene on chromosome 3 was detected. Given that *MLH1* is an essential component of the mismatch repair (MMR) pathway, its complete loss definitively establishes MMR deficiency in this sample. This explains the high tumour mutational burden observed in RT40, as well as the prominent SBS15 and ID7 mutational signatures, both of which are hallmarks of MMR-deficient cancers.

In RT45 (Myxofibrosarcoma), LOH was observed for multiple MMR genes, including *MLH1*, *MSH2*, *MSH6*, and *PMS2*, with one allele lost for each of these genes. While

not a complete deletion, this partial loss may have impaired MMR function, particularly given the functional dependence of *MLH1-PMS2* and *MSH2-MSH6* complexes. However, unlike RT40, where MMR deficiency is clearly explained by complete *MLH1* loss, the extent of MMR impairment in RT45 remains uncertain.

These findings confirm that the homozygous deletion of *MLH1* in RT40 is the key driver of its hypermutator phenotype, as MMR deficiency due to *MLH1* loss is well-established across multiple cancers. The LOH in RT45, while suggestive of MMR involvement, does not provide the same level of certainty. This difference between the two cases highlights the importance of complete gene loss in driving a fully MMR-deficient state.

It is important to note that the data for these two patient tumour samples were obtained from pre-radiotherapy biopsy specimens. The post-radiotherapy resection specimens showed extensive therapy-related necrosis. Histological examination of all tissue blocks revealed an insufficient number of viable tumour cells (indicative of a strong pathological response) for DNA and RNA sequencing studies. Both patients had similar staging of the resection specimen (ypT2b Nx Mx – TNM8). Upon reviewing the available clinical information, the patient with undifferentiated pleomorphic sarcoma in the thigh, diagnosed at age 76, remains disease-free 8 years post-treatment. Unfortunately, despite both patients showing an excellent pathological response to radiotherapy, the patient with myxofibrosarcoma in the shoulder, diagnosed at age 67, developed lung metastases 224 days post-radiotherapy and succumbed to the disease 522 days following treatment.

While not extensively investigated, there is growing evidence that mismatch repair-deficient tumours may be more sensitive to radiotherapy (Shin, Tut *et al.* 2013, Reijnen, Küsters-Vandeveld *et al.* 2019), as demonstrated here by the strong pathological response and absence of local recurrence in both cases. However overall disease-free survival can still be impacted by distant metastasis. In this instance, despite an excellent local response to radiotherapy, it is possible that metastasis in the patient with myxofibrosarcoma developed prior to treatment (albeit clinically undetectable on original staging), ultimately affecting survival outcomes.

3.5.4 Summary

In this section, I investigated the mutational signatures in pre- and post-radiotherapy sarcoma samples to identify any distinct mutational effects attributable to radiotherapy. Using SigProfiler, I identified 6 SBS signatures (SBS1, SBS2, SBS3, SBS5, SBS13, and SBS15) and 7 ID signatures (ID1, ID2, ID7, ID8, ID9, ID10, and ID23) across the samples.

SBS1 and SBS5, associated with endogenous aging processes, were consistently present in all samples. Conversely, SBS2 and SBS13, linked to APOBEC activity, and SBS3, associated with homologous recombination deficiency, were detected only in specific subtypes, indicating underlying genomic instabilities in certain sarcomas. Notably, SBS18, previously linked to radiotherapy, was absent.

For indel signatures, ID1 and ID2—related to aging and DNA mismatch repair deficiency—were observed in multiple subtypes, while ID8, associated with double-strand break repair, appeared in both pre- and post-radiotherapy samples.

These findings suggest that, within the sensitivity limits of whole-exome sequencing, radiotherapy does not introduce a new mutational profile or significantly alter the existing mutational landscape in soft tissue sarcomas. The stable prevalence of SBS and ID signatures pre- and post-radiotherapy suggests that radiotherapy has a limited impact on mutational signatures detectable by whole-exome sequencing. This stability may indicate that radiotherapy-induced changes in sarcomas are either minimal or require higher-resolution methods, such as whole-genome sequencing, or high-resolution duplex sequencing techniques such as NanoSeq to detect more subtle or subclonal alterations.

3.6 Using NanoSeq to more accurately interrogate the genomic response to radiotherapy.

In section 3.3, I presented the results of a comparison of the number of somatic mutations in soft tissue sarcomas pre- and post-radiotherapy. These results were generated using bulk Whole Exome Sequencing (WES) on formalin-fixed, paraffin-embedded (FFPE) tumour tissue. Based on the literature, an increase in small insertions and deletions (indels) was expected following radiotherapy due to its known effect on inducing DNA damage. However, the WES data did not show any significant changes in the number of SNVs or indels post-treatment.

The tumour mutational burden (TMB) in this cohort of WES samples was low (mean TMB 1.07 mutations/Mb), with a median of 3 indels identified in pre-radiotherapy samples and 2 indels post-radiotherapy.

I hypothesised that the limited resolution of bulk WES, especially when applied to FFPE tissue, might have missed subtle genomic alterations, particularly those present at low variant allele frequencies (VAFs) or in subclonal populations. Moreover, cancer genome sequencing studies have demonstrated that indels in coding regions of the genome are generally low because of a selection bias constraint that attempts to preserve protein function (de la Chaux, Messer *et al.* 2007, Martincorena, Raine *et al.* 2017). In bulk sequencing, signals from subclonal mutations are often diluted by the predominant clonal population, making it challenging to detect low-frequency variants such as indels. This limitation is further exacerbated by the degraded quality of DNA in FFPE samples. Consequently, key genomic changes induced by radiotherapy, particularly indels in minor subclonal populations, might have gone undetected with WES.

To address this, I collaborated with the Martincorena group at the Wellcome Trust Sanger Institute who developed the recently described NanoSeq method (Abascal, Harvey *et al.* 2021), a highly sensitive sequencing technology designed to detect low-frequency mutations, including indels and single nucleotide variants (SNVs), at a higher resolution. NanoSeq uses duplex sequencing, which reads both DNA strands and filters out sequencing errors, allowing for the detection of rare variants

even in low-purity tumour samples or those with a low mutational burden. The sequencing error rate is estimated to be less than five errors per billion base pairs which is reportedly two orders of magnitude lower than usual somatic mutation loads.

Using DNA extracted from fresh frozen tissue on 5 patients on both the pre- and post-radiotherapy samples they performed the sequencing technique and returned the aligned and processed VCF files along with counts of somatic mutations on which I performed the downstream analyses shown below.

NanoSeq was conducted on 5 patients with different sarcoma subtypes. These were patient 7 (pleomorphic liposarcoma), patient 19 (Myxofibrosarcoma), patient 27 (Dedifferentiated liposarcoma), patient 36 (Synovial sarcoma), and lastly patient 58 (Spindle cell sarcoma). The pre-radiotherapy biopsy for patient 19 (myxofibrosarcoma) showed evidence on contamination during quality control and so has been removed from the subsequent analyses.

3.6.1 Comparison of the frequency of somatic mutations pre- and post-radiotherapy.

The paired comparison of the number of indels and SNVs per cell before and after radiotherapy revealed distinct patterns (Figure 3.19). The indel and SNV mutation counts were normalised to the estimated cellular content of each sample to enable a more precise comparison of mutation rates.

The number of indels per cell demonstrated a significant increase following radiotherapy ($p = 0.01$), particularly in dedifferentiated liposarcoma and spindle cell sarcoma. The median number of indels per cell rose from 177 to 690 in these four patients after treatment, suggesting a pronounced effect of radiotherapy on the generation of small insertions and deletions.

In contrast, the number of SNVs per cell did not show a significant change post-radiotherapy ($p = 0.8$). The median number of SNVs decreased from 3813 to 2814 following radiotherapy. However, patient-specific responses varied: dedifferentiated liposarcoma, synovial sarcoma, and spindle cell sarcoma showed increases in SNVs of 19%, 18%, and 16%, respectively. Conversely, pleomorphic liposarcoma demonstrated a 52% reduction in SNVs. These findings demonstrate the heterogeneity in genomic responses to radiotherapy across different sarcoma subtypes.

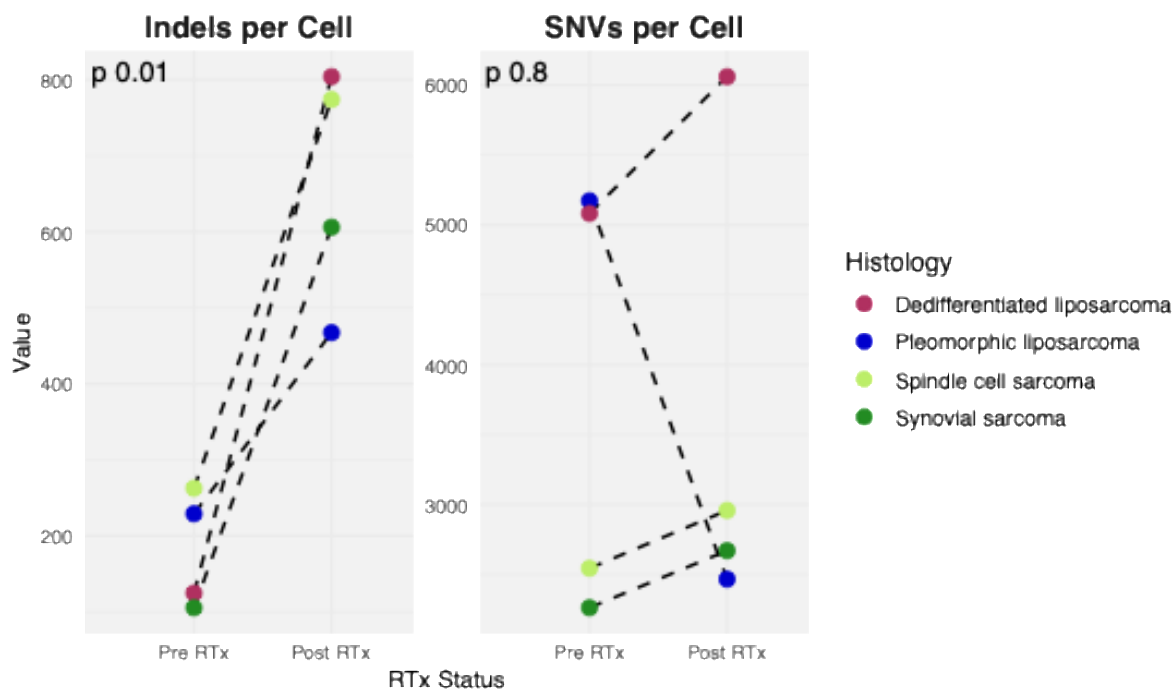


Figure 3.19. Indels and SNVs per Cell Pre- and Post-Radiotherapy.

Paired comparisons of indels per cell and SNVs per cell for each patient (PT). Indels per cell increased significantly post-radiotherapy ($p = 0.01$), while SNVs per cell did not show a significant difference ($p = 0.8$). Dashed lines connect pre- and post-radiotherapy values for individual patients across different histologies.

3.6.2 Assessment of the Indel-to-SNV and the Deletion-to-Insertion ratios.

The Indel-to-SNV and Deletion-to-Insertion ratios (see section 3.3.4), derived from the NanoSeq data, provide insight into the genomic impact of radiotherapy. The Indel-to-SNV ratio increased significantly from a median of 0.05 in pre-radiotherapy samples to 0.2 in post-radiotherapy samples ($p = 0.002$; Figure 3.20A). This finding suggests that radiotherapy elevates the frequency of double-strand breaks, leading to more indels through error-prone repair mechanisms such as non-homologous end joining (NHEJ).

To further deconstruct this increase, indels were separated into deletions and insertions. The Insertion-to-SNV ratio (Figure 3.20B) showed a slight rise from a median of 0.009 pre-radiotherapy to 0.01 post-radiotherapy ($p = 0.04$), indicating a mild increase in insertion events. By contrast, the Deletion-to-SNV ratio (Figure 3.20C) exhibited a more pronounced change, increasing from a median of 0.04 to 0.2 post-radiotherapy ($p = 0.005$). These results highlight deletions as the primary contributors to the elevated Indel-to-SNV ratio, consistent with observations from previous studies on papillary thyroid cancer and gliomas (Kocakavuk, Anderson *et al.* 2021, Morton, Karyadi *et al.* 2021), where deletions were prominent in radiation-induced and post-radiotherapy tumours respectively.

The Deletion-to-Insertion ratio also increased (Figure 3.20D), rising from a median of 3.6 before radiotherapy to 9.6 after treatment. However, this change did not reach statistical significance ($p = 0.08$). While this trend suggests a preference for deletions in resolving radiotherapy-induced DNA damage, there appears to be variability among samples.

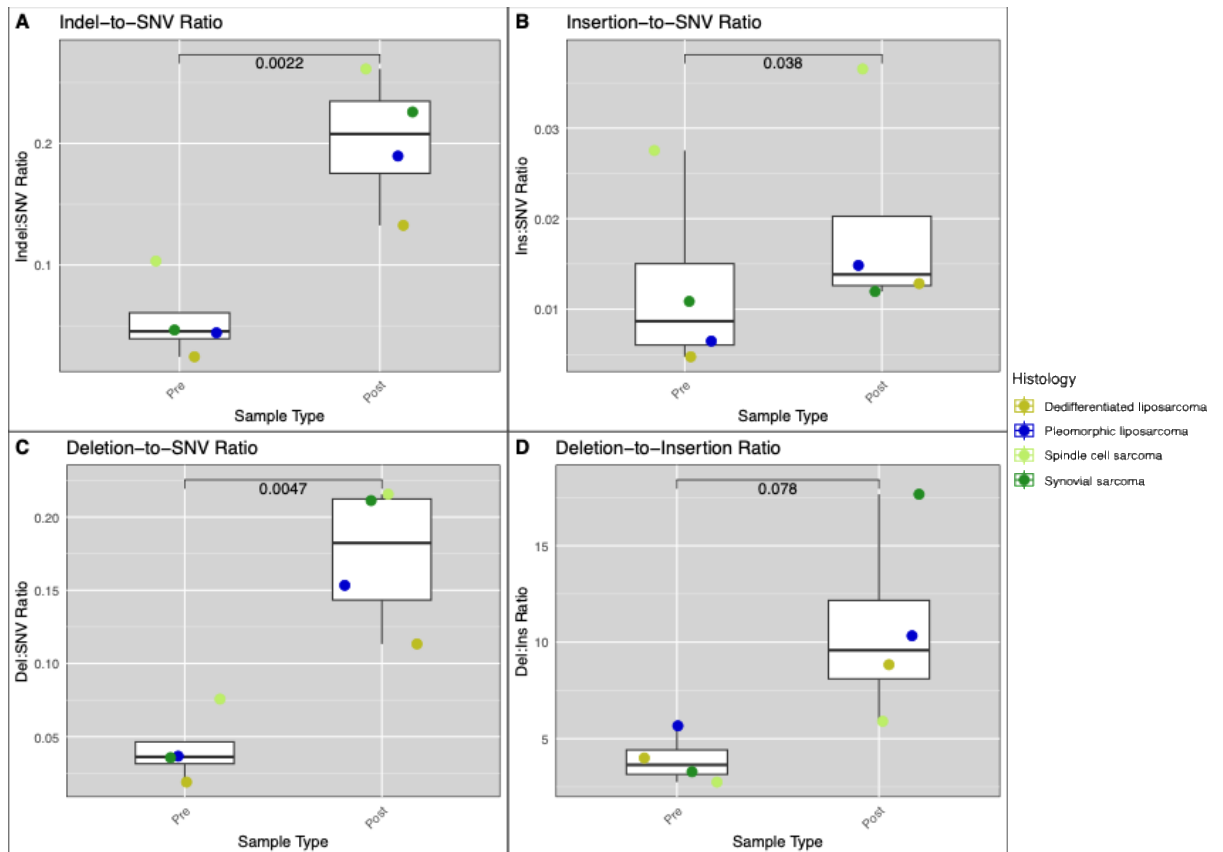


Figure 3.20. Comparison of Indel-to-SNV and Deletion-to-Insertion ratios

Boxplots of **(A)** Indel-to-SNV ratio, **(B)** Insertion-to-SNV ratio, **(C)** Deletion-to-SNV ratio, and **(D)** Deletion-to-Insertion ratio across pre-radiotherapy (Pre) and post-radiotherapy (Post) samples. Each point represents an individual tumour sample, coloured by histological subtype. Statistical comparisons between pre- and post-radiotherapy samples are displayed above each boxplot, with p-values calculated using paired t-tests. The centre line of the box plot represents the median value, with the edges of the box indicating the interquartile range, and the whiskers extending to 1.5 times the interquartile range.

3.6.3 Radiotherapy-induced shift towards microhomology-mediated DNA repair.

In light of the previous findings in this section, which demonstrated genomic changes post-radiotherapy, I aimed to investigate potential shifts in DNA damage repair pathways, specifically examining the reliance on microhomology-mediated end joining (MMEJ) for repair. MMEJ is an error-prone DNA repair pathway typically activated in response to DNA double-strand breaks, which are a known consequence of radiotherapy. This pathway introduces specific indel mutations, termed microhomology-mediated indels, that reflect reduced repair fidelity. Given the increased mutation burden observed in post-radiotherapy samples, I hypothesised that the MMEJ pathway might be more frequently utilised following radiotherapy.

Across all four patients, there was a significant increase in the proportion of microhomology-mediated indels following radiotherapy (paired t-test, $p = 0.002$). Specifically, the median proportion increased from 6% in pre-radiotherapy samples to 27% post-radiotherapy, and the median number of microhomology-mediated indels per cell rose from 9 to 151 (Figure 3.21). This marked increase suggests that post-radiotherapy, tumours are utilising the MMEJ pathway more frequently.

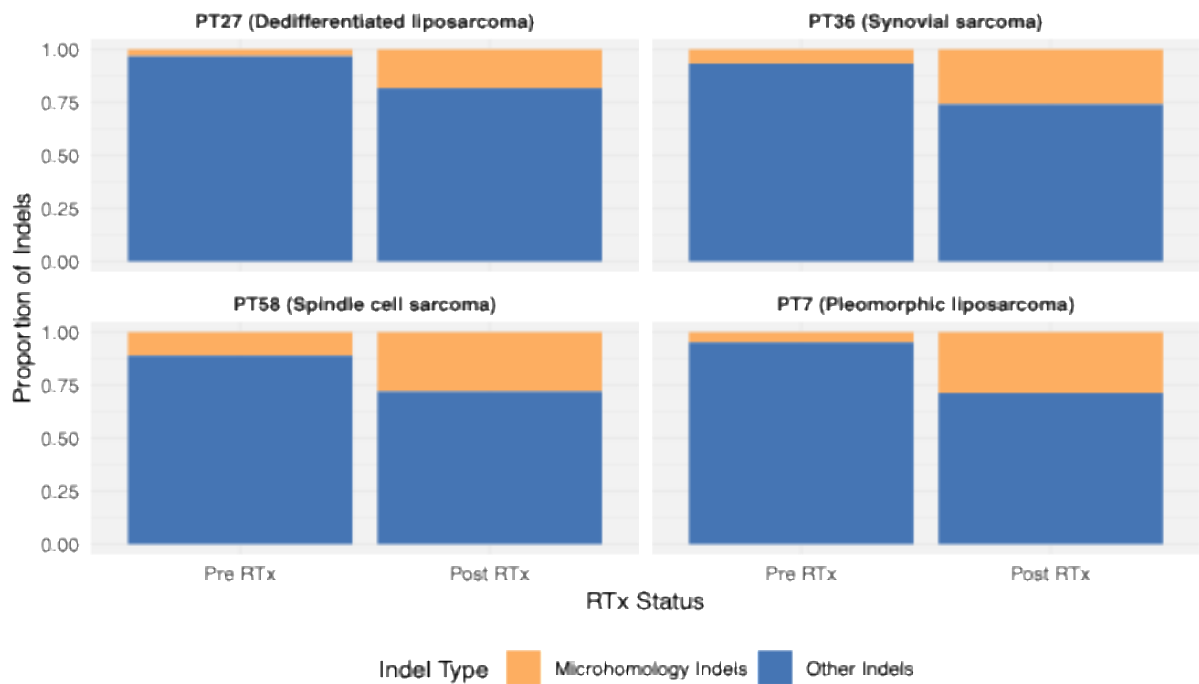


Figure 3.21. Proportion of microhomology Indels to total Indels pre- and post-radiotherapy.

Stacked bar plot showing the proportion of microhomology-mediated indels relative to other indels in four patients.

3.6.4 Mutational signature analysis of NanoSeq cohort

In this section, I used SigProfiler to decompose the single base substitution (SBS) and insertion/deletion (ID) mutations identified in VCF files from the NanoSeq cohort into COSMIC mutational signatures. This analysis aimed to explore any mutational signatures linked to radiotherapy in the cohort, as well as to understand the underlying mutational processes within the sarcoma subtypes.

Single Base Substitution Signatures

Three distinct SBS signatures—SBS1, SBS5, and SBS40a—were detected across all samples (Figure 3.22). The clock-like signatures SBS1 and SBS5 were detected across multiple samples. SBS1 was present in all subtypes except dedifferentiated liposarcoma. SBS5 was seen in all subtypes except the pre-radiotherapy biopsy of dedifferentiated liposarcoma. SBS40a, which is a signature of unknown aetiology, appeared in both pre- and post-radiotherapy samples of dedifferentiated liposarcoma as well as in the post-radiotherapy sample of pleomorphic liposarcoma.

A Wilcoxon signed-rank test was used to assess whether the mutational burden of these SBS signatures changed significantly following radiotherapy. No significant differences were observed for SBS1 ($p = 0.42$), SBS5 ($p = 0.88$), or SBS40a ($p = 1$), indicating that radiotherapy did not introduce substantial new SBS mutations.

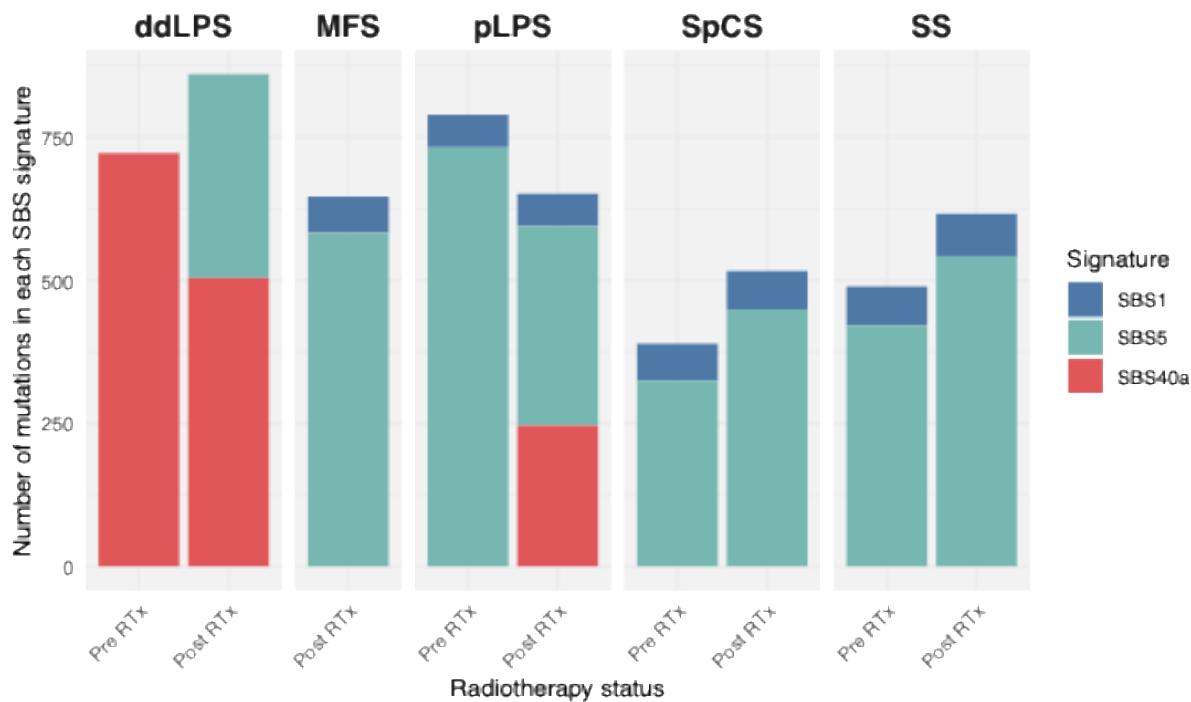


Figure 3.22. Single base substitution signatures pre- and post-radiotherapy samples across the NanoSeq cohort.

Single base substitution (SBS) mutational signatures in pre- and post-radiotherapy Stacked bar plots show the counts of COSMIC SBS signatures (SBS1, SBS5, and SBS40a) across five patients, with each bar representing the prevalence of each signature in pre- and post-radiotherapy samples. Dedifferentiated liposarcoma (ddLPS), myxofibrosarcoma (MFS), pleomorphic liposarcoma (pLPS), spindle cell sarcoma (SpCS), and synovial sarcoma (SS).

Indel Signatures

Five distinct ID signatures—ID1, ID2, ID5, ID8, and ID9—were identified (Figure 3.23), showing some variability across samples. ID1 and ID2, commonly associated with aging and DNA mismatch repair deficiencies, were consistently present. ID5 and ID9 have an unknown aetiology.

The Wilcoxon signed-rank test for paired pre- and post-radiotherapy samples showed no statistically significant differences for ID1 ($p = 0.20$), ID2 ($p = 0.42$), ID5 ($p = 0.25$), and ID9 ($p = 0.18$). However, ID8, which has been linked to double-strand break repair processes, was observed exclusively in all post-radiotherapy samples. This finding suggests a possible association between ID8 and radiotherapy-induced mutational processes, though statistical significance was not achieved ($p = 0.13$), likely due to the small sample size and absence of ID8 mutations in pre-radiotherapy samples, limiting the statistical power.

The lack of significant changes in SBS and most ID signatures indicates that radiotherapy, as detectable by NanoSeq in this cohort, does not dramatically alter the mutational signature profile. However, while not statistically significant the presence of ID8 identified exclusively in post-radiotherapy samples hints at a potential link to radiotherapy-induced mutagenic processes, warranting further investigation with larger cohorts.

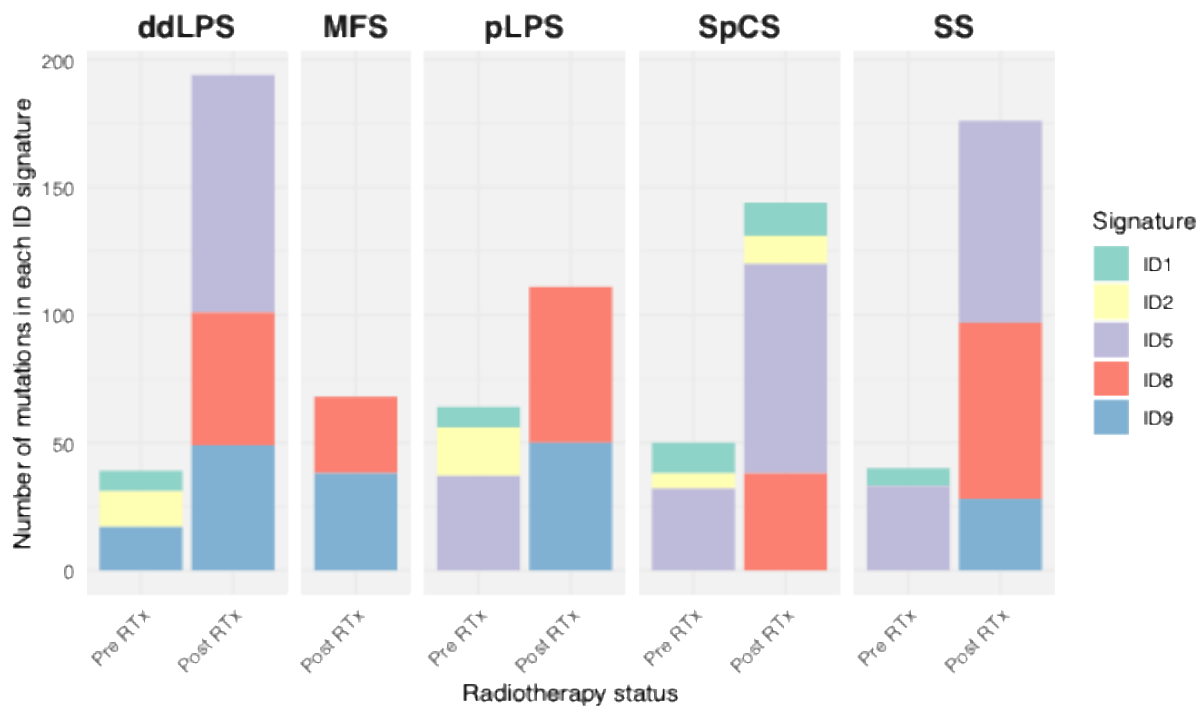


Figure 3.23. Indel (ID) mutational signatures in pre- and post-radiotherapy samples across the NanoSeq cohort.

Stacked bar plots illustrate the counts of COSMIC ID signatures (ID1, ID2, ID5, ID8, and ID9) across five patients, showing the variability of these signatures before and after radiotherapy. Dedifferentiated liposarcoma (ddLPS), myxofibrosarcoma (MFS), pleomorphic liposarcoma (pLPS), spindle cell sarcoma (SpCS), and synovial sarcoma (SS).

3.6.5 Summary

In this section, NanoSeq was applied to detect low-frequency mutations and subtle genomic changes in soft tissue sarcomas, overcoming the limitations of whole-exome sequencing in capturing radiotherapy-induced mutations. The increased sensitivity of NanoSeq allowed for a deeper examination of mutation types and DNA repair mechanisms post-radiotherapy.

NanoSeq analysis showed a statistically significant rise in indels per cell after radiotherapy, across all sarcoma subtypes examined. While SNVs per cell did not significantly increase overall, there were subtype-specific increases in SNVs suggested sarcoma-specific genomic responses. The Indel-to-SNV ratio also rose, mainly due to an increase in deletions, indicating a potential elevation in double-strand break repair events. Additionally, a notable increase in microhomology-mediated indels post-radiotherapy suggested a shift towards the error-prone microhomology-mediated end joining (MMEJ) repair pathway, potentially contributing to genomic instability.

Mutational signature analysis revealed the consistent presence of the clock-like SBS1 and SBS5 signatures across the cohort. These signatures were observed in both pre- and post-radiotherapy samples. Additionally, SBS40a, a signature of unknown aetiology, was detected in some samples pre- and post-radiotherapy, further highlighting the background mutational processes within these sarcomas.

Six indel signatures were identified: ID1, ID2, ID5, ID8, and ID9. ID1 and ID2, commonly associated with aging and DNA mismatch repair deficiencies, were observed consistently across samples, while ID5 and ID9, which are of unknown aetiology, showed some variability. Notably, ID8 - a signature associated with double-strand break repair — was uniquely present in post-radiotherapy samples, hinting at a potential radiotherapy-induced mutational effect. The emergence of ID8 post-treatment may reflect a radiotherapy-specific response, activating repair pathways associated with DNA double-strand breaks and potentially contributing to genomic instability.

3.7 Discussion of chapter 3

The findings in this chapter improve our understanding of the genomic consequences of radiotherapy in soft tissue sarcomas (STS), particularly in relation to tumour mutational burden (TMB), copy number alterations (CNAs), and mutational signatures. While radiotherapy is known to cause DNA damage, its effect on somatic mutation burden in STS has remained unclear. This study demonstrated that radiotherapy does not significantly increase overall TMB, aligning with previous reports of low baseline mutation rates (~1 mutation per Mb) in STS. However, subtle shifts in the Indel-to-SNV and Deletion-to-Insertion ratios suggest an increased reliance on error-prone DNA repair pathways following radiation exposure.

Copy Number Alterations and chromosomal instability

CNA analysis revealed that the fraction of genome altered (FGA) increased post-radiotherapy, suggesting a role for radiotherapy-induced chromosomal instability in STS. While CNAs have been linked to radiotherapy response in cancers such as lung adenocarcinoma and breast cancer, their role in sarcomas remains underexplored. Additionally, pre-existing genomic alterations must be distinguished from treatment-induced changes, as exemplified by a homozygous *MLH1* deletion in RT40, a pre-radiotherapy biopsy sample. These observations emphasise the need to differentiate between baseline chromosomal instability and radiotherapy-driven CNA evolution when evaluating post-treatment genomic changes.

Mutational signature and NanoSeq analysis

Mutational signature analysis was severely limited by the low number of mutations detected by WES, making it difficult to confidently attribute signatures or compare pre- and post-radiotherapy samples. The median number of SNVs and indels in the WES data was only 33 and 2 per sample, respectively, while the WGS data (available for only five samples) showed a median of 3873 SNVs and 145 indels per sample. Given these limitations, mutational signature attribution from WES carries a high degree of uncertainty, and comparisons between pre- and post-radiotherapy samples must be interpreted with caution. Similarly I would hypothesise that conventional WGS might also show a limited change in the detectable mutation rate post-radiotherapy. Due to the relatively short duration of exposure, there would be

insufficient time for significant clonal outgrowths that could be detectable by WGS unless undertaken at ultra-high depth (>1000x).

To overcome this, NanoSeq was employed, enabling the detection of low-frequency mutations with much higher sensitivity. This analysis revealed a significant increase in the number of indels following radiotherapy, with the median number of indels per cell rising from 177 pre-radiotherapy to 690 post-radiotherapy across four patients. This represents a nearly fourfold increase in indels, making it the most striking post-radiotherapy genomic change identified in this study. Additionally, the median number of SNVs decreased from 3813 to 2814 post-radiotherapy, suggesting potential shifts in mutational processes and repair pathway engagement.

NanoSeq provided a more reliable assessment of mutational processes than WES. SBS1 and SBS5, associated with endogenous aging, were consistently detected, while SBS40a (of unknown aetiology) appeared in some cases but was not linked to radiotherapy. No radiation-associated SBS signatures (e.g., SBS18) were detected post-radiotherapy, confirming that radiotherapy does not drive a distinct single base substitution mutational signature in STS.

In contrast, indel (ID) signature analysis provided evidence of potential radiotherapy-induced effects. Five ID signatures (ID1, ID2, ID5, ID8, and ID9) were detected, with ID8 emerging exclusively in all post-radiotherapy samples. ID8 has previously been linked to double-strand break repair via non-homologous end joining (NHEJ) (see Section 1.7.4), suggesting that radiotherapy-induced DNA damage may drive increased reliance on this error-prone repair pathway. The significant rise in overall indels post-radiotherapy, coupled with the presence of ID8 in all post-treatment samples, strongly suggests that radiotherapy promotes genomic instability through mutagenic repair processes. While statistical significance was not reached due to small sample size, these findings warrant further validation in larger cohorts.

Future Directions

While TMB remained stable post-radiotherapy, the significant rise in indels and copy number alterations (CNAs) suggests that radiotherapy alters DNA repair dynamics in STS, potentially influencing tumour evolution. Future studies should investigate

whether radiotherapy-induced shifts toward error-prone repair pathways (e.g., MMEJ) contribute to treatment resistance or recurrence. Longitudinal multi-omics profiling combined with functional validation (e.g., DNA repair assays, CRISPR-based studies) will be essential to determine whether these changes create targetable vulnerabilities for improving sarcoma treatment.

Chapter 4. The transcriptomic response to neoadjuvant radiotherapy.

4.1 Introduction

While Radiotherapy (RT) is known to cause DNA damage, its broader effects on tumour biology - particularly at the transcriptomic level - remain poorly understood. In Chapter 3, genomic analyses revealed no recurrent somatic variants or copy number alterations consistently induced by RT across subtypes, aligning with the known genomic heterogeneity of STS. However, NanoSeq analysis identified a significant increase in indels post-RT, with the emergence of ID8, a signature associated with double strand break repair via non-homologous end joining. These findings indicate that while RT does not generate a distinct SNV-based mutational signature, it likely induces genomic instability through increased reliance on error-prone DNA repair mechanisms.

Given the absence of recurrent RT-induced driver mutations, this raises a key question: do STS tumours instead exhibit shared transcriptional responses to RT? Exploring the transcriptomic landscape could reveal adaptive cellular programs that, despite underlying genomic heterogeneity, may be targetable for therapeutic intervention. If RT induces predictable changes in gene expression or pathway activity, these could represent potential vulnerabilities that could be leveraged for combination therapies to enhance treatment efficacy.

Despite the molecular diversity of STS, prior studies suggest that certain transcriptional programs, such as stress response and immune modulation, may be conserved across tumour types (McKelvey, Hudson *et al.* 2018, Wang, Lynch *et al.* 2024). However, little is known about how RT reshapes the transcriptome in STS. This chapter addresses this gap by identifying radiotherapy-induced gene expression changes, altered signalling pathways, and transcriptional responses that may contribute to tumour adaptation.

Objectives of this chapter

This chapter specifically addresses the following questions:

- Do STS tumours exhibit shared transcriptional responses to RT despite their genomic heterogeneity?
- Which genes and pathways are consistently altered following RT across sarcoma subtypes?
- How do transcriptomic responses vary by histological subtype?

By integrating differential expression analysis, pathway enrichment, and tumour phenotypic comparisons, this chapter aims to identify potentially targetable transcriptional responses to RT in STS.

Structure of this chapter

The chapter begins with a clinical overview of the cohort, outlining histological distribution and patient characteristics. Next, exploratory transcriptomic analyses, including principal component analysis (PCA) and UMAP clustering, assess global transcriptional shifts following RT.

Differential gene expression analysis is performed to identify genes that are significantly upregulated or downregulated post-RT across STS subtypes. To gain biological insight, gene set enrichment analysis (GSEA) and PROGENy pathway analysis are used to determine RT-induced changes in cellular signalling pathways. Additionally, Gene Ontology (GO) enrichment analysis characterises the functional roles of differentially expressed genes, while xCell immune deconvolution assesses changes in immune cell composition following radiotherapy.

4.2 Outline of dataset

To investigate the impact of radiotherapy on the transcriptome of soft tissue sarcomas, I identified 61 patients treated by the London Sarcoma Service who underwent neoadjuvant radiotherapy. Of these 61, 58 patients had FFPE tissue (pre- and postoperatively) and 1 had fresh frozen tissue available for RNA extraction and sequencing. For 58 patient samples, RNA extraction yielded sufficient quantities (minimum 265 ng, median 5.5 µg) and concentrations (minimum 5.3 ng/µl, median 111 ng/µl) for sequencing. The only exception was the pre-radiotherapy biopsy from patient 57, which had insufficient concentration for sequencing.

A total of 119 samples from 58 patients were sent for sequencing. One sample was repeated (patient 4 - post-radiotherapy sample) due to initial poor-quality sequencing. The repeated sample successfully passed quality control. Two samples (patient 42 and patient 56 - pre-radiotherapy biopsies) did not pass quality control after sequencing, and there was insufficient tissue to attempt a repeat extraction.

The median number of sequenced reads per sample was 118.3 million (range 90.3 – 131.1 million).

Summary of Samples and Patients

- **Total number of patients:** 57
- **Total number of sequenced samples that passed QC:** 117
- **Number of pre-radiotherapy biopsy samples:** 54 (from 54 separate patients)
- **Number of post-radiotherapy resection samples:** 55 (from 43 separate patients)
- **Number of paired pre- and post-radiotherapy patients:** 40
- **Number of recurrence samples:** 3 (from 3 separate patients)
- **Number of metastasis samples:** 5 (from 5 separate patients)

40 patients had matched pre- and post-radiotherapy tumour samples sequenced. Four patients (patients 4, 24, 32, and 61) had an additional metastasis sample sequenced. Patient 58 also had a tumour recurrence sample sequenced.

Additionally, 4 patients (patient 39, patient 46, patient 48, and patient 58) underwent multi-region sampling from the post-radiotherapy resection specimen.

Clinical Breakdown of all 57 patients

The histological breakdown of cases is shown in Table 4.1.

- **Age at diagnosis:** ranged between 17 to 89 years old (median 56 years, mean 53 years).
- **Gender distribution:** 38 males and 19 females, resulting in a male-to-female ratio of 2:1.

Table 4.1 Clinical characteristics of all patient samples

This table breaks down the patients in the study according to histological diagnosis.

Diagnosis	Symbol	No. of patients	Male:Female	Age range
Myxofibrosarcoma	MFS	21	14:7	46 – 89
Myxoid liposarcoma	mLPS	7	6:1	37 – 48
Synovial sarcoma	SS	7	4:3	22 – 56
Pleomorphic liposarcoma	PLS	4	3:1	31 – 57
Dedifferentiated liposarcoma	ddLPS	3	2:1	71 – 72
Pleomorphic leiomyosarcoma	pLMS	3	2:1	57 – 78
Undifferentiated pleomorphic sarcoma	UPS	3	2:1	29 – 76
Extraskeletal myxoid chondrosarcoma	EMC	2	2:0	56 – 63
Malignant peripheral nerve sheath tumour	MPNST	2	0:2	40 – 59
Spindle cell sarcoma NOS	SpCS	2	2:0	17 – 53
Alveolar soft part sarcoma	ASPS	1	0:1	27
Clear cell sarcoma	CCS	1	0:1	28
Malignant mixed tumour	MMT	1	1:0	17

Clinical Breakdown of the 40 paired patients

For some analyses data was restricted to matched patient samples. Below is the breakdown of the clinical characteristics of these samples (see Table 4.2).

- **Age at diagnosis:** ranged between 17 to 89 years old (median 56yrs, mean 51yrs).
- **Gender distribution:** 26 males and 14 females giving a male to female ratio of 1.9:1.

Table 4.2. Clinical characteristics of paired patient samples

Diagnosis	Symbol	No. of patients	Male:Female	Age range
Myxofibrosarcoma	MFS	13	8:5	46 – 89
Synovial sarcoma	SS	7	4:3	22 – 56
Pleomorphic liposarcoma	PLS	4	3:1	31 – 57
Dedifferentiated liposarcoma	ddLPS	3	2:1	71 – 72
Myxoid liposarcoma	MLS	2	2:0	38 – 42
Pleomorphic leiomyosarcoma	pLMS	2	1:1	57 – 78
Extraskeletal myxoid chondrosarcoma	EMC	2	2:0	56 – 63
Spindle cell sarcoma	SpCS	2	2:0	17 – 53
Undifferentiated pleomorphic sarcoma	UPS	1	1:0	29
Malignant peripheral nerve sheath tumour	MPNST	1	0:1	59
Alveolar soft part sarcoma	ASPS	1	0:1	27
Clear cell sarcoma	CCS	1	0:1	28
Malignant mixed tumour	MMT	1	1:0	17

4.2.1 Summary of dataset

This bulk RNAseq dataset comprises 117 tumour samples from 57 different patients, representing 13 distinct soft tissue sarcoma subtypes. This includes 54 pre-radiotherapy biopsies from 54 patients, 55 post-radiotherapy resection samples from 43 patients, 3 recurrence samples from 3 patients, and 5 metastasis samples from 5 patients.

Matched pre- and post-radiotherapy samples are available for 40 patients.

To the best of my knowledge, this represents the largest and most comprehensive soft tissue sarcoma bulk RNAseq dataset with matched human patient samples pre- and post-neoadjuvant radiotherapy. This clinically annotated dataset provides a solid foundation for examining transcriptomic changes induced by radiotherapy and for identifying potential biomarkers of treatment response and disease progression.

4.3 Exploratory analysis of transcriptomic data quality and clustering trends.

To explore the RNAseq data, I initially performed principal component analysis (PCA), a widely used dimensionality reduction technique for assessing variability within high-dimensional data. The PCA plot in Figure 4.1 shows that tumours of certain histological subtypes tend to cluster together. For example, synovial sarcoma samples are grouped in the lower-left area of the plot, while myxofibrosarcoma samples cluster to the right. This clustering indicates that the gene expression profiles within these subtypes are relatively homogeneous. Similar clustering patterns have been reported in previous studies on soft tissue sarcomas, which observed some clustering alongside overlap between histological subtypes (TCGA 2017, Lesluyes, Baud *et al.* 2019). To further visualise this pattern, I used Uniform Manifold Approximation and Projection (UMAP), as shown in Figure 4.2, which supports the PCA findings.

The effect of radiotherapy on the transcriptome, however, appears to be less pronounced than the differences attributable to histological subtype (Figure 4.3). The samples do not cluster distinctly by radiotherapy status, suggesting that the transcriptomic changes induced by radiotherapy are not as significant as the inherent gene expression profiles associated with each histological subtype.

Interestingly, the samples do cluster well by patient ID, showing a high degree of similarity between the pre-radiotherapy biopsy and the post-radiotherapy resection specimen(s) from the same patient. This consistency highlights the reliability and robustness of the FFPE RNA extraction and sequencing methodology, even in the absence of technical replicates.

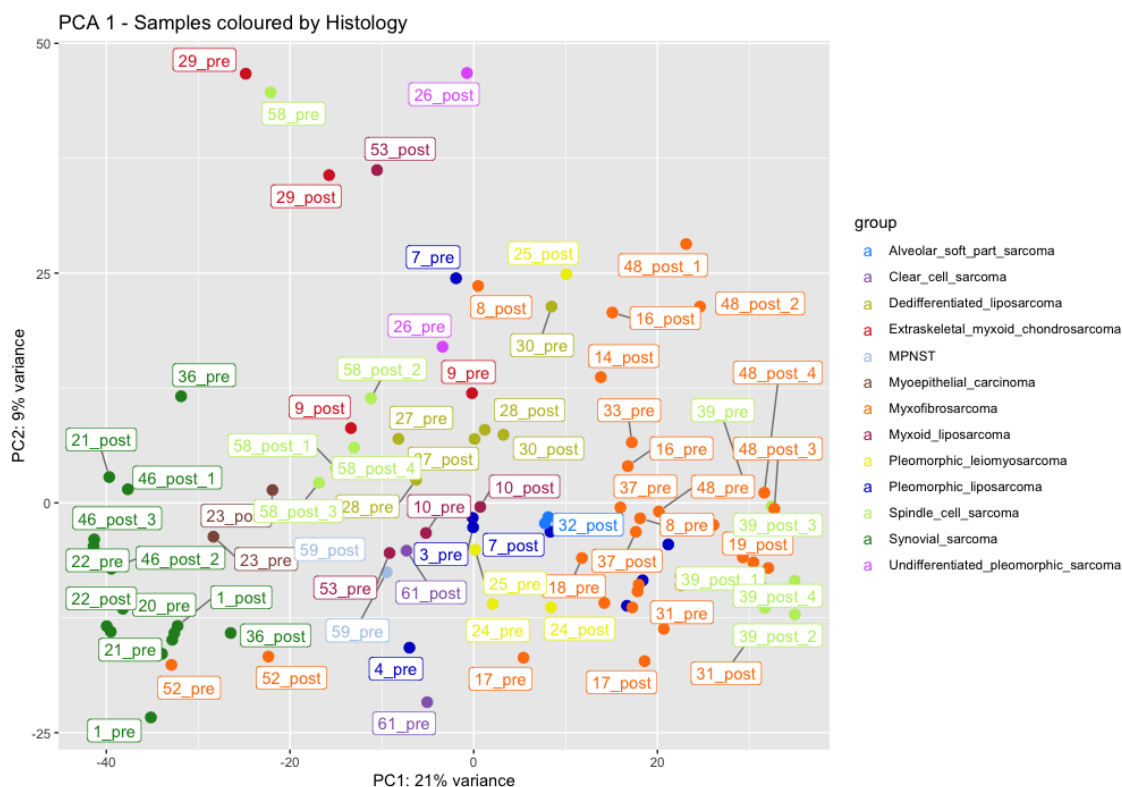


Figure 4.1. PCA plot demonstrating histological subtypes.

Samples are coloured by histological subtype. The sample labels show the patient ID number followed by whether the sample was pre- or post-radiotherapy. Where patients have had multiple post-radiotherapy samples sequenced this is indicated with a number at the end of the label.

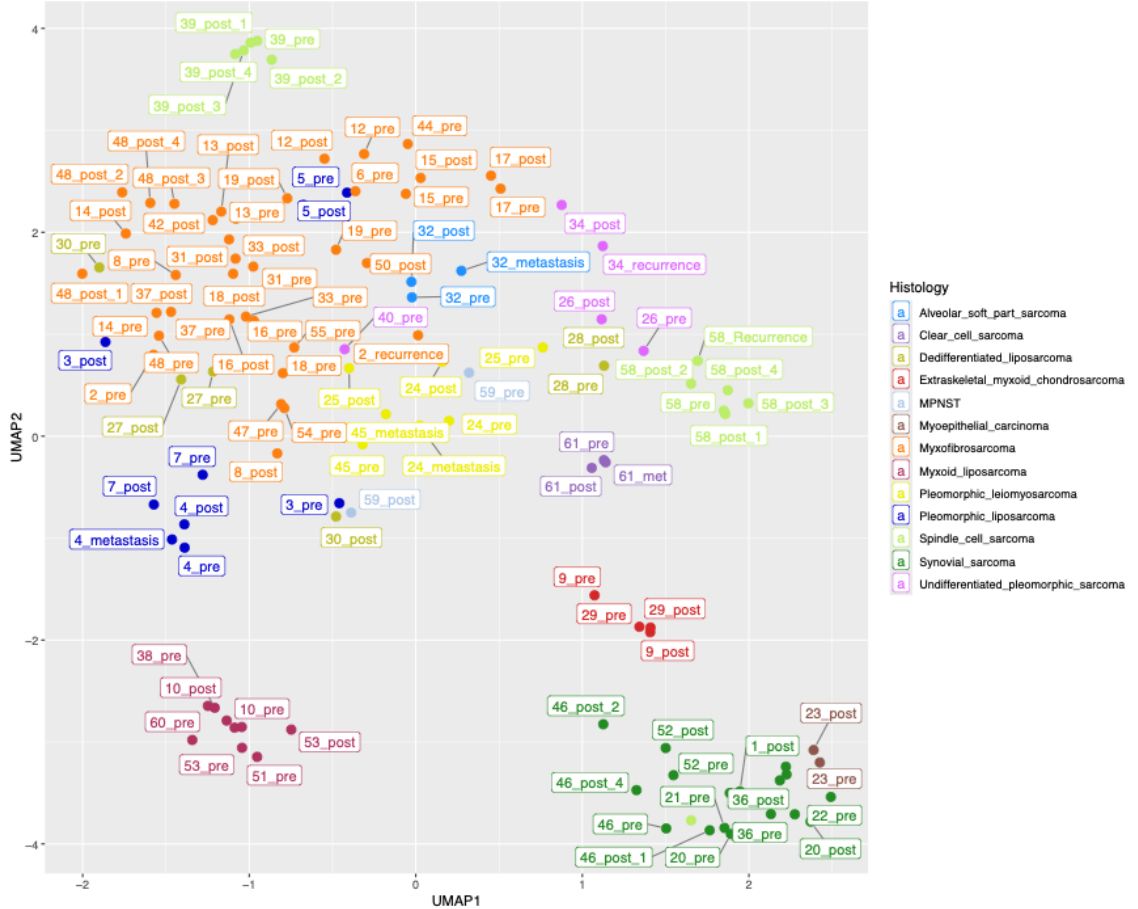


Figure 4.2 UMAP demonstrating clustering of samples according to histological subtype.

Samples are coloured by histological subtype. The sample labels show the patient ID number followed by whether the sample was pre- or post-radiotherapy. Where patients have had multiple post-radiotherapy samples sequenced this is indicated with a number at the end of the label.

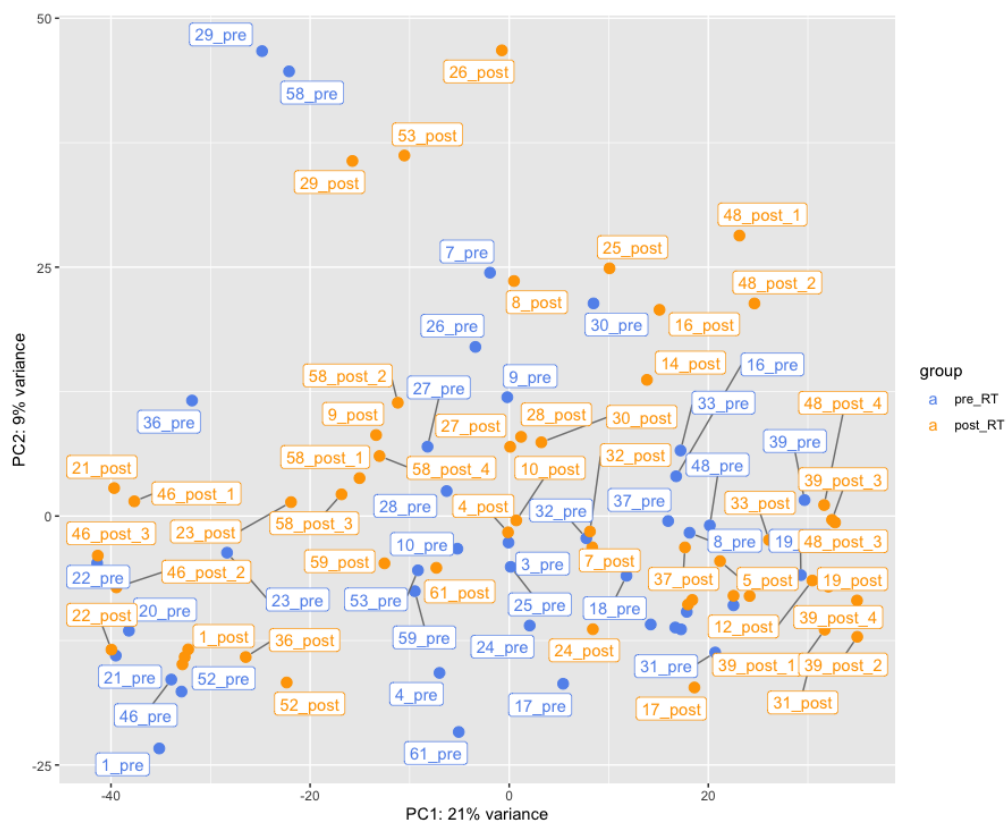


Figure 4.3. PCA plot demonstrating the radiotherapy status of the samples.

Samples are coloured by radiotherapy status. The sample labels show the patient ID number followed by whether the sample was pre- or post-radiotherapy. Where patients have had multiple post-radiotherapy samples sequenced this is indicated with a number at the end of the label.

4.3.1 Multiregional sampling

To ensure confidence in downstream analyses, I examined whether temporal and spatial sampling differences between biopsy and resection specimens would impact the results. To this end, I performed multiregional RNA sequencing on resection specimens from four patients, with each resection sampled from four separate FFPE blocks representing distinct tumour regions. PCA analysis of the RNAseq data showed that samples clustered by their patient ID, without clear separation between pre-radiotherapy biopsies and post-radiotherapy resection specimens (Figure 4.4).

The consistency observed in the transcriptomes of the five samples per patient, across all four patients, suggests limited transcriptional spatial heterogeneity within soft tissue sarcomas at the bulk RNAseq level. This finding supports the reliability of using a single sample for analysis and reinforces the reproducibility of the methodology used in this study.

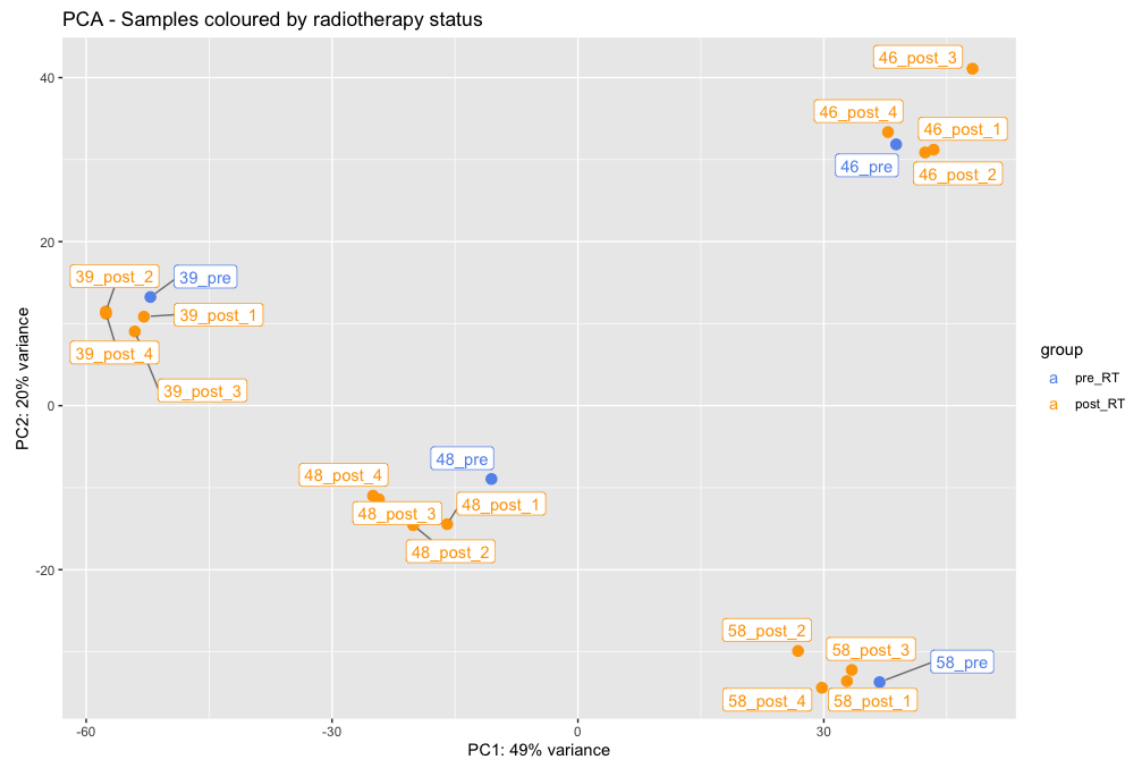


Figure 4.4. PCA plot demonstrating the effect of radiotherapy in patients with multi-region sampling of resection specimens.

Samples are coloured by radiotherapy status. The sample labels show the patient ID number followed by whether the sample was pre- or post-radiotherapy.

The samples in this study were sequenced in 4 batches. Batch 1 was the initial pilot study, batch 2 and 3 contained the rest of the FFPE samples. Batch 4 was the sequencing of the clear cell sarcoma case (patient 61). This was the only sample where the bulk RNA sequencing was performed on fresh frozen tissue. Principal component analysis shows no clustering by sequencing batch (Figure 4.5).

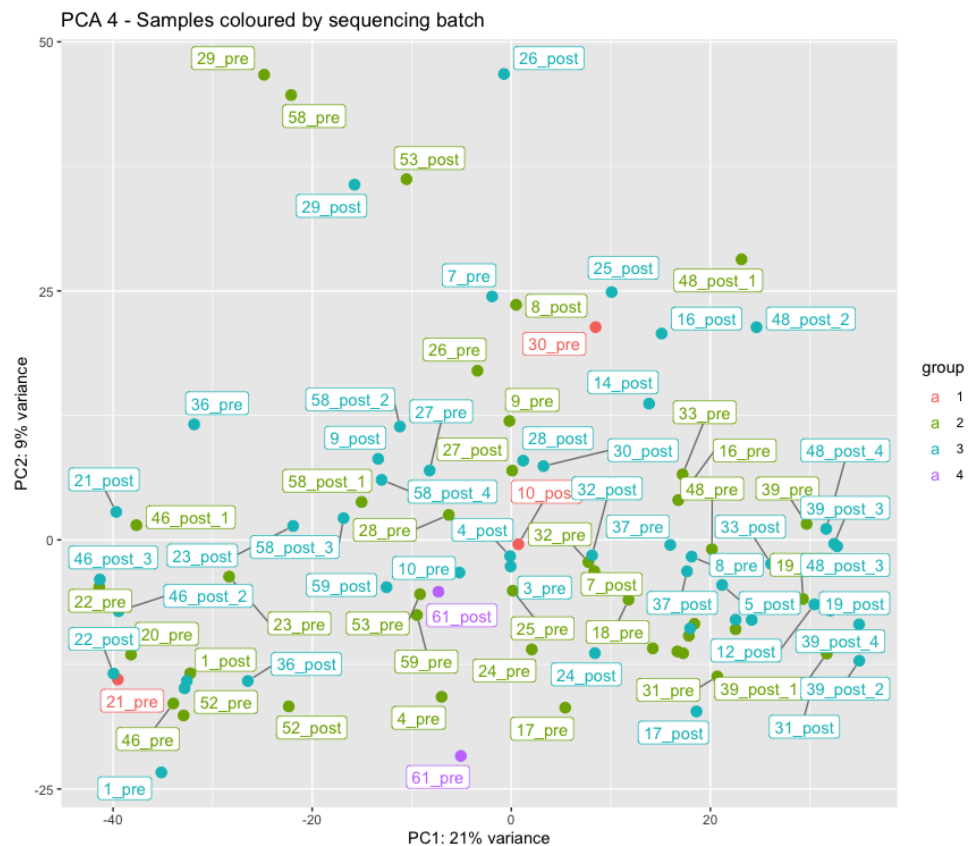


Figure 4.5. PCA plot demonstrating the effect of sequencing batch.

The samples are coloured by the sequencing batch. The sample labels show the patient ID number followed by whether the sample was pre- or post-radiotherapy. Where patients have had multiple post-radiotherapy samples sequenced this is indicated with a number at the end of the label.

4.3.2 Summary of transcriptome clustering analyses

The transcriptome analyses using PCA and UMAP show that while soft tissue sarcomas exhibit diverse gene expression profiles, samples tend to cluster primarily by histological subtype rather than by radiotherapy status. Some overlap between different histological diagnoses is observed.

The consistent clustering of multi-region samples from the same patient underscores the reliability and reproducibility of both the RNA extraction and sequencing methods. This is further supported by the lack of clustering according to sequencing batch.

4.4 Differential gene expression analysis of pre- and post-radiotherapy samples.

To investigate changes in gene expression following radiotherapy, I conducted differential gene expression analysis on 40 patients with matched pre- and post-radiotherapy samples. This analysis excluded 14 patients with only unpaired pre-radiotherapy biopsies and 3 patients with only unpaired post-radiotherapy resections. Using paired samples helps control for inter-patient variability, which enhances the reliability of the results, even if this comes at the expense of a reduced sample size and potentially lower statistical power.

Differential expression analysis was conducted in R using the DESeq2 package. Results are displayed in the volcano plot (Figure 4.6). Applying a threshold of Log_2 fold change > 1 and an adjusted p-value (padj) < 0.05 , I identified 140 differentially expressed genes. Among these, 107 genes were upregulated, and 33 genes were downregulated following radiotherapy. The top differentially expressed genes (though with Log_2 fold change > 1.5) are highlighted in Figure 4.7.

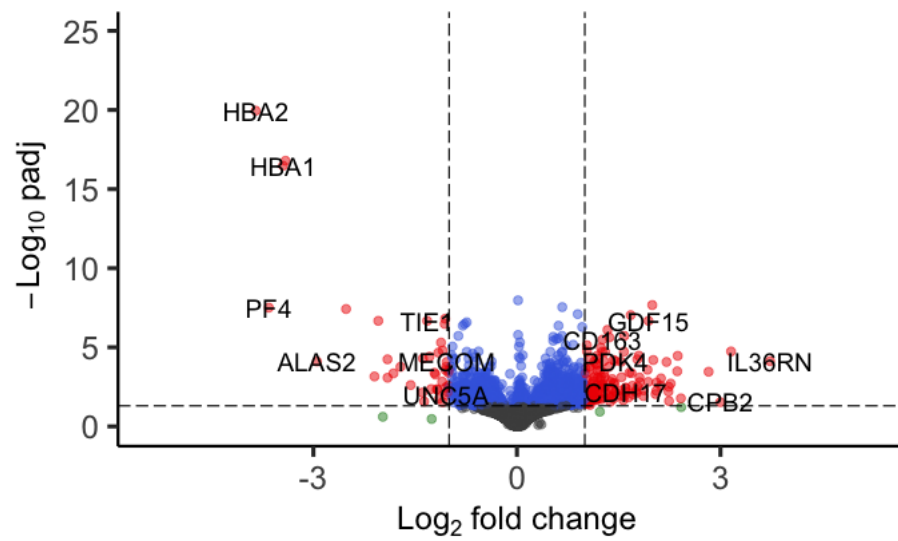


Figure 4.6. Volcano plot of differential gene expression following radiotherapy.

The volcano plot shows changes in gene expression between pre- and post-radiotherapy samples. The dashed vertical lines represent the significance threshold at $\log_2$ fold change of ± 1 , and the horizontal dashed line marks an adjusted p-value (padj) of <0.05 .

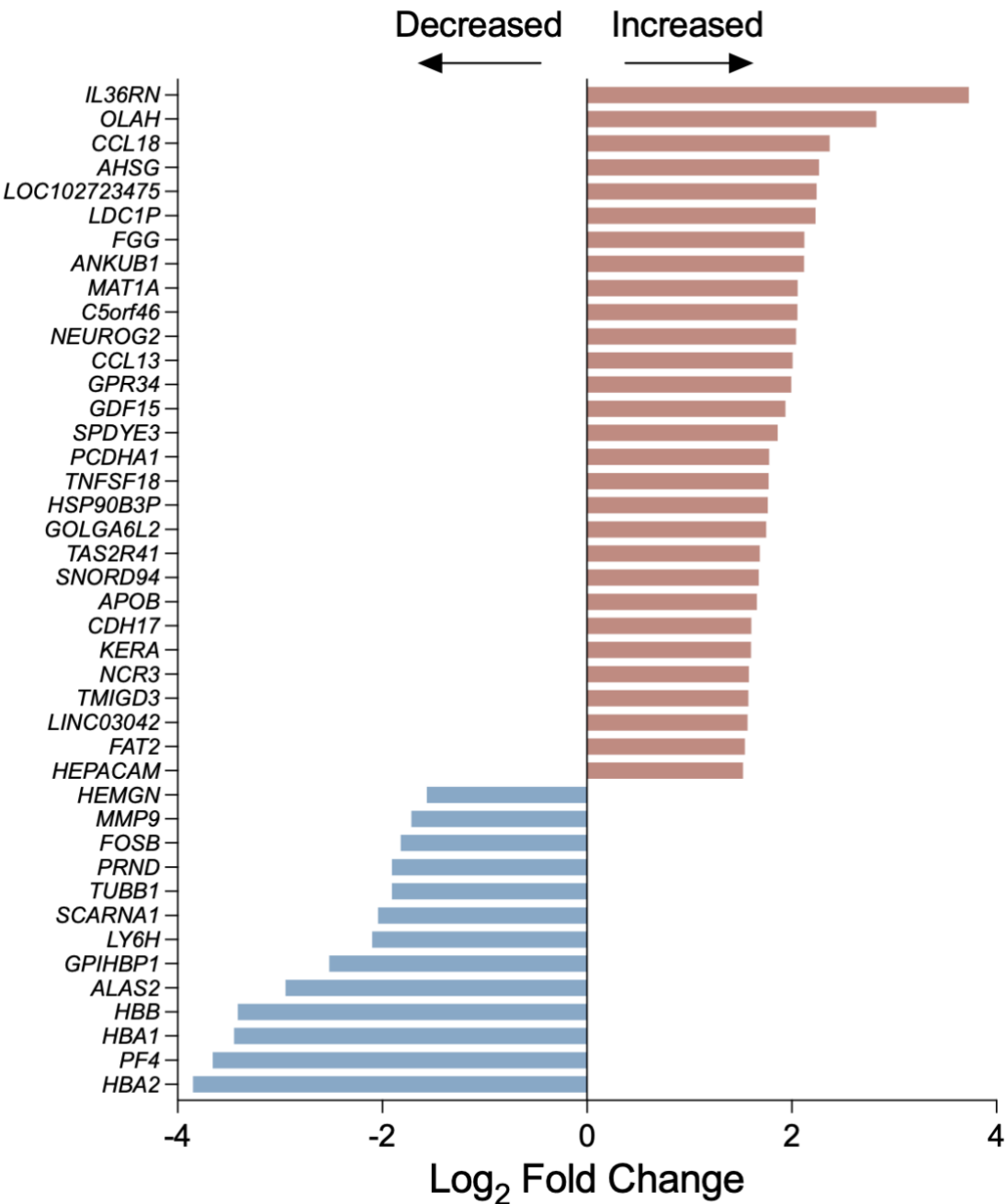


Figure 4.7. Top differentially expressed genes following radiotherapy.

This bar plot displays the top 42 differentially expressed genes with a Log₂ fold change > 1.5, show genes that exhibit the most substantial changes in expression post-radiotherapy.

4.5 Differential gene expression analysis of pre- and post-radiotherapy samples by histological subtype

The results present above in section 4.4 reflect the combined differential gene expression analysis across 13 different histological subtypes. To investigate potential subtype-specific changes and identify genes shared between subtypes, I next performed differential gene expression analyses on patients from subtypes with sufficient paired samples. This analysis included 35 patients across the following subtypes: myxofibrosarcoma (13), synovial sarcoma (7), pleomorphic liposarcoma (4), dedifferentiated liposarcoma (3), myxoid liposarcoma (2), pleomorphic leiomyosarcoma (2), extraskeletal myxoid chondrosarcoma (2), and spindle cell sarcoma (2).

Across this cohort, a total of 1403 genes were differentially expressed following radiotherapy. The majority of these genes (1216 genes, or 86.7%) were subtype specific. Each subtype showed between 68 and 503 differentially expressed genes, with 42% to 83% of these genes being unique to a specific subtype (Table 4.3).

Table 4.3. Numbers of differentially expressed genes following radiotherapy broken down by histological subtype.

Tumour type	Number of differentially expressed genes	Number of unique differentially expressed genes	Unique genes per subtype (%)
Myxofibrosarcoma	503	420	83
Synovial sarcoma	12	5	42
Pleomorphic liposarcoma	117	82	70
Dedifferentiated liposarcoma	68	40	58
Myxoid liposarcoma	445	346	78
Pleomorphic leiomyosarcoma	198	137	69
Extraskeletal myxoid chondrosarcoma	94	50	53
Spindle cell sarcoma	184	136	74

4.5.1 Shared differentially expressed genes between subtypes.

A total of 187 genes were differentially expressed in at least two subtypes. The most commonly observed shared genes were *HBA2* and *HBB*, which were differentially expressed in six subtypes: myxofibrosarcoma (MFS), myxoid liposarcoma (MLS), pleomorphic leiomyosarcoma (pLMS), dedifferentiated liposarcoma (ddLPS), spindle cell sarcoma (SpCS), and extraskeletal myxoid chondrosarcoma (EMC). The next most commonly shared gene was *F13A1*, differentially expressed in four subtypes (MFS, MLS, pLMS, and ddLPS). Additionally, 21 genes were differentially expressed in three subtypes, and 163 genes were differentially expressed in two subtypes. A complete list of shared genes and their respective tumour types is provided in Table 4.4.

The number of unique and shared differentially expressed genes varies across sarcoma subtypes, with relatively few genes overlapping between them. No two subtypes share more than 19 unique genes (notably between MFS and MLS, as well as MLS and pLMS). Most subtypes exhibit largely distinct transcriptional profiles, though some genes are shared across multiple subtypes in small numbers. These patterns are visualised in Figure 4.8, where an upset plot illustrates the extent of gene overlap, emphasizing the predominantly unique transcriptional landscapes of different sarcoma subtypes.

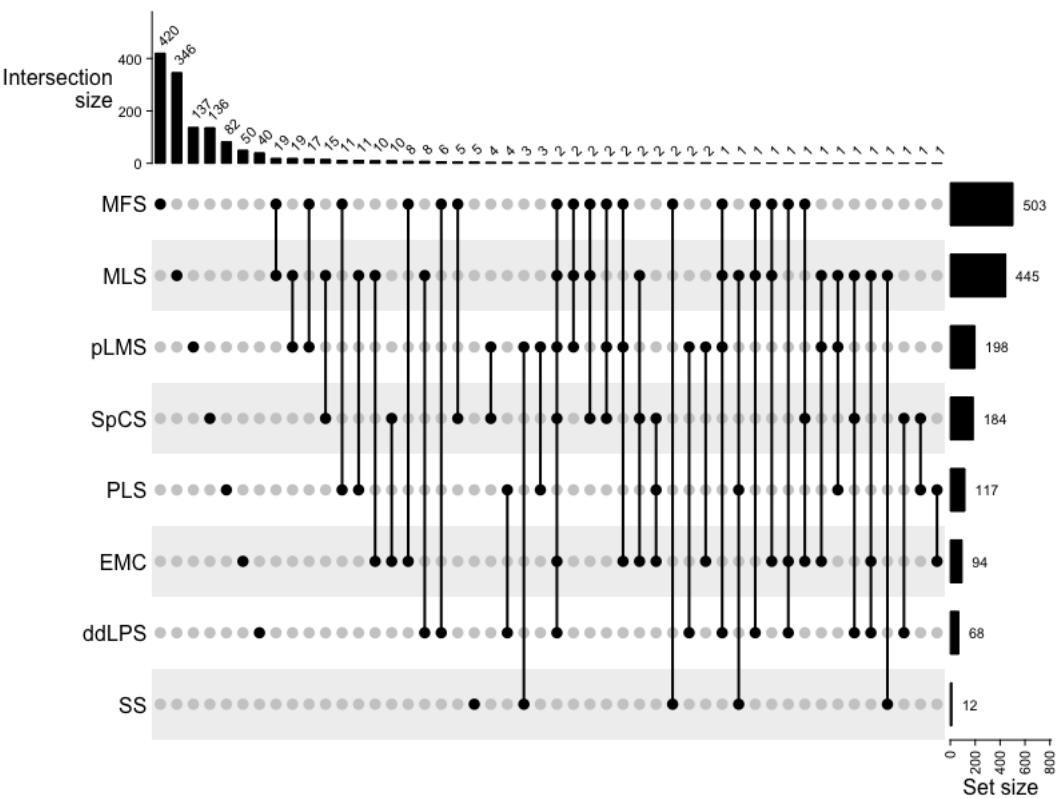


Figure 4.8. Upset plot displaying unique and shared differentially expressed genes following radiotherapy across histological subtypes.

Each subtype’s differentially expressed genes following radiotherapy are shown, highlighting both the number of unique and shared genes across these histologies. Myxofibrosarcoma (MFS), myxoid liposarcoma (MLS), pleomorphic leiomyosarcoma (pLMS), spindle cell sarcoma (SpCS), pleomorphic liposarcoma (PLS), extraskeletal myxoid chondrosarcoma (EMC), dedifferentiated liposarcoma (ddLPS), and synovial sarcoma (SS).

Table 4.4. Differentially expressed genes following radiotherapy that are shared between multiple soft tissue sarcoma subtypes.

Gene Name	Number of subtypes	Subtypes	Gene Name	Number of subtypes	Subtypes
ABC85	2	MFS, pLMS	IFITM1	2	MLS, PLS
ADAMTS16	2	pLMS, PLS	IGHV1-2	2	MLS, pLMS
ADORA3	2	MFS, ddLPS	ITGB8	2	MLS, SpCS
AFAP1L1	2	MLS, SpCS	JPH1	3	MLS, SpCS, EMC
ALAS2	2	MFS, MLS	JUNB	2	MFS, pLMS
ALPK3	2	MLS, SpCS	KCNH8	2	ddLPS, PLS
ANK3	2	MLS, SpCS	KLF6	2	MLS, EMC
ANKRD20A8	2	MLS, pLMS	UNC00632	2	SS, pLMS
ANOS	3	SpCS, EMC, PLS	LOXHD1	3	MFS, MLS, pLMS
APLN	2	MFS, pLMS	MARCO	2	MFS, PLS
APOA2	2	SS, MFS	MBP	2	SS, pLMS
ARL6IP5	2	MLS, pLMS	MCAM	2	pLMS, ddLPS
ASB2	2	MFS, EMC	MDM2	2	SpCS, EMC
B3GAT1	2	MFS, MLS	MECOM	2	MFS, PLS
BGN	2	MLS, pLMS	MEDAG	2	MLS, pLMS
BTNL9	2	MFS, MLS	MT1A	2	MFS, SpCS
C1QC	2	MFS, PLS	MT1E	2	MLS, SpCS
C3	2	MFS, MLS	MT1X	3	MFS, MLS, SpCS
C6	2	MLS, pLMS	MUC12	2	MLS, PLS
C7	2	MLS, pLMS	MUC17	2	MFS, PLS
CA3	2	MLS, EMC	MYH2	2	MFS, EMC
CACNA1E	2	SS, MLS	MYH7	2	MLS, ddLPS
CACNA2D1	2	MLS, EMC	MYO18B	2	SpCS, EMC
CASQ1	3	MFS, MLS, EMC	MYOM3	3	MLS, SpCS, EMC
CCDC102B	2	MFS, PLS	NID2	2	MFS, MLS
CCDC141	2	MFS, pLMS	NMNAT2	2	MLS, ddLPS
CCL13	2	MFS, PLS	NOTCH4	3	MFS, MLS, SpCS
CCL4	2	MFS, MLS	NR4A2	2	MFS, EMC
CCN1	3	MLS, ddLPS, EMC	OR4F8P	2	pLMS, PLS
CCN2	2	MLS, ddLPS	PADI2	2	SpCS, EMC
CD163L1	2	SS, MFS	PAPPA	2	MLS, EMC
CD209	2	MFS, PLS	PCDHA1	2	MFS, MLS
CD44	2	MLS, PLS	PDE4DIP	2	MLS, SpCS
CD53	2	MFS, MLS	PDK4	3	MFS, ddLPS, EMC
CDH5	2	MFS, EMC	PER1	2	MLS, EMC
CDKN1A	2	MLS, EMC	PLA2G2A	2	MLS, pLMS
CEST	2	MLS, ddLPS	PLA2G4B	2	SpCS, EMC
CFHR1	2	ddLPS, SpCS	PLCH1	2	MFS, pLMS
CHI3L1	2	MFS, MLS	PLTP	2	MLS, PLS
CHI3L2	2	MFS, MLS	PRSS12	3	MFS, pLMS, EMC
CLEC3B	2	MFS, EMC	PTGDS	2	MFS, EMC
CLEC4G	2	MFS, PLS	PYCR1	2	pLMS, ddLPS
CLEC7A	2	SS, pLMS	RARRES1	2	MFS, MLS
CLUC5	2	SpCS, EMC	RELN	2	MLS, PLS
CNTN1	2	MLS, pLMS	RGS1	3	MFS, MLS, ddLPS
COL11A1	2	MLS, pLMS	RIMS2	2	ddLPS, PLS
COL4A3	2	MLS, EMC	RN7SL1	2	MFS, PLS
COMMD6	2	MFS, SpCS	RN7SL2	2	MFS, PLS
CP	2	MLS, pLMS	RNASE1	2	MFS, ddLPS
CR1	2	MFS, pLMS	RYR1	2	MLS, SpCS
CRB1	2	MLS, ddLPS	RYR2	2	MLS, PLS
CRIP1	2	MLS, pLMS	RYR3	2	SpCS, EMC
CRK	3	MFS, pLMS, EMC	S100A1	2	MFS, EMC
CXADR	2	MLS, SpCS	SCARNA1	2	MFS, pLMS
DAB2	2	pLMS, SpCS	SCN4A	2	EMC, PLS
DCC	2	MLS, ddLPS	SELENOP	2	MFS, pLMS
DEFP1	2	MLS, ddLPS	SEMA3C	2	MLS, SpCS
DES	2	pLMS, EMC	SEPT5-GP1BB	2	MFS, SpCS
DMD	3	SpCS, EMC, PLS	SERPINA3	2	MFS, MLS
DSP	2	MLS, pLMS	SERPINE1	3	SS, MLS, PLS
DUSP1	2	MLS, EMC	SH3BGR	2	MLS, SpCS
DYSF	2	SpCS, EMC	SHOX2	2	MFS, MLS
EFEMP1	2	MLS, pLMS	SLAMF8	2	MFS, MLS
ESAM	2	MFS, MLS	SLC29A2	2	MLS, SpCS
F13A1	4	MFS, MLS, pLMS, ddLPS	SLCO5A1	2	MFS, MLS
FADS2	2	MFS, ddLPS	SMTNL2	2	MFS, ddLPS
FAM2E3	2	MFS, pLMS	SNORA49	2	MLS, pLMS
FAT2	2	MFS, ddLPS	SNRNP1	2	MFS, EMC
FCGR3A	2	MFS, PLS	SOX18	3	MFS, SpCS, EMC
FGF7	3	MLS, pLMS, PLS	SPAG17	2	MFS, pLMS
FGFR2	2	pLMS, SpCS	SPINK5	2	MFS, MLS
FIUP1	2	SpCS, EMC	SPRY4	2	MFS, pLMS
FLNC	2	SpCS, EMC	STAC3	2	SpCS, EMC
FLT1	2	MLS, EMC	STEAP1B	2	SpCS, PLS
FNDC1	2	MLS, pLMS	STRA6	2	pLMS, SpCS
FOSB	2	MFS, pLMS	SURF4	2	pLMS, EMC
FOSL2	2	MLS, PLS	SYNPO2L	2	MLS, SpCS
FRAS1	2	MFS, MLS	TAGLN	3	MLS, pLMS, EMC
GIPR	2	MLS, SpCS	THBS1	2	MLS, pLMS
GOLGA6L3	2	ddLPS, PLS	TIE1	2	MFS, SpCS
GOLGA6L4	2	MLS, PLS	TNC	2	MLS, pLMS
GIHBP1	2	MFS, MLS	TNFRSF11B	2	MLS, PLS
GPR34	2	MFS, pLMS	TNXB	2	MLS, PLS
GPR89B	2	MFS, SpCS	TRMT5	2	MFS, pLMS
H2AC13	2	ddLPS, PLS	TXNIP	2	pLMS, PLS
H2BC20P	2	MLS, ddLPS	TYRP1	2	MLS, EMC
HBA1	3	MFS, MLS, pLMS	UHRF1	2	MFS, ddLPS
HBA2	6	MFS, MLS, pLMS, ddLPS, SpCS, EMC	UNC5C	2	MFS, pLMS
HBB	6	MFS, MLS, pLMS, ddLPS, SpCS, EMC	USH2A	2	MFS, pLMS
HLA-DQA1	2	MLS, PLS	WDR62	3	MFS, pLMS, SpCS
HSPA2	2	MLS, SpCS	XIRP1	2	MFS, pLMS
HSPB8	2	MLS, SpCS	ZBTB16	3	MLS, ddLPS, SpCS
IFIT2	2	pLMS, SpCS	ZNF728	2	MLS, pLMS
IFIT3	3	MFS, pLMS, SpCS			

Differential gene expression analysis identified 24 genes with shared transcriptional changes across multiple sarcoma subtypes following radiotherapy. These genes were selected based on their differential expression in at least three subtypes, suggesting a common transcriptional response to treatment. The heatmap in Figure 4.9 illustrates these expression changes, highlighting patterns of radiotherapy-induced gene regulation across sarcoma subtypes.

Several genes exhibit consistently increased expression post-radiotherapy, including *SERPINE1*, *RGS1*, *MT1X*, *FGF7*, *IFIT3*, *CRK*, *F13A1*, *CCN1*, and *PDK4*. In contrast, *ANO5*, *LOXHD1*, *NOTCH4*, *DMD*, *SOX18*, *HBA1*, *HBA2*, and *HBB* show consistent decreases in expression across multiple subtypes.

Interestingly, some genes display varied expression changes depending on the subtype, suggesting subtype-specific responses. These genes include *CASQ1*, *PRSS12*, *JPH1*, *MYOM3*, *TAGLN*, and *ZBTB16*.

These findings suggest that while certain genes exhibit a common response to radiotherapy across sarcoma subtypes, others demonstrate subtype-specific expression patterns, highlighting the complexity of sarcoma transcriptomic responses to treatment.

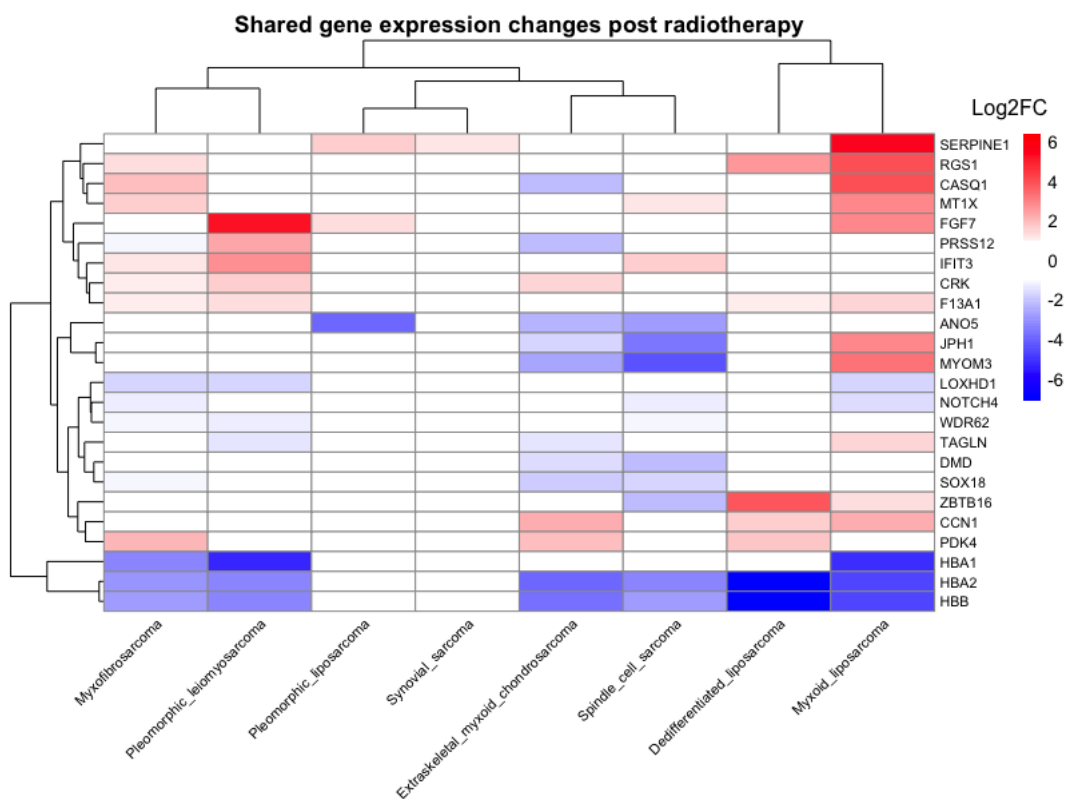


Figure 4.9. Heatmap of differentially expressed genes observed in at least three sarcoma subtypes following radiotherapy.

Genes with Log₂ fold changes of less than ±1 are coloured white to indicate minimal expression change. Red and blue denote upregulation and downregulation, respectively, with intensity reflecting the magnitude of the fold change.

4.5.2 Summary

This section examined gene expression changes following radiotherapy across various sarcoma subtypes. By analysing paired pre- and post-radiotherapy samples, I aimed to control for inter-patient variability and achieve more reliable results. Differential gene expression analysis on 40 paired samples from 13 sarcoma subtypes identified 140 significantly altered genes post-radiotherapy, with the majority (107) showing upregulation.

Subtype-specific analyses revealed considerable variability, with most differentially expressed genes being unique to individual subtypes. However, a subset of genes was consistently altered across multiple subtypes. For instance, *SERPINE1* displayed a consistent increase in expression, while the haemoglobin genes *HBA1*, *HBA2*, and *HBB* consistently decreased across subtypes. Other genes, such as *CASQ1*, exhibited subtype-specific responses, being upregulated in two subtypes but downregulated in another. These findings highlight both shared and subtype-specific mechanisms in response to radiotherapy.

4.6 Gene Set Enrichment Analysis (GSEA) of pre- vs post-radiotherapy Samples

To further investigate the biological processes influenced by radiotherapy, I performed Gene Set Enrichment Analysis (GSEA) on the differential gene expression (DGE) results from paired pre- and post-radiotherapy samples. This analysis aimed to identify enriched cellular pathways involved in the response to radiotherapy across various sarcoma subtypes.

I used the Hallmark gene sets which were developed by the Broad Institute (Liberzon, Birger *et al.* 2015). They represent a curated collection of 50 gene sets that cover essential biological processes and signalling pathways. These pathways are associated with fundamental aspects of cellular function, disease mechanisms, and developmental processes, providing a framework for interpreting large and complex gene expression datasets. By consolidating related genes into distinct biological pathways, the Hallmark gene sets reduce redundancy, making it easier to discern broader biological trends rather than isolated gene-level changes. Key pathways include those related to cell proliferation, inflammation, immune response, metabolism, and DNA repair; all processes potentially impacted by radiotherapy.

The GSEA results revealed both significantly upregulated and downregulated pathways following radiotherapy (Figure 4.10). Notably, several immune-related pathways were significantly upregulated, including interferon gamma response, TNFA signalling via NFkB, and inflammatory response, suggesting a strong immune activation in response to radiotherapy. Other pathways such as allograft rejection and interferon alpha response were also enriched, further supporting an immune-modulatory effect of the treatment.

Conversely, several proliferation pathways were significantly downregulated. These included MYC targets, G2M checkpoint, and E2F targets, indicating a potential reduction in tumour cell proliferation activity post-radiotherapy. This pattern aligns with the therapeutic goal of radiotherapy, which is to damage tumour cells and inhibit their growth.

These findings suggest that radiotherapy induces a complex response in soft tissue sarcomas, characterised by upregulation of immune pathways and downregulation of proliferative pathways.

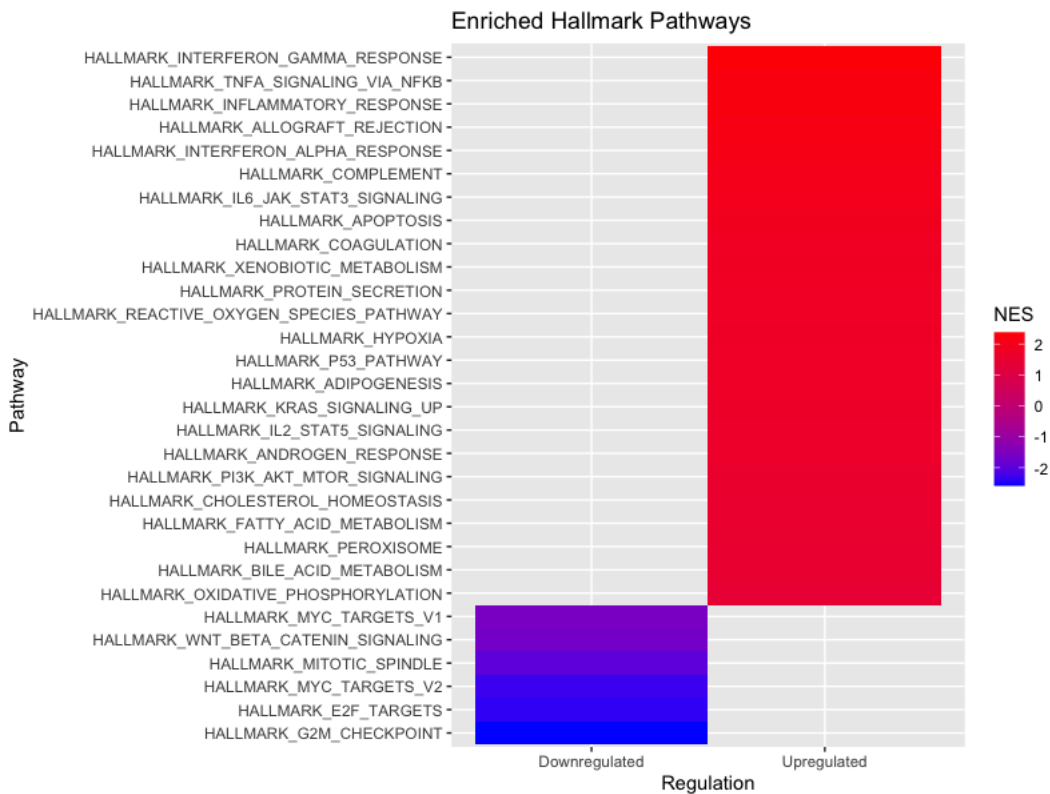


Figure 4.10. Heatmap of enriched Hallmark pathways in post-radiotherapy Samples.

This heatmap displays the normalised enrichment scores (NES) of significantly enriched Hallmark pathways following radiotherapy. The colour intensity reflects the NES, with red indicating upregulation and blue indicating downregulation post-radiotherapy.

4.7 Gene Ontology analysis

To further characterise the biological processes affected by radiotherapy, I performed a Gene Ontology (GO) enrichment analysis on the differentially expressed genes identified between pre- and post-radiotherapy samples (section 4.4). Unlike the Hallmark pathway-focused Gene Set Enrichment Analysis (GSEA), which aggregates lists of genes into broad curated pathways, GO analysis categorises genes based on specific biological processes, molecular functions, and cellular components (Ashburner, Ball *et al.* 2000, The Gene Ontology Consortium 2018).

The GO enrichment analysis of upregulated genes highlighted a strong enrichment of immune and inflammatory processes (Figure 4.11). Key biological processes significantly upregulated included T cell activation, cytokine-mediated signalling, and immune response-regulating signalling pathways. Additional processes, such as leukocyte migration, lymphocyte proliferation, and response to external stimuli, were also prominent among the top upregulated pathways. These results suggest that radiotherapy activates various immune-related processes, potentially enhancing the anti-tumour immune response within the tumour microenvironment.

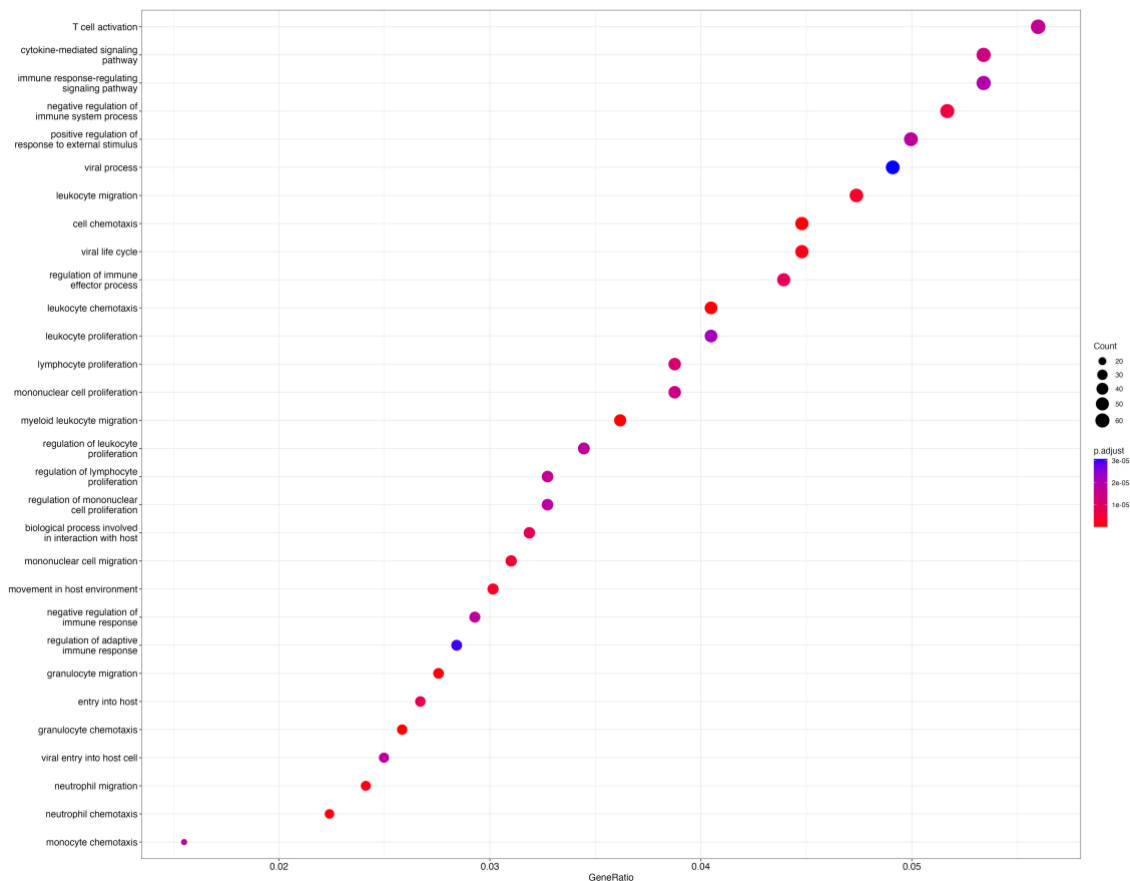


Figure 4.11. Top 30 Upregulated GO Biological Processes following radiotherapy.

This dot plot displays the top 30 upregulated Gene Ontology (GO) Biological Processes identified in post-radiotherapy samples. The size of each dot represents the number of genes involved in the process, while the colour intensity indicates the adjusted p-value, with red representing the most statistically significant processes. The x-axis represents the GeneRatio, defined as the ratio of genes associated with each GO term relative to the total number of upregulated genes.

Conversely, the GO enrichment analysis of downregulated genes revealed a distinct set of biological processes, primarily related to cell cycle regulation and DNA metabolism (Figure 4.12). Significantly downregulated processes included mRNA processing, RNA splicing, chromosome segregation, and DNA conformation changes. Additionally, pathways related to mitotic processes, such as spindle organisation, sister chromatid segregation, and mitotic nuclear division, were among the most significantly downregulated processes. These findings align with the observed reduction in proliferative pathways following radiotherapy seen in the GSEA analysis, suggesting a suppression of tumour cell growth and division.

Overall, these GO analysis results highlight the dual impact of radiotherapy on the transcriptome, with upregulated immune processes suggesting an activated immune response and downregulated cell cycle-related processes indicating decreased tumour proliferation.

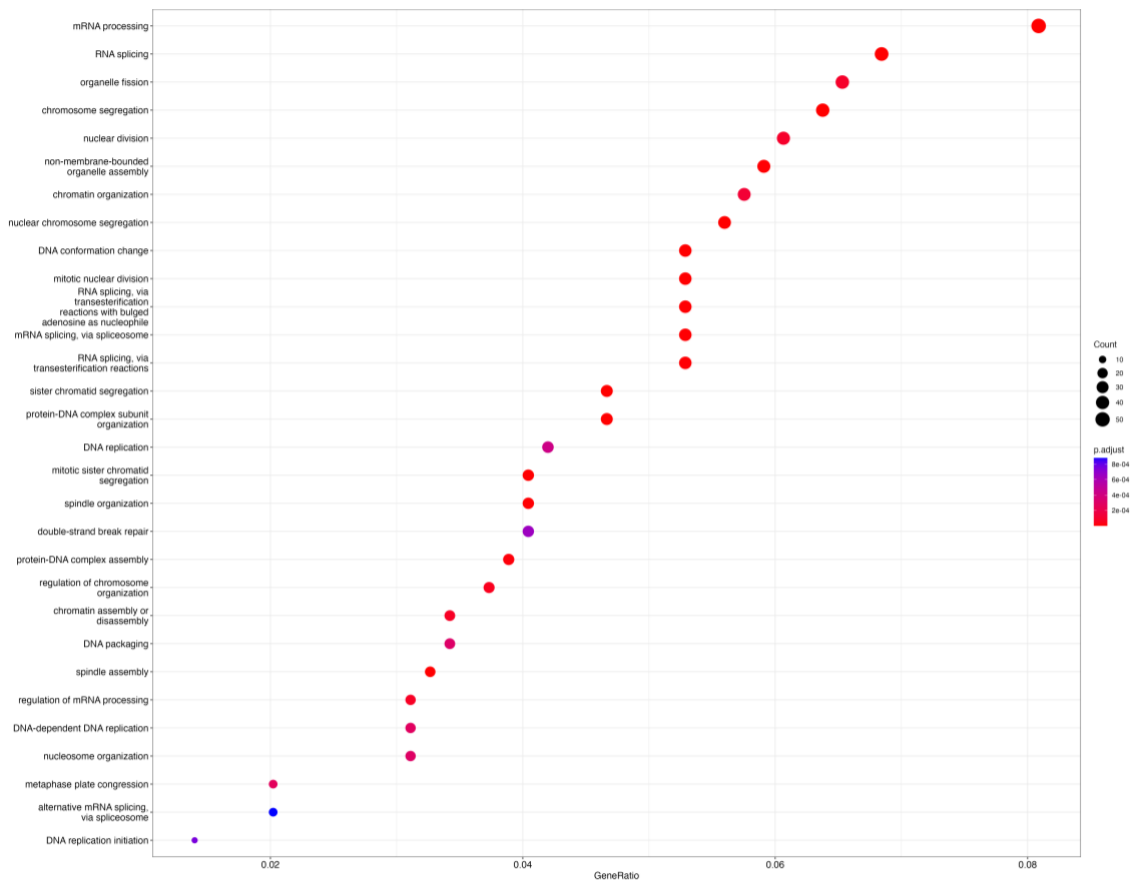


Figure 4.12. Top 30 Downregulated GO Biological Processes following radiotherapy.

This dot plot displays the top 30 downregulated Gene Ontology (GO) Biological Processes identified in post-radiotherapy samples. The size of each dot represents the number of genes involved in the process, while the colour intensity indicates the adjusted p-value, with red representing the most statistically significant processes. The x-axis represents the GeneRatio, defined as the ratio of genes associated with each GO term relative to the total number of upregulated genes.

4.8 PROGENy analysis

To investigate the impact of radiotherapy on the molecular pathways in soft tissue sarcomas, I used the PROGENy (Pathway RespOnsive GENes) R package (Schubert, Klinger *et al.* 2018). PROGENy is designed to infer pathway activity from bulk gene expression data. Unlike GSEA analysis (see section 4.5) which is designed to determine whether a predefined set of genes show a statistically significant differences between two biological states, PROGENy is designed to predict the activity of signalling pathways based on the expression of their downstream target genes, reportedly offering a more direct measure of pathway dynamics.

The development of PROGENy involved compiling a comprehensive list of pathway-responsive genes from experimental data, allowing the method to provide a more accurate and functional interpretation of pathway dynamics based on observed gene expression profiles. This data was curated from a variety of experimental platforms, including 581 perturbation experiments where specific pathways were activated or inhibited. The authors also used microarray gene expression and drug sensitivity data from the Cancer Cell Line Encyclopedia (CCLE), Gene expression data from The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO).

From these sources they derived pathway-specific signatures for 14 major signalling pathways (Androgen, EGFR, Estrogen, MAPK, Hypoxia, JAK-STAT, NF-kB, p53, PI3K, TGF- β , TNF α , TRAIL (apoptosis), VEGF, and Wnt.). These signatures are sets of genes that consistently respond to pathway activation or inhibition. These signatures are then used to score new gene expression data, providing an estimate of pathway activity. The scores are continuous values that reflect the level of pathway activity. Positive scores indicate higher pathway activity, while negative scores suggest reduced activity.

A comparison of pre- and post-radiotherapy pathway activity scores is shown in Figure 4.13. I performed paired t-tests to compare the pathway activity scores between matched pre- and post-radiotherapy samples. Notably, the Androgen ($P < 0.0001$), JAK-STAT ($P = 0.0001$), NF-kB ($P = 0.005$), p53 ($P < 0.0001$), and TNF α

($P = 0.0001$) pathways showed significant increase following radiotherapy. There was a statistically significant decrease in PI3K ($P = 0.00169$) pathway activity.

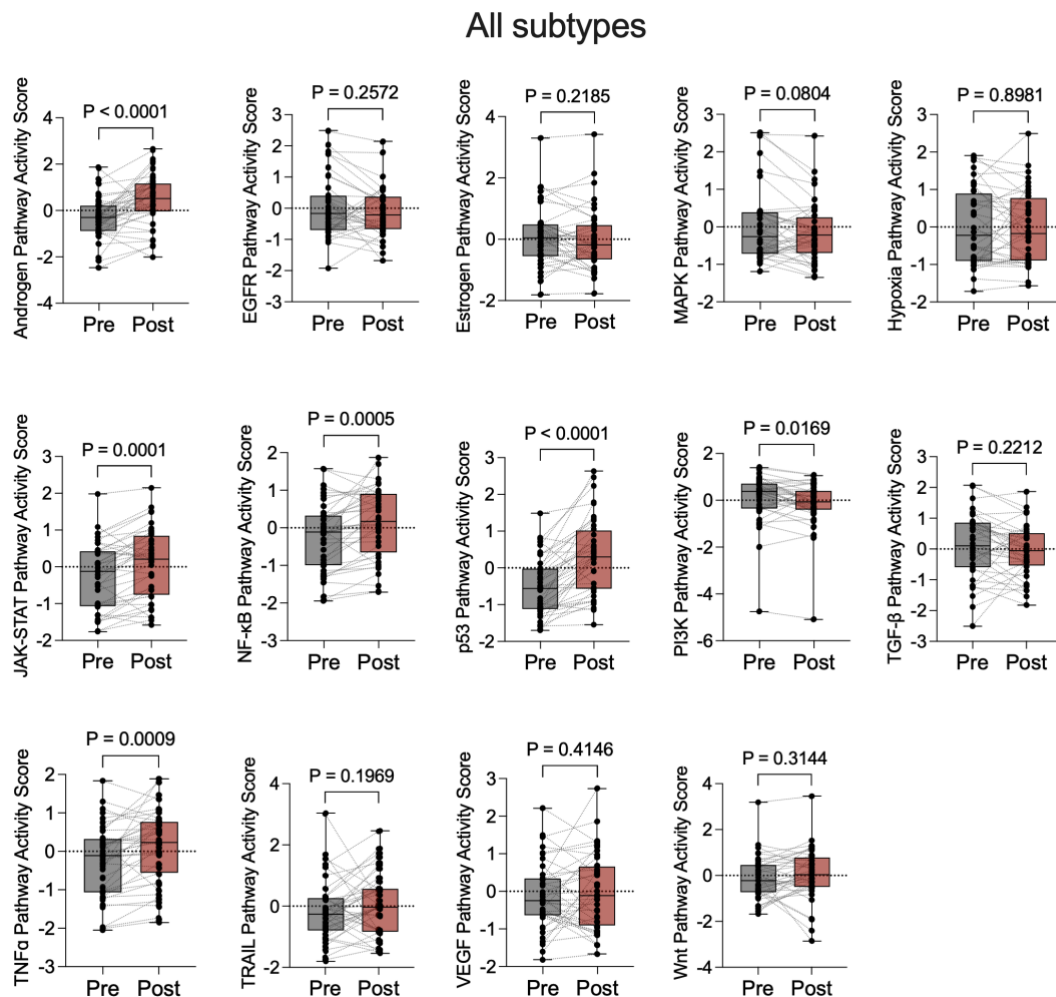


Figure 4.13. PROGENy pathway activity scores pre- and post-radiotherapy

PROGENy pathway activity scores comparing all pre- and post-radiotherapy paired samples (Paired t test).

4.8.1 Subtype specific analysis

I wanted to see if the changes in these pathways were seen across sarcoma subtypes or whether they were subtype specific. For samples where there were enough paired pre- and post-radiotherapy patient samples for statistical testing PROGENy scores were calculated. There were enough patients for the following subtypes: myxofibrosarcoma (n = 14 patients), synovial sarcoma (n = 7), and pleomorphic liposarcoma (n = 4).

When analysed by histological subtype, distinct patterns emerged (Figure 4.14). Statistically significant changes were observed in myxofibrosarcomas with increases seen in the Androgen (P = 0.0002), JAK-STAT (P = 0.0284), p53 (P < 0.0001), NF- κ B (P = 0.0391), TNF α (P = 0.0222), and Wnt (P = 0.0342) pathways. There was a decrease in the PI3K (P < 0.0001) pathways activity score. In the synovial sarcomas only the p53 pathway showed a statistically significant increase following radiotherapy (P = 0.0429). None of the other 14 pathways showed a statistically significant difference in the myxofibrosarcomas, synovial sarcomas, or the pleomorphic liposarcomas.

These findings suggest that radiotherapy causes distinct pathway activity changes in different histological subtypes of soft tissue sarcomas. The significant pathways identified may play important roles in the response to radiotherapy and could serve as potential targets for therapeutic intervention.

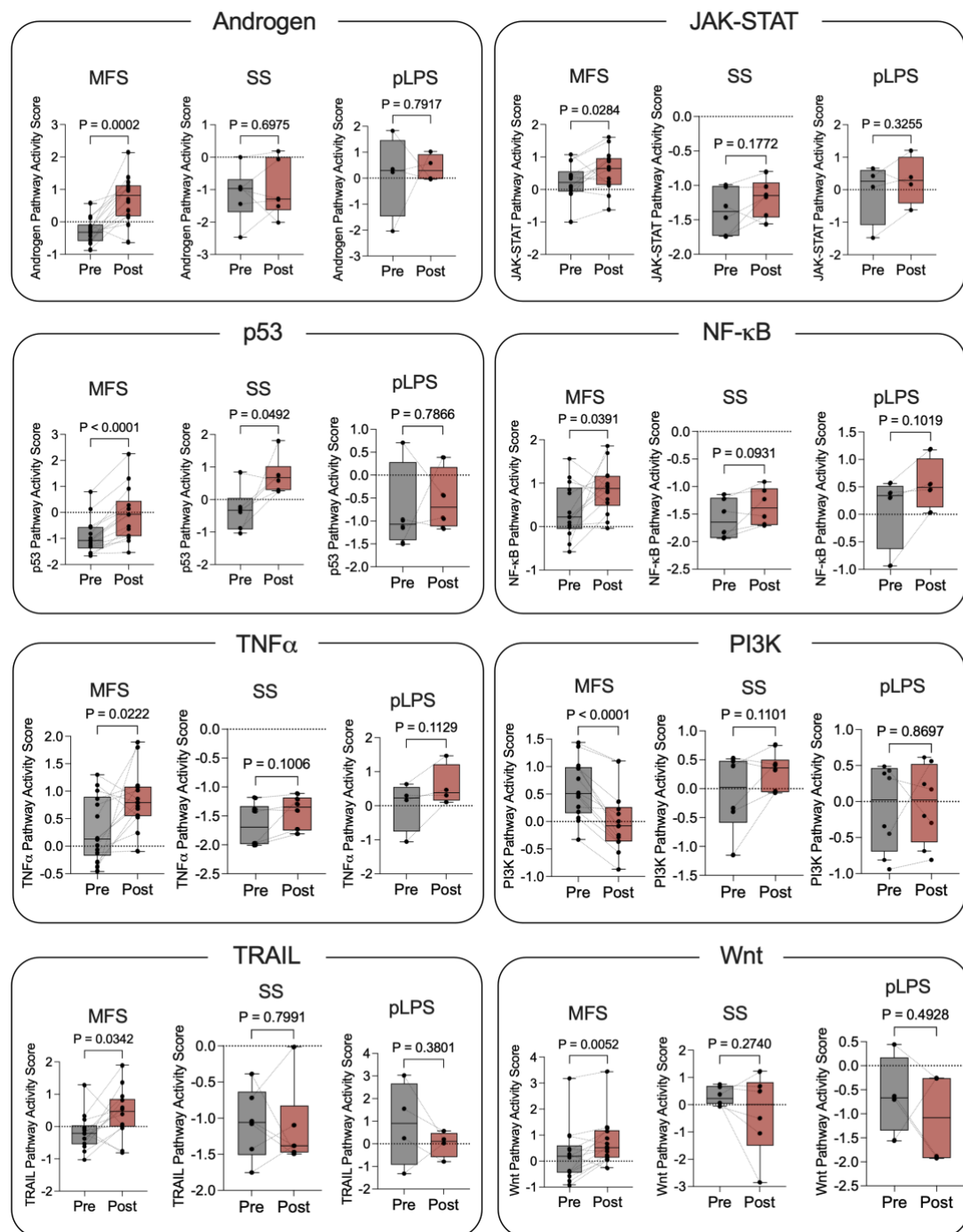


Figure 4.14. PROGENy pathway scores pre- and post-radiotherapy

PROGENy pathway scores for myxofibrosarcoma (MFS), synovial sarcoma (SS), pleomorphic liposarcoma (pLPS) pre- and post-radiotherapy (Paired t test).

4.8.2 Comparison to TCGA sarcoma samples

The sarcoma samples sequenced by the TCGA (TCGA 2017) were treatment naïve (both chemotherapy and radiotherapy). I wanted to see if the PROGENy pathway activity scores in the TCGA data resembled my pre-radiotherapy scores and to see if there was a difference between the TCGA scores and the post-radiotherapy scores.

There were 17 myxofibrosarcoma cases within the TCGA dataset. These were compared to the 14 patients paired pre- and post-radiotherapy dataset (Figure 4.15). There was a statistically significant increase in the Androgen ($P < 0.0001$), TNF α ($P = 0.0362$) pathways. There was a significant decrease in the PI3K ($P = 0.00499$) activity score between the TCGA and the post-radiotherapy samples. There was no statistically significant difference in the remainder of the other 14 pathways.

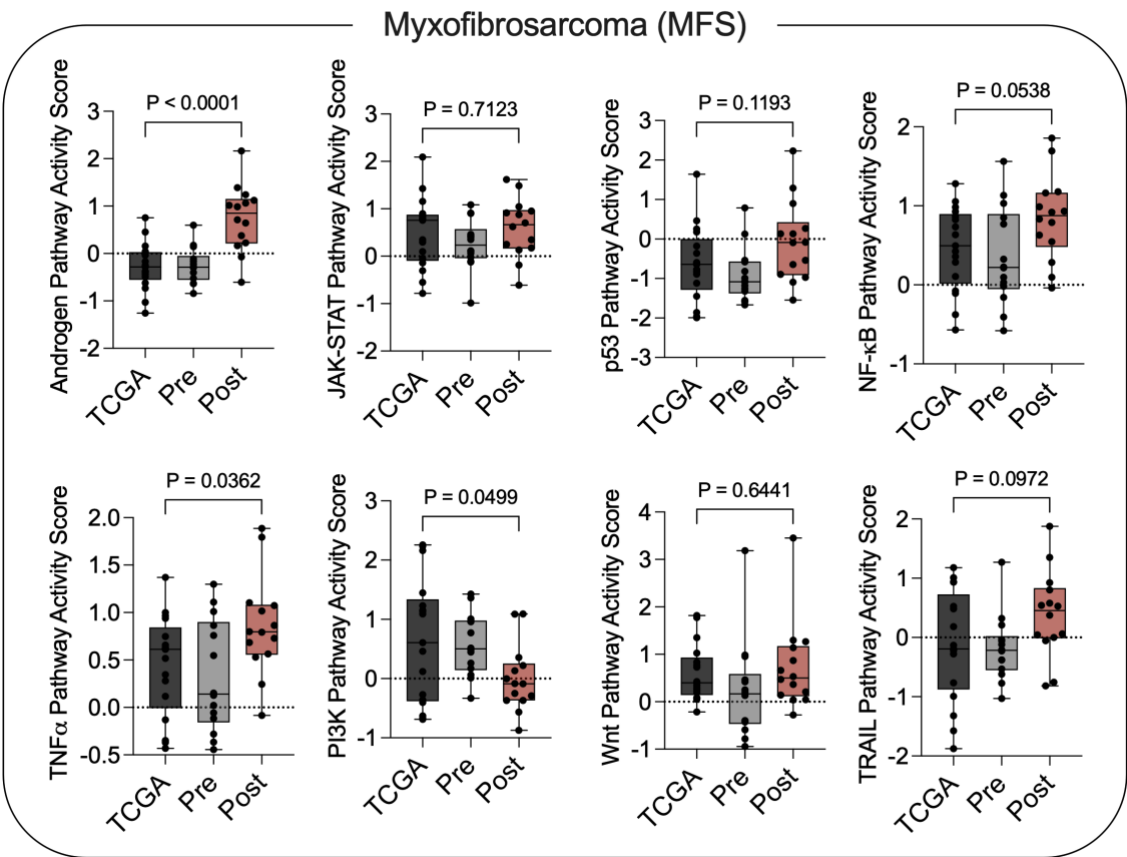


Figure 4.15. Comparison of PROGENy scores in myxofibrosarcomas between the TCGA and the radiotherapy datasets

PROGENy pathway activity scores for myxofibrosarcomas from the TCGA, pre- and post-radiotherapy datasets (Unpaired two-tailed t test between TCGA and post-radiotherapy samples). Only significant pathways

4.8.3 Summary

This PROGENy analysis revealed significant pathway activity changes in multiple soft tissue sarcomas following radiotherapy, suggesting potential molecular mechanisms underlying the response to treatment. Across all samples, radiotherapy significantly increased activity in the Androgen, JAK-STAT, NF- κ B, p53, and TNF α pathways, while PI3K pathway activity decreased.

When examined by histological subtype, distinct patterns emerged. Myxofibrosarcoma showed widespread pathway activation post-radiotherapy, while synovial sarcoma exhibited a notable increase only in the activity of the p53 pathway and pleomorphic liposarcoma showed no changes in any pathways.

Further comparison of my post-radiotherapy myxofibrosarcoma cases with treatment-naïve TCGA myxofibrosarcoma samples highlighted a similar increase in the Androgen and TNF α pathways and a decrease in PI3K activity post-radiotherapy, further underscoring these pathway-specific responses.

These findings overall indicate that radiotherapy elicits heterogeneous pathway responses across different sarcoma subtypes. This emphasises the need for subtype-specific therapeutic strategies. The observed increase in Androgen pathway activity post-radiotherapy raises the possibility that certain sarcomas could become more responsive to anti-androgen therapies as an adjuvant treatment. This suggests a potential avenue for targeted intervention that, while speculative, could be valuable to explore in future studies to enhance patient outcomes.

4.9 Cellular composition analysis using xCell.

Given the prominent immune signalling signatures identified in the GSEA and GO analyses, I conducted an xCell analysis (Aran, Hu *et al.* 2017) to further investigate the cellular composition including those of immune cells within the tumour microenvironment. Given that bulk RNAseq data captures gene expression averaged across all cell types in a sample, the resulting gene expression patterns may reflect not only changes intrinsic to tumour cells but also shifts in the cellular composition of the sample. This can be particularly relevant in the context of immune-related pathways, where an influx or reduction of specific immune cell populations could potentially influence the overall transcriptomic profile.

The primary aim of this analysis was to determine if changes in immune cell populations could be influencing the observed gene expression patterns, particularly those related to inflammation and immune signalling. By analysing the immune composition with xCell, I sought to clarify whether the immune-related pathways seen in the GSEA and GO results were, in part, a reflection of shifts in immune cell abundance rather than solely gene expression changes within tumour cells.

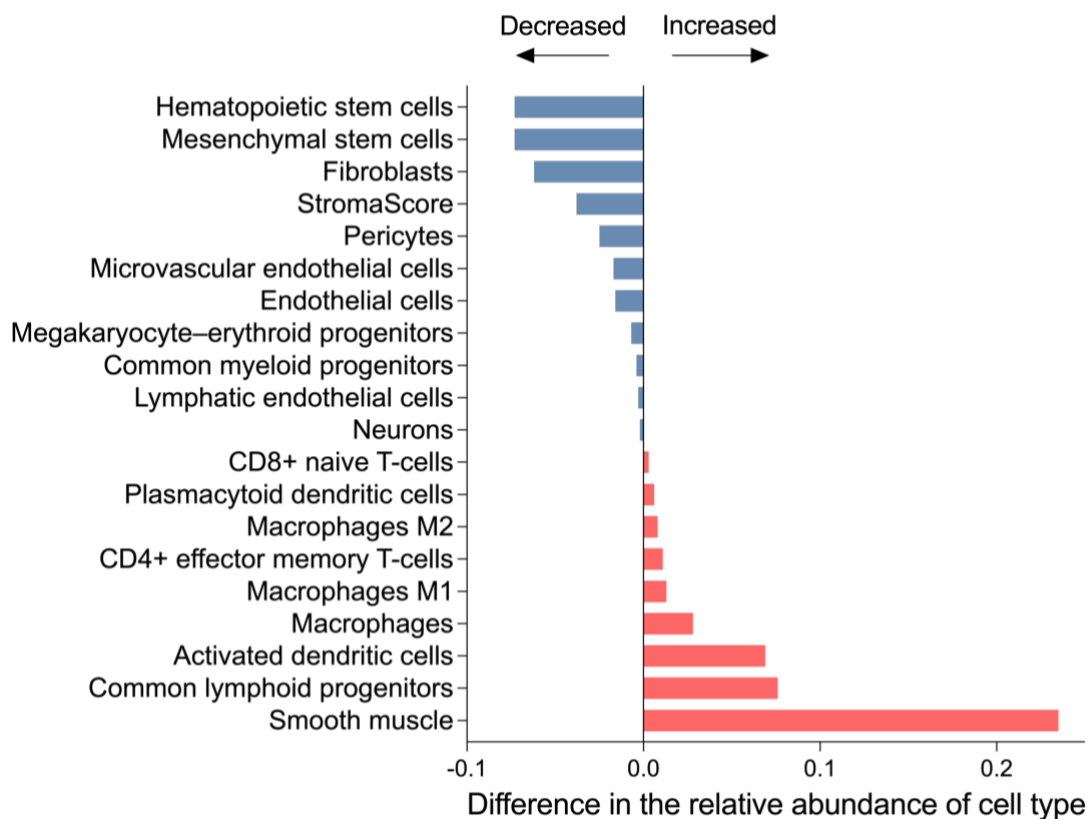


Figure 4.16. Changes in cellular composition identified by xCell analysis following radiotherapy.

This bar plot displays the significantly altered cell populations identified by xCell analysis, estimating the relative abundance of cell types within bulk RNAseq data. "Relative abundance" reflects the proportion of specific cell types within the overall tumour sample. Red bars indicate cell types with increased abundance post-radiotherapy, while blue bars show those with decreased abundance.

The xCell analysis (Figure 4.16) showed significant changes in 20 out of 64 cell types analysed following radiotherapy, highlighting a shift in the tumour's immune landscape. Notably, there was an increase in the abundance of macrophages and T cells, both of which play key roles in the immune response to tumours. The elevated presence of macrophages could indicate an immune response to radiotherapy-induced tumour cell damage. Depending on their polarisation state, these macrophages could either promote an anti-tumour immune response (in the case of pro-inflammatory M1 macrophages) or contribute to immune suppression and tumour progression (in the case of anti-inflammatory M2 macrophages) (Beach, MacLean *et al.* 2022). The increase in T cells, particularly cytotoxic T cells, suggests a potential enhancement of the anti-tumour immune response, which could support the therapeutic effects of radiotherapy (Sharabi, Lim *et al.* 2015).

The increase in activated dendritic cells (aDCs) following radiotherapy suggests that there may be an enhancement of antigen presentation within the tumour microenvironment. Dendritic cells are critical for capturing antigens, processing them, and presenting them to T cells, thereby initiating and shaping the adaptive immune response. The presence of more activated dendritic cells could mean that radiotherapy is promoting a pro-immunogenic environment, where tumour antigens released following radiotherapy are more readily presented to T cells, potentially enhancing their activation and function (Sharabi, Lim *et al.* 2015).

The shifts in immune cell populations seen in Figure 4.16 likely contribute to the upregulation of immune pathways observed in the GSEA and GO analyses, suggesting that some of the immune-related gene expression changes are a reflection of altered immune cell composition within the tumour microenvironment. However, it is important to note that while immune cell composition explains part of the observed transcriptomic shifts, it is unlikely to fully account for all the enriched pathways. The observed gene expression patterns are likely a combination of immune cell changes and intrinsic gene expression responses within tumour and stromal cells following radiotherapy.

The results of this xCell analysis highlights the challenge and complexity of interpreting bulk RNAseq data, where cellular composition changes can confound

differential gene expression analyses. These findings suggest that future studies might benefit from using spatial transcriptomic approaches, which would allow for more precise localisation of immune cells within the tumour tissue. By providing the spatial context, these techniques could enable the exclusion of immune cells from differential gene expression, GSEA, and GO analyses, leading to a more accurate characterisation of tumour-specific transcriptional responses to radiotherapy.

4.10 Discussion of chapter 4

This chapter examined the transcriptional response of soft tissue sarcomas to neoadjuvant radiotherapy, addressing whether shared transcriptional changes contribute to tumour adaptation. Given the absence of recurrent RT-induced driver mutations in Chapter 3, but evidence of genomic instability, this chapter explored whether RT elicits conserved gene expression changes that may be therapeutically relevant.

Transcriptional responses to radiotherapy

Clustering analyses using PCA and UMAP demonstrated that histological subtype exerted a stronger influence on gene expression than RT status, indicating that intrinsic tumour biology remains the dominant factor shaping transcriptomic profiles. However, differential gene expression analysis identified 140 significantly altered genes post-RT (107 upregulated, 33 downregulated), suggesting that despite intertumoural heterogeneity, RT induces specific transcriptional changes. Many of these genes were associated with cell stress, inflammation, and immune activation, while genes linked to proliferation and cell cycle progression were downregulated. This suggests that RT may suppress tumour growth while simultaneously triggering an adaptive immune response to cellular damage.

Pathway modulation and adaptive responses

Gene Set Enrichment Analysis (GSEA) and Gene Ontology (GO) analysis provided functional insights, highlighting upregulation of inflammatory pathways such as interferon signalling, TNF α response, and p53 activation, alongside downregulation of MYC targets and G2M checkpoint genes, suggesting reduced proliferative capacity. This dual effect of RT - enhancing immune-related processes while inhibiting tumour growth pathways aligns with prior studies demonstrating RT-induced stress responses and immune priming.

While individual transcriptional responses varied across histological subtypes, with different sets of upregulated and downregulated genes, many of the affected pathways were shared. This suggests that despite the molecular heterogeneity of STS, RT elicits common biological responses at the pathway level. For example,

inflammatory and immune-related pathways were consistently upregulated, while cell cycle and proliferation-related pathways were downregulated across multiple subtypes. However, histology-specific adaptations were also evident, with dedifferentiated liposarcoma and synovial sarcoma exhibiting distinct transcriptional profiles, reinforcing the importance of considering tumour context in evaluating treatment responses.

To further explore RT-induced pathway modulation, PROGENy analysis revealed increased activity in the Androgen, P53, JAK-STAT, NF- κ B, and TNF α pathways, consistent with stress and inflammatory responses. A notable decrease in PI3K activity was also detected, suggesting that RT may suppress this survival pathway post-treatment. Future research could explore whether targeting these upregulated pathways in high-risk patients in the form of adjuvant treatment could improve outcomes by preventing tumour adaptation or immune escape.

Immune activation: Direct RT effect or secondary adaptation?

xCell immune deconvolution provided further evidence of an immunogenic shift following RT, with significant increases in macrophages and T cells across multiple subtypes. The increase in cytotoxic T cells, in particular, aligns with the immune activation signatures detected in pathway enrichment analyses, suggesting that RT enhances tumour immunogenicity. However, whether these changes reflect direct RT-induced immune activation (e.g., cytokine release following DNA damage) or a secondary tumour-driven adaptation (e.g., immune evasion mechanisms) remains unclear. While RT has been shown to remodel the immune microenvironment in other cancers, its effects in STS remain poorly characterised. Further research is needed to determine whether these immune alterations promote sustained anti-tumour activity or contribute to immune suppression and recurrence.

Conclusions

These findings demonstrate that RT-induced transcriptional changes, while subtle compared to intertumoural differences, reflect key biological adaptations that could influence tumour progression or response to therapy. This chapter highlights how immune activation, stress adaptation, and altered pathway signalling may play a

more significant role in post-RT tumour behaviour than discrete genomic mutations alone.

However, the extent of these transcriptional changes varies between tumours, raising the question of whether they correlate with clinical outcomes. Chapter 5 directly investigates this by stratifying post-radiotherapy samples based on patient response, identifying gene expression patterns associated with RT resistance and progression. By determining whether specific transcriptomic features predict therapeutic response, the next chapter aims to uncover potential biomarkers or intervention targets for optimising STS treatment.

Chapter 5. Transcriptomic determinants of radiotherapy response in soft tissue sarcomas

5.1 Introduction

Neoadjuvant radiotherapy (RT) is an integral component of soft tissue sarcoma (STS) treatment, yet patient responses vary widely, with some achieving durable remission while others develop recurrence or metastasis. Identifying transcriptomic determinants of RT response is critical for refining treatment strategies, guiding post-treatment surveillance, and informing personalised therapeutic approaches.

Several transcriptomic prognostic models have been developed for STS, including CINSARC (Chibon, Lagarde *et al.* 2010), the Genomic Grade Index (GGI) (Bertucci, Finetti *et al.* 2012), and hypoxia-associated signatures (Yang, Forker *et al.* 2017). These signatures have demonstrated prognostic value by distinguishing tumours based on chromosomal instability, histological grade, and hypoxia-related transcriptional activity, all factors associated with metastatic potential and tumour aggressiveness. However, they have primarily been derived from treatment-naïve tumours and do not account for the transcriptional changes induced by RT. Given the significant differences between pre- and post-RT sarcomas observed in Chapter 4, this chapter aims to identify RT-specific transcriptomic determinants of treatment response and progression risk.

By stratifying tumours based on patient outcomes, this analysis seeks to identify gene expression signatures and pathways associated with RT resistance. While prior studies have shown that transcriptomic signatures can predict tumour aggressiveness, they have not specifically addressed whether RT-induced transcriptional changes influence disease progression. Given that RT has been shown to modulate the immune microenvironment in sarcomas, with increased CD8+ T-cell infiltration and altered expression of immune checkpoints (Sharma, Bode *et al.* 2013), the immune landscape may be a key determinant of progression post-RT. Integrating immune deconvolution and pathway enrichment analyses allows for a

deeper exploration of whether immune-related transcriptional programs distinguish responders from progressors.

The ultimate goal of this work is to identify biomarkers that can help and improve clinical decision making. A robust predictive signature could inform follow-up plans, enabling intensified surveillance and adjuvant therapy in high-risk patients while also guiding de-escalation strategies for those unlikely to benefit from or are responder particularly well to RT, thereby reducing unnecessary toxicity. Additionally, this work seeks to uncover potential therapeutic targets for post-RT interventions, with a focus on whether transcriptional programs activated in progressors reveal actionable vulnerabilities that could be leveraged for combination therapies.

Unlike previous signatures, which were largely developed in treatment-naïve STS, this work focuses specifically on post-RT tumour biology, addressing a critical gap in the field. By integrating transcriptomic data with patient outcomes and machine learning-based modelling, this chapter builds upon the findings of Chapters 3 and 4, providing a clinically relevant framework for predicting STS progression post-RT.

Objectives of this chapter

This chapter seeks to answer the following key questions:

- Are there specific gene expression changes associated with disease progression after RT?
- Which biological pathways are enriched in progressors, and do they reflect mechanisms of RT resistance?
- Does the tumour immune microenvironment differ between responders and progressors?
- Can a predictive model be developed to stratify patients based on post-RT transcriptomic profiles?

Structure of this chapter

The chapter begins with an overview of the clinical characteristics of responders and progressors, detailing differences in histological subtypes and disease progression timelines. This is followed by a differential gene expression analysis to identify genes associated with post-RT disease progression. Gene set enrichment and pathway analysis further contextualise these findings, highlighting biological processes linked to progression. Next, I assess differences in tumour immune composition using xCell, examining how immune features may contribute to disease progression. Finally, I present the development of a predictive model for disease progression based on post-RT gene expression signatures, aiming to establish a clinically relevant framework for patient stratification.

5.2 Clinical characteristics of post-radiotherapy responders vs progressors

The post-radiotherapy cohort consists of 43 patients with 52 available RNA samples, distributed across 13 different histological subtypes. Among these, 26 patients showed no evidence of disease progression, with a median follow-up time of 6.8 years (range: 3.8–10.3 years) from the end of intensity-modulated radiation therapy (IMRT). The median time from the end of IMRT to surgical resection for this group was 44 days (range 28 to 60 days).

In contrast, 17 patients experienced disease progression, defined as metastasis or recurrence, with a median time to progression following IMRT of 236 days (range: -31 days to 5.3 years). Of those who progressed, 14 patients ultimately died from their disease, with a median time from IMRT to death of 2.2 years (range: 223 days to 7.2 years). The median time from the end of IMRT to surgical resection for this group was 48 days (range 13 to 84 days).

Age differences between responders and progressors

Responders were generally older than progressors, with a median age of 60 years compared to 46 years. A two-sample t-test confirmed this difference was statistically significant ($p = 0.01$). While the precise role of age in influencing radiotherapy response was not explored in depth in this study, this difference should be taken into consideration in subsequent analyses. Future studies incorporating multivariable models may help determine whether age acts as a confounding factor in response to radiotherapy.

A breakdown of the histological subtypes and their progression status is shown in Table 5.1.

Table 5.1. Summary of histological subtypes and disease progression status in post-radiotherapy cohort.

This table presents the distribution of histological subtypes and disease progression status among 43 patients with post-radiotherapy RNA samples. Progression status is categorised into "Responders" (no evidence of disease progression) and "Progressors" (progression defined as metastasis or recurrence). The number of progressors who died of disease is shown in the final column. ASPS (Alveolar Soft Part Sarcoma), CCS (Clear Cell Sarcoma), ddLPS (Dedifferentiated Liposarcoma), EMC (Extraskeletal Myxoid Chondrosarcoma), MEC (Myoepithelial Carcinoma), MFS (Myxofibrosarcoma), mLPS (Myxoid Liposarcoma), MPNST (Malignant Peripheral Nerve Sheath Tumour), pLMS (Pleomorphic Leiomyosarcoma), pLPS (Pleomorphic Liposarcoma), SpCS (Spindle Cell Sarcoma), SS (Synovial Sarcoma), and UPS (Undifferentiated Pleomorphic Sarcoma).

Subtype	Total Number	Responders	Progressors	Died of disease
ASPS	1	0	1	0
CCS	1	0	1	1
ddLPS	3	2	1	0
EMC	2	2	0	0
MEC	1	0	1	1
MFS	15	13	2	2
mLPS	2	0	2	2
MPNST	1	1	0	0
pLMS	2	1	1	1
pLPS	4	0	4	3
SpCS	2	1	1	1
SS	7	5	2	2
UPS	2	1	1	1

5.3 Differential Gene Expression and Gene Set Enrichment Analysis.

In order to investigate the transcriptomic changes associated with disease progression in post-radiotherapy sarcomas, I performed a comprehensive differential gene expression and gene set enrichment analysis. This analysis sought to highlight both specific genes and pathways differentially regulated in patients with disease progression, potentially revealing mechanisms that contribute to tumour radio-resistance.

Differential expression analysis was conducted in R using the DESeq2 package. Results are displayed in the volcano plot (Figure 5.1). Applying a threshold of Log_2 fold change > 1 and an adjusted p-value (padj) < 0.05 , I identified 586 differentially expressed genes. Among these, 315 genes were upregulated, and 271 genes were downregulated in patients who went on to have subsequent disease progression. The top differentially expressed genes are highlighted in Figure 5.2. An in-depth discussion of the top differentially expressed genes is presented in the next chapter in section 6.3.2.

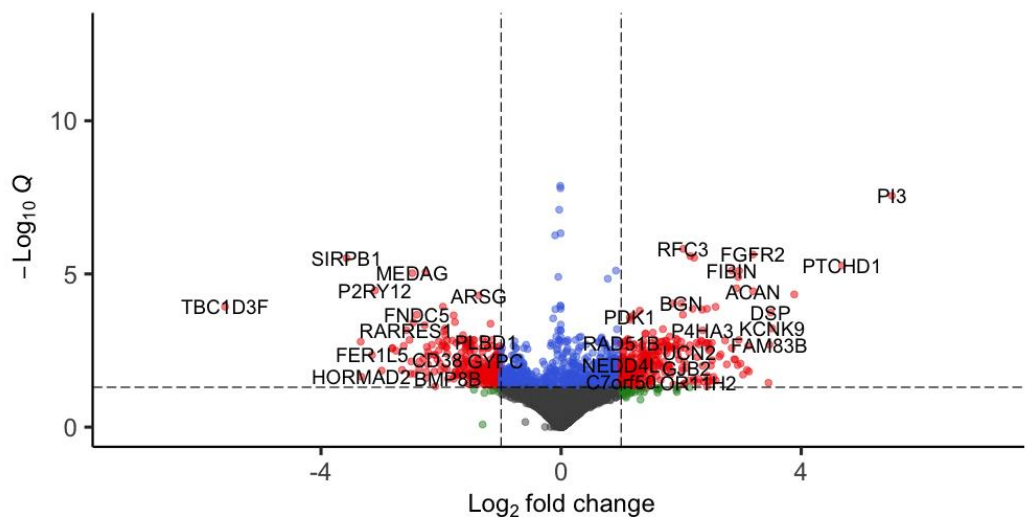


Figure 5.1. Volcano plot of differential gene expression between progressors and responders following radiotherapy.

The volcano plot illustrates changes in gene expression between patients with disease progression (progressors) and those without progression (responders) following radiotherapy. The dashed vertical lines represent the significance threshold at $\log_2$ fold change of ± 1 , and the horizontal dashed line marks an adjusted p-value (p_{adj}) of <0.05 .

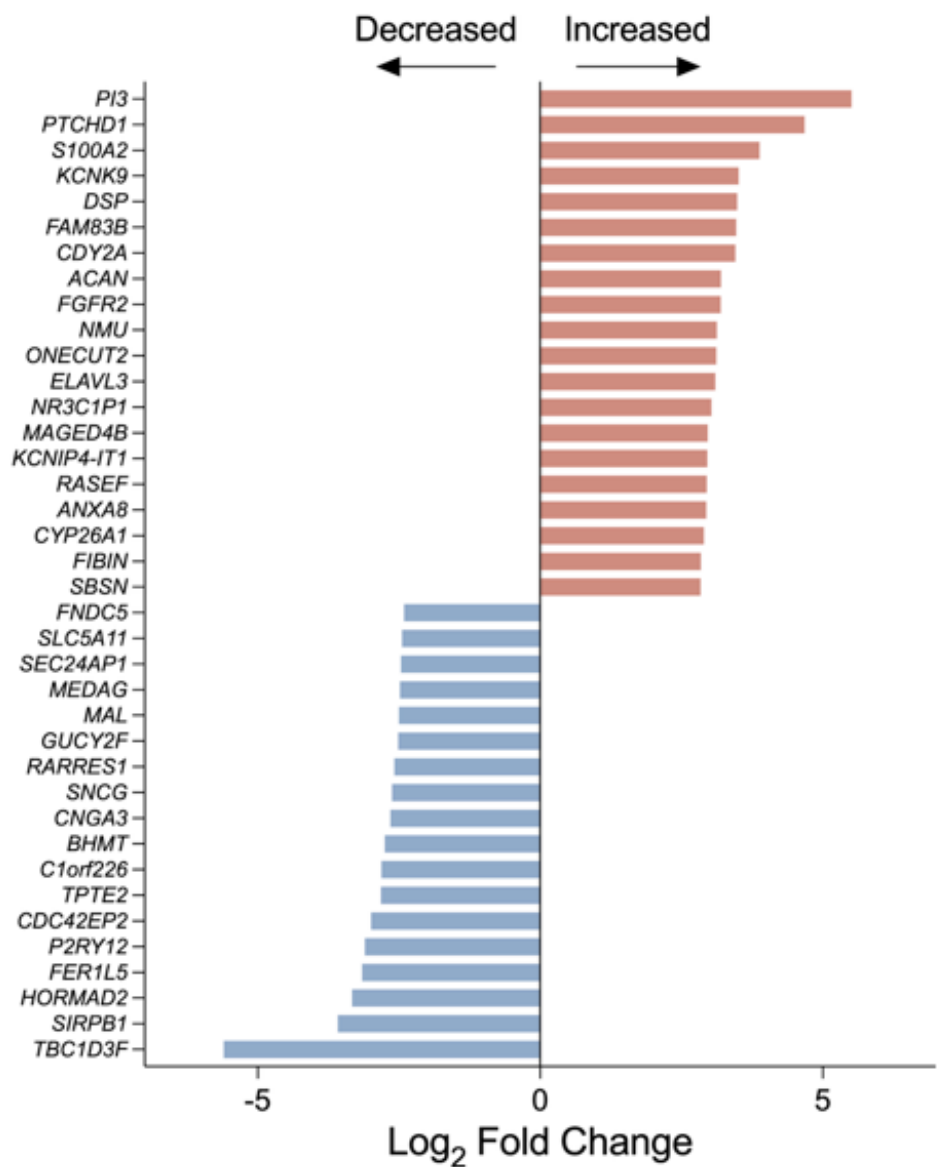


Figure 5.2. Top differentially expressed genes in patients with subsequent disease progression.

This bar plot displays the top differentially expressed genes between patients with disease progression (progressors) compared to those without progression (responders) following radiotherapy.

Gene Set Enrichment Analysis

Subsequently I performed Gene Set Enrichment Analysis (GSEA) to identify enriched biological pathways that distinguish progressors from responders. The hallmark pathways analysis revealed significant enrichment in several pathways (Figure 5.3). Upregulated pathways in progressors included E2F and MYC targets, G2M Checkpoint, and Epithelial-Mesenchymal Transition (EMT), highlighting a trend toward pathways associated with cell cycle progression, proliferation, and metastasis. These pathways are consistent with aggressive tumour behaviours, potentially correlating with the observed clinical progression in these patients (Schulze, Oshi *et al.* 2020, Huang, Hong *et al.* 2022, Chida, Oshi *et al.* 2023).

Conversely, in the progressors, several pathways commonly associated with immune response, inflammation, and cellular stress appear to be significantly downregulated. Notably, pathways such as Interferon Gamma and Alpha signalling and IL6-JAK-STAT3 signalling are reduced, potentially indicating a weakened immune response. Interferons, crucial components of the innate immune system, are typically involved in activating immune cells, promoting antigen presentation, and enhancing cytotoxicity against tumour cells. The downregulation of interferon pathways suggests a diminished capacity for immune cells to recognise and attack tumour cells effectively, potentially allowing for immune evasion by the tumour (Johnson, O'Keefe *et al.* 2018, Jorgovanovic, Song *et al.* 2020, Shi, Yao *et al.* 2022).

In addition, TNF α signalling via NF κ B and the Inflammatory Response pathways are also downregulated, which may indicate suppressed inflammation within the tumour microenvironment. TNF α and NF κ B play pivotal roles in orchestrating immune responses and promoting inflammation, which can limit tumour growth by activating immune cells. Reduced activity in these pathways may create a less hostile environment for tumour cells, reducing immune surveillance and potentially thereby facilitating tumour growth and progression without interference from inflammatory defences (Karin 2006, Wu and Zhou 2010, Alim, Keane *et al.* 2024).

Moreover, the downregulation of Reactive Oxygen Species (ROS) and Apoptosis pathways suggest that progressors may be less susceptible to cell death. The ROS pathway typically generates oxidative stress within cells, a state that can lead to

apoptosis when it becomes excessive. A reduction in ROS and apoptosis pathways might imply that tumour cells in progressors are better equipped to resist the damaging effects of oxidative stress, allowing them to survive longer under adverse the conditions induced by radiotherapy (Kumari, Badana *et al.* 2018).

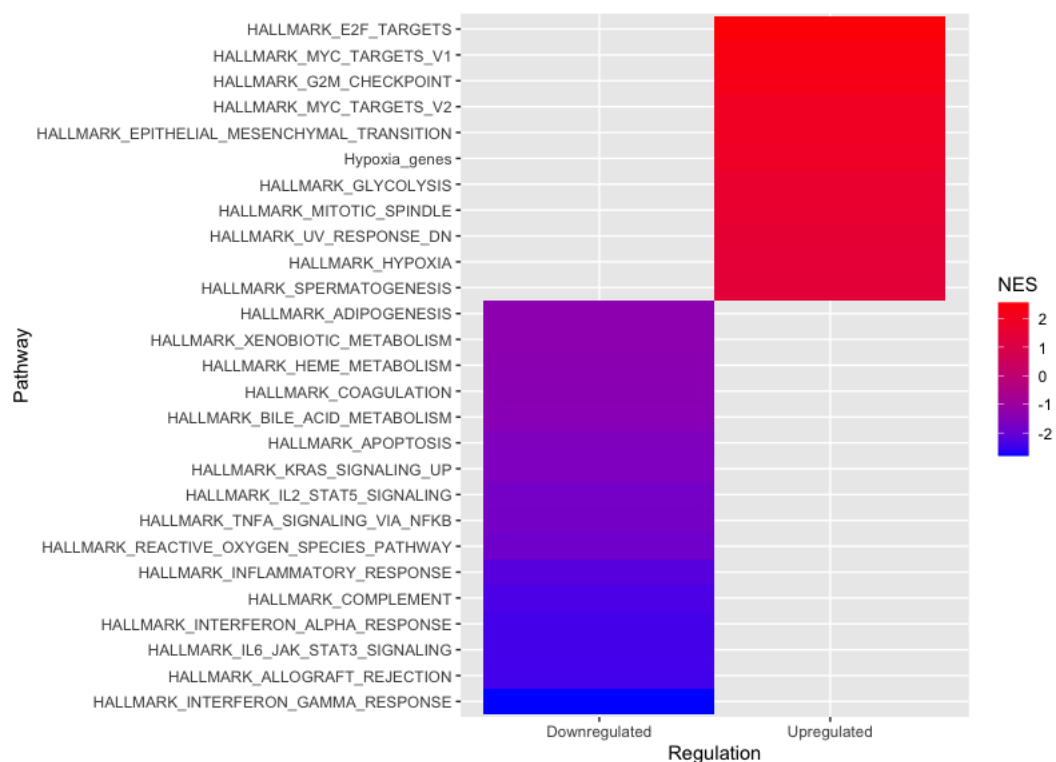


Figure 5.3. Enriched Hallmark pathways in progressors vs. responders following radiotherapy.

This heatmap displays the enriched hallmark pathways identified through Gene Set Enrichment Analysis (GSEA) comparing progressors (patients with disease progression) to responders (patients without progression) following radiotherapy. The colour scale represents the Normalised Enrichment Score (NES), with red indicating upregulation and blue indicating downregulation.

Summary

Overall, these findings suggest that tumours in progressors foster an environment conducive to both aggressive growth and immune evasion. Upregulated pathways, including E2F and MYC targets, G2M Checkpoint, Epithelial-Mesenchymal Transition, and Hypoxia, indicate enhanced cell cycle progression, proliferation, and metastatic potential. Concurrently, downregulated pathways related to Interferon signalling, IL6-JAK-STAT3, TNF α -NF κ B signalling, Inflammatory Response, ROS, and Apoptosis suggest reduced immune activity, inflammation, and apoptosis. Together, these transcriptomic changes suggest possible mechanisms that enable the tumour to evade immune detection and resist radiotherapy.

5.4 PROGENy analysis

To explore differences in signalling pathway activity between patients with disease progression (progressors) and those without progression (responders) following radiotherapy, I performed pathway activity analysis using PROGENy (Schubert, Klinger *et al.* 2018). This pathway-level analysis offers an additional layer of insight beyond gene-level changes identified in the differential gene expression (DGE) analysis and the broader enrichment patterns from gene set enrichment (GSEA) analyses. While DGE focuses on individual gene expression changes and GSEA identifies sets of genes linked to specific biological functions, PROGENy directly infers pathway activity, enabling a functional interpretation of the molecular data. This approach provides a targeted view of pathway dynamics following radiotherapy, aiming to identify differences that could shed light on mechanisms underlying radioresistance and disease progression.

The pathway activity scores for the 14 signalling pathways assessed by PROGENy were compared between progressors and responders, as shown in Figure 5.4.

Hypoxia was the only pathway with a statistically significant difference between the two groups (p-value 0.03). Progressors demonstrated higher hypoxia pathway activity scores than responders, suggesting that hypoxia-related signalling might support survival and progression in the post-radiotherapy microenvironment. All other pathways, including NF- κ B, p53, JAK-STAT, and PI3K, did not reach statistical significance.

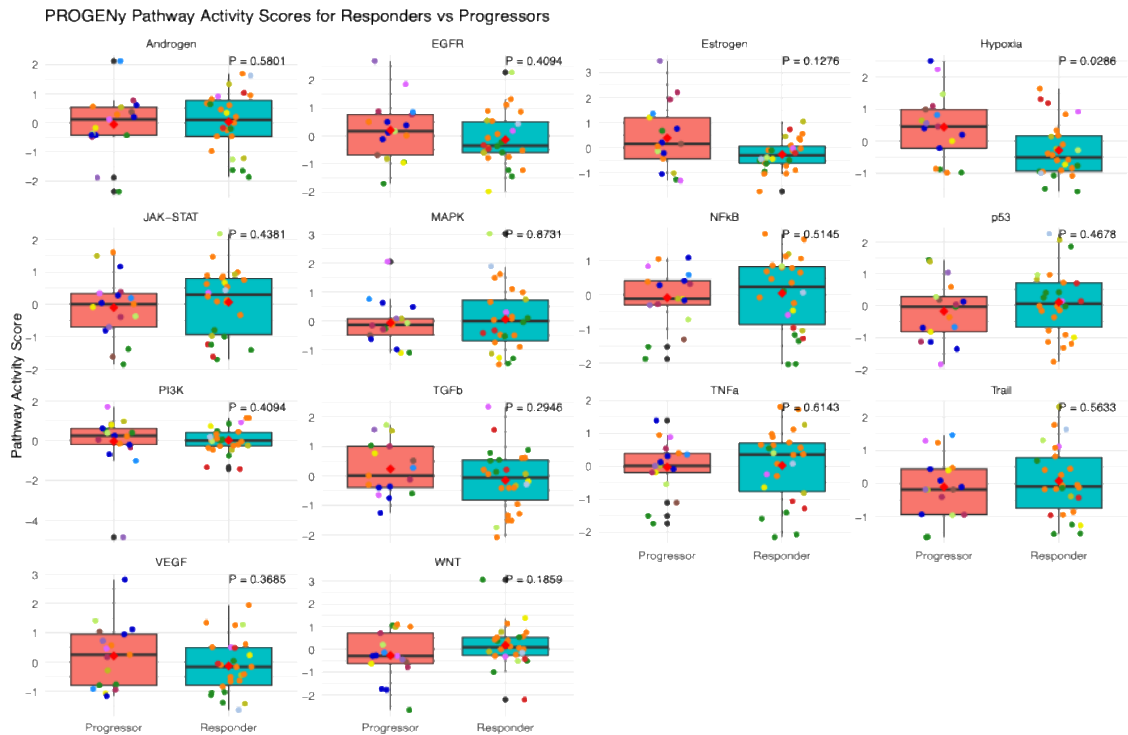


Figure 5.4. PROGENy pathway activity scores for responders vs. progressors.

Box plots showing pathway activity scores for responders and progressors across 14 major signalling pathways, derived from PROGENy analysis. Each box represents the interquartile range (IQR) of pathway activity scores, with the central line indicating the median value. Whiskers extend to 1.5 times the IQR.

5.5 Cellular composition analysis

To conclude the analysis of differential gene expression and pathway activity in progressors post-radiotherapy, an xCell analysis was performed to assess shifts in cellular composition within the tumour microenvironment (Figure 5.5). This analysis builds on findings from Chapter 4, where xCell was used to investigate cellular composition in pre- versus post-radiotherapy samples. Here, the focus is on identifying changes specific to progressors to understand how cellular dynamics might contribute to disease progression following radiotherapy.

The xCell analysis revealed significant alterations in the relative abundance of specific cell types. Notably, smooth muscle cells and class-switched memory B-cells were more abundant in the progressors post-radiotherapy. This increased smooth muscle cell presence might reflect tissue remodelling or fibrotic responses within the tumour microenvironment, potentially contributing to a pro-tumourigenic environment. Class-switched memory B-cells may indicate an adaptive immune response, although their precise role in progression remains to be clarified. In contrast, there was a decrease in CD4⁺ memory T-cells and microvascular endothelial cells (mv Endothelial cells), which could signify an impaired immune response and reduced vascular integrity, possibly facilitating immune evasion and tumour survival.

These changes in cellular composition, coupled with the upregulation of hypoxia-related pathways observed in the PROGENy analysis, suggest that progressors might experience a microenvironment shift towards hypoxic, fibrotic, and potentially immunosuppressive conditions. This altered tumour microenvironment could enhance survival and resilience against radiotherapy effects, emphasising the importance of considering cellular composition in understanding tumour response to treatment.

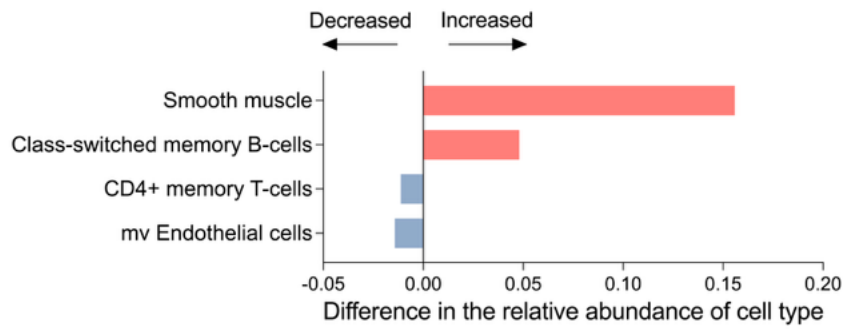


Figure 5.5. Changes in cellular composition identified by xCell analysis in progressors following radiotherapy.

This bar plot displays the significantly altered cell populations identified by xCell analysis, estimating the relative abundance of cell types within bulk RNAseq data. "Relative abundance" reflects the proportion of specific cell types within the overall tumour sample. Red bars indicate cell types with increased abundance in progressors post-radiotherapy, while blue bars show those with decreased abundance.

5.6 Modelling disease progression in soft tissue sarcomas following radiotherapy.

Predicting disease progression in patients with soft tissue sarcomas after radiotherapy could enable personalised follow-up strategies and guide adjuvant treatment decisions to improve patient outcomes. While previous chapters have characterised the genomic and transcriptomic effects of RT, it remains unclear whether post-RT transcriptional profiles can distinguish patients who experience disease progression from those who remain progression-free.

This section describes the development and evaluation of a predictive model for disease progression using bulk RNA sequencing data from post-radiotherapy tumour resection specimens. The goal is to determine whether specific gene expression patterns are associated with disease progression and could be used to refine risk stratification. Identifying transcriptional features linked to poor outcomes may also help prioritise patients for intensified surveillance or future studies investigating post-RT targeted interventions or adjuvant therapy.

Data preparation and normalisation

Normalised gene expression profiles were generated from Salmon quantification using tximport and DESeq2. Counts were normalised with DESeq2 size factors and then mapped from Ensembl IDs to HGNC symbols.

Where multiple Ensembl IDs mapped to the same HGNC symbol, expression was averaged (mean) per gene symbol. Values were $\log_2(x+1)$ transformed to stabilise variance.

For patients with multiple post-RT tumour samples, expression was averaged across samples so that each patient contributed a single profile (one row per patient, one column per gene).

Candidate gene discovery (training cohort only)

To avoid information leakage between the testing and training data, the list of candidate genes for the model was defined using the training set only. Within the training patients, a differential expression analysis using DESeq2 was performed

comparing progressors vs non-progressors (the design included histology as a model covariate). Genes meeting the significance threshold ($p_{adj} < 0.05$) were carried forward as the starting feature set. Test set patients were not used for this step.

Model development

To build a progression classifier, I used a two-stage pipeline:

1. **select features with LASSO**
2. **train a Random Forest** on those features.

All modelling decisions (feature selection and tuning) were made inside the training set only. The testing set was used once at the end for unbiased evaluation.

Feature selection with LASSO

LASSO (Least Absolute Shrinkage and Selection Operator) is a statistical technique used to identify which features in the dataset are most associated with disease progression following radiotherapy. This method helps in selecting a subset of features (in this case genes) that are most predictive of an outcome (Tibshirani 2018). It works by adding a penalty to a regression model that shrinks the coefficients of less important features to zero, effectively removing them from the model. This process helps in preventing overfitting - where the model fits the training data too closely, including its noise and outliers. This can lead to poor performance on new unseen data. By focusing on the most relevant features, LASSO enhances the model's predictive power and ensures better generalisation in new data.

I fit a binomial LASSO model on the training patients and chose the penalty by stratified 3-fold cross-validation optimising AUC (small k was used to keep ~10 samples per fold for stable AUC estimates). Using the *lambda.min* solution, LASSO selected a 20-gene panel:

ADGRG6, ALDH1A2, APLN, CAP2P1, CCNE2, CLEC12B, CYP11A1, FAT2, KRTAP5-7, LINC03033, LRRC1, MAGED4, OR1L8, POLD4, RFXAP, TBC1D3F, TIAM2, TP53TG1, UCN2, ZNF454.

Cross-validation and selection of optimal Lambda

Cross-validation is a method used to evaluate how well a predictive model generalises to an independent dataset. It is an important step because it helps ensure the model's reliability. The process involves partitioning the data into subsets, training the model on some subsets, and validating it on the remaining ones. This is repeated multiple times to ensure that the model performs consistently across different subsets of data. By doing this, cross-validation helps to avoid overfitting, ensuring that the model's performance is not just a result of peculiarities in the training data.

In the context of LASSO, the penalty term applied to the regression coefficients is controlled by a parameter called lambda. The choice of lambda is critical because it determines the strength of the penalty, which in turn influences the number of features (genes) selected and thus model's complexity. Cross-validation is used to find the best lambda value, which strikes the right balance between bias and variance. This optimal lambda minimises prediction error by ensuring the model is neither too simple (high bias) nor too complex (high variance), providing the best trade-off between model accuracy and generalisability.

The selection of which lambda value to use is done after reviewing the cross-validation plot (Figure 5.6). Choosing the minimum lambda (lambda.min) value meant that the model was created using the 20 genes mentioned above.

Alternatively, there is a choice of using the lambda.1se value which represents the largest lambda value that is within one standard error of the minimum cross-validated error (lambda.min). This often results in a simpler model with fewer features (genes), which can help to prevent overfitting and improve generalisation to new data. In this case, using the lambda.1se value, generated a model using fewer genes ($n = 16$) but this performed worse on the data, so I proceeded with lambda.min.

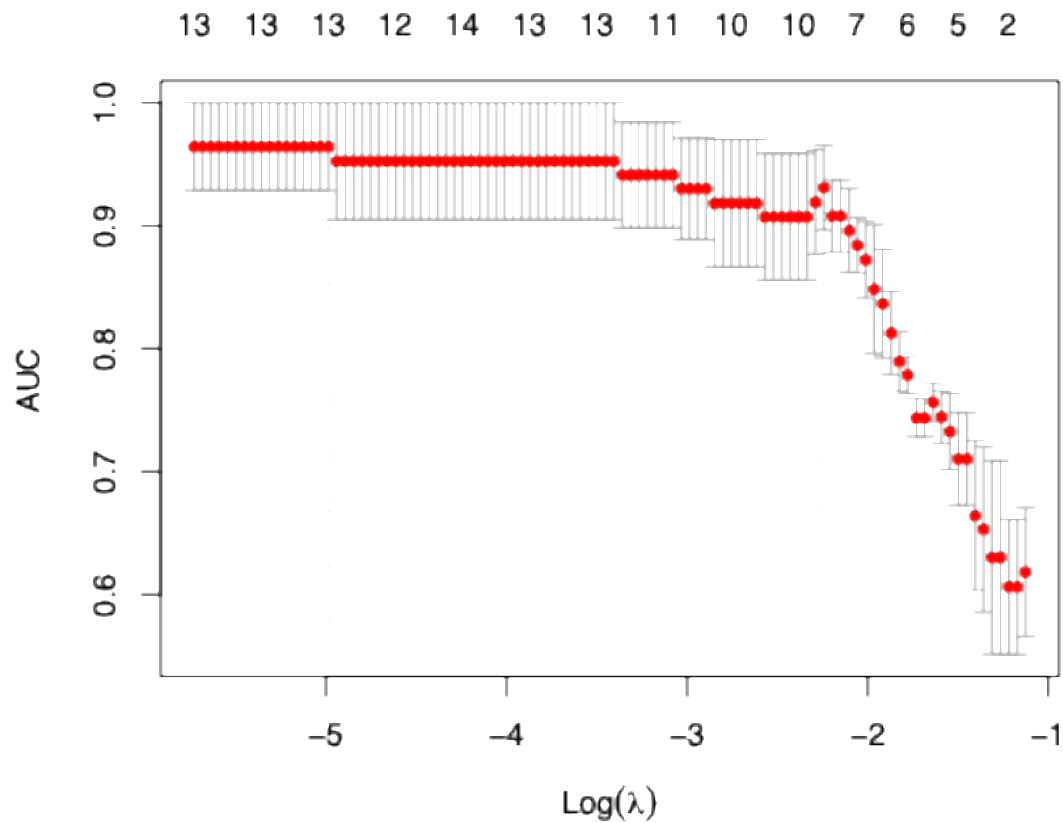


Figure 5.6. LASSO Cross-validation plot.

The x-axis shows $\log(\lambda)$, the regularisation strength. The y-axis shows mean cross-validated AUC (higher is better). Vertical bars are ± 1 SE. The dashed lines mark $\lambda_{\min}$ (highest mean AUC) and λ_{1se} (largest λ within one SE of the best). Numbers above the curve indicate the count of genes at each λ .

Data Splitting

The data was split into training and testing sets using a method called stratified sampling. Stratified sampling maintains the same proportion of progressed and non-progressed cases in both training and testing sets, making the evaluation more reliable. The data was split into training (75%) and testing (25%) sets.

Summary of LASSO model training

The LASSO model was trained on the training set using cross-validation to determine the best lambda value. Cross-validation involved repeatedly splitting the training data into subsets, training the model on some subsets, and validating it on the remaining ones to find the lambda that provided the best performance.

The data partitioning for stratified sampling, the cross-validation, the model training, and the model evaluation was performed using the R package caret.

Generating a Random Forest model

To predict disease progression, a Random Forest model was built using the features (genes) selected by the LASSO model. Random Forest (Breiman 2001) is a very popular machine learning algorithm that is well-suited for high-dimensional data. It is commonly used in classification tasks (e.g., predicting disease presence) and regression tasks (e.g., predicting house prices).

The Random Forest algorithm works by building numerous decision trees and then aggregating the results to arrive at a prediction (Rigatti 2017).

- 1. Building multiple trees:** During training, Random Forest constructs numerous decision trees using different subsets of the training data and features. Each tree is trained on a bootstrap sample (random subset with replacement) and uses a random subset of features for making splits, ensuring each tree is unique.
- 2. Aggregating predictions:** The final prediction combines all individual tree predictions. For classification tasks, this involves majority voting; for regression tasks, it involves averaging predictions. This aggregation enhances model accuracy and reduces overfitting.

I trained a Random Forest classifier on the LASSO-selected genes. Model hyperparameters were tuned using stratified 10-fold cross-validation within the training set, with AUC as the optimisation target and class balance preserved in each fold. After tuning, I refit the model on the full training data and evaluated it once on the held-out test set, converting predicted probabilities to class labels at a 0.5 threshold.

Model evaluation

The final model was applied once to the held-out test set (10 patients: 4 progressed, 6 non-progressed). The confusion matrix is shown in Figure 5.7, and summary performance metrics are reported in Table 5.2 (Sensitivity, Specificity, PPV, NPV, Accuracy, and AUC).

Discrimination was summarised by ROC AUC (Figure 5.8). The model outputs a probability of progression for each patient (0–1). To turn this into a yes/no call for progression, a threshold T is chosen: if $p \geq T$ the patient is labelled “progressed”; if $p < T$ they are labelled “non-progressed”. I used the conventional $T = 0.5$ (i.e., call “progressed” only when the estimated risk is $\geq 50\%$). I also derived a data-driven threshold from the training cross-validation using the Youden index (the point that maximises sensitivity + specificity – 1 on the ROC curve), which gave $T = 0.454$. Both thresholds produced the same classifications on the test set — none of the test probabilities fell between 0.454 and 0.50 — so the results are not driven by a finely tuned cut-point.

		Actual values	
		Progressor	Responder
Predicted values	Progressor	TP = 2	FP = 1
	Responder	FN = 2	TN = 5

Figure 5.7. Confusion matrix evaluating the model on the testing data

This matrix demonstrates the predictions of the model.

Summary of model performance

On the test set the model showed modest performance overall. Discrimination was fair (AUC = 0.67). At a 0.5 threshold it prioritised specificity over sensitivity —correctly identifying 50% of progressors but 83% of non-progressors (PPV 67%, NPV 71%, Accuracy 70%).

In practical terms, as calibrated here the model would miss about half of those who later progressed, limiting immediate clinical utility. Given that the testing set was small ($n = 10$) these estimates are imprecise, but taken together the results are best viewed as proof-of-concept rather than a currently deployable risk tool.

Table 5.2. Performance metrics of the Random Forest progression prediction model

Test	Definition	Results
Sensitivity	The ability of the model to correctly identify patients who have progressed. $TP / (TP + FN)$	50%
Specificity	The ability of the model to correctly identify patients who have not progressed $TN / (TN + FP)$	83%
Positive Predictive Value	The proportion of positive predictions that are actually true positives $TP / (TP + FP)$	67%
Negative Predictive Value	The proportion of negative predictions that are actually true negatives. $TN / (TN + FN)$	71%
Accuracy	The proportion of correct predictions (both true positives and true negatives) $(TP + TN) / (TP + TN + FP + FN)$	70%
AUC	A measurement to quantify the overall ability of the model to discriminate between positive and negative classes.	0.67

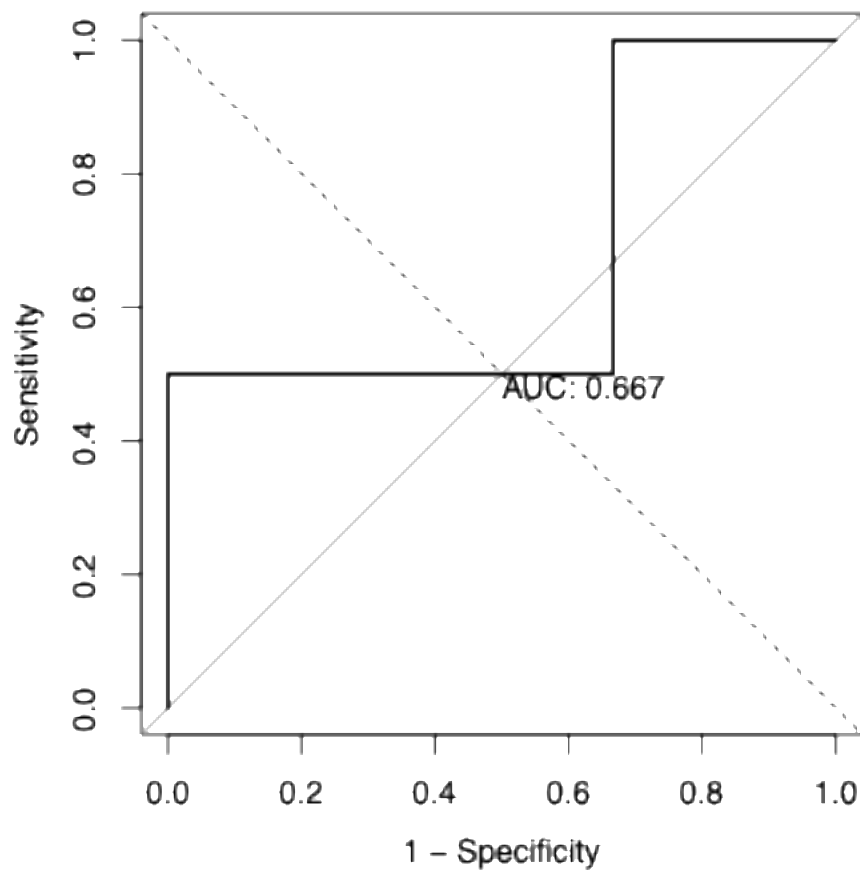


Figure 5.8. ROC curve for the post-RT progression classifier on the held-out test set.

Curve derived from random-forest probabilities on the **independent test cohort** ($n = 10$; progressed = 4, non-progressed = 6). AUC = 0.67. The diagonal line shows chance. Threshold 0.50 (Youden 0.454 gave identical calls).

5.6.1 External validation of model in the TCGA Sarcoma cohort

The classifier was developed on post radiotherapy tumour resections. To understand whether it captures general prognosis or is specifically related to radiotherapy, I applied the model to an independent sarcoma cohort of 206 patients from the TCGA (TCGA 2017) and asked two questions:

1. Is the score prognostic for overall survival across all patients in the cohort?
2. Does its association with outcome differ by receipt of adjuvant radiotherapy (i.e. a predictive signal)?

Data preparation and normalisation

I used TCGA-SARC raw RNAseq counts (GRCh38) and the accompanying clinical meta data. To mirror the training pipeline I retained primary tumour samples only (sample-type code “01” in the TCGA barcode). When a patient had multiple RNAseq aliquots (technical replicates), I combined them to a single patient-level profile (summing counts per gene before normalisation). Counts were then normalised using DESeq2 size factors, Ensembl IDs were mapped to HGNC symbols (and as in my model I averaged expression when multiple Ensembl IDs mapped to the same symbol), and values were subsequently $\log_2(x+1)$ transformed. This produced a final matrix (arranged as patients $\times$ genes) to which I applied the model.

Applying the trained classifier to TCGA SARC dataset

The 20-gene random-forest model derived from the internal training set was carried forward unchanged and applied to TCGA SARC dataset. After harmonising gene symbols, 18/20 features (genes) were available. 2 Genes - *CYP11A1* and *TBC1D3F* were absent. To keep the model structure intact without adding information, these two columns were inferred and filled with a neutral value (the median of the available signature features across patients), which is a conservative choice that tends to dilute rather than inflate signal.

For each TCGA patient the model outputs a probability of progression between 0 and 1 (hereafter called the “RF score”), where higher values indicate a higher predicted risk of disease progression.

Clinical information used for the model

From the available TCGA SARC clinical information I used information on overall survival (“OS_event”) and days (“OS days”) to create the model. The OS survival event was recorded as ‘alive’ or ‘dead’ and of the 206 patients there were 78 recorded deaths.

Radiotherapy exposure was taken from the “radiation treatment adjuvant” field and coded as a two-level factor (RT or No RT). It is important to note that the TCGA RNAseq data is all derived from treatment naïve patients and that the history is of adjuvant and not neoadjuvant radiotherapy.

Results of prognostic association

Using a Cox proportional hazards regression model where $OS \sim RF \text{ score}$, found no evidence that the score is prognostic for overall survival where the Hazard ratio = 0.89 per 1-unit increase in the 0 – 1 RF score (95% CI 0.09 – 8.52, $p = 0.92$). Splitting the patients at the median RF score into 2 group – RF score high and low – did. Not show any statistically significant difference in outcome (Figure 5.9).

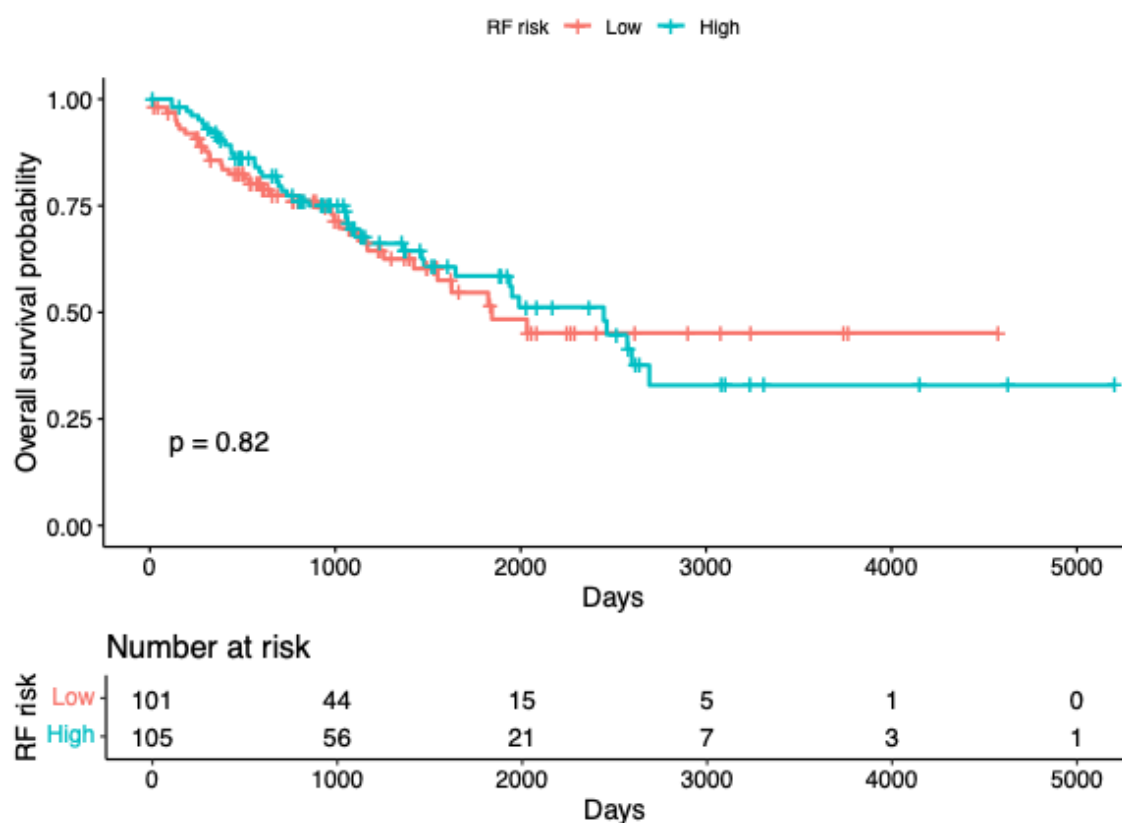


Figure 5.9. Kaplan–Meier overall survival by RF score

Patients were split at the median RF progression probability (High vs Low). Curves show overall survival (days). The number at risk and the log-rank p-value are displayed.

Results of radiotherapy predictive association

To see whether the model scores are different depending on the clinical history of radiotherapy, I tested the model using the interaction: OS ~ RF score × RT_any. This model was restricted to the 197 patients for whom there was recorded information in the “radiation treatment adjuvant” field. Of the 197 patients, 58 patients received adjuvant radiotherapy. Of the 58 patients who received radiotherapy, there were 20 recorded deaths. Of the 139 patients with no history of adjuvant radiotherapy there were 57 deaths.

The results showed that there was no evidence of a differential association by RT status: the interaction HR = 1.32 per 1-SD increase in the RF score (95% CI 0.75–2.34; $p = 0.33$). So, while for each one–standard deviation higher score, the relative

increase in hazard was 32% greater in the RT group than in the No RT group, this difference was not statistically significant.

Looking within groups gave the same message – the No RT group (n=139; 57 deaths) results showed the HR = 0.53 per 1-SD (95% CI 0.04–7.04; p = 0.63) and in the RT group (n=58; 20 deaths) the HR = 8.24 per 1-SD (95% CI 0.05–1315; p = 0.42). The very wide confidence intervals reflect the small RT subgroup and limited number of events.

Summary of external validation results

In TCGA-SARC the score was neither prognostic nor predictive. That said, these null findings should be read with caution. The radiotherapy subgroup was still small (58 patients; 20 deaths), which limits power for interaction testing. There is also an endpoint mismatch: the model was trained to predict progression, whereas TCGA provides overall survival data only. The data types differ (TCGA fresh-frozen RNA-seq vs our FFPE-derived profiles), and 2 of the 20 genes in the signature (*CYP11A1*, *TBC1D3F*) were absent and had to be imputed, which would tend to dilute signal rather than create it.

Most importantly, the model was built on post-radiotherapy resection specimens, while TCGA samples are treatment-naïve. In light of the pre- versus post-RT expression changes shown in Chapter 4, limited generalisability of this post-RT model to untreated samples is biologically reasonable. Overall, the TCGA analysis suggests the signature is not a general prognostic marker, and if it has value, it is likely to only be specific to post-RT biology and should be tested in the future in a matched, post-RT external cohort with progression endpoints.

5.7 Discussion of chapter 5

This chapter investigated transcriptomic differences between soft tissue sarcomas that responded to radiotherapy and those that progressed despite treatment. By stratifying post-RT tumour samples based on patient outcomes, I aimed to identify gene expression changes and pathway alterations associated with treatment resistance. These findings I hope can provide potential biomarkers and therapeutic targets that could inform risk stratification and post-RT management strategies.

Transcriptomic signatures of disease progression

Differential gene expression analysis revealed that progressors exhibited upregulation of pathways associated with cell cycle progression (E2F targets, G2M checkpoint) and epithelial-mesenchymal transition (EMT), suggesting an increased proliferative and invasive phenotype. These findings are consistent with prior studies in STS that have linked tumour aggressiveness to transcriptomic programs associated with genomic instability and dedifferentiation (Chibon, Lagarde *et al.* 2010). In contrast, responders showed higher expression of immune-related pathways, particularly interferon signalling and inflammatory response genes, reinforcing the hypothesis that immune activation may contribute to durable disease control post-RT.

The observation that immune pathways were downregulated in progressors is of particular interest, given the emerging role of the immune microenvironment in RT response. While RT can enhance tumour immunogenicity, leading to T-cell recruitment and immune priming, some tumours may evade immune surveillance post-treatment through mechanisms such as PD-L1 upregulation or recruitment of immunosuppressive cells (McKelvey, Hudson *et al.* 2018, Wang, Lynch *et al.* 2024). These findings suggest that further investigation into the role of immune modulation in post-RT STS is warranted.

Pathway enrichment analysis reveals hypoxia-driven radioresistance

Pathway-level analysis using PROGENy identified hypoxia as the only significantly enriched pathway in progressors, supporting the well-established role of hypoxia in radioresistance and tumour progression. Hypoxia-related gene expression

signatures have been linked to aggressive tumour phenotypes, poor prognosis, and treatment resistance in multiple cancers (Aggerholm-Pedersen, Sørensen *et al.* 2016, Yang, Forker *et al.* 2017). These signatures highlight the role of HIF1 α -mediated transcriptional programs, which drive tumour adaptation through metabolic reprogramming, angiogenesis, immune evasion, and enhanced DNA damage repair. Furthermore, CAIX (carbonic anhydrase IX), a well-characterised endogenous marker of tumour hypoxia, has been validated as a prognostic immunohistochemical biomarker in STS and could potentially serve as a screening tool to stratify patients (Forker, Gaunt *et al.* 2018).

The enrichment of hypoxia in progressors within this study suggests that STS tumours that fail RT may exhibit a transcriptomic profile consistent with previously identified hypoxia signatures. This finding aligns with evidence that hypoxic tumours are more likely to metastasise and resist treatment. Notably, hypoxia-driven gene expression changes in STS have not been extensively explored in the post-radiotherapy setting, reinforcing the novelty of these results.

Targeting hypoxia-mediated resistance mechanisms represents a promising therapeutic strategy. HIF1 α inhibitors, metabolic interventions, and hypoxia-activated prodrugs have shown preclinical efficacy in reversing hypoxia-driven resistance (Bui, Nguyen *et al.* 2022, Kao, Bai *et al.* 2023). However, their role in the post-radiotherapy setting in STS remains unexplored. Given the increasing availability of functional imaging techniques such as hypoxia PET scans (Gouel, Decazes *et al.* 2023) and molecular biomarkers, future studies should investigate whether integrating hypoxia-targeting therapies post-RT could improve patient outcomes.

These findings underscore the importance of considering tumour oxygenation status when evaluating response heterogeneity in STS. By identifying hypoxia-driven transcriptional changes in progressors, this study highlights a potential biomarker for treatment resistance and provides a rationale for incorporating hypoxia-targeted interventions in post-RT therapeutic strategies.

Cellular composition and the tumour microenvironment

Using xCell deconvolution, I identified changes in the tumour microenvironment between responders and progressors. In particular, progressors exhibited increased smooth muscle cell and memory B-cell signatures, which may reflect stromal remodelling and immune adaptation. The presence of class-switched memory B-cells in progressors raises questions regarding the role of B-cell responses in STS immune evasion, as prior studies have suggested that B-cell infiltration can both promote and suppress tumour progression depending on context. Conversely, the relative decrease in CD4⁺ memory T-cells and microvascular endothelial cells in progressors suggests that loss of adaptive immune function and reduced vascular integrity that may contribute to poor outcomes.

Modelling disease progression

The final section of this chapter explored whether post-RT gene expression could predict subsequent disease progression. I used LASSO to select features and trained a Random Forest classifier. On the test data the model showed modest discrimination (AUC = 0.67) with Accuracy 70%, Sensitivity 50%, Specificity 83%, PPV 67%, and NPV 71%.

I then attempted external validation in TCGA-SARC. The model did not demonstrate prognostic or predictive ability there. This is not entirely surprising: the TCGA radiotherapy subset is small (58 patients; 20 events), the endpoint differs (my model was trained for progression, whereas TCGA provides overall survival), the tissue/platforms are not matched (FFPE RNA-seq vs fresh-frozen), two signature genes — *CYP11A1* and *TBC1D3F* — were absent from TCGA, and, crucially, my model was derived from post-RT specimens whereas TCGA is largely treatment-naïve. Chapter 4 shows clear pre- vs post-RT transcriptional shifts, so limited transfer from a post-RT model to a treatment-naïve cohort is biologically plausible.

One key challenge in predictive modelling is distinguishing causative transcriptional changes from those that may be passenger effects of tumour progression. Future studies should validate this model in an independent post-RT cohort with disease progression endpoints; integrate additional clinical variables (e.g., age, histological subtype, grade, and resection margins) and assess whether combining

transcriptomic signatures with genomic alterations (e.g., copy number variations or mutational burden) improves predictive power.

Future directions and clinical implications

These results highlight potential avenues for post-RT risk stratification and therapeutic targeting:

- **Refining follow-up frequency:** If validated, gene expression signatures associated with progression could inform more personalised post-RT monitoring, identifying patients who may benefit from intensified imaging surveillance or earlier intervention.
- **Targeting hypoxia-associated resistance:** Given the enrichment of hypoxia-related pathways in progressors, exploring whether hypoxia-targeted therapies (e.g., HIF1 α inhibitors, anti-angiogenic agents) could improve outcomes in high-risk patients may be a worthwhile avenue for future studies.
- **Immune-based interventions:** The immune-related differences observed in responders vs. progressors suggest that immune checkpoint inhibitors or therapies aimed at restoring immune surveillance could be evaluated in post-RT STS settings.

Conclusions

This chapter builds upon the findings of Chapters 3 and 4 by integrating clinical outcomes with transcriptomic profiling to explore predictors of post-RT disease progression in STS. While RT-induced transcriptional changes were previously characterised in Chapter 4, this chapter demonstrates that progressors exhibit distinct gene expression patterns compared to responders, particularly in pathways related to cell cycle progression, hypoxia adaptation, and immune suppression. The predictive modelling approach suggests that transcriptomic features may hold promise for stratifying post-RT patients based on progression risk, though further validation is needed.

Chapter 6. Discussion

This thesis investigates the genomic and transcriptomic effects of neoadjuvant radiotherapy in soft tissue sarcomas, aiming to identify molecular biomarkers predictive of treatment response and disease progression. Through a multi-omics approach integrating whole exome sequencing, RNA sequencing, pathway analysis, and predictive modelling, this work provides a deeper understanding of RT-induced alterations across multiple sarcoma subtypes.

Key findings from each chapter have been discussed in detail at the end of their respective sections. Rather than reiterating all results, this chapter focuses on a few select themes that hold relevance for understanding RT responses in STS. Specifically, I will explore:

- The impact of radiotherapy on genomic instability and mutational burden (Chapter 3)
- The transcriptional landscape of post-radiotherapy tumours (Chapter 4)
- The potential for transcriptomic biomarkers to predict disease progression (Chapter 5)

These discussions will be framed within the broader context of sarcoma biology and therapeutic response, evaluating how these findings contribute to existing knowledge and identifying areas for future translational research. The chapter will also critically assess the study's limitations and implications for clinical application, particularly in refining patient risk stratification and treatment personalisation.

6.1 Genomic alterations and mutational signatures following radiotherapy

This section discusses the key genomic findings detailed in Chapter 3, focusing on the mutational landscape, somatic mutation burden, and mutational signatures within the London Sarcoma Service (LSS) cohort.

6.1.1 The mutational landscape of the LSS cohort

6.1.1.1 Tumour mutational burden and hypermutated samples

My analysis using WES on the LSS cohort demonstrated a mean TMB of 1.07 mut/Mb. This aligns with data from a 2017 TCGA analysis of 206 soft tissue sarcomas representing six different subtypes, which found an average TMB of 1.06 mutations/Mb (TCGA 2017). In the TCGA cohort, two hypermutated cases (defined as a TMB ≥ 10 mutations/Mb) were identified, representing approximately 1% of the dataset. These hypermutated tumours (histological subtypes not given) were characterised by mismatch repair deficiency and associated with a COSMIC SBS 6 Signature, reflecting the mutational processes driven by defective DNA mismatch repair. These tumours had a frameshift mutation in *MSH6* and low expression of *MSH2*, highlighting known causes of DNA mismatch repair.

A more recent study from Memorial Sloan Kettering (MSK) analysed 2138 sarcomas representing 45 different histological subtypes and reported a median TMB of 2.4 mutations/Mb (Gounder, Agaram *et al.* 2022). They found, 3.9% of samples in the MSK cohort were hypermutated (TMB ≥ 10 mutations/Mb). Hypermutation was predominantly observed in undifferentiated pleomorphic sarcoma (UPS) and angiosarcoma. Notably, hypermutation in cutaneous angiosarcoma was attributed to UV-induced mutational processes, reflecting the environmental exposure associated with this subtype. Interestingly, this same UV signature was seen in several of the hypermutated UPS samples while other UPS samples showing a predominance of “aging” related signatures.

In this LSS cohort, two hypermutated cases were identified: one myxofibrosarcoma (MFS) and one UPS. These tumours exhibited TMB values exceeding 10 mutations/Mb, consistent with the proportion reported by MSK. Both hypermutated tumours in the LSS cohort demonstrated mutational signatures associated with mismatch repair deficiency (SBS15 and ID7), despite the absence of mutations in key MMR genes (*MLH1*, *MSH2*, *MSH6*, *PMS2*). Instead, reduced expression of MMR genes (e.g., low *MLH1* expression in UPS and low *MSH2* expression in MFS) was found. Examination of their copy number status revealed loss of heterozygosity or homozygous deletions of MMR genes, likely contributed to the observed hypermutation phenotype. This contrasts with the TCGA cohort, where hypermutation was directly linked to mutations in the MMR genes.

It should be noted however that the method of TMB calculation differs between these studies and this may account for some variability in the findings. The MSK study employed the MSK-IMPACT targeted sequencing panel (Cheng, Mitchell *et al.* 2015), which captures 341–468 cancer-associated genes, whereas the TCGA cohort and the LSS cohort calculated TMB using whole exome and whole genome sequencing. Targeted sequencing panels like MSK-IMPACT may underestimate TMB by excluding non-coding and intergenic regions, yet they provide clinically validated approximations for large-scale analyses. The broader scope of WES/WGS, while more comprehensive, introduces variability due to differences in sequencing depth between samples and studies as well as different analytical pipelines used in calling mutations. Discrepancies in the proportion of hypermutated cases may also reflect differences in cohort composition. The MSK cohort encompassed 45 subtypes, with a larger representation of hypermutation-prone subtypes like UPS and angiosarcoma. In contrast, the LSS cohort contains fewer subtypes, and the TCGA fewer still.

The repeated observation of hypermutated tumours seen in this LSS cohort underscores the genomic heterogeneity of sarcomas and raises questions about the mechanisms underlying hypermutation beyond MMR deficiency. While hypermutation is often associated with improved responses to immune checkpoint inhibitors (Graham, Pritchard *et al.* 2021), clinical outcomes in the LSS cohort were variable, with one hypermutated case demonstrating long-term disease-free survival

and the other developing metastases (see section 3.5.3) post-radiotherapy. These findings suggest that hypermutation alone may not predict therapeutic outcomes and must be interpreted in the context of tumour subtype, immune microenvironment, and clinical factors like grade, stage, and resection margin status.

6.1.1.2 Novel mutations identified in the LSS cohort

Novel somatic mutations were identified in the LSS cohort (Section 3.2.2), some of which had high REVEL scores (Ioannidis, Rothstein *et al.* 2016), indicating a strong likelihood of pathogenicity. These include *GNAQ*, *CBLB*, *FGFR1*, *MSN*, and *HOXC13* (see Table 3.1). These mutations are noteworthy not only because of their rarity but also due to their potential functional significance, either as oncogenic drivers or as contributors to key biological processes underlying the development of disease progression.

Mutations in metastasis and recurrence samples

The *GNAQ* mutation, identified in a metastatic alveolar soft part sarcoma (ASPS) sample, and the *MSN* mutation, found in a recurrence spindle cell sarcoma (NOS) sample, were not detected in the corresponding pre-radiotherapy biopsies. This strongly suggests that these mutations arose as a consequence of radiotherapy-induced mutagenesis or it could represent sampling bias with subclones containing the mutations not being represented in the sequenced tissue. Although these mutations could represent passenger events, their emergence post-treatment raises the possibility that they may contribute to tumour recurrence or metastasis, potentially through clonal selection or by driving further tumour evolution.

The *GNAQ* gene encodes a G-protein subunit involved in cell signalling and has been implicated in the development of uveal melanoma (Silva-Rodríguez, Fernández-Díaz *et al.* 2022) and vascular tumours (Jansen, Müller *et al.* 2021). Its role in metastatic ASPS is less clear, but its activation could theoretically promote signalling pathways associated with cell proliferation or migration. Similarly, *MSN* encodes Moesin, a cytoskeletal linker protein involved in maintaining cell shape and

motility which has previously been shown to be overexpressed in colorectal cancer and associated with poor survival (Huang, Wei *et al.* 2023).

These novel mutation findings emphasise the dual impact of radiotherapy: its primary purpose of eliminating tumour cells and its unintended potential to contribute to genomic changes that drive recurrence or metastasis. Understanding the functional consequences of such mutations could lead to development of strategies to mitigate these effects.

Potential driver mutations in pre-radiotherapy samples

In contrast, the mutations identified in the pre-radiotherapy samples from myxofibrosarcoma (*CBLB* and *FGFR1*) and undifferentiated pleomorphic sarcoma (*HOXC13*) are more likely to represent primary oncogenic drivers. These mutations were present before the onset of treatment, making it more plausible that they contributed to tumour initiation or progression rather than arising because of radiotherapy.

6.1.2 Somatic mutation burden pre- and post-radiotherapy

Comparative analyses of pre- and post-radiotherapy sarcoma samples using WES revealed no significant differences in the number of SNVs ($p = 0.88$) or indels ($p = 0.17$) across the cohort (Figure 3.3). Similarly, paired analyses of seven patient samples found no significant changes in SNVs ($p = 0.49$) or indels ($p = 0.36$) following radiotherapy (Figure 3.5). These findings suggest that radiotherapy does not immediately induce substantial mutational changes detectable by WES.

However, the higher-resolution NanoSeq WGS data provided additional insights into mutational changes induced by radiotherapy at a subclonal level. NanoSeq revealed a significant ($p = 0.01$), increase in the indel burden with the median number of indels per cell rising from 177 in pre-radiotherapy samples to 690 in post-radiotherapy samples (Figure 3.19). Furthermore, the indel-to-SNV ratio increased (Figure 3.20) from 0.05 to 0.2 ($p = 0.002$). This increase was primarily driven by deletions, as indicated by a higher deletion-to-insertion ratio observed in post-radiotherapy

samples. These findings underscore the sensitivity of NanoSeq in detecting low-frequency mutations that may not be captured by WES, suggesting that radiotherapy-induced mutational effects could be more pronounced than initially apparent from the bulk WES data and are induced in individual cells or small clones of cells.

This pattern becomes even more pronounced in recurrence samples, where the Del:Ins ratio rose significantly compared to pre-radiotherapy samples (median ratio: 4.0 vs. 1.0, $p = 2e-05$; Figure 3.6A). In contrast, post-radiotherapy samples had a median ratio of 1.0, indicating no immediate increase in deletion prevalence following treatment. Metastasis samples showed an intermediate median Del:Ins ratio of 1.89, further supporting the hypothesis of time-dependent mutational dynamics. This delayed emergence of deletions may reflect the time taken for clonal selection of radiation-induced mutations or ongoing genomic instability in tumour cells surviving treatment.

The longitudinal case study of patient 58 (see section 3.3.5) provides additional support for this hypothesis. Tumour samples collected pre-radiotherapy, post-radiotherapy (32 days after treatment), and at recurrence (186 days after treatment) showed stable SNV numbers between pre- and post-radiotherapy samples (22 vs. 28), with 86% of mutations shared across both time points. However, the recurrence sample contained 26 novel SNVs and 7 additional indels absent from earlier samples, suggesting that radiotherapy-induced mutations may initially exist at low variant allele frequencies (VAF) and become detectable only after clonal expansion. The 32-day interval between radiotherapy and post-treatment sample collection may have been insufficient to capture these mutational dynamics, while the 186-day interval allowed sufficient time for clonal growth and enrichment of deleterious mutations.

Findings from other cancers, such as those linked to radiation exposure, provide additional context. Studies of papillary thyroid carcinomas following the Chernobyl disaster (Morton, Karyadi *et al.* 2021) revealed a dose-dependent enrichment of small deletions and structural variants in radiation-exposed tumours, which the number of mutation inversely correlating with distance from the Chernobyl disaster site. Many of these deletions were clonal and reflected non-homologous end-joining

(NHEJ) repair, a mechanism activated in response to double-strand breaks. These mutations were detected in tumour samples taken several decades following radiation exposure.

Similarly, analyses of radiation-associated second malignancies (Behjati, Gundem *et al.* 2016) identified high proportions of deletions, particularly in breast and sarcoma cases, further reinforcing the contribution of NHEJ to the genomic landscape of radiation-induced tumours. These findings align with the elevated Del:Ins ratio observed in both NanoSeq data and recurrence samples from this LSS cohort.

Radiotherapy-associated mutational signatures, such as COSMIC ID8, have also been detected more prominently in recurrent tumours. A study comparing primary and recurrent gliomas (Kocakavuk, Anderson *et al.* 2021) demonstrated an enrichment of ID8 in recurrence samples, emphasising the delayed mutational effects of radiotherapy through clonal evolution and tumour progression.

In the NanoSeq data from this study, COSMIC ID8 was uniquely detected in all five post-radiotherapy samples (Figure 3.23), whereas it was absent in all pre-radiotherapy samples. This finding highlights the impressive sensitivity of NanoSeq in detecting subtle mutational changes induced by radiotherapy. ID8, associated with non-homologous end joining (NHEJ) repair mechanisms, is characterised by deletions ≥ 5 bp without microhomology, reflecting the error-prone nature of this pathway. Its consistent presence across post-radiotherapy samples suggests that radiotherapy activates specific DNA repair processes that leave a distinct genomic signature. Importantly, the absence of ID8 in pre-radiotherapy samples supports its association with radiotherapy rather than pre-existing tumour biology.

Integrating these WES and NanoSeq findings demonstrates the complementary nature of these techniques. Bulk WES provides an overview of the mutational landscape by capturing SNVs and indels across the coding genome. However, its limited sensitivity for detecting low-frequency mutations or subtle subclonal changes, such as small deletions with low VAFs, which underscores the need for higher-resolution approaches. NanoSeq, a WGS “single molecule” technique by contrast, revealed significant increases in indels, particularly deletions, that likely contribute to

tumour evolution over time. These findings suggest that radiotherapy-induced mutagenesis may not be fully apparent immediately after treatment but becomes more evident during recurrence as subclonal mutations clonally expand.

Future studies could leverage the sensitivity of NanoSeq and similar high-resolution techniques to monitor subclonal dynamics and pinpoint the precise timing of radiotherapy-induced mutagenesis. For example, spatially mapped longitudinal sampling of tumours at multiple time points post-radiotherapy, coupled with single-cell sequencing, could provide a clearer understanding of how radiation-induced mutations evolve and clonally expand throughout the tumours.

Additionally, integrating multi-omics approaches, such as transcriptomics to explore changes in gene expression or epigenomics to identify shifts in chromatin accessibility, could reveal pathways implicated in immune evasion or therapy resistance. Targeting dysregulated pathways with DNA repair inhibitors, such as PARP inhibitors (Sun, Chu *et al.* 2023) or ATM inhibitors (García, Kirsch *et al.* 2022), may offer a strategy to mitigate the effects of radiotherapy-induced mutagenesis and improve clinical outcomes in sarcomas.

6.1.3 Critical appraisal of methods and limitations

This study represents a comprehensive effort to investigate the genomic effects of neoadjuvant radiotherapy in soft tissue sarcomas. However, several methodological limitations and areas for potential improvement in hindsight deserve discussion.

Tissue availability and sample size

A significant limitation of this study was the small number of paired pre- and post-radiotherapy samples, with only seven patients included in the paired analysis. This restricted sample size limited the statistical power of the comparisons and may have hindered the detection of subtle genomic changes induced by radiotherapy. A larger cohort of paired samples would have provided a more robust dataset to validate these findings. However, the availability of frozen tissue, particularly for both pre- and post-radiotherapy samples, was a major constraint. Most samples were formalin-fixed paraffin-embedded (FFPE), which despite the optimisation of high-quality nucleic acid extraction through use of the Covaris machine still poses challenges for high-quality DNA extraction and downstream analyses.

Choice of sequencing approach

Whole-exome sequencing (WES) was selected for this study due to its cost-effectiveness and focus on coding regions. However, in hindsight, whole-genome sequencing (WGS) might have been a more suitable approach, especially given the low tumour mutational burden (TMB) observed in this LSS cohort. WGS could have captured a broader spectrum of genomic alterations, including non-coding mutations and structural variants, providing additional insights into radiotherapy-induced mutagenesis. The higher resolution of WGS would likely have increased the number of mutations detected, allowing for more statistically significant comparisons and a more nuanced understanding of the genomic response to radiotherapy. Whilst I had the opportunity to investigate WGS of sarcoma through Genomics England, that dataset was unsuitable due to a lack of paired pre-and post-radiotherapy samples.

Sensitivity to low-frequency mutations

The low TMB observed in most samples presented challenges for detecting radiotherapy-induced genomic changes by WES. While NanoSeq was employed for increased sensitivity to low-frequency mutations, this technique was applied to only a subset of samples. Expanding the use of high-sensitivity techniques like NanoSeq across all samples might have enhanced the ability to identify subclonal mutations and other subtle genomic alterations associated with radiotherapy.

Tumour heterogeneity

Another limitation was the inherent tumour heterogeneity within and between samples. The genomic differences observed may partly reflect sampling variation rather than true treatment-induced effects. This is particularly relevant as biopsies were compared with resection specimens, as these may not entirely represent the same tumour regions. An ideal experiment would use multi-regional sampling of both pre- and post-radiotherapy specimens, combined with both radiological mapping of samples and spatial transcriptomic approaches, to provide a more comprehensive assessment of tumour heterogeneity and radiotherapy-induced changes.

Recommendations for future studies

Based on these limitations, future studies should prioritise:

1. Expanding paired cohorts

Efforts should be made to collect larger numbers of paired pre- and post-radiotherapy samples, with a focus on preserving matched frozen tissue wherever feasible.

2. Using Whole-Genome Sequencing

Transitioning to WGS of pre and post-radiotherapy samples at high depth could allow for a more comprehensive analysis of genomic alterations, including non-coding and structural changes.

3. Implementing High-sensitivity sequencing techniques

Employing NanoSeq or other high-sensitivity methods more broadly across cohorts could enhance detection of subclonal mutations.

4. Addressing tumour heterogeneity

Integrating multi-regional sampling and spatially resolved techniques would better account for intra-tumoural variability and improve the interpretability of results.

6.2 Transcriptomic responses to neoadjuvant radiotherapy

This section discusses the transcriptomic changes in soft tissue sarcomas following neoadjuvant radiotherapy, as detailed in Chapter 4. Key findings discussed include the global clustering trends in transcriptomic data, differential gene expression analysis, pathway enrichment analyses, and immune composition changes.

6.2.1 Transcriptomic clustering trends

Principal Component Analysis (PCA) (Figure 4.1) and Uniform Manifold Approximation and Projection (UMAP) (Figure 4.2) demonstrated that transcriptomic data clustered more strongly by histological subtype than by radiotherapy status. This stresses the dominant role of intrinsic tumour characteristics in shaping global gene expression profiles, consistent with previous studies in sarcomas, including findings from the TCGA (TCGA 2017) and a study comparing post-radiation versus sporadic sarcomas (Lesluyes, Baud *et al.* 2019). In both datasets, sarcoma subtypes, such as leiomyosarcoma (LMS) and synovial sarcoma (SS), formed distinct clusters, emphasising the influence of intrinsic features specific to the histological subtype over external factors like treatment.

Interestingly, in this study, patient-specific clustering highlighted the reproducibility and robustness of sequencing data while suggesting that radiotherapy-induced transcriptomic effects are subtle or overshadowed by inter-patient variability (Figure 4.3 and Figure 4.4). Similarly, the Lesluyes study (Lesluyes, Baud *et al.* 2019) observed significant overlap between sporadic and post-radiation sarcomas, reinforcing the notion that sarcoma biology, rather than treatment, drives transcriptomic clustering patterns. However, some overlaps between histological

subtypes were also observed in all studies, reflecting likely shared molecular features or pathways between subtypes, particularly those with similar biological origins.

The TCGA study further demonstrated the value of integrating transcriptomic data with other omics layers, such as methylation and copy-number variation, to refine clustering and reveal molecular subsets within histological subtypes (TCGA 2017). For instance, the separation of uterine and soft tissue leiomyosarcomas clusters using integrated data highlights the potential for multi-omic approaches to uncover finer stratifications not apparent in transcriptomics alone.

Given the limitations of transcriptomics in fully capturing radiotherapy-induced changes, epigenetic profiling could offer a complementary approach for sarcoma classification and understanding treatment responses. DNA methylation-based classifiers, for example, have been effective in stratifying sarcomas into molecular subtypes with strong prognostic correlations, even resolving cases previously deemed unclassifiable (Koelsche, Schimpf *et al.* 2021). Importantly, evidence from other cancers highlights that radiotherapy can induce significant epigenetic alterations, including changes in DNA methylation and histone modifications. These alterations can influence gene expression, DNA repair, and tumour behaviour, as demonstrated by studies showing radiotherapy-induced hypermethylation of tumour suppressor genes like *TP53* and changes in histone methylation that modulate radiosensitivity (Wang, Han *et al.* 2022).

To build on the foundation of this work, future studies should prioritise multi-omic integration to capture the interplay between intrinsic tumour characteristics and treatment-induced changes. Combining transcriptomic and epigenomic data, alongside advanced techniques like spatial transcriptomics and single-cell RNA sequencing, could provide a more nuanced understanding of radiotherapy's effects on the tumour microenvironment. These approaches hold the potential to reveal spatial and cellular heterogeneity, uncover subtle radiotherapy-induced changes, and better characterise tumour-stromal interactions far better than the bulk RNAseq approach taken here. Furthermore, integrating multi-omic layers may refine molecular stratification in sarcomas and identify novel biomarkers or therapeutic targets to enhance radiosensitivity and overcome treatment resistance.

6.2.2 Differential Gene Expression Analysis

6.2.2.1 *Global transcriptomic changes post-radiotherapy*

Global analysis of 40 paired pre- and post-radiotherapy samples across 13 sarcoma subtypes revealed 140 significantly differentially expressed genes, with 107 upregulated and 33 downregulated post-radiotherapy (Figure 4.6). Key findings include the significant downregulation of haemoglobin-related genes (*HBA1*, *HBA2*, *HBB*, *ALAS2*, and *HEMGN*) and notable upregulation of genes involved in immune modulation as well as genes involved in tissue repair and remodelling (Figure 4.7).

6.2.2.2 *Haemoglobin-related genes*

The downregulation of haemoglobin-related genes observed in this study represents an intriguing finding, given their well-established roles in oxygen transport via erythrocytes and emerging evidence of non-canonical functions in cancer biology (Lerebours, Vacher *et al.* 2021, Han, Zhang *et al.* 2022, Kim, Choi *et al.* 2023).

Pre-radiotherapy, haemoglobin genes such as *HBA1*, *HBA2*, and *HBB*—encoding the alpha and beta globin chains—were highly expressed. Although their functional role in sarcomas remains unclear, studies in other cancers suggest that haemoglobin genes can play a role in tumour cell survival under stress conditions.

Haemoglobin gene expression was reported in 2009 in tumour cells, independent of stromal contamination, using fluorescence-activated cell sorting (FACS) purification of colorectal adenocarcinomas (Smith, Culhane *et al.* 2009). Subsequent studies have highlighted the non-canonical roles of haemoglobin genes in cancer biology. For example, *HBB* (haemoglobin beta) is overexpressed in inflammatory breast cancer, where it mitigates oxidative stress and supports an aggressive phenotype (Lerebours, Vacher *et al.* 2021). Similarly, in cervical carcinoma, *HBB* acts as an antioxidant by scavenging free radicals, aiding cell survival in oxidative environments (Li, Wu *et al.* 2013).

In clear cell renal cell carcinoma (ccRCC), *HBB* expression is associated with poor prognosis, shorter recurrence-free survival, and increased tumour aggressiveness. Functional studies revealed that *HBB* suppresses reactive oxygen species (ROS) levels, promoting cell proliferation and invasion under hypoxic conditions. Notably, oxidative stress itself can upregulate *HBB*, indicating its role in the redox adaptation of tumour cells (Kurota, Takeda *et al.* 2023). These findings suggest that *HBB* may act as both a survival mechanism in hypoxic microenvironments and a driver of malignancy.

Interestingly, haemoglobin genes have been implicated in hypoxic tumour environments beyond ccRCC. In lung adenocarcinoma, elevated expression of *HBQ1*, a related haemoglobin gene, reduces ROS levels and supports cell proliferation, emphasising the antioxidant and homeostatic roles of haemoglobin in cancer progression (Kim, Choi *et al.* 2023).

This twin role of haemoglobin-related genes in both oxidative stress management and tumour aggressiveness highlights their importance in tumour biology and potential as therapeutic targets. The functional implications of their expression in sarcomas, particularly post-radiotherapy, remain to be elucidated.

In this study, post-radiotherapy downregulation of these genes may reflect changes in the tumour microenvironment, such as reduced cellular or metabolic demands for haemoglobin-like activity. Alternatively, the suppression could indicate broader shifts in transcriptional regulation following treatment, consistent with radiotherapy-induced disruption of tumour homeostasis. However, the precise implications of this downregulation for vascular remodelling or oxygen delivery remain speculative in the absence of direct experimental evidence in these sarcomas.

To further support this idea, additional haemoglobin-related genes, such as *ALAS2* and *HEMGN*, were also significantly downregulated post-radiotherapy in this study. These genes play critical roles in haem biosynthesis, oxidative stress management, and erythroid lineage survival, highlighting potential disruptions in these tumour microenvironment functions.

6.2.2.3 *ALAS2 and vascular remodelling*

ALAS2, an enzyme critical for haem biosynthesis, is typically upregulated under hypoxic conditions via TGF- β signalling in erythroid cells, facilitating terminal differentiation and intracellular haem production essential for oxidative stress management and iron homeostasis. Experimental studies in erythroid cell lines have demonstrated that hypoxia-induced TGF- β signalling enhances ALAS2 expression, supporting adaptive metabolic functions during stress (Kaneko, Furuyama *et al.* 2009). Its significant suppression post-radiotherapy, as observed here, may reflect the tumour's diminished ability to counteract oxidative stress, further limiting metabolic flexibility and resilience in the hypoxic microenvironment.

This aligns with broader disruptions in vascular integrity following radiotherapy. xCell analysis revealed a significant decrease in endothelial cell abundance in post-radiotherapy samples (Figure 4.16), consistent with vascular remodelling driven by radiation-induced endothelial cell damage. Supporting this, previous studies have shown that doses exceeding 10 Gy can induce severe vascular damage in tumours, leading to reduced perfusion and increased hypoxia (Park, Griffin *et al.* 2012). Radiation-induced endothelial cell apoptosis has been directly linked to impaired oxygen delivery and microenvironmental collapse, contributing to tumour cell death indirectly by disrupting vascular supply.

Such vascular damage likely amplifies the metabolic stress experienced by tumours, further diminishing their capacity for haem biosynthesis. Evidence from prior studies suggests that the functional vascular volume in irradiated tumours declines rapidly, with profound decreases even at moderate radiation doses (Park, Griffin *et al.* 2012). This reduction in vascular support not only limits oxygen delivery but may also impair iron metabolism, an essential factor for haem biosynthesis, compounding the suppression of ALAS2 observed here. While the precise mechanisms remain unclear, disrupted hypoxia-responsive pathways and reduced endothelial cell survival represent plausible factors driving the observed transcriptional changes.

These vascular changes are accompanied by significant transcriptional shifts revealed by PROGENy (section 4.8) and GSEA (section 4.6) analyses. PROGENy

identified a significant decrease in PI3K activity ($P = 0.002$), a pathway critical for vascular integrity and angiogenesis, aligning with the observed vascular remodelling and potential endothelial dysfunction (Kobialka and Graupera 2019). Conversely, increased activity in TNF α ($P = 0.0001$) and NF- κ B ($P = 0.005$) pathways suggests heightened inflammatory responses, which may contribute to microenvironmental stress and further impair vascular function (Balkwill 2009, Hoesel and Schmid 2013). Similarly, GSEA identified enrichment of inflammatory pathways, including TNF α signalling via NF- κ B and IL6-JAK-STAT3, reinforcing the role of inflammation in these radiotherapy-induced changes.

These disruptions likely exacerbate metabolic stress, further impairing the tumour's capacity for haem biosynthesis. GSEA also highlighted enrichment of hallmark pathways such as hypoxia and reactive oxygen species, reflecting increased oxidative and hypoxic stress post-radiotherapy. Alongside the observed downregulation of oxidative phosphorylation and related metabolic pathways, these findings suggest a potential link between vascular damage, oxidative stress, and disrupted hypoxia-responsive pathways that may contribute to the suppression of *ALAS2*. While direct causation cannot be established, these pathways represent plausible contributors to the observed transcriptional changes.

6.2.2.4 *HEMGN* and erythroid differentiation

Similarly, *HEMGN* (hemogen), a transcriptional regulator involved in hematopoietic stem cell survival and erythroid differentiation, has been shown to modulate responses to stress, including irradiation and hypoxia. It plays a protective role by negatively regulating interferon-gamma (IFN- γ) signalling, thereby limiting apoptosis and promoting survival and regeneration of hematopoietic stem and progenitor cells (HSPCs) during hematopoietic stress (Zhao, Liu *et al.* 2022). Studies in murine models have demonstrated that *HEMGN* expression is significantly induced under conditions of irradiation and transplantation stress (Jiang, Yu *et al.* 2010). This induction is critical for maintaining HSPC engraftment and functionality, with *HEMGN*-deficient cells exhibiting impaired engraftment, increased apoptosis, and heightened oxidative stress following transplantation. Furthermore, *HEMGN*

supports erythroid differentiation by facilitating chromatin modifications through histone acetyltransferases, playing a key role in hematopoietic lineage fidelity.

In the context of this study, the observed decreased expression of *HEMGN* post-radiotherapy raises important questions about the underlying mechanisms driving this change. One possibility is that radiotherapy-induced vascular damage and increased oxidative stress may disrupt erythroid-like lineage functions within the tumour microenvironment, leading to reduced *HEMGN* expression. Alternatively, the decreased expression may reflect changes in the cellular composition of the tumour, with selective loss of tumour or stromal cell populations that predominantly express *HEMGN*.

Radiotherapy's impact on transcriptional regulation could also play a role, with stress-induced signalling pathways, such as those involving interferon-gamma (IFN- γ) or inflammatory cytokines, potentially suppressing *HEMGN* transcription as part of a broader shift in gene expression. Another explanation could be that the decrease in *HEMGN* represents an adaptive response by the tumour to mitigate excessive oxidative stress.

Future directions

These findings collectively highlight the intricate interplay between haemoglobin-related gene expression, vascular integrity, inflammation, and oxidative stress in the tumour microenvironment. While the precise mechanisms underlying these changes remain speculative, the data suggest plausible contributors that warrant further investigation. To deepen our understanding of the role of haemoglobin genes in sarcomas and their potential impact on radiotherapy outcomes, several areas of future research are proposed:

Spatial and cellular expression analysis

To delineate the spatial and cellular context of haemoglobin gene expression, future studies could employ spatial transcriptomics and single-cell RNA sequencing. These technologies would enable precise mapping of the localised expression of genes such as *HBA1*, *HBA2*, and *HBB* within the tumour microenvironment. By identifying the specific tumour or stromal cell populations expressing these genes and their

spatial relationships to hypoxic regions and vasculature, we could start to understand the functional significance of these genes within sarcomas.

In addition, immunohistochemistry could be used to validate protein-level expression and confirm that the detected RNA transcripts are translated into functional proteins. Such approaches would provide critical evidence linking haemoglobin gene expression to the metabolic and structural adaptations of the tumour microenvironment both pre- and post-radiotherapy.

Functional analysis

To assess the functional roles of haemoglobin genes in sarcoma biology, knockdown studies targeting *HBA1*, *HBA2*, and *HBB* could be performed. These experiments would help determine the impact of silencing these genes on tumour growth, reactive oxygen species (ROS) levels, and radiotherapy sensitivity. Such studies could reveal whether these genes play a direct role in modulating tumour progression and treatment responses.

Furthermore, direct measurements of tissue oxygen levels in sarcomas with high versus low haemoglobin gene expression could provide valuable insights into their role in tumour oxygenation, metabolic activity, and radiotherapy efficacy. This line of research would help clarify whether haemoglobin gene expression contributes to radiotherapy resistance or sensitivity. This could be context dependent and so coupling this information with spatial transcriptomic data would be key to address this question.

Radiotherapy models

Developing preclinical radiotherapy models with altered haemoglobin gene expression could offer direct evidence of the influence of these genes on treatment outcomes. For instance, models engineered to overexpress or suppress *HBA1*, *HBA2*, or *HBB* could be used to test the hypothesis that haemoglobin gene expression enhances radiotherapy by sustaining oxygen levels or exacerbating oxidative stress. These experiments could help establish causality and provide a framework for understanding the relationship between haemoglobin gene expression and radiotherapy efficacy.

Mechanistic studies

Mechanistic investigations should focus on elucidating the pathways that link haemoglobin gene expression with oxidative stress, hypoxia adaptation, and tumour cell survival. For example, exploring interactions with hypoxia-inducible factors (HIFs) or ROS-regulating proteins could provide insight into the adaptive responses of tumour cells in hypoxic environments.

Additionally, it would be valuable to investigate whether haemoglobin gene expression correlates with cancer-associated fibroblast (CAF) infiltration or vascular remodelling. These factors are known to influence the tumour microenvironment and may play a role in modulating treatment responses (Ansems and Span 2020). By connecting haemoglobin gene expression with broader aspects of sarcoma biology, these future studies could uncover novel therapeutic strategies for improving response to radiotherapy.

6.2.3 Subtype-specific transcriptomic changes post-radiotherapy

Further analysis of transcriptomic responses to radiotherapy revealed both shared and subtype-specific gene expression changes across soft tissue sarcomas, emphasising the heterogeneity of their biological responses to treatment. A total of 187 genes were differentially expressed in at least two subtypes, with some genes, such as the haemoglobin genes *HBA2* and *HBB*, displaying consistent downregulation across six subtypes: myxofibrosarcoma (MFS), myxoid liposarcoma (MLS), pleomorphic leiomyosarcoma (pLMS), dedifferentiated liposarcoma (ddLPS), spindle cell sarcoma (SpCS), and extraskeletal myxoid chondrosarcoma (EMC). While the reasons for this downregulation remain unclear, it may reflect broader changes in the tumour microenvironment following radiotherapy. One possibility is that tumour cells expressing haemoglobin genes may have been more sensitive to radiotherapy-induced oxidative stress, leading to their preferential elimination. Alternatively, this suppression could result from radiotherapy-induced transcriptional reprogramming, potentially driven by changes in vascular integrity, hypoxia, or shifts in cellular composition within the tumour microenvironment.

6.2.3.1 *F13A1*

F13A1, encoding coagulation factor XIII subunit A, was the most commonly upregulated gene post-radiotherapy, differentially expressed in four sarcoma subtypes (MFS, MLS, pLMS, and ddLPS) (Figure 4.9). Beyond its well-known role in blood clot stabilisation, *F13A1* contributes to extracellular matrix (ECM) remodelling, tissue repair, and angiogenesis - processes that may be activated in response to RT-induced damage (Lehrer, Dembitzer *et al.* 2018, Peltier, Roperch *et al.* 2018, Ercan, Mauracher *et al.* 2021). In glioblastoma, increased *F13A1* copy number correlates with improved survival, possibly by influencing tumour-associated coagulation pathways. Similarly, in lung cancer, *F13A1* processing was linked to hypercoagulability and a pro-metastatic microenvironment, while in colorectal cancer, decreased serum AP-F13A1 suggests sequestration within tumours, facilitating ECM remodelling and tumour growth. The consistent upregulation of *F13A1* in sarcomas post-RT may reflect its involvement in tissue repair and adaptive responses, but its role in coagulation raises the possibility of promoting tumour resilience and immune evasion. Given its diverse tumour-associated functions, further investigation into *F13A1*'s role in sarcoma biology and its potential as a therapeutic target is warranted, particularly in the context of mitigating tumour regrowth and enhancing RT efficacy.

6.2.3.2 *SERPINE1*

SERPINE1 (plasminogen activator inhibitor-1, PAI-1) was upregulated in myxoid liposarcoma, pleomorphic liposarcoma, and synovial sarcoma post-radiotherapy, suggesting a role in tumour adaptation. A known regulator of fibrinolysis and cell adhesion, *SERPINE1* prevents plasmin-mediated ECM degradation, stabilising the extracellular matrix (ECM) and contributing to post-RT tissue repair. In head and neck squamous cell carcinoma, *SERPINE1* is part of a radioresistance-associated gene signature linked to poor prognosis, angiogenesis, and DNA damage repair, as well as increased macrophage and CD4+ T cell infiltration (Zhang, Wang *et al.* 2022). In triple-negative breast cancer, *SERPINE1* facilitates DNA double-strand break repair following RT, enhancing tumour survival, and its inhibition resensitised tumours to RT in preclinical models (Su, Wu *et al.* 2023). In sarcomas, its post-RT upregulation may reflect a similar adaptive response, supporting ECM remodelling

and DNA repair to promote tumour survival. Targeting *SERPINE1*-driven pathways could enhance RT efficacy by disrupting both ECM dynamics and mechanisms of radioresistance.

6.2.3.3 *RGS1*

RGS1 was upregulated following radiotherapy in myxoid liposarcoma, dedifferentiated liposarcoma, and myxofibrosarcoma, suggesting a conserved role in immune regulation and tumour adaptation. As a key regulator of G-protein-coupled receptor (GPCR) signalling, *RGS1* influences immune cell migration and inflammatory responses. In melanoma, its overexpression is linked to increased tumour thickness, mitotic rate, and lymph node metastasis, where it facilitates immune evasion through calcium influx regulation and activation of ERK and AKT signalling (Yang, Zhang *et al.* 2023). The impact of radiotherapy on *RGS1* expression appears cell-type dependent, with studies showing its downregulation in Jurkat cells but upregulation in TK6 and HFL1 cells, indicating a variable response to radiation-induced stress (Chaudhry 2008). The consistent upregulation of *RGS1* in post-RT sarcomas suggests a potential role in remodelling the tumour microenvironment, modulating immune cell infiltration, and contributing to tissue repair. Whether this confers a protective or pro-tumorigenic effect remains to be further investigated.

6.2.4 Genes with variable expression patterns following radiotherapy

While the genes discussed above highlight shared transcriptional responses across sarcoma subtypes, the results of this study (Figure 4.9) also revealed genes with variable expression patterns depending on the subtype. *CASQ1* and *TAGLN* for example exhibited subtype specific expression patterns and are discussed below.

6.2.4.1 *Calsequestrin 1 (CASQ1) and calcium signalling in radiotherapy response*

Calsequestrin 1 (*CASQ1*) is a calcium-binding protein primarily located in the sarcoplasmic reticulum of skeletal muscles, where it plays a critical role in calcium homeostasis by buffering calcium ions and modulating their release during muscle contraction. *CASQ1* interacts with proteins such as triadin, junctin, and ryanodine receptors to regulate calcium release channels and maintain efficient excitation-contraction coupling. It also directly influences store-operated calcium entry (SOCE) by interacting with STIM1 to regulate calcium influx under conditions of sarcoplasmic reticulum depletion (Rossi, Gamberucci *et al.* 2021). In this study, *CASQ1* expression was found to increase in myxoid liposarcoma (MLS) and myxofibrosarcoma (MFS) but decrease in extraskeletal myxoid chondrosarcoma (EMC) following radiotherapy. These changes in expression suggest that *CASQ1* may play a role in the cellular response to radiotherapy, potentially through its established function in regulating intracellular calcium levels. Calcium signalling is critical for various cellular processes, including proliferation, apoptosis, and stress responses (Patergnani, Danese *et al.* 2020), and the observed differential expression of *CASQ1* may reflect subtype-specific differences in how these sarcomas manage calcium homeostasis under the stress of radiotherapy.

Additional differentially expressed genes in post-radiotherapy samples further highlight the role of calcium signalling in these sarcoma subtypes. For example, *CACNA1E* (Phan, Wang *et al.* 2017), a voltage-gated calcium channel subunit, and *CACNA2D1*, a calcium channel auxiliary subunit, were differentially expressed in MLS, EMC, and synovial sarcoma (SS) (Table 4.4). These genes are essential for calcium entry into cells and are involved in numerous calcium-dependent processes,

including signalling cascades that regulate cell survival and apoptosis. Similarly, *RYR1* and *RYR2*, which encode ryanodine receptors responsible for calcium release from the sarcoplasmic reticulum (Wang, Yu *et al.* 2022), were differentially expressed in sarcoma subtypes such as MLS, EMC, spindle cell sarcoma (SpCS), and pleomorphic liposarcoma (pLPS) (Table 4.4). These receptors are key regulators of intracellular calcium homeostasis, and their dysregulation has been linked to altered calcium signalling, genomic instability, and impaired apoptotic pathways, which may contribute to subtype-specific responses to radiotherapy-induced stress (Wang, Yu *et al.* 2022).

Further evidence for the role of calcium signalling in radiotherapy responses comes from a study on undifferentiated pleomorphic sarcomas (UPS) (Blomain, Soudi *et al.* 2025). Blomain *et al.* demonstrated that radiotherapy imposes significant selective pressures on tumour subclones, leading to dynamic changes in their abundance. Specifically, subclones that contracted—meaning their prevalence significantly decreased—were enriched for mutations in calcium signalling pathways, suggesting that disruptions in calcium homeostasis may render these subclones more radiosensitive. These findings provide a compelling link between calcium signalling and the cellular mechanisms that determine radiotherapy outcomes. Mutations in genes involved in calcium transport, storage, and signalling likely impair the ability of tumour cells to buffer radiotherapy-induced calcium stress, ultimately leading to their elimination.

Blomain *et al.* also identified significant upregulation of calcium ATPase genes, including *ATP2A1*, *ATP2A3*, *ATP2B1*, and *ATP2B2*, in post-radiotherapy samples. Similarly, they also observed the increased expression of *CACNA1* and *RYR*-related genes that I have seen in the LSS cohort. These genes overall play critical roles in exporting calcium ions from the cytoplasm and maintaining intracellular calcium balance. Their upregulation likely reflects a mechanism to mitigate the increased calcium flux and cellular stress caused by radiotherapy. Functional experiments further demonstrated that inhibiting plasma membrane calcium ATPases using Caloxin 2A1 sensitised sarcoma cell lines to radiotherapy.

Taken together, these findings emphasise the role of calcium signalling, including the function of *CASQ1*, in mediating cellular responses to radiotherapy. The differential expression of *CASQ1*, *CACNA1E*, *CACNA2D1*, *RYR1*, and *RYR2*, combined with the enrichment of calcium signalling mutations in radiotherapy-sensitive subclones, highlights how sarcoma subtypes adapt to the unique stress conditions imposed by radiotherapy. These findings suggest that targeting calcium signalling pathways, either through modulation of calcium transport or inhibition of adaptive responses, could enhance radiosensitivity and improve therapeutic outcomes.

6.2.4.2 Transgelin (*TAGLN*) and vascular remodelling

TAGLN (transgelin), also known as SM22 α , is a cytoskeletal protein predominantly involved in actin filament stabilisation, with emerging roles in angiogenesis and endothelial cell dynamics. In this study, *TAGLN* was found to be upregulated in myxoid liposarcoma (MLS) but downregulated in extraskeletal myxoid chondrosarcoma (EMC) and pleomorphic leiomyosarcoma (pLMS) following radiotherapy. Such differential expression suggests that *TAGLN* may contribute to the varying vascular responses across sarcoma subtypes under radiotherapy-induced stress.

TAGLN plays a critical role in endothelial cell elongation, a key process in angiogenesis. Under angiogenic stimuli such as VEGF, *TAGLN* expression supports endothelial sprouting and vessel morphogenesis. Conversely, its disruption enhances angiogenic behaviours like excessive cord-like structure formation, indicating that *TAGLN* can act as a regulator of endothelial elongation depending on the cellular context (Tsuji-Tamura, Morino-Koga *et al.* 2021).

Radiotherapy-induced vascular remodelling involves complex processes, including changes in angiogenesis, immune infiltration, and extracellular matrix (ECM) composition (Ahmed, Malachowska *et al.* 2025) (see section 6.2.2.3). The downregulation of *TAGLN* in EMC and pLMS may reflect a shift towards greater endothelial plasticity, facilitating angiogenic remodelling and vessel sprouting in

response to vascular damage. In contrast, its upregulation in MLS could indicate an effort to stabilise endothelial structures and counteract radiotherapy-induced stress. The biology behind the unique “chicken-wire” vascular pattern commonly observed in MLS may also influence its reliance on *TAGLN* to maintain vascular integrity post-radiotherapy.

Radiotherapy has been shown to induce significant vascular damage, including endothelial cell death, reduced perfusion, and increased tumour hypoxia. High-dose radiation can lead to endothelial apoptosis, impairing vascular support and contributing to tumour control (Park, Griffin *et al.* 2012). However, the extent of vascular damage varies across tumour types, with some restoring perfusion through angiogenesis while others experience vascular collapse. *TAGLN*'s differential expression may mirror these processes, with MLS favouring stabilisation mechanisms through increased *TAGLN*, while EMC and pLMS adapt by employing alternative angiogenic pathways that are less dependent on *TAGLN*.

These findings are consistent with evidence suggesting that radiotherapy-induced tissue regeneration involves molecular pathways such as Wnt and p53 signalling, which regulate cytoskeletal and vascular dynamics (Ahmed, Malachowska *et al.* 2025). *TAGLN*-mediated cytoskeletal remodelling may interact with these pathways, influencing how tumour microenvironments respond to radiotherapy. The differential expression of *TAGLN* across sarcoma subtypes stresses the role of tumour-specific vascular remodelling in determining radiotherapy outcomes.

6.2.4.3 Summary

The limited overlap of differentially expressed genes across sarcoma subtypes, as shown in the Upset plot (Figure 4.8) demonstrates the complexity and heterogeneity of transcriptomic responses to radiotherapy. While certain subtypes, such as MFS and MLS, exhibit the greatest gene overlap, their overall transcriptional responses remain largely distinct, highlighting the necessity for therapeutic approaches tailored

to each subtype's unique biology. This complexity is further illustrated by the coexistence of shared vulnerabilities — such as the consistent suppression of haemoglobin-related genes and the upregulation of tissue remodelling pathways like *F13A1* — and unique adaptations, including subtype-specific alterations in calcium signalling and cytoskeletal dynamics.

Future research should prioritise functional studies to unravel the roles of these genes in mediating radiotherapy outcomes. Investigating whether shared pathways, such as *F13A1*-mediated ECM remodelling, can be targeted to enhance treatment efficacy or whether subtype-specific adaptations, such as *CASQ1*-driven calcium signalling, provide exploitable vulnerabilities will be key to advancing precision medicine for sarcomas. Additionally, integrating spatial and single-cell transcriptomic techniques will be critical for mapping the cellular origins of these changes and understanding their interactions with the tumour microenvironment.

6.2.5 Pathway enrichment analysis highlights immune activation post-radiotherapy

The findings from this study strongly suggest that radiotherapy stimulates immune activation, a phenomenon well-documented in the literature (Deloch, Derer *et al.* 2016, Rückert, Flohr *et al.* 2021). Radiotherapy not only induces direct DNA damage to tumour cells but also initiates secondary effects that modify the tumour microenvironment, thereby promoting immunogenic cell death (ICD). ICD is characterised by the release of damage-associated molecular patterns (DAMPs) such as calreticulin, HMGB1, and ATP, which enhance the recruitment and activation of antigen-presenting cells (APCs) like dendritic cells. This process facilitates T-cell priming and subsequent systemic anti-tumour immune responses.

One of the most intriguing immune-mediated effects of radiotherapy is the abscopal effect, where irradiation of a local tumour site results in tumour regression at distant, untreated sites. This phenomenon is believed to occur due to radiation-induced immune activation, which enhances systemic anti-tumour immunity. Although the

abscopal effect was historically considered a rare occurrence, increasing evidence suggests that combining RT with immunotherapies, particularly immune checkpoint inhibitors, can enhance its frequency and potency. Mechanistically, RT triggers the release of tumour antigens, activates the cGAS-STING pathway, and promotes the maturation of DCs, leading to a systemic T-cell response against cancer cells at both irradiated and non-irradiated sites (Janopaul-Naylor, Shen *et al.* 2021).

Recent work has also emphasised that the immunological effects of radiotherapy are complex and context-dependent, with both immunostimulatory and immunosuppressive components (Rückert, Flohr *et al.* 2021). While radiotherapy enhances antigen presentation, cytokine release, and immune cell infiltration, it can also create an immunosuppressive tumour microenvironment. This occurs through mechanisms such as the recruitment of regulatory T cells (Tregs), upregulation of immunosuppressive cytokines (e.g., TGF- β , IL-10), and increased expression of immune checkpoint molecules such as PD-L1. Thus, radiotherapy alone may not always generate a strong systemic immune response, and combination strategies with immunotherapies are increasingly being explored to overcome these barriers (Deloch, Derer *et al.* 2016, Janopaul-Naylor, Shen *et al.* 2021, Rückert, Flohr *et al.* 2021).

In this study, GSEA and Gene Ontology (GO) analyses revealed significant enrichment of immune-related pathways, such as interferon signalling, TNF α signalling, and inflammatory response pathways. These results align with the documented role of radiotherapy in promoting immune-mediated effects through the upregulation of pro-inflammatory cytokines (e.g., IL-6, IL-8, TNF- α) and the enhancement of T-cell infiltration (Ahmed, Malachowska *et al.* 2025). The observed activation of pathways such as JAK-STAT (Johnson, O'Keefe *et al.* 2018) and NF- κ B (Hoesel and Schmid 2013) further supports this, as these are pivotal in modulating immune responses.

Conversely, the downregulation of pathways associated with cell proliferation, including MYC targets and the G2M checkpoint, highlights the twin therapeutic effect of radiotherapy. By suppressing tumour growth and division while concurrently activating the immune system, radiotherapy offers a multifaceted approach to cancer treatment.

6.2.6 PROGENy analysis identifies androgen signalling as a potential target for adjuvant therapy.

Analysis of the PROGENy results, a pathway analysis method that infers pathway activity from gene expression data, revealed distinct patterns of pathway activation and suppression among different sarcoma subtypes following radiotherapy (see section 4.8). These subtype-specific differences highlight the diverse biological responses to treatment. In myxofibrosarcoma, radiotherapy was associated with a significant increase in the activity of pathways involved in stress and immune signalling, including the Androgen, JAK-STAT, p53, NF- κ B, TNF α , and Wnt pathways (Figure 4.14). At the same time, there was a pronounced decrease in PI3K signalling, which may reflect diminished cell survival signalling. These changes suggest following radiotherapy there is enhancement of immune and stress-related responses while suppression of tumour-promoting pathways.

In contrast, synovial sarcoma exhibited a more focused response, with a significant increase observed only in the p53 pathway. While this suggests a more limited activation of stress response pathways, it is also possible that the reduced sample size ($n = 7$) limited the power to detect other significant changes in pathway activity. Similarly, pleomorphic liposarcoma, with only four paired samples available for analysis, showed no statistically significant changes in the pathways analysed. This lack of statistical findings may reflect the challenges of achieving sufficient power in such a small cohort rather than an absence of biological response. Future studies with larger cohorts are warranted to validate and further explore these observations.

To further contextualise these findings, pathway activity scores from radiotherapy-treated samples were compared with treatment-naïve sarcoma samples from the TCGA dataset. This comparison (Figure 4.15) showed that in myxofibrosarcoma, the post-radiotherapy increases in Androgen and TNF α pathway activity and the decrease in PI3K signalling were consistent with the differences observed between TCGA samples and post-radiotherapy samples. These findings add evidence to support the transformative role of radiotherapy in reshaping tumour biology, inducing distinct transcriptomic changes that are not present in untreated sarcomas.

The observed post-radiotherapy increase in the activation of the androgen pathway, raises the possibilities for therapeutic intervention. Preclinical models of desmoplastic small round cell tumours (DSRCTs), have shown that anti-androgen therapies, including enzalutamide and AR-directed antisense oligonucleotides (AR-ASOs), effectively reduce tumour growth by suppressing AR activity and its transcriptional program (Lamhamedi-Cherradi, Maitituoheti *et al.* 2022).

Combining radiotherapy with anti-androgen therapies could provide a novel adjuvant approach to suppress this pathway and limit tumour progression. Specifically, targeting AR could disrupt both direct androgen-mediated tumourigenic signals and the crosstalk between AR and other pathways, such as PI3K, that are modulated by radiotherapy.

6.3 Transcriptomic determinants of radiotherapy resistance

This section explores the findings presented in Chapter 5, which focuses on the differences in gene expression, pathway activity, and tumour microenvironment composition between responders and progressors following radiotherapy. By identifying molecular and cellular mechanisms associated with disease progression, this analysis provides evidence for new biomarkers and therapeutic targets to improving outcomes for patients with soft tissue sarcomas.

6.3.1 Clinical characteristics of responders vs. progressors

Significant differences in disease trajectories were observed between responders and progressors. Responders, who remained recurrence-free, had a median follow-up of 6.8 years post-radiotherapy. In contrast, progressors had a median progression-free interval of 236 days, with most succumbing to their disease within 787 days of completing RT. These findings demonstrate the urgency of identifying predictive biomarkers to better stratify patients for intensified surveillance and treatment to improve these outcomes.

The distribution of histological subtypes differed between responders and progressors, reflecting known biological behaviour. Pleomorphic liposarcoma (pLPS) and myxoid liposarcoma (mLPS), had a 100% progression rate (4/4 and 2/2 cases, respectively), while clear cell sarcoma (CCS), myoepithelial carcinoma (MEC), and alveolar soft part sarcoma (ASPS) were found exclusively in progressors, aligning with their typically poor prognosis. In contrast, myxofibrosarcoma (MFS) and dedifferentiated liposarcoma (ddLPS), which are often considered genomically complex yet potentially radiosensitive, had higher responder rates (MFS: 13/15; ddLPS: 2/3). Extraskeletal myxoid chondrosarcoma (EMC) and malignant peripheral nerve sheath tumour (MPNST) only had responders, though their small sample size limits interpretation.

This variability suggests that histology-specific factors may strongly influence RT response, reinforcing the need for subtype-specific treatment considerations. However, the unequal distribution of subtypes between responders and progressors highlights a potential dataset imbalance, particularly the overrepresentation of pLPS

and mLPS in progressors and the absence of EMC and MPNST from the progressor group. Future studies incorporating larger cohorts or stratifying by histology are needed to determine whether these trends reflect genuine biological differences or sampling bias.

Age differences were also observed, with responders generally older than progressors. The median age at diagnosis was 60 years for responders versus 46 years for progressors ($p = 0.0109$, t-test). While the impact of age on RT response was not explored in depth, this difference should be considered in future analyses, particularly in multivariable models assessing predictors of disease progression.

6.3.2 Differential gene expression analysis identifies potential biomarkers of disease progression.

Differential gene expression (DGE) analysis revealed distinct transcriptional profiles distinguishing responders from progressors. Among the most significant findings, *PI3* and *PTCHD1* were the two most upregulated genes, while *TBC1D3* and *SIRPB1* were the most downregulated (Figure 5.1). These genes and their potential roles will be discussed in detail below. Their expression or lack thereof could serve as a biomarker to predict disease progression.

PI3 (Elafin) was significantly upregulated in progressors post-radiotherapy (46-fold increase), suggesting a role in modulating inflammation and protease activity to enhance tumour resilience. Elevated PI3 expression has been associated with aggressive tumour behaviour, therapy resistance, and immune evasion across multiple cancer types (Saidi, Javerzat *et al.* 2008, Verbovšek, Motaln *et al.* 2014, Tromp, Boerman *et al.* 2020).

PTCHD1, primarily studied in neurodevelopmental disorders, showed a 26-fold increase in progressors. Though originally linked to Hedgehog signalling, recent findings suggest *PTCHD1* interacts with cholesterol rather than Sonic Hedgehog, indicating a role in lipid metabolism. Its upregulation may reflect metabolic adaptation to radiotherapy-induced stress, but conflicting evidence exists regarding its role in

treatment response, with some studies linking higher *PTCHD1* expression to better chemotherapy outcomes (Hiltunen, Timmis *et al.* 2023, Pastore, Muhammad *et al.* 2023).

TBC1D3, an oncogene involved in EGFR and Ras signalling, was 46-fold downregulated in progressors. Its overexpression has been linked to aggressive tumour phenotypes in various cancers (Wainszelbaum, Charron *et al.* 2008, Wang, Chen *et al.* 2021). However, its suppression post-radiotherapy suggests a context-dependent role, potentially reflecting a shift in tumour survival strategies that prioritise DNA repair over proliferation in response to radiation-induced stress.

SIRPB1, a regulator of immune responses, was downregulated 12-fold in progressors. In other cancers, it has been associated with tumour-associated macrophage activation and inflammatory cytokine release (Geng, Zhao *et al.* 2024). Its reduced expression post-RT may contribute to immune evasion, weakening anti-tumour immune responses and promoting tumour persistence (Cerchione, Guadagnuolo *et al.* 2019, Song, Qin *et al.* 2020).

These findings highlight key transcriptional differences in progressors post-radiotherapy, with upregulated genes potentially driving tumour adaptation and survival, while downregulated genes may reflect loss of immune surveillance or alterations in proliferative signalling. The consistent dysregulation of these genes suggests they may play a role in tumour persistence and progression following RT. Further investigation is needed to determine whether these genes could serve as biomarkers for post-RT disease progression or as therapeutic targets to mitigate treatment resistance.

6.4 Summary and future directions

This study provides a comprehensive assessment of transcriptomic responses to radiotherapy in soft tissue sarcomas, identifying key gene expression changes and pathway alterations associated with disease progression. However, several challenges remain, and further work is needed to validate these findings and explore their clinical relevance.

Validation in independent cohorts

Larger, multi-institutional studies are required to confirm the molecular signatures and pathway activities identified in this study. Currently, no publicly available datasets exist that match the specific cohort characteristics and treatment context analysed here. Collaborative efforts to generate and share transcriptomic datasets from post-radiotherapy sarcomas would be essential for external validation.

Spatial transcriptomics and single-cell analysis

Bulk RNA sequencing provides valuable insights but does not resolve the spatial heterogeneity of tumour and immune cell interactions. Spatial transcriptomics and single-cell RNA sequencing could help distinguish whether transcriptomic shifts post-radiotherapy arise from tumour-intrinsic changes or alterations in the tumour microenvironment. These techniques would also enable a more detailed exploration of immune cell infiltration and its role in treatment response.

Multi-Omics integration

Beyond transcriptomics, integrating proteomic and epigenomic data could uncover additional mechanisms of disease progression and therapy resistance. Proteomics may reveal post-transcriptional modifications affecting tumour behaviour, while epigenomic profiling could identify regulatory elements driving differential gene expression post-radiotherapy. Such multi-omics approaches could help refine molecular classifications and identify novel therapeutic vulnerabilities.

In summary, this thesis contributes to our understanding of radiotherapy's molecular impact in soft tissue sarcomas, identifying genomic instability, immune modulation, and transcriptional predictors of disease progression as key factors shaping post-radiotherapy tumour evolution. Integrating these findings into future clinical and translational efforts could advance biomarker-driven risk stratification, optimise patient selection for combination therapies, and maximise radiotherapy efficacy while minimising unnecessary toxicity. Identifying patients for increased surveillance or therapy de-escalation is a goal in radiation oncology and one I hope that this thesis will contribute to the effort of.

Chapter 7. Appendix

7.1 Publications

The following publications have arisen from work carried out during my PhD:

Nottley, S. W. G. and N. Pillay (2024). "Clear cell sarcoma: a rare cause of a lump in the foot." *Diagnostic Histopathology* 30(1): 81-85.

Haefliger, S., O. Chervova, C. Davies, **S. Nottley**, S. Hargreaves, V. P. Sumathi, . . . S. Beck (2023). "Subclassification of epithelioid sarcoma with potential therapeutic impact." *The Journal of Pathology* 260(4): 368-375.

Hames-Fathi, S., **S. W. Nottley** and N. Pillay (2022). "Unravelling undifferentiated soft tissue sarcomas: insights from genomics." *Histopathology* 80(1): 109-121.

7.2 Statement of contributions

Unless otherwise stated, all work in this thesis was carried out by me. I am grateful to the following individuals for their contributions:

Dr Akanksha Farswan provided the variant call format (VCF) files for the patient with Clear Cell Sarcoma (PT61), performing both alignment and variant calling for this case.

Christopher Davies provided the salmon counts from the RNA sequencing data for PT61. He also extracted DNA from the patients' blood for sequencing of the normal samples and from fresh frozen tissue for the samples sent for NanoSeq analysis.

The NanoSeq samples were sequenced at the Wellcome Trust Sanger Institute, who generously conducted the sequencing. VCF files and mutation counts were returned for downstream analysis. I had useful discussions with Dr Federico Abascal regarding this work. I subsequently performed mutational signature analysis on the identified mutations.

Dr Shadi Hames-Fathi provided the processed TCGA RNA sequencing counts data, which I used for comparative analysis with my cohort in the PROGENy analysis.

Reference List

- Abaricia, S. and A. C. Hirbe (2018). "Diagnosis and treatment of myxoid liposarcomas: histology matters." Current treatment options in oncology **19**(12): 64.
- Abascal, F., L. M. R. Harvey, E. Mitchell, A. R. J. Lawson, S. V. Lensing, P. Ellis, . . . I. Martincorena (2021). "Somatic mutation landscapes at single-molecule resolution." Nature **593**(7859): 405-410.
- Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, . . . S. R. Sunyaev (2010). "A method and server for predicting damaging missense mutations." Nature Methods **7**(4): 248-249.
- Aggerholm-Pedersen, N., B. S. Sørensen, J. Overgaard, K. Toustrup, S. Baerentzen, O. S. Nielsen, . . . A. Safwat (2016). "A prognostic profile of hypoxia-induced genes for localised high-grade soft tissue sarcoma." British journal of cancer **115**(9): 1096-1104.
- Ahmed, M. M., B. Malachowska and C. Guha (2025). "Radiation-Induced Tissue Regeneration: Pathways, Mechanisms, and Therapeutic Potential." Hematology/Oncology Clinics.
- Alexandrov, L. B., J. Kim, N. J. Haradhvala, M. N. Huang, A. W. Tian Ng, Y. Wu, . . . P. Consortium (2020). "The repertoire of mutational signatures in human cancer." Nature **578**(7793): 94-101.
- Alexandrov, L. B., S. Nik-Zainal, D. C. Wedge, S. A. Aparicio, S. Behjati, A. V. Biankin, . . . A.-L. Børresen-Dale (2013). "Signatures of mutational processes in human cancer." nature **500**(7463): 415-421.
- Alim, L. F., C. Keane and F. Souza-Fonseca-Guimaraes (2024). "Molecular mechanisms of tumour necrosis factor signalling via TNF receptor 1 and TNF receptor 2 in the tumour microenvironment." Current Opinion in Immunology **86**: 102409.
- Anderson, W. J. and V. Y. Jo (2019). Pleomorphic liposarcoma: updates and current differential diagnosis. Seminars in Diagnostic Pathology, Elsevier.
- Angel, M., M. Zarba and J. P. Sade (2021). "PARP inhibitors as a radiosensitizer: a future promising approach in prostate cancer?" Ecancermedicalscience **15**: ed118.
- Ansems, M. and P. N. Span (2020). "The tumor microenvironment and radiotherapy response; a central role for cancer-associated fibroblasts." Clinical and Translational Radiation Oncology **22**: 90-97.
- Aran, D., Z. Hu and A. J. Butte (2017). "xCell: digitally portraying the tissue cellular heterogeneity landscape." Genome Biology **18**(1): 220.
- Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, . . . G. Sherlock (2000). "Gene Ontology: tool for the unification of biology." Nature Genetics **25**(1): 25-29.

- Bacon, A., K. Wong, M. S. Fernando, B. Rous, R. J. W. Hill, S. D. Collins, . . . S. J. Strauss (2023). "Incidence and survival of soft tissue sarcoma in England between 2013 and 2017, an analysis from the National Cancer Registration and Analysis Service." Int J Cancer **152**(9): 1789-1803.
- Bailey, C., O. Pich, K. Thol, T. B. Watkins, J. Luebeck, A. Rowan, . . . R. Bentham (2024). "Origins and impact of extrachromosomal DNA." Nature **635**(8037): 193-200.
- Balkwill, F. (2009). "Tumour necrosis factor and cancer." Nat Rev Cancer **9**(5): 361-371.
- Barazzuol, L., R. P. Coppes and P. van Luijk (2020). "Prevention and treatment of radiotherapy - induced side effects." Molecular oncology **14**(7): 1538-1554.
- Barretina, J., B. S. Taylor, S. Banerji, A. H. Ramos, M. Lagos-Quintana, P. L. Decarolis, . . . S. Singer (2010). "Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy." Nat Genet **42**(8): 715-721.
- Baskar, R., K. A. Lee, R. Yeo and K. W. Yeoh (2012). "Cancer and radiation therapy: current advances and future directions." Int J Med Sci **9**(3): 193-199.
- Beach, C., D. MacLean, D. Majorova, J. N. Arnold and M. M. Olcina (2022). "The effects of radiation therapy on the macrophage response in cancer." Front Oncol **12**: 1020606.
- Becquerel, H. (1896). "Sur les radiations émises par phosphorescence." Comptes rendus de l'Academie des Sciences, Paris **122**: 420-421.
- Behjati, S., G. Gundem, D. C. Wedge, N. D. Roberts, P. S. Tarpey, S. L. Cooke, . . . P. J. Campbell (2016). "Mutational signatures of ionizing radiation in second malignancies." Nat Commun **7**: 12605.
- Benelli, M., C. Pescucci, G. Marseglia, M. Severgnini, F. Torricelli and A. Magi (2012). "Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript." Bioinformatics **28**(24): 3232-3239.
- Benjamin, D., T. Sato, K. Cibulskis, G. Getz, C. Stewart and L. Lichtenstein (2019). "Calling somatic SNVs and indels with Mutect2." BioRxiv: 861054.
- Bertucci, F., P. Finetti, J. Ostrowski, W. Kim, H. Kim, M. A. Pantaleo, . . . D. Birnbaum (2012). "Genomic Grade Index predicts postoperative clinical outcome of GIST." British journal of cancer **107**(8): 1433-1441.
- Betgen, A., R. L. Haas and J. J. Sonke (2013). "Volume changes in soft tissue sarcomas during preoperative radiotherapy of extremities evaluated using cone-beam CT." J Radiat Oncol **2**(1): 55-62.
- Blomain, E., S. Soudi, Z. Wang, A. Somani, A. Subramanian, S. C. L. Nouth, . . . E. J. Moding (2025). "Evolutionary Pressures Shape Undifferentiated Pleomorphic Sarcoma Development and Radiotherapy Response." Cancer Res.

- Borrego-Soto, G., R. Ortiz-Lopez and A. Rojas-Martinez (2015). "Ionizing radiation-induced DNA injury and damage detection in patients with breast cancer." Genet Mol Biol **38**(4): 420-432.
- Brahmi, M., H. Vanacker, N. Macagno, F. Tirode and A. Dufresne (2022). "CIC-DUX4 sarcomas." Curr Opin Oncol **34**(4): 342-347.
- Breiman, L. (2001). "Random Forests." Machine Learning **45**(1): 5-32.
- Bryant, J. M., J. Weygand, E. Keit, R. Cruz-Chamorro, M. L. Sandoval, I. M. Oraiqat, . . . V. Feygelman (2023). "Stereotactic magnetic resonance-guided adaptive and non-adaptive radiotherapy on combination MR-linear accelerators: current practice and future directions." Cancers **15**(7): 2081.
- Bui, B. P., P. L. Nguyen, K. Lee and J. Cho (2022). "Hypoxia-Inducible Factor-1: A Novel Therapeutic Target for the Management of Cancer, Drug Resistance, and Cancer-Related Pain." Cancers (Basel) **14**(24).
- Casali, P., N. Abecassis, S. Bauer, R. Biagini, S. Bielack, S. Bonvalot, . . . J. Broto (2018). "Soft tissue and visceral sarcomas: ESMO–EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up." Annals of Oncology **29**: iv51-iv67.
- Caso, R., F. Sanchez-Vega, K. S. Tan, B. Mastrogiacomio, J. Zhou, G. D. Jones, . . . D. R. Jones (2020). "The Underlying Tumor Genomics of Predominant Histologic Subtypes in Lung Adenocarcinoma." J Thorac Oncol **15**(12): 1844-1856.
- Cerami, E., J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, . . . N. Schultz (2012). "The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data." Cancer Discov **2**(5): 401-404.
- Cerchione, C., V. Guadagnuolo, C. Papayannidis, I. Iacobucci, A. Padella, G. Simonetti, . . . F. Volpato (2019). "PS997 A NEW BIOMARKER OF RESPONSE TO 5 - AZACITIDINE THERAPY IN MDS AND AML PATIENTS: SIRPB1." HemaSphere **3**: 448-449.
- Chapman, B., R. Kirchner, L. Pantano, S. Naumenko, M. De Smet, L. Beltrame, . . . J. Yoon (2021). "bcbio/bcbio-nextgen: (v1.2.9)." Zenodo.
- Chaudhry, M. A. (2008). "Analysis of gene expression in Normal and Cancer cells exposed to γ -radiation." Journal of Biomedicine and Biotechnology **2008**.
- Cheng, D. T., T. N. Mitchell, A. Zehir, R. H. Shah, R. Benayed, A. Syed, . . . M. F. Berger (2015). "Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology." J Mol Diagn **17**(3): 251-264.
- Chibon, F., P. Lagarde, S. Salas, G. Pérot, V. Brouste, F. Tirode, . . . A. Aurias (2010). "Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity." Nature Medicine **16**(7): 781-787.

- Chida, K., M. Oshi, A. M. Roy, T. Yachi, M. Nara, K. Yamada, . . . K. Takabe (2023). "E2F target score is associated with cell proliferation and survival of patients with hepatocellular carcinoma." Surgery **174**(2): 307-314.
- Cho, B. (2018). "Intensity-modulated radiation therapy: a review with a physics perspective." Radiat Oncol J **36**(1): 1-10.
- Chung, P. W., B. M. Deheshi, P. C. Ferguson, J. S. Wunder, A. M. Griffin, C. N. Catton, . . . B. O'Sullivan (2009). "Radiosensitivity translates into excellent local control in extremity myxoid liposarcoma: a comparison with other soft tissue sarcomas." Cancer **115**(14): 3254-3261.
- Connell, P. P. and S. Hellman (2009). "Advances in radiotherapy and implications for the next century: a historical perspective." Cancer research **69**(2): 383-392.
- Correa, H. (2016). "Li–Fraumeni syndrome." Journal of pediatric genetics **5**(02): 084-088.
- Cortés-Ciriano, I., J. J.-K. Lee, R. Xi, D. Jain, Y. L. Jung, L. Yang, . . . D. S. Pellman (2020). "Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing." Nature genetics **52**(3): 331-341.
- Craparotta, I., L. Mannarino, R. Zadro, S. Ballabio, S. Marchini, G. Pavesi, . . . R. Frapolli (2024). "Mechanism of efficacy of trabectedin against myxoid liposarcoma entails detachment of the FUS-DDIT3 transcription factor from its DNA binding sites." J Exp Clin Cancer Res **43**(1): 309.
- Curie, P. (1898). "Sur une nouvelle substance fortement radioactive, contenue dans la pechblende." Comptes rendus de l'Academie des Sciences, Paris **127**: 1215-1217.
- Cyril Fisher, K. T. (2022) "Dataset for histopathological reporting of soft tissue sarcomas."
- de la Chaux, N., P. W. Messer and P. F. Arndt (2007). "DNA indels in coding regions reveal selective constraints on protein evolution in the human lineage." BMC Evol Biol **7**: 191.
- Deloch, L., A. Derer, J. Hartmann, B. Frey, R. Fietkau and U. S. Gaipl (2016). "Modern radiotherapy concepts and the impact of radiation on immune activation." Frontiers in oncology **6**: 141.
- Dermawan, J. K., P. Chi, W. D. Tap, E. Rosenbaum, S. D'Angelo, K. M. Alektiar and C. R. Antonescu (2023). "Distinct genomic landscapes in radiation-associated angiosarcoma compared with other radiation-associated sarcoma histologies." J Pathol **260**(4): 465-477.
- Despeignes, V. (1896). "Observation concernant un cas de cancer de l'estomac traite par les rayons Rontgen." Lyon méd. **82**: 428-430; 503-506.
- Díaz Casas, S. E., J. M. Villacrés, C. Lehmann Mosquera, M. García Mora, I. Mariño Lozano, J. Ángel Aristizábal, . . . R. Sánchez Pedraza (2024). "Prognostic Factors Associated with Tumor Recurrence and Overall Survival in Soft Tissue Sarcomas of

the Extremities in a Colombian Reference Cancer Center." Current Oncology **31**(4): 1725-1738.

Dickson, M. A. (2014). "Systemic Treatment Options for Radiation-Associated Sarcomas." Current Treatment Options in Oncology **15**(3): 476-481.

Dona Lemus, O. M., M. Cao, B. Cai, M. Cummings and D. Zheng (2024). "Adaptive radiotherapy: next-generation radiotherapy." Cancers **16**(6): 1206.

Dong, Y., Q. He, X. Chen, F. Yang, L. He and Y. Zheng (2023). "Extrachromosomal DNA (ecDNA) in cancer: mechanisms, functions, and clinical implications." Frontiers in oncology **13**: 1194405.

Doyle, L. A., E. Möller, P. Dal Cin, C. D. Fletcher, F. Mertens and J. L. Hornick (2011). "MUC4 is a highly sensitive and specific marker for low-grade fibromyxoid sarcoma." The American journal of surgical pathology **35**(5): 733-741.

Ercan, H., L.-M. Mauracher, E. Grilz, L. Hell, R. Hellinger, J. A. Schmid, . . . M. Zellner (2021). "Alterations of the platelet proteome in lung cancer: accelerated F13A1 and ER processing as new actors in hypercoagulability." Cancers **13**(9): 2260.

Ernestos, B., P. Nikolaos, G. Koulis, R. Eleni, B. Konstantinos, G. Alexandra and K. Michael (2010). "Increased Chromosomal Radiosensitivity in Women Carrying BRCA1/BRCA2 Mutations Assessed With the G2 Assay." International Journal of Radiation Oncology, Biology, Physics **76**(4): 1199-1205.

Ewels, P., M. Magnusson, S. Lundin and M. Käller (2016). "MultiQC: summarize analysis results for multiple tools and samples in a single report." Bioinformatics **32**(19): 3047-3048.

Foray, N. (2012). "[Claudius Regaud (1870-1940): A pioneer of radiobiology and radiotherapy]." Cancer Radiother **16**(4): 315-321.

Forker, L., P. Gaunt, S. Sioletic, P. Shenjere, R. Potter, D. Roberts, . . . A. Hughes (2018). "The hypoxia marker CAIX is prognostic in the UK phase III Vortex-Biobank cohort: an important resource for translational research in soft tissue sarcoma." British journal of cancer **118**(5): 698-704.

Gambella, A., L. Bertero, M. Rondón-Lagos, L. Verdun Di Cantogno, N. Rangel, C. Pitino, . . . P. Cassoni (2023). "FISH Diagnostic Assessment of MDM2 Amplification in Liposarcoma: Potential Pitfalls and Troubleshooting Recommendations." Int J Mol Sci **24**(2).

Garcia, F. A. O., E. S. de Andrade and E. I. Palmero (2022). "Insights on variant analysis in silico tools for pathogenicity prediction." Front Genet **13**: 1010327.

García, M. E. G., D. G. Kirsch and Z. J. Reitman (2022). "Targeting the ATM Kinase to Enhance the Efficacy of Radiotherapy and Outcomes for Cancer Patients." Seminars in Radiation Oncology **32**(1): 3-14.

Gazendam, A. M., S. Popovic, S. Munir, N. Parasu, D. Wilson and M. Ghert (2021). "Synovial sarcoma: a clinical review." Current Oncology **28**(3): 1909-1920.

- Geng, R., Y. Zhao, W. Xu, X. Ma, Y. Jiang, X. Han, . . . Y. Li (2024). "SIRPB1 regulates inflammatory factor expression in the glioma microenvironment via SYK: functional and bioinformatics insights." Journal of Translational Medicine **22**(1): 338.
- Gerlinger, M., A. J. Rowan, S. Horswell, M. Math, J. Larkin, D. Endesfelder, . . . C. Swanton (2012). "Intratumor heterogeneity and branched evolution revealed by multiregion sequencing." N Engl J Med **366**(10): 883-892.
- Gouel, P., P. Decazes, P. Vera, I. Gardin, S. Thureau and P. Bohn (2023). "Advances in PET and MRI imaging of tumor hypoxia." Front Med (Lausanne) **10**: 1055062.
- Gounder, M. M., N. P. Agaram, S. E. Trabucco, V. Robinson, R. A. Ferraro, S. Z. Millis, . . . D. X. Jin (2022). "Clinical genomic profiling in the management of patients with soft tissue and bone sarcoma." Nat Commun **13**(1): 3406.
- Graham, L. S., C. C. Pritchard and M. T. Schweizer (2021). "Hypermutation, Mismatch Repair Deficiency, and Defining Predictors of Response to Checkpoint Blockade." Clin Cancer Res **27**(24): 6662-6665.
- Grégoire, V., M. Guckenberger, K. Haustermans, J. J. W. Lagendijk, C. Ménard, R. Pötter, . . . D. Zips (2020). "Image guidance in radiation therapy for better cure of cancer." Mol Oncol **14**(7): 1470-1491.
- Grigoriadis, K., A. Huebner, A. Bunkum, E. Colliver, A. M. Frankell, M. S. Hill, . . . N. McGranahan (2024). "CONIPHER: a computational framework for scalable phylogenetic reconstruction with error correction." Nature Protocols **19**(1): 159-183.
- Gronchi, A. (2015). "Individualizing the use/non-use of radiation therapy (RT) in soft tissue sarcoma (STS): When abstention is better than care." J Surg Oncol **111**(2): 133-134.
- Guillou, L., J. M. Coindre, F. Bonichon, B. B. Nguyen, P. Terrier, F. Collin, . . . J. Costa (1997). "Comparative study of the National Cancer Institute and French Federation of Cancer Centers Sarcoma Group grading systems in a population of 410 adult patients with soft tissue sarcoma." J Clin Oncol **15**(1): 350-362.
- Guo, Q., J. Zhao, X. Du and B. Huang (2022). "Survival outcomes of surgery for retroperitoneal sarcomas: A systematic review and meta-analysis." PLOS ONE **17**(7): e0272044.
- Gutmann, D. H., R. E. Ferner, R. H. Listernick, B. R. Korf, P. L. Wolters and K. J. Johnson (2017). "Neurofibromatosis type 1." Nature Reviews Disease Primers **3**(1): 17004.
- Hames-Fathi, S., S. W. Nottley and N. Pillay (2022). "Unravelling undifferentiated soft tissue sarcomas: insights from genomics." Histopathology **80**(1): 109-121.
- Han, D., B. Zhang, W. Chang, A. Song, Y. Wang, P. Zhao, . . . P. Shang (2022). "A pan-cancer analysis of the hemoglobin subunit beta (hbb) in human tumors." Available at SSRN 4245565.
- Hastie, T. and J. Qian (2014). "Glmnet vignette." Retrieved June 9(2016): 1-30.

- Hayes, A. J., I. F. Nixon, D. C. Strauss, B. M. Seddon, A. Desai, C. Benson, . . . A. Dangoor (2024). "UK guidelines for the management of soft tissue sarcomas." British Journal of Cancer.
- Hiltunen, M. K., A. J. Timmis, M. Thomsen, D. S. Gkotsi, H. Iwai, O. M. Ribeiro, . . . N. A. Riobo-Del Galdo (2023). "PTCHD1 binds cholesterol but not sonic hedgehog, suggesting a distinct cellular function." International Journal of Molecular Sciences **24**(3): 2682.
- Hoesel, B. and J. A. Schmid (2013). "The complexity of NF- κ B signaling in inflammation and cancer." Molecular cancer **12**: 1-15.
- Hopkins, J. J., M. N. Wakeling, M. B. Johnson, S. E. Flanagan and T. W. Laver (2023). "REVEL Is Better at Predicting Pathogenicity of Loss - of - Function than Gain - of - Function Variants." Human Mutation **2023**(1): 8857940.
- Hornick, J. L. and C. D. Fletcher (2003). "Myoepithelial tumors of soft tissue: a clinicopathologic and immunohistochemical study of 101 cases with evaluation of prognostic parameters." The American journal of surgical pathology **27**(9): 1183-1196.
- Huang, C.-Y., P.-L. Wei, U. Batzorig, P. T. Makondi, C.-C. Lee and Y.-J. Chang (2023). "Identification of Moesin (MSN) as a Potential Therapeutic Target for Colorectal Cancer via the β -Catenin-RUNX2 Axis." International Journal of Molecular Sciences **24**(13): 10951.
- Huang, Y., W. Hong and X. Wei (2022). "The molecular mechanisms and therapeutic strategies of EMT in tumor progression and metastasis." Journal of Hematology & Oncology **15**(1): 129.
- Ibrahim, R. M., S. S. Jensen and J. Juel (2018). "Clear cell sarcoma-A review." Journal of orthopaedics **15**(4): 963-966.
- Inchaustegui, M. L., K. Kon-Liao, K. Ruiz-Arellanos, G. A. E. Silva, M. R. Gonzalez and J. Pretell-Mazzini (2023). "Treatment and Outcomes of Radiation-Induced Soft Tissue Sarcomas of the Extremities and Trunk—A Systematic Review of the Literature." Cancers **15**(23): 5584.
- Ioannidis, N. M., J. H. Rothstein, V. Pejaver, S. Middha, S. K. McDonnell, S. Baheti, . . . D. Karyadi (2016). "REVEL: an ensemble method for predicting the pathogenicity of rare missense variants." The American Journal of Human Genetics **99**(4): 877-885.
- Izumi, T., P. Rychahou, L. Chen, M. H. Smith and J. Valentino (2023). "Copy number variation that influences the ionizing radiation sensitivity of oral squamous cell carcinoma." Cells **12**(20): 2425.
- Jamal-Hanjani, M., G. A. Wilson, N. McGranahan, N. J. Birkbak, T. B. Watkins, S. Veeriah, . . . R. Rosenthal (2017). "Tracking the evolution of non–small-cell lung cancer." New England Journal of Medicine **376**(22): 2109-2121.
- Janopaul-Naylor, J. R., Y. Shen, D. C. Qian and Z. S. Buchwald (2021). "The abscopal effect: a review of pre-clinical and clinical advances." International journal of molecular sciences **22**(20): 11061.

- Jansen, P., H. Müller, G. C. Lodde, A. Zaremba, I. Möller, A. Sucker, . . . K. G. Griewank (2021). "GNA14, GNA11, and GNAQ Mutations Are Frequent in Benign but Not Malignant Cutaneous Vascular Tumors." Frontiers in Genetics **12**.
- Jiang, J., H. Yu, Y. Shou, G. Neale, S. Zhou, T. Lu and B. P. Sorrentino (2010). "Hemgn is a direct transcriptional target of HOXB4 and induces expansion of murine myeloid progenitor cells." Blood, The Journal of the American Society of Hematology **116**(5): 711-719.
- Johnson, D. E., R. A. O'Keefe and J. R. Grandis (2018). "Targeting the IL-6/JAK/STAT3 signalling axis in cancer." Nature Reviews Clinical Oncology **15**(4): 234-248.
- Jorgovanovic, D., M. Song, L. Wang and Y. Zhang (2020). "Roles of IFN- γ in tumor progression and regression: a review." Biomarker Research **8**(1): 49.
- Kane, M. F., M. Loda, G. M. Gaida, J. Lipman, R. Mishra, H. Goldman, . . . R. Kolodner (1997). "Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines." Cancer Res **57**(5): 808-811.
- Kaneko, K., K. Furuyama, H. Aburatani and S. Shibahara (2009). "Hypoxia induces erythroid - specific 5 - aminolevulinate synthase expression in human erythroid cells through transforming growth factor - β signaling." The FEBS Journal **276**(5): 1370-1382.
- Kao, T.-W., G.-H. Bai, T.-L. Wang, I.-M. Shih, C.-M. Chuang, C.-L. Lo, . . . Y.-A. Shen (2023). "Novel cancer treatment paradigm targeting hypoxia-induced factor in conjunction with current therapies to overcome resistance." Journal of Experimental & Clinical Cancer Research **42**(1): 171.
- Karin, M. (2006). "Nuclear factor- κ B in cancer development and progression." Nature **441**(7092): 431-436.
- Keung, E. Z., Y.-J. Chiang, R. K. Voss, J. N. Cormier, K. E. Torres, K. K. Hunt, . . . C. L. Roland (2018). "Defining the incidence and clinical significance of lymph node metastasis in soft tissue sarcoma." European Journal of Surgical Oncology **44**(1): 170-177.
- Kim, H., N.-P. Nguyen, K. Turner, S. Wu, A. D. Gujar, J. Luebeck, . . . S. Namburi (2020). "Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers." Nature genetics **52**(9): 891-897.
- Kim, K., E.-Y. Choi, H.-M. Ahn, D.-G. Kim and Y.-J. Kim (2023). "Hemoglobin Subunit Theta 1 Promotes Proliferation by Reducing Reactive Oxygen Species in Lung Adenocarcinoma." Cancers **15**(23): 5504.
- Kim, S., K. Scheffler, A. L. Halpern, M. A. Bekritsky, E. Noh, M. Källberg, . . . P. Krusche (2018). "Strelka2: fast and accurate calling of germline and somatic variants." Nature methods **15**(8): 591-594.
- Kleinerman, R. A., S. J. Schonfeld, B. S. Sigel, J. R. Wong-Siegel, E. S. Gilbert, D. H. Abramson, . . . L. M. Morton (2019). "Bone and Soft-Tissue Sarcoma Risk in Long-

Term Survivors of Hereditary Retinoblastoma Treated With Radiation." J Clin Oncol **37**(35): 3436-3445.

Kobialka, P. and M. Graupera (2019). "Revisiting PI3-kinase signalling in angiogenesis." Vascular Biology **1**(1): H125-H134.

Kocakavuk, E., K. J. Anderson, F. S. Varn, K. C. Johnson, S. B. Amin, E. P. Sulman, . . . R. G. W. Verhaak (2021). "Radiotherapy is associated with a deletion signature that contributes to poor outcomes in patients with cancer." Nat Genet **53**(7): 1088-1096.

Koelsche, C., D. Schrimpf, D. Stichel, M. Sill, F. Sahm, D. E. Reuss, . . . K. Beck (2021). "Sarcoma classification by DNA methylation profiling." Nature communications **12**(1): 498.

Korotkevich, G., V. Sukhov, N. Budin, B. Shpak, M. N. Artyomov and A. Sergushichev (2016). "Fast gene set enrichment analysis." bioRxiv: 060012.

Kou, F., L. Wu, Y. Guo, B. Zhang, B. Li, Z. Huang, . . . L. Yang (2021). "Somatic copy number alterations are predictive of progression-free survival in patients with lung adenocarcinoma undergoing radiotherapy." Cancer Biol Med **19**(5): 685-695.

Kuhn, M., J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, . . . R. C. Team (2020). "Package 'caret'." The R Journal **223**(7): 48.

Kumar, P., S. Henikoff and P. C. Ng (2009). "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." Nature Protocols **4**(7): 1073-1081.

Kumari, S., A. K. Badana, M. M. G, S. G and R. Malla (2018). "Reactive Oxygen Species: A Key Constituent in Cancer Survival." Biomark Insights **13**: 1177271918755391.

Kurota, Y., Y. Takeda, O. Ichiyanagi, S. Saitoh, H. Ito, S. Naito, . . . N. Tsuchiya (2023). "Hemoglobin β expression is associated with poor prognosis in clear cell renal cell carcinoma." Biomedicines **11**(5): 1330.

Lai, Z., A. Markovets, M. Ahdesmaki, B. Chapman, O. Hofmann, R. McEwen, . . . J. R. Dry (2016). "VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research." Nucleic acids research **44**(11): e108-e108.

Lamhamedi-Cherradi, S.-E., M. Maitituoheti, B. A. Menegaz, S. Krishnan, A. M. Vetter, P. Camacho, . . . J. A. Ludwig (2022). "The androgen receptor is a therapeutic target in desmoplastic small round cell sarcoma." Nature Communications **13**(1): 3057.

Landrum, M. J., J. M. Lee, M. Benson, G. Brown, C. Chao, S. Chitipiralla, . . . J. Hoover (2016). "ClinVar: public archive of interpretations of clinically relevant variants." Nucleic acids research **44**(D1): D862-D868.

Lebas, A., C. Le Fèvre, W. Waissi, I. Chambrelant, D. Brinkert and G. Noël (2023). "Prognostic Factors in Extremity Soft Tissue Sarcomas Treated with Radiotherapy: Systematic Review of the Literature." Cancers **15**(18): 4486.

- Lee, D. W., H. S. Kim and I. Han (2021). "Actual long-term survival after resection of stage III soft tissue sarcoma." BMC Cancer **21**(1): 21.
- Lehrer, S., F. R. Dembitzer, P. H. Rheinstein and K. E. Rosenzweig (2018). "In primary glioblastoma fewer tumor copy number segments of the F13A1 gene are associated with poorer survival." Thromb Res **167**: 12-14.
- Lerebours, F., S. Vacher, J. Guinebretiere, S. Rondeau, M. Caly, D. Gentien, . . . X. Liang (2021). "Hemoglobin overexpression and splice signature as new features of inflammatory breast cancer?" Journal of advanced research **28**: 77-85.
- Lesluyes, T., J. Baud, G. Pérot, C. Charon-Barra, A. You, I. Valo, . . . S. Renard-Oldrini (2019). "Genomic and transcriptomic comparison of post-radiation versus sporadic sarcomas." Modern Pathology **32**(12): 1786-1794.
- Lesluyes, T., J. Baud, G. Pérot, C. Charon-Barra, A. You, I. Valo, . . . F. Chibon (2019). "Genomic and transcriptomic comparison of post-radiation versus sporadic sarcomas." Mod Pathol **32**(12): 1786-1794.
- Li, H. (2013). "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM." arXiv preprint arXiv:1303.3997.
- Li, X., Z. Wu, Y. Wang, Q. Mei, X. Fu and W. Han (2013). "Characterization of adult α - and β -globin elevated by hydrogen peroxide in cervical cancer cells that play a cytoprotective role against oxidative insults." PLoS One **8**(1): e54342.
- Liaw, A. and M. Wiener (2002). "Classification and regression by randomForest." R news **2**(3): 18-22.
- Liberzon, A., C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov and P. Tamayo (2015). "The Molecular Signatures Database (MSigDB) hallmark gene set collection." Cell Syst **1**(6): 417-425.
- Lin, B., F. Gao, Y. Yang, D. Wu, Y. Zhang, G. Feng, . . . X. Du (2021). "FLASH radiotherapy: history and future." Frontiers in oncology **11**: 644400.
- Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome biology **15**: 1-21.
- Martincorena, I., K. M. Raine, M. Gerstung, K. J. Dawson, K. Haase, P. Van Loo, . . . P. J. Campbell (2017). "Universal Patterns of Selection in Cancer and Somatic Tissues." Cell **171**(5): 1029-1041.e1021.
- Mavroeidis, L., A. Napolitano, P. Huang and R. L. Jones (2024). "Novel Therapeutics in Soft Tissue Sarcoma." Cancers **17**(1): 10.
- McKelvey, K. J., A. L. Hudson, M. Back, T. Eade and C. I. Diakos (2018). "Radiation, inflammation and the immune response in cancer." Mammalian genome **29**(11): 843-865.
- McLaren, W., L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, . . . F. Cunningham (2016). "The Ensembl Variant Effect Predictor." Genome Biology **17**(1): 122.

- Messiou, C., S. Bonvalot, A. Gronchi, D. Vanel, M. Meyer, P. Robinson, . . . R. L. Haas (2016). "Evaluation of response after pre-operative radiotherapy in soft tissue sarcomas; the European Organisation for Research and Treatment of Cancer-Soft Tissue and Bone Sarcoma Group (EORTC-STBSG) and Imaging Group recommendations for radiological examination and reporting with an emphasis on magnetic resonance imaging." Eur J Cancer **56**: 37-44.
- Moelans, C. B., C. M. G. van Maldegem, E. van der Wall and P. J. van Diest (2018). "Copy number changes at 8p11-12 predict adverse clinical outcome and chemo- and radiotherapy response in breast cancer." Oncotarget **9**(24): 17078-17092.
- Mohan, R. (2022). "A review of proton therapy–Current status and future directions." Precision radiation oncology **6**(2): 164-176.
- Monsky, W. L., B. Jin, C. Molloy, R. J. Canter, C. S. Li, T. C. Lin, . . . A. J. Chaudhari (2012). "Semi-automated volumetric quantification of tumor necrosis in soft tissue sarcoma using contrast-enhanced MRI." Anticancer Res **32**(11): 4951-4961.
- Morgan, M. A. and T. S. Lawrence (2015). "Molecular Pathways: Overcoming Radiation Resistance by Targeting DNA Damage Response Pathways." Clin Cancer Res **21**(13): 2898-2904.
- Morton, L. M., D. M. Karyadi, C. Stewart, T. I. Bogdanova, E. T. Dawson, M. K. Steinberg, . . . S. J. Chanock (2021). "Radiation-related genomic profile of papillary thyroid carcinoma after the Chernobyl accident." Science **372**(6543).
- Nacev, B. A., F. Sanchez-Vega, S. A. Smith, C. R. Antonescu, E. Rosenbaum, H. Shi, . . . W. D. Tap (2022). "Clinical sequencing of soft tissue and bone sarcomas delineates diverse genomic landscapes and potential therapeutic targets." Nat Commun **13**(1): 3405.
- Nichelli, L. and S. Casagrande (2021). "Current emerging MRI tools for radionecrosis and pseudoprogression diagnosis." Curr Opin Oncol **33**(6): 597-607.
- Nicolas, M. M., P. Tamboli, J. A. Gomez and B. A. Czerniak (2010). "Pleomorphic and dedifferentiated leiomyosarcoma: clinicopathologic and immunohistochemical study of 41 cases." Human pathology **41**(5): 663-671.
- Noeueglise, M., W. Tessier, M. Barthoulot, G. Decanter, A. Cayeux, H. Marin, . . . A. Cordoba (2024). "Preoperative versus postoperative radiotherapy for localized soft tissue sarcoma treated with curative intent in a French tertiary center "SARCLOC". " BMC Cancer **24**(1): 1550.
- O'Donnell III, E., M. Muñoz, R. Davis, J. Bergonio, R. L. Randall, C. Tepper and J. Carr-Ascher (2025). "Genetic and epigenetic characterization of sarcoma stem cells across subtypes identifies EZH2 as a therapeutic target." npj Precision Oncology **9**(1): 7.
- O'Sullivan, B., A. M. Davis, R. Turcotte, R. Bell, C. Catton, P. Chabot, . . . B. Zee (2002). "Preoperative versus postoperative radiotherapy in soft-tissue sarcoma of the limbs: a randomised trial." Lancet **359**(9325): 2235-2241.

- Paoluzzi, L. and R. G. Maki (2019). "Diagnosis, prognosis, and treatment of alveolar soft-part sarcoma: a review." JAMA oncology **5**(2): 254-260.
- Pariyar, M., A. Johns, R. F. Thorne, R. J. Scott and K. A. Avery-Kiejda (2021). "Copy number variation in triple negative breast cancer samples associated with lymph node metastasis." Neoplasia **23**(8): 743-753.
- Park, H. J., R. J. Griffin, S. Hui, S. H. Levitt and C. W. Song (2012). "Radiation-induced vascular damage in tumors: implications of vascular damage in ablative hypofractionated radiotherapy (SBRT and SRS)." Radiation research **177**(3): 311-327.
- Pasquali, S., E. Palmerini, V. Quagliuolo, J. Martin - Broto, A. Lopez - Pousa, G. Grignani, . . . R. Diaz - Beveridge (2022). "Neoadjuvant chemotherapy in high - risk soft tissue sarcomas: a Sarculator - based risk stratification analysis of the ISG - STS 1001 randomized trial." Cancer **128**(1): 85-93.
- Pastore, S. F., T. Muhammad, C. Stan, P. W. Frankland, P. A. Hamel and J. B. Vincent (2023). "Neuronal transcription of autism gene PTCHD1 is regulated by a conserved downstream enhancer sequence." Scientific Reports **13**(1): 20391.
- Patergnani, S., A. Danese, E. Bouhamida, G. Aguiari, M. Previati, P. Pinton and C. Giorgi (2020). "Various aspects of calcium signaling in the regulation of apoptosis, autophagy, cell proliferation, and cancer." International journal of molecular sciences **21**(21): 8323.
- Patro, R., G. Duggal, M. I. Love, R. A. Irizarry and C. Kingsford (2017). "Salmon provides fast and bias-aware quantification of transcript expression." Nat Methods **14**(4): 417-419.
- Pećina-Šlaus, N., A. Kafka, I. Salamon and A. Bukovac (2020). "Mismatch Repair Pathway, Genome Stability and Cancer." Front Mol Biosci **7**: 122.
- Peltier, J., J.-P. Roperch, S. Audebert, J.-P. Borg and L. Camoin (2018). "Activation peptide of the coagulation factor XIII (AP-F13A1) as a new biomarker for the screening of colorectal cancer." Clinical Proteomics **15**: 1-11.
- Pennacchioli, E., G. Tosti, M. Barberis, T. M. De Pas, F. Verrecchia, C. Menicanti, . . . G. Mazzarol (2012). "Sarcoma spreads primarily through the vascular system: are there biomarkers associated with vascular spread?" Clinical & experimental metastasis **29**: 757-773.
- Phan, N. N., C. Y. Wang, C. F. Chen, Z. Sun, M. D. Lai and Y. C. Lin (2017). "Voltage-gated calcium channels: Novel targets for cancer therapy." Oncology letters **14**(2): 2059-2074.
- Phipps, W., B. Bhinder, A. Towler, P. Mooka, J. Kafeero, M. Fitzgibbon, . . . E. Cesarman (2025). "Exome sequencing reveals a sparse genomic landscape in Kaposi sarcoma." Mol Cancer Res.
- Pisters, P. W., L. B. Harrison, D. H. Leung, J. M. Woodruff, E. S. Casper and M. F. Brennan (1996). "Long-term results of a prospective randomized trial of adjuvant brachytherapy in soft tissue sarcoma." J Clin Oncol **14**(3): 859-868.

- Prakash, A., A. Kumar Upadhyay and A. Kumar (2024). "Henri Coutard (1876-1950): A Pioneer of Fractionated Radiotherapy Regimens." Cureus **16**(9): e69371.
- Reijnen, C., H. V. N. Küsters-Vandeveld, C. F. Prinsen, L. Massuger, M. Snijders, S. Kommos, . . . J. M. A. Pijnenborg (2019). "Mismatch repair deficiency as a predictive marker for response to adjuvant radiotherapy in endometrial cancer." Gynecol Oncol **154**(1): 124-130.
- Requena, D. O., T. A. Longacre, A. E. Rosenberg, J. M. V. Torres, N. Yanchenko, M. T. Garcia-Buitrago, . . . E. A. Montgomery (2024). "Synovial Sarcoma of the Gastrointestinal Tract." Modern Pathology **37**(1): 100383.
- Reva, B., Y. Antipin and C. Sander (2011). "Predicting the functional impact of protein mutations: application to cancer genomics." Nucleic Acids Res **39**(17): e118.
- Ricardi, U., S. Badellino and A. R. Filippi (2015). "Stereotactic body radiotherapy for early stage lung cancer: History and updated role." Lung Cancer **90**(3): 388-396.
- Rigatti, S. J. (2017). "Random forest." Journal of Insurance Medicine **47**(1): 31-39.
- Rizvi, H., F. Sanchez-Vega, K. La, W. Chatila, P. Jonsson, D. Halpenny, . . . M. D. Hellmann (2018). "Molecular Determinants of Response to Anti-Programmed Cell Death (PD)-1 and Anti-Programmed Death-Ligand 1 (PD-L1) Blockade in Patients With Non-Small-Cell Lung Cancer Profiled With Targeted Next-Generation Sequencing." Journal of Clinical Oncology **36**(7): 633-641.
- Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz and J. P. Mesirov (2011). "Integrative genomics viewer." Nature biotechnology **29**(1): 24-26.
- Rodriguez, R., R. Rubio and P. Menendez (2012). "Modeling sarcomagenesis using multipotent mesenchymal stem cells." Cell research **22**(1): 62-77.
- Röntgen, W. C. (1896). "On a new kind of rays." Science **3**(59): 227-231.
- Rossi, D., A. Gamberucci, E. Pierantozzi, C. Amato, L. Migliore and V. Sorrentino (2021). "Calsequestrin, a key protein in striated muscle health and disease." Journal of Muscle Research and Cell Motility **42**(2): 267-279.
- Roy, P., A. Das, A. Sarkar, A. Dwari and S. Datta (2012). "A primary synovial sarcoma of lung." North American Journal of Medical Sciences **4**(5): 241.
- Rückert, M., A.-S. Flohr, M. Hecht and U. S. Gaipl (2021). "Radiotherapy and the immune system: More than just immune suppression." Stem Cells **39**(9): 1155-1165.
- Saidi, A., S. Javerzat, A. Bellahcène, J. De Vos, L. Bello, V. Castronovo, . . . M. Hagedorn (2008). "Experimental anti-angiogenesis causes upregulation of genes associated with poor survival in glioblastoma." Int J Cancer **122**(10): 2187-2198.
- Sardenberg, R. A., L. P. Figueiredo, F. J. Haddad, J. L. Gross and R. N. Younes (2010). "Pulmonary metastasectomy from soft tissue sarcomas." Clinics (Sao Paulo) **65**(9): 871-876.

- Sbaraglia, M., E. Bellan and A. P. Dei Tos (2021). "The 2020 WHO Classification of Soft Tissue Tumours: news and perspectives." Pathologica **113**(2): 70-84.
- Sbaraglia, M. and A. P. Dei Tos (2019). "The pathology of soft tissue sarcomas." Radiol Med **124**(4): 266-281.
- Schlegel, B. P., F. M. Jodelka and R. Nunez (2006). "BRCA1 promotes induction of ssDNA by ionizing radiation." Cancer Res **66**(10): 5181-5189.
- Schubert, M., B. Klinger, M. Klünemann, A. Sieber, F. Uhlitz, S. Sauer, . . . J. Saez-Rodriguez (2018). "Perturbation-response genes reveal signaling footprints in cancer gene expression." Nature Communications **9**(1): 20.
- Schulz, R. A., J. A. Stein and N. J. Pelc (2021). "How CT happened: the early development of medical computed tomography." J Med Imaging (Bellingham) **8**(5): 052110.
- Schulze, A., M. Oshi, I. Endo and K. Takabe (2020). "MYC Targets Scores Are Associated with Cancer Aggressiveness and Poor Survival in ER-Positive Primary and Metastatic Breast Cancer." Int J Mol Sci **21**(21).
- Senkin, S., S. Moody, M. Díaz-Gay, B. Abedi-Ardekani, T. Cattiaux, A. Ferreiro-Iglesias, . . . P. Brennan (2024). "Geographic variation of mutagenic exposures in kidney cancer genomes." Nature **629**(8013): 910-918.
- Seol, J. H., E. Y. Shim and S. E. Lee (2018). "Microhomology-mediated end joining: Good, bad and ugly." Mutat Res **809**: 81-87.
- Serrano, C. and S. Bauer (2022). "New tyrosine kinase inhibitors for the treatment of gastrointestinal stromal tumors." Current oncology reports **24**(2): 151-159.
- Sharabi, A. B., M. Lim, T. L. DeWeese and C. G. Drake (2015). "Radiation and checkpoint blockade immunotherapy: radiosensitisation and potential mechanisms of synergy." Lancet Oncol **16**(13): e498-509.
- Sharma, A., B. Bode, G. Studer, H. Moch, M. Okoniewski, A. Knuth, . . . M. Van Den Broek (2013). "Radiotherapy of human sarcoma promotes an intratumoral immune effector signature." Clinical Cancer Research **19**(17): 4843-4853.
- Sherry, S. T., M.-H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski and K. Sirotkin (2001). "dbSNP: the NCBI database of genetic variation." Nucleic acids research **29**(1): 308-311.
- Shi, W., X. Yao, Y. Fu and Y. Wang (2022). "Interferon- α and its effects on cancer cell apoptosis." Oncol Lett **24**(1): 235.
- Shin, J. S., T. G. Tut, T. Yang and C. S. Lee (2013). "Radiotherapy response in microsatellite instability related rectal cancer." Korean J Pathol **47**(1): 1-8.
- Shiravand, Y., F. Khodadadi, S. M. A. Kashani, S. R. Hosseini-Fard, S. Hosseini, H. Sadeghirad, . . . A. Kulasinghe (2022). "Immune checkpoint inhibitors in cancer therapy." Current Oncology **29**(5): 3044-3060.

- Sicinska, E., V. S. R. Kola, J. A. Kerfoot, M. L. Taddei, A. Al-Ibraheemi, Y.-H. Hsieh, . . . M. L. Hemming (2024). "ASPSCR1::TFE3 Drives Alveolar Soft Part Sarcoma by Inducing Targetable Transcriptional Programs." Cancer Research **84**(14): 2247-2264.
- Silva-Rodríguez, P., D. Fernández-Díaz, M. Bande, M. Pardo, L. Loidi and M. J. Blanco-Teijeiro (2022). "GNAQ and GNA11 Genes: A Comprehensive Review on Oncogenesis, Prognosis and Therapeutic Opportunities in Uveal Melanoma." Cancers **14**(13): 3066.
- Smith, M., A. Culhane, M. Donovan, J. Coffey, B. Barry, M. Kelly, . . . T. Cotter (2009). "Analysis of differential gene expression in colorectal cancer and stroma using fluorescence-activated cell sorting purification." British journal of cancer **100**(9): 1452-1464.
- Sondka, Z., S. Bamford, C. G. Cole, S. A. Ward, I. Dunham and S. A. Forbes (2018). "The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers." Nature Reviews Cancer **18**(11): 696-705.
- Song, Q., S. Qin, L. E. Pascal, C. Zou, W. Wang, H. Tong, . . . M. Morrell (2020). "SIRPB1 promotes prostate cancer cell proliferation via Akt activation." The Prostate **80**(4): 352-364.
- Stacchiotti, S., G. G. Baldi, C. Morosi, A. Gronchi and R. Maestro (2020). "Extraskeletal myxoid chondrosarcoma: state of the art and current research on biology and clinical management." Cancers **12**(9): 2703.
- Stacchiotti, S., P. Verderio, A. Messina, C. Morosi, P. Collini, A. Llombart-Bosch, . . . A. Gronchi (2012). "Tumor response assessment by modified Choi criteria in localized high-risk soft tissue sarcoma treated with chemotherapy." Cancer **118**(23): 5857-5866.
- Steele, C. D., A. Abbasi, S. M. A. Islam, A. L. Bowes, A. Khandekar, K. Haase, . . . N. Pillay (2022). "Signatures of copy number alterations in human cancer." Nature **606**(7916): 984-991.
- Steele, C. D., M. Tarabichi, D. Oukrif, A. P. Webster, H. Ye, M. Fittall, . . . G. Collord (2019). "Undifferentiated sarcomas develop through distinct evolutionary pathways." Cancer Cell **35**(3): 441-456. e448.
- Steele, C. D., M. Tarabichi, D. Oukrif, A. P. Webster, H. Ye, M. Fittall, . . . N. Pillay (2019). "Undifferentiated Sarcomas Develop through Distinct Evolutionary Pathways." Cancer Cell **35**(3): 441-456.e448.
- Stephens, P. J., C. D. Greenman, B. Fu, F. Yang, G. R. Bignell, L. J. Mudie, . . . L. A. Stebbings (2011). "Massive genomic rearrangement acquired in a single catastrophic event during cancer development." cell **144**(1): 27-40.
- Stinson, S. F., T. F. Delaney, J. Greenberg, J. C. Yang, M. H. Lampert, J. E. Hicks, . . . E. J. Glatstein (1991). "Acute and long-term effects on limb function of combined modality limb sparing therapy for extremity soft tissue sarcoma." International Journal of Radiation Oncology* Biology* Physics **21**(6): 1493-1499.

- Su, Y.-H., Y.-Z. Wu, D. K. Ann, J. L.-Y. Chen and C.-Y. Kuo (2023). "Obesity promotes radioresistance through SERPINE1-mediated aggressiveness and DNA repair of triple-negative breast cancer." Cell Death & Disease **14**(1): 53.
- Sun, C., A. Chu, R. Song, S. Liu, T. Chai, X. Wang and Z. Liu (2023). "PARP inhibitors combined with radiotherapy: are we ready?" Frontiers in Pharmacology **14**: 1234973.
- TCGA, T. (2017). "Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas." Cell **171**(4): 950-965.e928.
- The Gene Ontology Consortium (2018). "The Gene Ontology Resource: 20 years and still GOing strong." Nucleic Acids Research **47**(D1): D330-D338.
- Thway, K. (2019). Well-differentiated liposarcoma and dedifferentiated liposarcoma: an updated review. Seminars in diagnostic pathology, Elsevier.
- Thway, K., J. Wang, T. Mubako and C. Fisher (2014). "Histopathological diagnostic discrepancies in soft tissue tumours referred to a specialist centre: reassessment in the era of ancillary molecular diagnosis." Sarcoma **2014**(1): 686902.
- Tibshirani, R. (2018). "Regression Shrinkage and Selection Via the Lasso." Journal of the Royal Statistical Society: Series B (Methodological) **58**(1): 267-288.
- Tjalma, W., T. Van Waes, L. Van den Eeden and J. Bogers (2005). "Role of human papillomavirus in the carcinogenesis of squamous cell carcinoma and adenocarcinoma of the cervix." Best practice & research Clinical obstetrics & gynaecology **19**(4): 469-483.
- Trojani, M., G. Contesso, J. M. Coindre, J. Rouesse, N. B. Bui, A. de Mascarel, . . . C. Lagarde (1984). "Soft-tissue sarcomas of adults; study of pathological prognostic variables and definition of a histopathological grading system." Int J Cancer **33**(1): 37-42.
- Tromp, J., L. M. Boerman, I. E. Sama, S. Maass, J. H. Maduro, Y. M. Hummel, . . . P. van der Meer (2020). "Long-term survivors of early breast cancer treated with chemotherapy are characterized by a pro-inflammatory biomarker profile compared to matched controls." Eur J Heart Fail **22**(7): 1239-1246.
- Tsuji-Tamura, K., S. Morino-Koga, S. Suzuki and M. Ogawa (2021). "The canonical smooth muscle cell marker TAGLN is present in endothelial cells and is involved in angiogenesis." Journal of cell science **134**(15): jcs254920.
- Turc-Carel, C., P. Dal Cin, J. Limon, F. Li and A. A. Sandberg (1986). "Translocation X; 18 in synovial sarcoma." Cancer Genetics and Cytogenetics **23**(1): 93-93.
- Van Loo, P., S. H. Nordgard, O. C. Lingjærde, H. G. Russnes, I. H. Rye, W. Sun, . . . B. Naume (2010). "Allele-specific copy number analysis of tumors." Proceedings of the National Academy of Sciences **107**(39): 16910-16915.
- Van Loo, P., S. H. Nordgard, O. C. Lingjærde, H. G. Russnes, I. H. Rye, W. Sun, . . . V. N. Kristensen (2010). "Allele-specific copy number analysis of tumors." Proc Natl Acad Sci U S A **107**(39): 16910-16915.

- Vanni, S., A. De Vita, L. Gurrieri, V. Fausti, G. Miserocchi, C. Spadazzi, . . . A. Bongiovanni (2022). "Myxofibrosarcoma landscape: Diagnostic pitfalls, clinical management and future perspectives." Therapeutic Advances in Medical Oncology **14**: 17588359221093973.
- Verbovšek, U., H. Motaln, A. Rotter, N. A. Atai, K. Gruden, C. J. Van Noorden and T. T. Lah (2014). "Expression analysis of all protease genes reveals cathepsin K to be overexpressed in glioblastoma." PLoS One **9**(10): e111819.
- Vogelstein, B., N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, Jr. and K. W. Kinzler (2013). "Cancer genome landscapes." Science **339**(6127): 1546-1558.
- Wah, N. W., Y. Mok, N. Omar, K. T. E. Chang, T. K. Y. Tay, S. S. Hue and V. K. M. Lee (2023). "Clinicopathologic and Molecular Characteristics of Epstein-Barr Virus-Associated Smooth Muscle Tumor Compared With Those of Leiomyoma and Leiomyosarcoma." Mod Pathol **36**(6): 100127.
- Wainszelbaum, M. J., A. J. Charron, C. Kong, D. S. Kirkpatrick, P. Srikanth, M. A. Barbieri, . . . P. D. Stahl (2008). "The hominoid-specific oncogene TBC1D3 activates Ras and modulates epidermal growth factor receptor signaling and trafficking." Journal of Biological Chemistry **283**(19): 13233-13242.
- Wang, B., D. Chen and H. Hua (2021). "TBC1D3 family is a prognostic biomarker and correlates with immune infiltration in kidney renal clear cell carcinoma." Molecular Therapy-Oncolytics **22**: 528-538.
- Wang, F., J. Yu, P. Lin, C. Sigalas, S. Zhang, Y. Gong, . . . L. Song (2022). "The ryanodine receptor mutational characteristics and its indication for cancer prognosis." Scientific Reports **12**(1): 16113.
- Wang, L., C. Lynch, S. P. Pitroda, A. Piffkó, K. Yang, A. K. Huser, . . . R. R. Weichselbaum (2024). "Radiotherapy and immunology." Journal of Experimental Medicine **221**(7): e20232101.
- Wang, W.-L., E. Mayordomo, W. Zhang, V. S. Hernandez, D. Tuvin, L. Garcia, . . . D. Lopez-Terrada (2009). "Detection and characterization of EWSR1/ATF1 and EWSR1/CREB1 chimeric transcripts in clear cell sarcoma (melanoma of soft parts)." Modern Pathology **22**(9): 1201-1209.
- Wang, Y., Y. Han, Y. Jin, Q. He and Z. Wang (2022). "The Advances in Epigenetics for Cancer Radiotherapy." Int J Mol Sci **23**(10).
- Wardelmann, E., R. L. Haas, J. V. Bovee, P. Terrier, A. Lazar, C. Messiou, . . . S. Bonvalot (2016). "Evaluation of response after neoadjuvant treatment in soft tissue sarcomas; the European Organization for Research and Treatment of Cancer-Soft Tissue and Bone Sarcoma Group (EORTC-STBSG) recommendations for pathological examination and reporting." Eur J Cancer **53**: 84-95.
- WHO (2020). WHO classification of tumours soft tissue and bone tumours. Lyon, France, IARC Press.
- Wu, Y. and B. P. Zhou (2010). "TNF- α /NF- κ B/Snail pathway in cancer cell migration and invasion." British Journal of Cancer **102**(4): 639-644.

- Xi, R., S. Lee, Y. Xia, T.-M. Kim and P. J. Park (2016). "Copy number analysis of whole-genome data using BIC-seq2 and its application to detection of cancer susceptibility variants." Nucleic Acids Research **44**(13): 6274-6286.
- Yang, C., X. Zhang, X. Yang, F. Lian, Z. Sun, Y. Huang and W. Shen (2023). "Function and regulation of RGS family members in solid tumours: a comprehensive review." Cell Communication and Signaling **21**(1): 316.
- Yang, J. C., A. E. Chang, A. R. Baker, W. F. Sindelar, D. N. Danforth, S. L. Topalian, . . . S. A. Rosenberg (1998). "Randomized prospective study of the benefit of adjuvant radiation therapy in the treatment of soft tissue sarcomas of the extremity." J Clin Oncol **16**(1): 197-203.
- Yang, L., L. Forker, J. J. Irlam, N. Pillay, A. Choudhury and C. M. West (2017). "Validation of a hypoxia related gene signature in multiple soft tissue sarcoma cohorts." Oncotarget **9**(3): 3946.
- Yao, C., H. Zhou, Y. Dong, A. Alhaskawi, S. Hasan Abdullah Ezzi, Z. Wang, . . . H. Lu (2023). "Malignant peripheral nerve sheath tumors: latest concepts in disease pathogenesis and clinical management." Cancers **15**(4): 1077.
- Yu, G., L.-G. Wang, Y. Han and Q.-Y. He (2012). "clusterProfiler: an R package for comparing biological themes among gene clusters." Omics: a journal of integrative biology **16**(5): 284-287.
- Zaborowski, M., A. C. Vargas, J. Pulvers, A. Clarkson, D. de Guzman, L. Sioson, . . . A. J. Gill (2020). "When used together SS18-SSX fusion-specific and SSX C-terminus immunohistochemistry are highly specific and sensitive for the diagnosis of synovial sarcoma and can replace FISH or molecular testing in most cases." Histopathology **77**(4): 588-600.
- Zack, T. I., S. E. Schumacher, S. L. Carter, A. D. Cherniack, G. Saksena, B. Tabak, . . . R. Beroukhi (2013). "Pan-cancer patterns of somatic copy number alteration." Nature Genetics **45**(10): 1134-1140.
- Zafarana, G., A. S. Ishkanian, C. A. Malloff, J. A. Locke, J. Sykes, J. Thoms, . . . V. R. Ramnarine (2012). "Copy number alterations of c - MYC and PTEN are prognostic factors for relapse after prostate cancer radiotherapy." Cancer **118**(16): 4053-4062.
- Zehir, A., R. Benayed, R. H. Shah, A. Syed, S. Middha, H. R. Kim, . . . M. F. Berger (2017). "Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients." Nat Med **23**(6): 703-713.
- Zhang, Y., H. Wang, F. Wang, W. Ma, N. Li, C. Bo, . . . M. Liu (2022). "A four - gene signature associated with radioresistance in head and neck squamous cell carcinoma identified by text mining and data analysis." Computational and Mathematical Methods in Medicine **2022**(1): 5693806.
- Zhao, K., J. F. Liu, Y. X. Zhu, X. M. Dong, R. H. Yin, X. Liu, . . . Q. Wang (2022). "Hemgn Protects Hematopoietic Stem and Progenitor Cells Against Transplantation Stress Through Negatively Regulating IFN - γ Signaling." Advanced Science **9**(5): 2103838.

Zhu, A., J. G. Ibrahim and M. I. Love (2018). "Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences." Bioinformatics **35**(12): 2084-2092.