

**Molecular Characterisation of Muscle-Invasive
Bladder Cancer**

Melissa Pei-Yi Tan

The Institute of Cancer Research

MD(Res)

I, Melissa Pei-Yi Tan, 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

Background:

Muscle-invasive bladder cancer (MIBC) is a heterogeneous disease with a poor prognosis. There is a clinical need for biomarkers in MIBC to guide treatment strategy for individual patients and improve outcomes. In this work, I established a protocol to allow the collection of clinical material and data from patients with MIBC to support 3 pilot studies exploring potential clinically relevant biomarkers at a genomic and transcriptomic level.

Method:

In the first pilot study, plasma DNA was extracted from pre-treatment blood samples of 52 patients. In a subset of 10 patients receiving first-line chemotherapy, plasma DNA extracted from sequential samples underwent low-pass whole genome sequencing, with resultant copy number data being used to estimate tumour fraction. In the 2nd pilot, multiregion whole-exome sequencing was performed on DNA extracted from formalin-fixed paraffin-embedded (FFPE) tumour tissue of 9 patients to identify potential aberrations to interrogate in plasma. In the 3rd pilot study, RNA extracted from FFPE tumour tissue of 43 patients treated with radical radiotherapy was profiled using the Nanostring platform. Non-negative matrix factorisation was applied to the gene expression data to explore molecular subtypes, with a focus on potential association with radiotherapy response.

Results:

Key findings include that baseline plasma DNA concentration is a prognostic factor for bladder cancer-specific survival. Plasma tumour fraction is correlated with radiological disease burden and may be an early indicator of response or resistance to chemotherapy. At a transcriptomic level, MIBC can be divided

into 5 new molecular subtypes which appear to be associated with radiotherapy response.

Conclusion:

I have established a successful protocol to facilitate the collection of clinical material and data to support translational research in MIBC. Pilot studies have demonstrated the feasibility of applying a range of approaches and techniques to these samples, and my results have highlighted areas worthy of further investigation.

Acknowledgements

I would firstly like to thank my supervisors, Dr Anguraj Sadanandam, Professor Gert Attard and Professor Robert Huddart, for their valued expertise, inspiration, support and advice throughout this project. I am indebted also to Dr Sadanandam and Professor Attard's teams for welcoming me so warmly into their laboratories. Special mention must go to Erica Little, Dr Elisa Fontana, Ania Wingate, Dr Inma Spiteri and Dr Anu Jayaram who all taught me the essential laboratory skills I needed to be able to work independently in the lab, and encouraged me when, as is often true in science, things did not at first go to plan! I would also like to thank Dr Gift Nyamundanda for his invaluable help and support with the gene expression work, and Dr Marina Parry who introduced me to pivot tables and helped me develop a framework for reviewing my whole-exome sequencing data.

I am grateful to the BRC for financially supporting my role and also to the many contributors towards the department's Bladder Cancer Research Fund which has supported this work.

This project would not have been possible without the establishment of the CoMB protocol and I would like to thank Kelly Jones for her help and support in getting CoMB set up. I am also grateful to all those in the Bob Champion Unit who supported this protocol and helped it to succeed. Thank you also to the RMH Biobank team who have tirelessly supported CoMB and my requests to access samples.

I would also like to thank my family who have supported me throughout this MD(Res) and provided many hours of additional childcare to allow me to escape to the library to work on it!

Finally, my heartfelt thanks goes to all the patients who kindly took part in CoMB. The generosity of patients never ceases to amaze me, and I hope that this and future work using CoMB samples do justice to them.

Table of Contents

Chapter 1. Introduction	12
1.1. Overview.....	12
1.2. Bladder Cancer.....	12
1.2.1. Overview and epidemiology	12
1.2.2. Current management of MIBC	13
1.3. Molecular Biology of MIBC.....	18
1.3.1. DNA aberrations in MIBC	18
1.3.2. Gene expression patterns in MIBC.....	23
1.4. Current Biomarker Candidates in MIBC.....	31
1.4.1. Potential genomic biomarkers	31
1.4.2. Potential transcriptomic biomarkers	35
1.4.3. Immunohistochemical biomarkers	36
1.4.4. Biomarkers of radiosensitivity.....	38
1.5. Approaches to Interrogating Tumour Samples	42
1.5.1. Limitations of current approaches	42
1.5.2. Alternative approaches to interrogating genomic aberrations in MIBC	45
1.5.3. Alternative approaches to interrogating molecular subtypes in MIBC	58
1.6. Conclusion	64
1.7. References	65
 Chapter 2. Establishing a Framework for Biomarker Discovery & Development in MIBC: the Collection of Clinical Material for Molecular Stratification in Muscle-Invasive Bladder Cancer (CoMB) protocol	 74

2.1.	Introduction	74
2.1.1.	Rationale for a MIBC biomarker protocol	74
2.1.2.	Aim	75
2.2.	Methods	75
2.2.1.	Framework for collection of human material.....	75
2.3.	Results.....	80
2.3.1.	Key milestones.....	80
2.3.2.	Ethical approval.....	80
2.3.3.	Laboratory work undertaken to establish workflows for tissue processing standard operating protocols	82
2.3.4.	Recruitment.....	86
2.4.	Use of samples in planned work	86
2.4.1.	Overview	86
2.4.2.	Plasma DNA pilot	87
2.4.3.	Multiregion whole exome sequencing pilot.....	87
2.4.4.	Molecular subtyping pilot.....	87
2.5.	Conclusion	88
2.6.	References	89
2.7.	Appendix.....	90
Chapter 3.	Plasma DNA and Circulating Tumour DNA in MIBC	131
3.1.	Introduction	131
3.2.	Hypotheses & Aims	131
3.2.1.	Hypotheses	131
3.2.2.	Aims	132
3.3.	Method.....	133
3.3.1.	Patient cohort	133

3.3.2.	Plasma DNA extraction & quantification	133
3.3.3.	Measuring disease burden	134
3.3.4.	Whole-genome library preparation	136
3.3.5.	Low-pass whole-genome sequencing (LP-WGS).....	136
3.3.6.	Estimation of tumour fraction.....	136
3.3.7.	Analysis.....	137
3.4.	Results.....	138
3.4.1.	Patient characteristics	138
3.4.2.	Plasma DNA concentration	141
3.4.3.	Tumour fraction	150
3.5.	Discussion	159
3.5.1.	Plasma DNA concentration	159
3.5.2.	Tumour fraction	166
3.6.	Conclusion	170
3.7.	References	172
3.8.	Appendix.....	174

Chapter 4.	Exploring the Use of Multiregion Whole Exome- Sequencing of FFPE Tumour Tissue and Publicly Available Data to Identify Plasma ctDNA Targets in MIBC	190
4.1.	Introduction	190
4.1.1.	Overview	190
4.1.2.	Using a commercially available panel	191
4.2.	Hypotheses and Aims	191
4.2.1.	Hypotheses	191
4.3.	Aims.....	192
4.4.	Materials and Methods.....	192
4.4.1.	Cohort selection	192

4.4.2.	Processing of samples	193
4.4.3.	Foundation Medicine FoundationOne Panel	198
4.4.4.	Data analysis.....	199
4.5.	Results.....	201
4.5.1.	FoundationOne Panel	201
4.5.3.	Whole-exome sequencing.....	213
4.6.	Discussion	225
4.6.1.	Whole-exome sequencing in MIBC	225
4.6.2.	Targeted sequencing in MIBC: A commercial panel.....	229
4.6.3.	Designing a targeted panel for plasma.....	233
4.6.4.	Future work	237
4.7.	Conclusion	239
4.8.	References	240
4.9.	Appendix.....	243

Chapter 5. Exploring Molecular Subtype as a Biomarker of Radiosensitivity in Muscle-Invasive Bladder Cancer 288

5.1.	Introduction	288
5.1.1.	Molecular subtyping in muscle-invasive bladder cancer.....	288
5.2.	Hypotheses and Aims	289
5.2.1.	Hypotheses:	289
5.2.2.	Aims	289
5.3.	Methods.....	290
5.3.1.	Application of CRCAssigner-38 to Publicly Available Data.....	290
5.3.2.	Patient cohort	291
5.3.3.	Processing of FFPE blocks	292
5.3.4.	Nanostring Protocol.....	294

5.3.5. Statistics and analysis	295
5.4. Results:.....	298
5.4.1. Applying CRCAssigner-38 to GSC dataset	298
5.4.2. Radiotherapy patient cohort	302
5.4.3. A new subtyping classifier with clinical relevance in MIBC? ...	316
5.4.4. The NMF71 classifier	327
5.4.5. Inpatient heterogeneity within RT cohort.....	341
5.4.6. Differentially expressed genes	341
5.4.7. Comparing gene expression and genomic aberration data	344
5.5. Discussion	346
5.5.1. Molecular subtype as a biomarker of radiosensitivity	346
5.5.2. Molecular features of radiosensitivity	352
5.5.3. Molecular subtype stratification for treatment.....	358
5.5.4. Heterogeneity of molecular subtype.....	362
5.5.5. Radiosensitivity Index.....	363
5.5.6. Differentially expressed genes	364
5.5.7. Limitations and future directions.....	364
5.6. Conclusion	368
5.7. References	369
5.8. Appendix.....	372
Chapter 6. Discussion.....	438
6.1. Overview of findings.....	438
6.2. Cross-platform analysis	439
6.2.1. Data available within CoMB	439
6.2.2. ATM.....	442
6.2.3. DNA damage repair genes.....	443

6.3. Further areas of interest.....	444
6.3.1. Immunotherapy	444
6.3.2. Tumour microenvironment and tumour-infiltrating lymphocytes 444	
6.3.3. Consensus molecular subtypes.....	445
6.4. The CoMB Protocol: Future opportunities.....	446
6.5. Conclusion	449
6.6. References	450

Chapter 1. Introduction

1.1. Overview

Muscle-invasive bladder cancer (MIBC) is a heterogeneous disease associated with marked variation in its behaviour and clinical outcomes. Despite surgical and oncological advances, 5-year overall survival has not significantly changed and remains at approximately 50%².

There is a real clinical need for further translational research in MIBC to identify predictive and prognostic biomarkers to guide treatment strategy for individual patients.

This research project therefore proposes to establish a collection of clinical samples and data from patients with MIBC, and to utilise these samples in pilot studies with the overarching aim of identifying potential clinically relevant biomarkers at a genomic and transcriptomic level.

I will now discuss the current management of MIBC and give an overview of the current understanding of its molecular biology. I will also highlight potential candidate biomarkers of clinical relevance, and explore alternative approaches to interrogating patient samples at a genomic and transcriptomic level.

1.2. Bladder Cancer

1.2.1. Overview and epidemiology

With over 10 000 new cases in 2015, bladder cancer is the 7th most common cancer in the UK³. Incidence increases with age and the majority of patients are aged 75 or over at diagnosis³. Coupled with the fact that bladder cancer is strongly associated with smoking, the typical bladder cancer patient is an older patient with age and smoking-associated medical co-morbidities. With this in mind, it could therefore be suggested that this is a patient population where

predictive and prognostic biomarkers would be particularly advantageous to ensure prompt delivery of treatment with likely benefit, and perhaps more importantly, to minimise unnecessary toxicity from a treatment likely to fail.

At a clinical level, bladder cancer is divided by histological assessment into non-muscle invasive disease and muscle-invasive disease. Non-muscle invasive bladder cancer (NMIBC) is usually treated with local resection and intravesical agents to reduce the risk of recurrence. While risk of recurrence is high, the majority do not progress further and generally carry a good prognosis. Muscle-invasive bladder cancer (MIBC) however has an aggressive natural history with a poor prognosis.

1.2.2. Current management of MIBC

1.2.2.1. Radical treatment- bladder preservation vs surgery

Definitive radical management of MIBC includes cystectomy with pelvic lymph dissection, or bladder preservation with radiotherapy and a radiosensitiser. Neoadjuvant platinum-based combination chemotherapy should also be considered in those suitable as it has been shown to confer a 5% survival advantage at 5 years⁴⁻⁶, and international guidelines recommend it is considered for all surgical patients with T2-4N0M0 disease^{7,8}. UK-based guidelines⁹ recommend offering neoadjuvant chemotherapy (NAC) before radical cystectomy or radical radiotherapy, and this is in keeping with data from the BA06 study¹⁰, which demonstrated that the choice of local treatment i.e. radiotherapy or surgery, did not impact on the effect of NAC. While response rates to neoadjuvant chemotherapy (NAC) of over 80%¹¹ are reported with pathological complete response in 30-49%^{5,6,11}, a subset of patients do not respond and may progress during treatment. A predictive biomarker of response to NAC is therefore needed in order to avoid unnecessary toxicity in non-responders, and also to minimise delay to definitive treatment, something which has been shown to compromise overall outcome¹². It is important to also note the potential effect of overall treatment time in MIBC with prolonged timelines potentially facilitating repopulation and so impacting on outcomes. Recent data has suggested that using a hypofractionated radiotherapy

schedule of 55Gy in 20 fractions compared to 64Gy in 32 fractions results in improved invasive locoregional control¹³.

Bladder preservation with combined modality treatment (CMT) is now increasingly recognised as an alternative to surgery in the radical setting. 5-year overall survival rates of 50-57% have been reported with CMT^{14,15}, which was defined as including a maximal transurethral resection (TURBT) and chemoradiation. Locoregional relapse rates at 2 years of 67% are reported, with over half due to NMIBC disease only¹⁶. In those with locoregional invasive relapse, salvage cystectomy may be performed and rates of 10-30%^{8,10} are documented. However, this cohort of patients has been subjected to the toxicity of both radiation and surgery, with the delay to effective treatment potentially compromising overall outcome.

The decision between surgery and radiotherapy is currently based upon patient factors and disease parameters. Patients who achieve a complete response to NAC are noted to have better overall and recurrence-free survival¹⁷, and this has been used to guide the selection of patients suitable for bladder preservation, with non-responders being counselled towards surgery. However, response to NAC is most likely a prognostic rather than predictive factor of response to radiotherapy. Likewise, patients with carcinoma-in-situ or hydronephrosis have traditionally been counselled towards surgery as previous studies have suggested a poorer response to radiotherapy¹⁴, but again the evidence to support this is not clear and such features should be considered prognostic rather than predictive. Immunohistochemical studies have put forward *MRE11* and *TIP60* as potential predictive markers of response to radiotherapy and surgery respectively^{18,19}, but have not been prospectively validated. Furthermore, recent work from 2 groups were not able to reproduce these findings and found no association between *MRE11* and outcomes following radical radiotherapy^{20,21}.

Therefore, in current clinical practice, there are no validated biomarkers to guide the decision between radiotherapy or surgery.

1.2.2.2. Palliative treatment

Regardless of whether a surgical or bladder-preserving approach is taken, distant metastatic relapse rates following radical treatment of 22-52%² are reported. Treatment options include palliative platinum-based chemotherapy with second line options including taxanes with response rates of 5-10%²², or the more recently approved immune-checkpoint agents with response rates of 21% reported²³.

With regards to potential biomarkers in the palliative setting, interesting retrospective work looking at DNA damage repair (DDR) gene aberrations have suggested that patients with DDR alterations receiving platinum-based chemotherapy have a better progression-free and overall survival compared to those without²⁴. Further work is needed to validate this but this may represent a potential predictive biomarker to help guide treatment options in the advanced disease setting.

There has been much interest in investigating potential biomarkers associated with response to immunotherapy. There is some evidence to suggest that high tumour mutational burden, PD1 or PDL-1 staining, or interferon-gamma gene expression²⁵ may be predictors of response to immunotherapy in urothelial cancer^{23,26} although, again, further work is required to fully explore this. There has been much attention on the role of the tumour microenvironment with data suggesting that 'inflamed' tumours i.e. those heavily infiltrated with immune cells and pro-inflammatory cytokines, respond better to immunotherapy²⁷.

In a recent large pan-cancer study which included bladder cancer, *ARID1A* status has recently been suggested to be associated with response to immunotherapy with improved overall survival seen in patients receiving checkpoint inhibitors²⁸. *ARID1A* aberrations were also noted to be associated with increased immune infiltrate in endometrial, colon and gastric cancer but interestingly, immune infiltrates were in contrast much lower in renal cell carcinoma, suggesting a tumour site-specific effect.

Inflammatory infiltrate has also been of interest in the context of predicting response to systemic treatment and have recently been considered in the

context of molecular subtypes in bladder cancer. These will be further discussed in further detail in 1.4.2.

1.2.2.3. *Clinical dilemmas in MIBC- the need for personalised medicine*

Biomarkers are needed in MIBC to personalise treatment and improve outcomes. Figure 1.1 illustrates the patient pathway and the potential applications of biomarkers in MIBC. As can be seen, biomarkers could have significant clinical impact at each stage of a patient's treatment pathway. In this thesis, I will be focussing on biomarkers of response to chemotherapy and those that may predict response to radical chemoradiation in MIBC at a genomic and transcriptomic level. I will therefore now give an overview of the relevant molecular biology of MIBC and highlight the candidate biomarkers put forward with a focus on work related to DNA aberrations and gene expression-based subtypes.

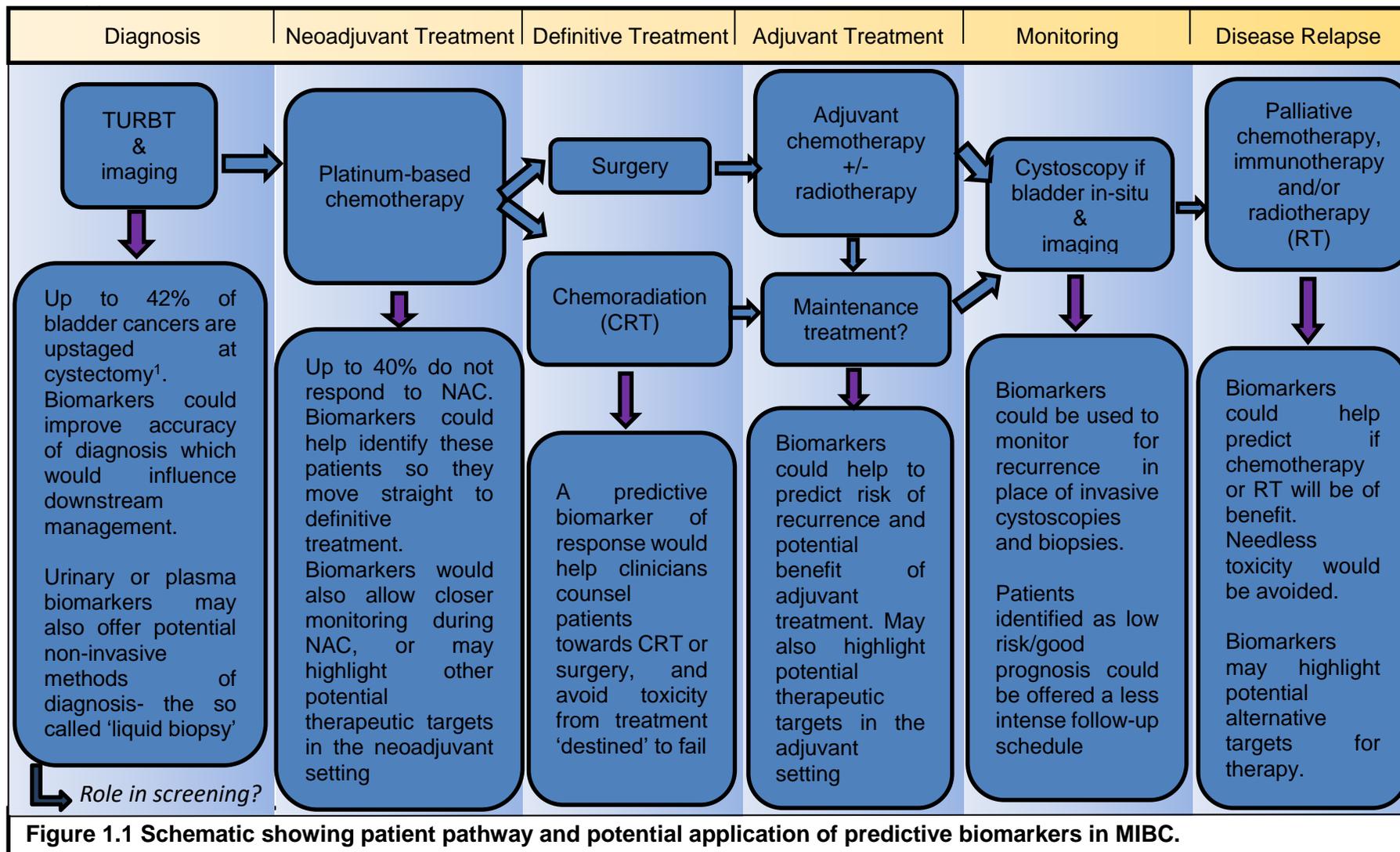


Figure 1.1 Schematic showing patient pathway and potential application of predictive biomarkers in MIBC.

1.3. Molecular Biology of MIBC

1.3.1. DNA aberrations in MIBC

1.3.1.1. Somatic genomic mutations in MIBC

Multiple studies have reported on the range of aberrations seen in MIBC. The largest single cohort studied at the time of starting this MD(Res) was that from The Cancer Genome Network (TCGA). Their analysis included whole exome sequencing (WES) on 130 MIBC samples and RNA sequencing on 129 samples, but did not report on the association of results with clinical data or outcomes. At a genomic level, MIBC demonstrated a greater number of DNA aberrations than most of the other adult malignancies studied within the project, and was third only to melanoma and lung cancers. Thirty-two statistically significant gene mutations were found and potential therapeutic targets were noted in 69% of the samples. The most frequently aberrant genes included *TP53* and *MLL2*. Table 1.1 overleaf lists the top 10 significantly mutated genes identified by the TCGA, and summarises data available at the time of commencing this MD(Res) regarding its use as a potential biomarker. The studies listed were all retrospective in design with relatively low numbers, and have not been further validated meaning that there remains a stark absence of validated biomarkers in MIBC.

Of note, the TCGA reported that MIBC had the greatest number of chromatin regulating gene aberrations when compared to other solid malignancies they had studied. This is in keeping with other work demonstrating the prominence of aberrations in chromatin regulatory and modifying genes in bladder cancer^{29,30}. Given the dominance of chromatin regulatory gene aberrations in MIBC, further work looking at their role is warranted as they could present a potential therapeutic target. The TCGA report has since been updated to include results now from 412 patients with MIBC. In the updated report³¹, a total of 58 statistically significant genes were reported.

Table 1.1 Top 10 significantly mutated genes reported by TCGA³²

Gene	Frequency	Notes
<i>TP53</i>	49%	<p>17p13. Tumour suppressor gene. Key role in cell cycle arrest and apoptosis.</p> <p>Evidence demonstrating that <i>TP53</i> mutations more commonly seen in patients with extravesical disease and/or lymph node involvement; associated with a poorer outcome in radical cystectomy cohorts³⁰. Aberrations mutually exclusive with those in <i>CDKN2A</i> and involving <i>MDM2</i>³¹</p>
<i>MLL2</i> / <i>KMT2D</i>	27%	<p>12q13. Chromatin modifying enzyme. No published data on association with clinical outcomes.</p>
<i>ARID1A</i>	25%	<p>1p36. Chromatin modifying enzyme. No published data on association with clinical outcomes specifically in bladder cancer. Recent pan-cancer data suggesting association with immunotherapy outcomes²⁸</p>
<i>PIK3CA</i>	22%	<p>3q26. Codes of catalytic subunit of PI3K enzyme. Cell proliferation, migration and survival pathways. Conflicting reports regarding association of mutation with survival^{30,33}</p>
<i>EP300</i>	15%	<p>22q13. Chromatin modifying enzyme. Role in cell proliferation and differentiation. No published data on association with clinical outcomes.</p>

Table 1.1 (continued)

Gene	Frequency	Notes
<i>CDKN1A</i>	14%	6p21. Codes for p21. Cell cycle gene. Put forward as possible driver mutation ³⁴
<i>RB1</i>	13%	13q14.2. Tumour suppressor gene. Key role in cell cycle and DNA repair. Mutations reported to be associated with poorer disease-specific survival ³⁵ and poorer overall survival ³⁶
<i>ERCC2</i>	12%	19q13. Nucleotide excision repair. Possible biomarker of response to cisplatin chemotherapy ³⁷ . Improved recurrence-free survival reported in presence of mutation in MIBC cystectomy cohort ³⁸
<i>FGFR3</i>	12%	4p16. Mitogenesis and differentiation. Mutations more commonly seen in superficial disease. Some evidence in small cohort to suggest presence of <i>FGFR</i> mutation in MIBC associated with better prognosis ³⁹
<i>STAG2</i>	11%	Xq25. Part of cohesion complex regulating separation of sister chromatids during cell division. No published data on association with clinical outcomes.

1.3.1.2. Copy number alterations in MIBC

Large scale mutations are also common in MIBC. Common changes include chromosome 9 deletion, loss of 8p, deletion of *RB1*, and loss of heterozygosity of 15q^{24,25}.

Table 1.2 overleaf summarises the most common copy number alterations seen in MIBC according to the TCGA report³², and the data available at the time regarding potential associations with clinical outcomes.

Table 1.2 Amplifications/deletions in regions with significant copy number alterations identified by the TCGA in MIBC³²

Gene	Frequency	Notes
<i>CDKN2A</i>	47%	9q21. Deletion. Cell-cycle regulatory gene coding for p14 and p16. . No published data on association with clinical outcomes.
<i>NCOR1</i>	25%	17p11. Deletion. Promotes chromatin condensation. No published data on association with clinical outcomes.
<i>YWHAZ</i>	22%	8q22. Amplification (peak contained several genes). No published data on association with clinical outcomes.
<i>E2F3/ SOX4</i>	20%	Amplification. Transcription factors both on 6p22.3 (amplified in up to 25% advanced MIBC ⁴⁰ and put forward as marker of aggressive disease ⁴¹).
<i>PVRL4</i>	19%	1q23. Amplification (peak contained several genes). No published data on association with clinical outcomes.
<i>PPARG</i>	17%	3p25. Amplification. No published data on association with clinical outcomes.
<i>RB1</i>	14%	13q14. Deletion. Tumour suppressor. Cell cycle. <i>RB1</i> heterozygous deletion associated higher T-stage disease (T3-4) ⁴²
<i>MYC</i>	13%	8q24. Amplification. Transcription factor. No published data on association with clinical outcomes.

Table 1.2 (continued)

Gene	Frequency	Notes
<i>PTEN</i>	13%	10q23. Deletion. Tumour suppressor. No published data on association with clinical outcomes.
<i>CREBBP</i>	13%	16p13. Deletion. Chromatin modifying. No published data on association with clinical outcomes.

1.3.2. Gene expression patterns in MIBC

1.3.2.1. Using gene expression profiles to determine molecular subtype

With regards to analysis at a transcriptomic level in MIBC, there has been much interest in exploring potential molecular subtypes. Molecular subtyping refers to the classification of a disease based on gene expression profiles, where samples with similar gene expression features are clustered together into a subgroup. Molecular subtyping of cancer was first reported in lymphoma⁴³ but has perhaps been most associated with breast cancer after Perou's landmark paper described the existence of luminal and basal subtypes⁴⁴, and in colorectal cancer^{45,46}.

1.3.2.2. Molecular subtyping in MIBC

Several groups have explored molecular subtypes in MIBC and the number of subtypes reported ranges from 2 to 7^{31,32,47-54}. Earlier cohorts included NMIBC cases and overall, cases were seen to separate into distinct clusters according to whether muscle invasion was present or not⁴⁷⁻⁵¹. This is in keeping with the acknowledged fact that in the vast majority of cases, NMIBC and MIBC are separate entities, developing along separate distinct pathways⁵⁵.

Despite the differing number of subtypes reported within MIBC, a broad division into 'basal' and 'luminal' subgroups is consistently present, and has

been noted to be similar to the basal and luminal A profiles seen in breast cancer⁵³. Luminal MIBC tends to express higher levels of markers of differentiation such as uroplakins (expressed in terminal umbrella cells)⁵⁶ and urothelial differentiation markers e.g. *FOXA1*, *PPARG*³¹. Luminal subtypes are also enriched for *FGFR3* alterations⁵⁶. Basal tumours in contrast are characterised by expression of basal keratins e.g. *KRT 5/6*, and tend to have squamous or sarcomatoid features. High levels of *EGFR* expression⁵¹⁻⁵³ have also been demonstrated in basal bladder cancers, and have been associated with poorer disease-specific survival⁵⁷. This raises the possibility that *EGFR* targeted therapies could be a valid strategy in such patients.

In recent years, there have been increasing efforts to come to a consensus on the molecular subtypes present in MIBC. A consensus meeting in 2015 agreed the presence of basal-squamous-like tumours which are characterised by high expression of *KRT5/6* and *KRT14*, and low/undetectable levels of expression of *FOXA1* and *GATA3*⁵⁸. It was agreed that the definition of a luminal subtype needed to be optimised. ,

Table 1.3 summarises some of the subtype classifications put forward by the different groups. Of note, some of the subtypes appeared to be associated with clinical outcome and these will be further explored in the next section.

Table 1.3 Summary of main molecular subtyping studies in MIBC

Group	No. samples		No. subtypes identified	Details
	MIBC	NMIBC		
Dyrskjot et al ⁴⁷	10	30	3	Broadly correlated with pathological tumour (“T”) stage ie. subtypes clustered Ta, T1 & T2-4 tumours
Blaveri et al ⁴⁸	53	27	5	2 main clusters overall with one containing 26/27 superficial cases, and the other 48/53 MIBC. Superficial cluster subdivided into pTa and pT1. MIBC cluster divided into 3. One group contained tumours from 11 patients with median survival 5.4 months.
Sanchez Carbayo et al ⁴⁹	72	33	3	Two clusters of MIBC and one of superficial disease
Lindgren et al ⁵⁰	44	97 (+3Tx)	2	MS1 & MS2. MS1 dominated by <i>FGFR3/PIK3CA</i> activating mutations. MS2 demonstrated <i>RB1</i> loss and genomic instability

Table 1.3 (continued)

Group	No. samples		No. subtypes identified	Details
	MIBC	NMIBC		
Sjodahl et al ⁵¹	93	213	2 main clusters: MS1&MS2 as above) Subdivided into total of 7 subtypes	<p>Urobasal A (subdivided into MS1a and MS1b): Mostly superficial cancers (95% of group). Characterised by <i>FGFR3</i>, <i>CCND1</i>, <i>TP63</i> expression. Good prognosis</p> <p>Urobasal B (MS2b2.1): Similar to Urobasal A but frequent <i>TP53</i> mutations and expression of SCC-like keratins. 50% were MIBC. Poor prognosis</p> <p>Genomically unstable (divided into MS2a1 & MS2a2): 40% cases MIBC. Frequent <i>TP53</i> mutations, <i>CCNE</i> & <i>ERBB2</i> expression. Low <i>PTEN</i> expression. Intermediate prognosis</p> <p>Squamous cell carcinoma like (MS2b2.2): High expression basal keratins. Poor prognosis</p> <p>Infiltrated (MS2b1): Infiltrated by nontumour cells. Intermediate prognosis</p>

Table 1.3 (continued)

Group	No. samples		No. subtypes identified	Details
	MIBC	NMIBC		
Choi et al ⁵²	73	0	3	<p>Luminal: <i>FGFR3</i> mutations, and active <i>PPAR</i> and <i>ER</i> transcription</p> <p><i>P53</i>-like (subset of luminal): active <i>p53</i>-like gene expression signature, low <i>Myc</i> activation, low cell cycle and proliferation markers. Associated with resistance to chemotherapy.</p> <p>Basal: <i>p63</i> activation and squamous differentiation</p>
Damrauer et al ⁵³	262	0	2	<p>Luminal: <i>FGFR3</i> alterations. Improved disease specific survival and overall survival compared to basal-like subtype. Basal-like: <i>RB1</i>, <i>E2F3</i></p>

Table 1.3 (continued)

Group	No. samples		No. subtypes identified	Details
	MIBC	NMIBC		
TCGA ³² (2014)	129	0	4	<p>Cluster I- characterised by <i>FGFR3</i> mutations, copy gain & overexpression. High levels <i>ERBB2</i> & <i>ESR2</i> oestrogen receptor. Similar to luminal A breast cancer</p> <p>Cluster II: High levels <i>ERBB2</i> and oestrogen receptor (<i>ESR2</i>). Similar to luminal A breast cancer as above.</p> <p>Cluster III: similar to basal-like breast cancers, and squamous cell carcinomas of the lung, and head and neck region with expression of basal markers such as <i>KRT5</i> and <i>KRT6</i>.</p> <p>Cluster IV: no details given</p>
TCGA ³¹ (2017)	408	0	5	<p>Luminal subtypes showed high expression of uroplakins and urothelial differentiation markers.</p>

Group	No. samples		No. subtypes identified	Details
	MIBC	NMIBC		
TCGA (2017) cont.				<p>Luminal papillary: enriched for papillary morphology, <i>FGFR3</i> aberrations; lower T-stage and higher tumour purity</p> <p>Luminal infiltrated: lower tumour purity; higher lymphocytic infiltration, higher expression of smooth muscle markers and myofibroblasts; majority shared similarities with Choi et al's <i>p53</i>-like subtype. Increased expression of immune markers</p> <p>Luminal: Highest expression of uroplakins and urothelial differentiation markers</p> <p>Basal-squamous: Basal, squamous and stem-like markers; strongest immune expression signature</p> <p>Neuronal: Neuroendocrine markers but also included tumours with no histopathological neuroendocrine features</p>

Group	No. samples		No. subtypes identified	Details
	MIBC	NMIBC		
Seiler et al ⁵⁴	305	0	5	<p>Genomic single classifier (GSC) based on consensus of different subtyping classifiers:</p> <p>Basal-squamous: Basal and squamous markers</p> <p>Luminal: Markers of urothelial differentiation; best prognosis</p> <p>Luminal infiltrated: EMT and immune infiltration</p> <p>Claudin-low: High expression of basal, immune and epithelial to mesenchymal (EMT) markers</p>

1.3.2.3. Does cohort size matter?

At the time of setting up this project in 2014/15, the largest cohort size used for subtyping in MIBC was 129³². While this was later expanded to include data from 408 patients³¹, this remains a relatively small number when compared to other cancer sites such as colorectal cancer, where gene expression profiles from 1290 tumours have been used to establish subtyping classifications⁴⁵.

As larger cohort sizes are recognised to allow more robust identification of molecular subtypes⁵⁹, I plan to explore whether publicly available data from different groups and across different platform can be successfully combined to create a larger, potentially more robust, dataset for analysis of subtypes in MIBC. This is discussed further in Chapter 5.

1.4. Current Biomarker Candidates in MIBC

1.4.1. Potential genomic biomarkers

1.4.1.1. TP53

TP53 is the most commonly mutated gene in MIBC and has been put forward as a prognostic biomarker. It is recognised to be more commonly found in MIBC than NMIBC⁵⁵ and so perhaps has potential utility in the setting of monitoring NMIBC to potentially identify those patients at risk of progression or a poorer outcome. Aberrations in *TP53* have been reported to be associated with poorer prognosis in retrospective radical cystectomy cohorts, and more commonly seen in patients with extra-vesical and/or regional lymph node involvement³⁰. It was also put forward as a predictive biomarker of response to adjuvant chemotherapy⁶⁰ but p53 immunohistochemical status was not shown to be a prognostic or predictive biomarker in the phase III SWOG 4B951 study. This study was limited by several factors including a failure to randomise 'p53-negative' patients, and only 2/3s of patients receiving the 3 cycles of chemotherapy as intended.

TP53 therefore has yet to be validated for use as a prognostic biomarker in clinical practice and has not been shown to be of relevance as a predictive marker in MIBC.

1.4.1.2. *PIK3CA*

PIK3CA mutations have been associated with an improved recurrence-free survival and cancer-specific survival in cystectomy cohorts³⁰ but there are conflicting reports³³. *PIK3CA* mutations are the most common aberration of the *PIK3CA/AKT/mTOR* pathway in bladder cancer³², and thus are an attractive target for therapy. There remains however insufficient evidence currently to support its use as a clinical biomarker in bladder cancer.

1.4.1.3. *CDKN2A*

CDKN2A aberrations were associated with poorer recurrence-free survival and cancer-specific survival in cystectomy cohort³⁰. Again, there is insufficient data currently to support its use in clinical practice as a biomarker.

1.4.1.4. *Chromatin regulating genes*

Despite the dominance of chromatin regulating gene aberrations (e.g. *ARID1A*, *MLL2*, *KDM6A*), no association with clinical outcomes has been reported.

1.4.1.5. *TERT promoter*

One limitation of the TCGA dataset is the fact that the *TERT* promoter was not included. *TERT* promoter aberrations have been shown to be common alterations on in MIBC with frequencies of 66.3%⁶¹ reported. The vast majority of alterations occur at position 124 and 146 from the transcription start site. In a study which included primarily non-muscle-invasive bladder cancer, the impact of a *TERT* promoter mutation on survival was seen to be affected by the presence of a specific polymorphism (i.e. a normal variant occurring in at least 1% of the population; rs2853669); a *TERT* promoter mutation was associated with poorer survival in the absence of the polymorphism but not in its presence⁶¹.

1.4.1.6. DNA Damage Repair Genes

Various publications have looked at the impact of somatic alterations in the DNA damage repair (DDR) genes in MIBC, as both potential prognostic and predictive biomarkers. In a cohort of 81 patients with MIBC treated with radical cystectomy³⁸, those with an alteration in at least one of 6 DDR genes (*ATM*, *ERCC2*, *FANCD2*, *PALB1*, *BRCA1*, *BRCA2*) had a significantly better recurrence-free survival compared to those with none, suggesting a potential role for DDR alterations as a prognostic biomarker. Tumours with mutations in at least one of these genes had significantly higher overall numbers of somatic mutations.

Further work focusing on DDR alterations came from Plimack et al⁶² who interrogated DNA extracted from formalin-fixed paraffin-embedded (FFPE) blocks using a targeted sequencing panel. In their cohort of 34 patients receiving neoadjuvant chemotherapy for MIBC, 13/15 patients responding to NAC (i.e. downstaged to T1N0 or less at cystectomy) had an alteration in at least one of either *ATM*, *RB1* or *FANCC*. In contrast, none of the 19 patients classified as non-responders had an aberration in the above mentioned genes. The authors validated their findings in a cohort of 24 patients.

ERCC2 has been put forward as a promising potential biomarker of response to platinum-based neoadjuvant chemotherapy. As a nucleotide excision repair gene, it plays a key role in DNA repair, including the repair of cisplatin-induced crosslinks. It has been hypothesised that loss of function or overexpression may confer sensitivity or resistance to cisplatin respectively. Van Allen et al³⁷ performed whole exome sequencing with a mean target coverage of 121x on 59 tumour samples from patients with MIBC undergoing neoadjuvant chemotherapy. Of the 50 successfully sequenced, half had achieved downstaging with NAC to pT0/is and were considered 'responders'. The remaining 25 'non-responders' had residual muscle-invasive disease post NAC. 'Responders' were noted to have a significantly higher median mutation rate compared to 'non-responders'. Enrichment analysis was performed to identify genes selectively mutated in responders compared to non-responders. *ERCC2* was the only gene significantly enriched in the responder cohort with

all non-synonymous mutations in *ERCC2* (n=9) occurring in 'responders'. However, in this study, 16/25 responders to NAC did not demonstrate an *ERCC2* mutation. With an overall reported mutation frequency rate of only 12%³², it could be argued that as present in only a minority of patients, clinical use as a biomarker would be limited.

In a study focussing on 48 MIBC patients treated with radiotherapy +/- chemotherapy, tissue samples were subjected to screening using a targeted next generation sequencing (NGS) panel²⁰. Twenty-six patients had non-synonymous alterations in at least one DDR gene although 64% of the mutations seen had not been previously reported and were of unclear significance. The authors therefore focussed on those likely to have functional impact and this resulted in a cohort of 12 patients. Alterations were seen in *ATM*, *BRCA1*, *BRIP1*, *ERCC2*, *FANCD2*, *PALB2* and *RAD50*. On univariable analysis, the presence of a DDR deleterious alteration was associated with a non-significant trend towards reduced incidence of bladder recurrence or any recurrence. In a multivariable model with stage and complete TURBT as covariates, a trend towards reduced recurrence (local and distant) was reported again²⁰. They also reported that *ERCC2* mutations were associated with lower 2-year distant metastatic recurrence, and that only 1 of 6 patients with an *ERCC2* missense aberration had a local recurrence. The subset of patients with an *ERCC2* aberration in their cohort was very small and further work is needed to explore the role of *ERCC2* in the context of radiotherapy response. There is however a current project looking at the potential of *ERCC2* and other DNA damage repair gene aberrations as biomarkers in identifying patients suitable for a bladder preservation approach⁶³.

In other work looking at DDR genes, Teo et al²⁴ used the MSK-IMPAKT gene panel to profile 100 patients with advanced urothelial cancer, including 82 with bladder cancer, all treated with either cisplatin or carboplatin chemotherapy. 47/100 had at least 1 aberration in the 34 DDR genes included in the panel. The cohort was dichotomised into those with or without DDR aberration. The authors reported a significantly better survival in the subgroup with a DDR aberration than those without (23.7 vs 13 months). On multivariate analysis,

only performance status and DDR status were significant independent prognostic factors.

1.4.2. Potential transcriptomic biomarkers

At a transcriptomic level, there has been much interest in determining whether molecular subtypes could be a useful biomarker in bladder cancer.

As illustrated in section 1.3.2.2, numerous molecular subtyping classifications have been put forward by different groups, with some also exploring whether subtype could be a prognostic biomarker. Blaveri et al⁴⁸ identified 3 clusters within their subcohort of 47 MIBC cases. One cluster had a poor median survival of 5.4 months compared to 16.7 and 19.3 months in the other two clusters. Unfortunately, details regarding the gene expression profile of this poor prognosis group are not available. Sjodahl et al⁵¹ identified clusters associated with good, intermediate and poor prognoses, with the good prognosis group consisting primarily of NMIBC cases. Damrauer et al⁵³ reported basal and luminal subtypes in MIBC similar to those seen in breast cancer. Luminal tumours had better disease-specific and overall survival compared to basal tumours.

In 2014, Choi et al⁵⁶ published a paper exploring molecular subtypes in MIBC and suggested that a 'p53-like' subtype could be associated with chemoresistance. The chemoresistant 'p53-like' subtype they reported was associated with a 100% rate of non-response to NAC (defined as downstaging to pT0 or pT1). However, only 7 of the 26 patients within this group received chemotherapy and so the numbers involved are too small to draw any definitive conclusions. However, such a finding would have significant clinical impact if validated and their work introduced the concept that molecular subtype could be associated with chemoresistance, and so be used as a predictive biomarker. Work by Seiler et al⁵⁴ suggested that patients with basal-squamous subtypes gained the most benefit from neoadjuvant chemotherapy in terms of overall survival, whereas claudin-low tumours did not see any improvement in survival with NAC.

With regards to the use of molecular subtype to predict radiotherapy response, there is very little published data. One group reported on the use of molecular subtype in predicting response to bladder-sparing trimodality treatment⁶⁴. In their work, gene expression microarray data was generated for 189 patients with MIBC undergoing bladder-sparing trimodality treatment. 108 cases (57.1%) passed quality checks and were classified into one of 4 subtypes- basal, basal claudin-low, infiltrated luminal or luminal. The authors reported a trend towards improved overall survival in the luminal subtype (HR 0.63, p = 0.1), and a trend for patients with infiltrated luminal cancers to require less salvage surgery (p=0.08). It is interesting that only 57.1% of their cases with microarray data generated passed quality checks for inclusion in the final analysis. The reasons for this are not clear and are not addressed within the published abstract.

A 24-gene hypoxia signature for MIBC has also been developed from publicly available transcriptomic data. This has been applied to independent datasets and shown to be predictive of benefit from the addition of carbogen and nicotinamide (CON) to radiotherapy, and prognostic of outcomes in surgical and radiotherapy + CON cohorts⁶⁵.

More recently, work has also looked at whether immune and stromal signatures are prognostic in patients who have undergone bladder preservation with trimodality treatment (TMT) i.e. maximal TURBT and chemoradiation⁶⁴. The authors reported that patients with high immune infiltration had improved disease-specific survival after TMT but this was not seen in a surgical cohort, suggesting that immune signature might be a predictive biomarker of response to TMT.

1.4.3. Immunohistochemical biomarkers

Multiple immunohistochemical targets have been put forward as potential biomarkers including p53^{19,66,67}, Rb1^{19,66}, ERCC1^{68,69} and XRCC1⁶⁹. Those of particular interest in the context of response to chemotherapy and/or radiotherapy include MRE11, AIMP3 and hypoxia markers.

1.4.3.1. MRE11

MRE11 is part of the MRN complex which plays a key role in the detection of DNA damage. *MRE11* expression was shown to be predictive of cause-specific survival following radical radiotherapy for MIBC^{18,19} with lower levels associated with poorer outcomes. This was noted to be counter-intuitive as the hypothesis had been that lower expression of *MRE11* would result in lower detection of DNA damage with subsequent reduced repair capacity leading to radiosensitivity. Further work to understand the underlying mechanisms suggested that *MRE11* is modified in the post-translational setting given that *MRE11* protein and mRNA levels did not correlate⁷⁰. Of note, work by the original reporting team to standardise testing across 3 UK centres failed to reproduce the same result²¹, and the authors concluded that this was more likely due to issues with standardising the assay and interpretation, rather than there necessarily not being a biological basis for their initial result.

1.4.3.2. AIMP3

AIMP3 is a tumour suppressor gene and upstream regulator of DNA damage. Loss of expression as seen on immunohistochemistry was reported to be associated with poorer overall survival⁷¹ when explored in a phase III cohort of 217 patients with MIBC randomised to radiotherapy alone or with the addition of carbogen and nicotinamide. No difference in survival was seen in a comparison cystectomy group indicating *AIMP3* expression status could be a predictive biomarker of survival following radiotherapy in MIBC. This data has not yet been validated in an independent dataset.

1.4.3.3. Hypoxia markers

HIF1- α

HIF1- α is considered a marker of hypoxia. Data on *HIF1- α* expression on immunohistochemistry was available for 137/217 patients treated within the BCON study^{72,73}. Authors reported that patients with high *HIF1- α* status receiving carbogen and nicotinamide had better outcomes compared to those receiving radiotherapy alone, suggesting that *HIF1- α* status could be used to help identify those patients most likely to benefit from the addition of these hypoxia modifying agents. Carbogen and nicotinamide however are used in a

limited number of centres, with most using chemotherapy as a radiosensitiser as per the BC2001 study¹⁶.

CAIX & GLUT1

CAIX and *GLUT1* are regulated by *HIF1- α* and have been shown to be intrinsic markers of hypoxia, and independent predictors of overall and cause-specific survival in MIBC⁷⁴ in a small retrospective cohort of 21 patients. However, these results were not reproduced in a larger retrospective cohort by the same group⁷⁵.

Necrosis

Using 231 samples from the BCON study, Eustace et al⁷⁵ demonstrated that necrosis, as assessed on histopathological features, was an independent prognostic factor in patients with bladder cancer receiving radiotherapy alone or radiotherapy plus nicotinamide and carbogen. They also showed that necrosis was a predictive indicator of benefit from the addition of carbogen and nicotinamide versus radiotherapy alone. Conversely, necrosis was not found to be a predictive or prognostic in the BC2001 cohort who were treated with radiotherapy +/- concurrent chemotherapy (5-fluorouracil and mitomycin C)⁷⁶.

1.4.4. Biomarkers of radiosensitivity

1.4.4.1. Biology of Radiosensitivity

In this MD(Res), I plan to explore potential biomarkers of radiosensitivity. I will therefore give a brief overview of some of the relevant radiobiology underpinning radiosensitivity and discuss further candidate biomarkers.

Radiotherapy is the use of ionising radiation to induce DNA damage and subsequent cell death. Ionising radiation may directly hit the DNA causing direct damage, or may cause indirect damage via the formation of free radicals including electrons. Most biological damage is caused by ejected electrons i.e. indirectly.

DNA damage is detected and repair attempted via several pathways. Those of particular relevance in radiotherapy are base excision repair (BER), homologous recombination (HR) and non-homologous end-joining (NHEJ).

Figure 1.2 summarises the different types of DNA repair of particular relevance to radiosensitivity, and also includes the nucleotide excision pathway given its relevance to neoadjuvant chemotherapy (as discussed in section 1.4.1.6).

The BER pathway repairs most single strand breaks and involves molecules such as PARP and XRCC1. Expression of XRCC1 and ERCC1 was put forward as a possible prognostic biomarker of improved outcomes after trimodality treatment in a retrospective study, but these findings have not been independently validated⁶⁹.

Double strand DNA breaks (DSB) are considered the most 'lethal' and their presence is detected by the ATM protein and MRN complex, which is formed of the MRE11, NBS1 and RAD50⁷⁷. The importance of ATM and NBS1 are demonstrated by the clinical syndromes of ataxia telangiectasia and Nijmegen breakage syndrome where patients have mutations in *ATM* and *NBS1* respectively, resulting in characteristic features and radiosensitivity.

In MIBC, genomic aberrations in *ATM* have been shown to be associated with response to neoadjuvant chemotherapy⁶², and overall improved outcomes³⁸. In the context of chemoradiation for MIBC, Desai et al²⁰ reported that deleterious aberrations in DNA damage repair genes (including *ATM*) may be associated with a lower risk of recurrence. However, within their cohort of 48 patients, only 5 had *ATM* aberrations and so it is not possible to comment on the significance of these in the context of chemoradiation. Interestingly, no association between *ATM* expression and outcomes has been noted at an immunohistochemical level^{18,19}. This may reflect post-translational modification and emphasises the importance of further exploring the underlying mechanisms and biology.

Double strand DNA breaks can be repaired by one of two pathways- HR or NHEJ. NHEJ is the primary mode of DSB repair. Members of this pathway

include *DNA-PK*, *XRCC4* and *DNA ligase IV*. Members of the HR pathway include *BRCA1/2*, *RAD54L*, *BRIP1* and *SLX4*.

The oxygen fixation hypothesis puts forward that DNA damage is enhanced by the presence of oxygen⁷⁸, hence hypoxia is thought to be a contributing factor towards radioresistance.

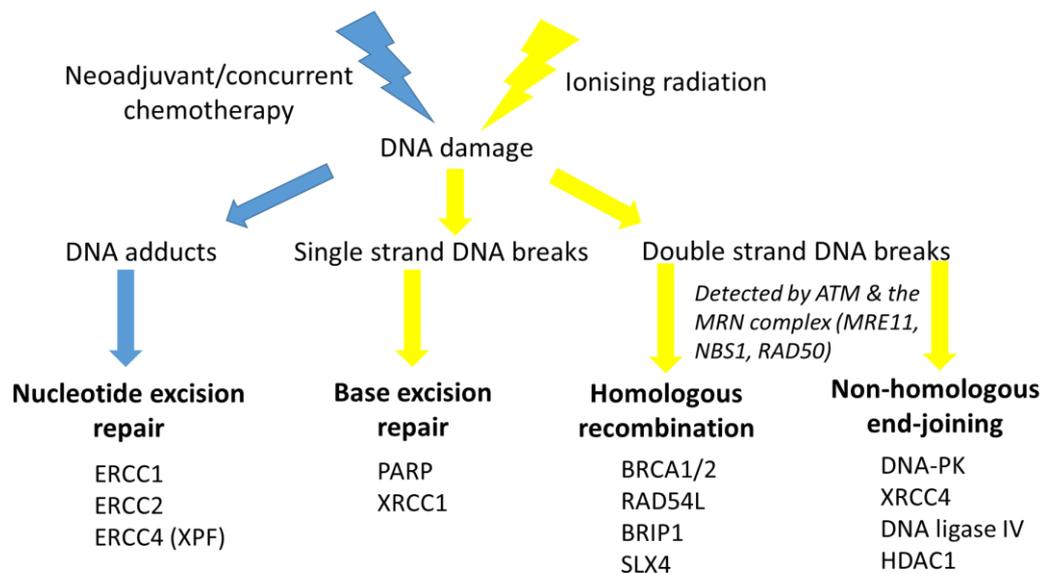


Figure 1.2 Diagram summarising primary DNA pathways for type of DNA damage.

Several groups have also explored gene signatures associated with radiosensitivity, and one that has been perhaps more explored is the radiosensitivity index.

1.4.4.2. The Radiosensitivity Index (RSI)

The radiosensitivity index (RSI) was derived from work investigating the SF2 in 48 cell lines from the National Cancer Institute panel of 60 using an Affymetrix HU-6800 platform⁷⁹. The final model consists of 10 genes which are ranked according to their expression level (highest 10 and lowest 1). The following rank-based linear algorithm is then applied to calculate the RSI. The lower the RSI value, the more radiosensitive the tumour is proposed to be.

Rank-based linear algorithm

$$(-0.0098009 \times AR) + (0.0128283 \times cJun) + (0.0254552 \times STAT1) - (0.0017589 \times PKC) - (0.0038171 \times RelA) + (0.1070213 \times cABL) - (0.0002509 \times SUMO1) - (0.0092431 \times CDK1) - (0.0204469 \times HDAC1) - (0.0441683 \times IRF1)$$

In this paper⁷⁹, the algorithm was applied to 2 prospective pilot cohorts of patients with rectal and oesophageal cancer undergoing pre-operative chemoradiation (n = 14 and 12 respectively). The authors reported that the model distinguished responders from non-responders with mean RSI values of 0.34 vs 0.48 (p=0.002) respectively. Response was defined as either a pathological/clinical complete response or decrease in at least 1 or 2 T-stages in rectal and oesophageal cancer respectively. Of note, cohort sizes were small and the oesophageal group was very heterogeneous in terms of chemoradiation received.

In 92 patients with head and neck cancer (HNC) treated within randomised trials with concurrent chemoradiation, the average RSI was lower when compared to the rectal and oesophageal cohorts (0.06 vs 0.39 vs 0.43)⁷⁹. The authors suggested this was likely secondary to platform differences i.e. Affymetrix vs NKL arrays. However, the predicted radiosensitive group had an improved 2-year locoregional control rate of 86% vs 61%. Of note, they hypothesised that as their model would not take into account the radiosensitising effect of chemotherapy in this cohort, they expected the model to be most accurate in the most radiosensitive quartile and so chose a RSI cut off point for HNC patients at 25th percentile.

Further work in 2 retrospective breast cancer cohorts totalling 503 subjects⁸⁰ reported that patients receiving adjuvant radiotherapy classified as radiosensitive using the RSI had an improved 5-year recurrence-free survival compared to those labelled as radioresistant. This difference was not seen in patients who did not receive radiotherapy, in keeping with RSI being a predictive rather than prognostic tool.

The RSI has also been reported to predict clinical outcome in patients with glioblastoma multiforme⁸¹ and pancreatic cancer⁸². Criticisms of the RSI

include that it is based upon cell line data and therefore does not take into account the contribution of microenvironment and stromal cells on radiosensitivity. At the time of project design, the RSI had not been applied to MIBC. I therefore felt it would be worthwhile exploring this further and explore this chapter 5.

1.5. Approaches to Interrogating Tumour Samples

1.5.1. Limitations of current approaches

1.5.1.1. FFPE

The vast majority of translational work performed in MIBC at the time of setting up the MD(Res), has been on tumour tissue in the form of either formalin-fixed paraffin-embedded (FFPE) blocks or fresh frozen (FF) tissue.

FFPE blocks form the backbone of histopathological diagnosis in the clinical setting and so form a valuable resource in translational work, given the large numbers of samples potentially available for study. However, while FFPE are ideal for preserving the tissue architecture to allow histological diagnosis, the fixation process greatly affects the integrity of DNA and RNA extracted from these samples. Formalin damages nucleic acids by creating cross-links and increasing fragmentation⁸³. Deamination of cytosine also results in C>T/G>A artefacts, which are characteristic of DNA extracted from FFPE⁸⁴ and may result in false positive or false negative results⁸³, particularly in the 1-10% allelic frequency where such artefact has been demonstrated to be disproportionately greater compared to that in the 10-25% allelic frequency range⁸⁴.

Other factors such as ischaemic time, fixation duration and duration of block storage are also thought to potentially affect downstream analysis of nucleic acids extracted from FFPE tissue⁸⁵. Unfortunately these factors can vary quite significantly between samples and are impossible to control in the setting of retrospective sample collection.

Some groups however have suggested that although differences exist in technical parameters of next generation sequencing between FF- and FFPE-derived nucleic acids (e.g. coverage variability, library insert size), there is overall concordance with regards to sequencing data obtained, and concluded therefore that FFPE-derived nucleic acids can be reliably used in NGS^{86,87}. These studies however have involved relatively small numbers of matched FFPE and FF samples.

Potential strategies to reduce the risk of false positives would therefore include using a higher threshold for variant allele frequency to call aberrations, particularly if high levels of coverage are being sought where there is likely to be increased noise and artefact. Technical replicates (i.e. testing the same nucleic acid sample repeatedly) would give confidence in labelling a variant as a true aberration if found to be reproducible. Biological replicates (i.e. testing multiple samples from the same tumour), might also be useful but discordant results here could be either due to artefact or intratumour heterogeneity.

1.5.1.2. Intratumour heterogeneity

Intratumour heterogeneity describes the observation that individual cancer cells in a tumour may display different genomic aberrations and biological features. In a landmark paper, Gerlinger et al⁸⁸ performed whole-exome sequencing on multiple spatially-separated regions of the same renal cell tumour, and demonstrated that not all mutations seen in one region were necessarily identified in another. This concept of intratumour heterogeneity was put forward as a likely contributing cause of treatment failure and drug resistance. Intratumour heterogeneity is thought to arise by the acquisition of mutations as a cancer grows over time and through tumour evolution⁸⁹, with expansion of subclones that offer a selective advantage.

There has been little published looking at intratumour heterogeneity in muscle-invasive bladder cancer, with the majority of work exploring intratumour heterogeneity in superficial disease⁹⁰. Diaz-Cano et al⁹¹ took 44 cases of MIBC and tested a deep and superficial sample from each tumour. Heterogeneity in loss of tumour suppressor genes (*RB1*, *WT1*, *TP53*, *NF1*) was reported between the superficial and deep samples in just over 31% of cases.

More recently, Thomsen et al⁹² performed whole exome sequencing at mean target coverage of 68x (range 27-215) on a total of 22 tissue samples taken from 3 treatment-naïve patients with regional nodal disease +/- distant metastases. The majority of samples were fresh frozen tissue. They reported significant heterogeneity between the primary tumour and metastases, but not within the primary itself (3 samples taken from each primary). They also concluded there was a lack of therapeutic targets common to all cancer cells. They went on to use WES data on samples from four patients (mean coverage 65x; range 35-113x) to design two amplicon panels composed of the aberrations with predicted functional impact (n = 1435 and n = 802)⁹³. The original samples then underwent laser microdissection and regions dissected were sequenced using the panels to a mean depth of 3049x. In two patients, they reported the presence of subclones in the primary tumour and concluded that there was evidence for subclonal populations and intratumour heterogeneity. However it should be noted that in one of the patients, samples from MIBC and NMIBC were compared and concluded to show heterogeneity. I would suggest that the different aberration profiles seen here are to be expected given that MIBC and NMIBC are considered two separate entities with divergent molecular pathways⁵⁵. More interesting however is the second patient who had a large multifocal T3b tumour; sequencing of 7 samples taken from the primary tumour showed shared *STAG2*, *PIK3CA*, *CLTC* and *CHEK2* aberrations across all 7 samples, but an otherwise shared *TP53* mutation appeared to be absent in one sample and subclonal in a second. Multiregion whole-exome sequencing data to be generated within this project will hopefully offer some further insight into this question in MIBC, and this will be reported in chapter 4.

1.5.1.3. Subtype heterogeneity

In the same way that intratumour heterogeneity has been noted at a genomic level in cancers, it is not unreasonable to question whether such heterogeneity would also be present with regards to molecular subtype i.e. will different tumours from the same patient or even different regions within a primary tumour be allocated different molecular subtypes due to heterogeneity in gene

expression profiles? I think this is an interesting question that has not been specifically studied in any cancer until recently.

In 2017, Thomsen et al performed multiregion expression profiling in 4 patients with MIBC and classified samples into either basal or luminal subtypes⁹³. In two of the patients, they reported co-existence of basal and luminal subtypes. It is worth noting however that in one patient, the tumour assigned a basal subtype was a MIBC whilst the luminal subtype tumour was a synchronous NMIBC. The 'heterogeneity' reported here is then not so surprising given that MIBC and NMIBC are recognised to develop along separate pathways. The co-existence however of the basal and luminal subtypes in a single invasive tumour in the second patient is of more significance and interest.

More recently, Warrick et al⁹⁴ explored intratumoural heterogeneity of molecular subtypes in bladder cancer patients with more than one tumour histology seen on examination of cystectomy specimens from 309 patients. They used the Lund immunohistochemistry classification and reported that out of 83 cases with variant (i.e. non TCC) histology, 41/83 had at least 2 histologies present and 16 (39% had molecular heterogeneity). It is a pity that they limited their analysis to those with variant histology as with their larger cohort, it would have been interesting to see if the IHC classification differentiated between those with the same morphology.

I will now discuss alternative approaches to address the limitations discussed above, and will introduce the work I have planned within this MD(Res).

1.5.2. Alternative approaches to interrogating genomic aberrations in MIBC

1.5.2.1. Plasma DNA

What is plasma DNA?

It has long been recognised that blood and other bodily fluids, such as urine and cerebrospinal fluid⁹⁵, contain fragments of DNA⁹⁶. The mechanism by which such fragments are released into the circulation has not been fully

elucidated but likely involves apoptosis, necrosis and active secretion⁹⁷. DNA fragments may therefore be derived from a variety of cells including normal (wild-type) cells, tumour cells or even fetal cells⁹⁸. These fragments have collectively been termed 'cell-free DNA'.

Cell-free DNA isolated from blood plasma is referred to as plasma DNA. Part of my work in Chapter 3 will look at the role of plasma DNA as a potential biomarker in MIBC.

There has been great interest in the potential use of cell-free DNA in cancer research for many years. In 1977, it was demonstrated that blood serum cell-free DNA levels were raised in cancer patients compared to healthy controls, and that levels were higher in patients with metastatic disease compared to those with non-metastatic disease⁹⁹. Furthermore, this paper reported that where decreases in serum DNA levels were seen after radiation treatment, they appeared to be associated with a decrease in tumour size or reduction of pain.

Later work has since shown that in patients with cancer with detectable plasma DNA, a proportion of the fragments are derived not only from the primary tumour¹⁰⁰, but also from macro- and potentially micro-metastatic deposits^{101,102}. These tumour-specific DNA fragments are termed circulating tumour DNA (ctDNA).

Advantages of plasma DNA

As touched upon in section 1.5.1.1, traditional tissue biopsies are inevitably subject to spatial bias, and the invasive nature of obtaining such biopsies often limits the potential to collect sequential samples. Plasma DNA offers a possible way to overcome these limitations, and has the further advantage of potentially encompassing tumour heterogeneity¹⁰⁰⁻¹⁰². As samples are obtained via a simple blood draw, sequential samples can be taken conveniently and safely, thereby overcoming the practical considerations of repeated invasive biopsies, including that of patient safety, patient willingness and acceptance. Furthermore it may be that this approach allows sampling of tumour cells when

a tissue biopsy might otherwise not be possible e.g. tumour at an inaccessible site or biopsy precluded by patient comorbidities.

Challenges associated with plasma DNA

In the majority of patients with malignancies, plasma DNA is mainly composed of wildtype i.e. normal DNA. Therefore in order to detect the tumour component of plasma DNA i.e. ctDNA, it is necessary to identify tumour-specific somatic aberrations. Quantification of such aberrations is used as a surrogate for tumour fraction. Although some patients with advanced disease have demonstrated high tumour fractions above 50%¹⁰², these are the minority and in numerous studies of metastatic disease, tumour fractions as low as 0.04%^{103,104} have been reported. Therefore, in order to distinguish ctDNA from wildtype DNA, ultra-sensitive techniques have been required to identify and detect somatic aberrations harboured by tumour DNA fragments.

Broadly speaking, two approaches have been employed in detecting ctDNA. One strategy is to first sequence a patient's tumour tissue before selecting identified aberrations to interrogate in plasma. Tumour tissue may be subjected to a broad *de novo* sequencing approach such as whole-exome (WES) or whole-genome sequencing (WGS), or a more targeted approach using a pre-defined set of aberrations via assays or a panel. The latter approach requires *a priori* knowledge of the relevant aberrations in the cancer of interest.

The advantage of a targeted panel over whole genome or whole exome sequencing is that lower DNA inputs are required, and a higher depth is achievable for much less cost. This means that aberrations present at low abundance (i.e. subclones or rare variants) can be more readily detected. This is a particular advantage when sequencing plasma DNA where in the majority of cases, the tumour component is already 'diluted' by wildtype DNA. Furthermore, as the workflow with a targeted panel is quicker than that of WES or WGS, results should be more quickly available.

The alternative strategy is to sequence plasma upfront, again either with a broad or more targeted approach as described above. While broad

approaches can give an overview of aberrations present and information regarding copy number aberrations, a minimum tumour fraction, in the order of at least 10-20%^{105,106}, is required. This approach will therefore not be achievable in a significant proportion of cases where tumour fraction does not meet this threshold, and is a particular barrier in the setting of early disease, where ctDNA levels may be in region of 0.01%¹⁰⁷. In these scenarios, PCR-based specific assays or a focussed NGS approach such as droplet digital PCR (dd-PCR), BEAMing or TAM-Seq can achieve sensitivity levels down to 0.01%^{107,108}. The disadvantage of these techniques however is that the assays are designed for specific mutations and only a few can be interrogated at a time, thus greatly limiting the breadth of the panel. In cancers such as melanoma where the TCGA dataset demonstrates the presence of a *BRAF* V600E aberration in 206/240 (85.5%) of patients¹⁰⁹, such small panels may be able to capture the majority of patients, or at least those with targetable aberrations. However, this is more challenging in MIBC where there is no such single or even group of documented specific mutations that can encompass the majority of patients with MIBC.

In Chapter 3, I will be exploring the role of both plasma DNA and ctDNA as a biomarker in MIBC, and will now review the current knowledge base of both these in turn.

1.5.2.2. Plasma DNA as a potential biomarker in cancer

Several groups have looked at whether plasma DNA levels may be of clinical relevance with regards to the diagnosis and prognosis of various cancer types. There have been conflicting results both between and within tumour sites. Table 1.4 summarises a selection of relevant publications, and I will focus on those relevant to bladder cancer in the next section.

Of note, some early work reported on DNA fragments isolated from blood serum ('serum DNA'). However, it has since been demonstrated that cell-free DNA levels are higher in serum than in plasma¹¹⁰, due to lysis of white blood cells during clotting¹¹¹. Plasma DNA is therefore generally preferred to avoid the further 'dilution' of ctDNA with wild-type DNA fragments from lysed white blood cells.

Table 1.4 Selection of relevant papers exploring plasma (or serum) DNA as a diagnostic and prognostic tool in cancer

Reference	Cancer Site	N	Key findings
Schwarzenbach et al ¹¹²	Colorectal	55 patients 20 healthy volunteers	Median serum DNA concentration in CRC patients was 868ng/ml. Healthy volunteer concentration ranged from 5-16ng/ml
Gormally et al ¹¹³	Multiple including lung, bladder, leukaemia	359 patients- (Bladder= 89) 776 controls 49 with COPD	Samples from 9 countries in Europe analysed. The amount of plasma DNA showed great variation between centres; healthy volunteer mean = 67ng/ml (SD = 403!). Plasma DNA was not significantly associated with development of bladder cancer (note- not specified whether NMIBC or MIBC)
Mehra et al ¹¹⁴	Advanced prostate	571	Baseline plasma DNA levels were an independent prognostic variable of radiological progression-free survival and overall survival in 1 st and 2 nd line chemotherapy settings

Table 1.4 (continued)

Reference	Cancer Site	N	Key findings
Perkins et al ¹¹⁵	Advanced solid cancers- multiple types included eg. (CRPC, breast)	104 patients 20 healthy volunteers	Dichotomised patients into 'high' and 'low' plasma DNA concentration levels using healthy volunteer data; median OS was 10.5 vs 6.5 months for low and high respectively. Plasma DNA level was an independent predictor of survival in a multivariate model.
Gangadha et al ¹¹⁶	Advanced Melanoma	25	Reported a positive correlation between plasma DNA level and radiological disease burden (R = 0.54, p = 0.0363)
Ellinger et al ¹¹⁷	Bladder cancer	45 (33/45- MIBC)	Cell-free serum DNA fragment concentration was significantly higher in patients with bladder cancer compared to healthy controls (16.9ng/ml vs 6.5ng/ml). Levels did not correlate with clinicopathological features or disease progression

Table 1.4 (continued)

Reference	Cancer Site	N	Key findings
Hauser et al ¹¹⁸	Bladder Cancer	95 with BCa 132 subjects with no BCa (benign urological disease, TURBT or healthy controls)	Significant increase in serum small DNA fragments in cancer patients, but no significant difference seen when comparing specifically with patients with benign urological conditions.
Tissot et al ¹¹⁹	Non-small cell lung cancer	218	Baseline plasma DNA concentration was an independent prognostic factor for overall survival on multivariate analysis, but was not significantly associated with patient or tumour characteristics, or response to chemotherapy
Valpione et al ¹²⁰	Metastatic Melanoma	38	Baseline plasma DNA concentration correlated with pre-treatment radiological tumour burden, and was significantly correlated with overall survival.

Plasma DNA in MIBC- as a diagnostic tool

Overall, there has been relatively little work exploring plasma or cell-free DNA as a diagnostic tool in bladder cancer. Ellinger et al¹¹⁷ looked at a cohort of 45 patients undergoing cystectomy for bladder cancer (including 33 with MIBC). They demonstrated that serum DNA concentration was significantly higher in patients with bladder cancer compared to healthy controls with medians of 16.9ng/ml and 6.5ng/ml respectively. They reported however that levels did not correlate with clinicopathological features or disease progression.

Hauser et al¹¹⁸ went on to expand on this work and performed a multicentre prospective study, recruiting 95 patients with bladder cancer and 132 patients without bladder cancer (51 healthy controls, 31 with benign urological disease and 48 patients undergoing TURBT but with no bladder cancer). Of note, only 20/95 (21.2%) of the bladder cancer cohort had MIBC, although subgroup analysis reported similar serum DNA concentration in MIBC and NMIBC patients. Using quantitative real-time PCR to measure short and long beta-actin serum DNA fragments, higher levels of short DNA fragments were documented in patients with bladder cancer when compared to non-cancer subjects. However, further analysis found that patients undergoing TURBT without bladder cancer had similar serum DNA levels to patients with bladder cancer, thus limiting the diagnostic applicability of serum DNA levels in bladder cancer.

Given that plasma and serum cell-free DNA levels are raised in non-malignant conditions such as inflammation, auto-immune conditions, trauma and in physiological states (e.g. after exercise)⁹⁷, and the fact that there is overlap in the range of levels observed in individuals with and without cancer, I would suggest that plasma DNA levels alone are currently unlikely to be specific enough in the diagnostic setting. However, of more interest is whether plasma DNA levels may be of clinical relevance once a diagnosis of MIBC has been made i.e. is it prognostic of overall outcome, do levels reflect disease burden, and could levels be used to monitor patients on treatment or over a period of time?

Plasma DNA in MIBC- as a prognostic tool

Baseline plasma DNA levels have been shown to be associated with tumour burden and prognosis¹²⁰⁻¹²² in some cancer types including breast, prostate and colorectal cancer.

To the best of my knowledge, no studies have looked specifically at whether plasma DNA levels may be of potential interest as a prognostic tool or a surrogate of disease burden in MIBC. This question forms the basis of my first hypothesis in Chapter 3. My second hypothesis will look at the potential role of ctDNA fraction as a biomarker of disease status and response to treatment.

1.5.2.3. Circulating tumour DNA as a potential biomarker in cancer

ctDNA in cancer research

ctDNA has been extensively explored across many cancer types. One of the key papers from Bettegowda et al¹²³ looked at 640 patients with localised and metastatic disease. Like many papers that followed, they reported that the proportion of ctDNA fragments has been shown to increase with disease burden^{103,107,123} and also to vary between tumour types¹²³.

It has also been shown that detection of ctDNA may pre-date clinical diagnosis of progression^{103,124,125} and therefore could be a useful tool in monitoring patients for disease progression or relapse. Studies have also demonstrated the emergence of subclones associated with disease progression and treatment resistance^{102,126,127} thus furthering our understanding of mechanisms of disease progression and allowing exploration of potential therapeutic targets.

With regards to application to the clinical setting, in the UK, ctDNA testing for *EGFR* mutations is currently available from 7 NHS centres¹²⁸ for non-small cell lung cancer patients suitable for treatment with a tyrosine-kinase inhibitor (TKI) but with no biopsy sample available or failed tissue testing, or to identify a T790M mutation in those who have progressed on a TKI. At the time of writing, this is however not routinely offered and work is ongoing to evaluate its introduction into routine NHS care.

ctDNA in MIBC

At the time of setting up this MD(Res), there was very little published on ctDNA in MIBC. Sidransky et al¹²⁹ performed the first work over 30 years ago showing that *TP53* mutations were detectable in urine from patients with MIBC. There then followed a hiatus in ctDNA work, mainly hindered by the lack of sensitive detection methods. Bettgowda's landmark paper¹²³ in 2014 studied ctDNA in 640 patients with various tumour types. Bladder cancer patients were included in the analysis of patients with metastatic disease but only accounted for 2.2% (3/164) of the cohort. Of note however, ctDNA was detectable in all 3 cases. Aberrations were first identified in tumour tissue using a targeted next generation sequencing (NGS) panel then quantified in plasma using SafeSeq. In all 3 patients, a *TP53* mutation was detected (mutant fragments/5ml plasma = 2.0, 308 and 655).

However, there has since been huge interest in ctDNA in MIBC. In 2016, Sonpavde et al published an abstract reporting the use of a commercially available panel (Guardant360) to sequence 29 plasma samples from patients with advanced urothelial cancer¹³⁰. The panel consisted of 68 cancer-related genes and aberrations were detected in 86.2% of patients, with the most frequently detected being *TP53* and *BRCA1*. Later work using an alternative 62-gene commercial panel (FoundationACT) was reported to detect at least one aberration in plasma of 48/66 (73%) patients with metastatic urothelial cancer¹³¹. Of note, both of these commercial panels were not bladder cancer-specific and included at most only 9 of the most frequent 23 gene mutations documented by the TCGA³¹. The detection rate therefore is impressive compared to other publications using custom, albeit much smaller bladder cancer-specific panels, and is most likely due to the fact that the commercial panel cohorts included only patients with advanced disease where ctDNA levels would be expected to be higher than in those with localised disease only.

Using custom panels, work has demonstrated that by using a patient-specific approach where aberrations have first been identified in an individual's tumour tissue, it is possible to identify the same aberrations in plasma and urinary

DNA in patients with localised and advanced disease¹³²⁻¹³⁵. Groups have used a combination of whole exome sequencing and highly focused panels to interrogate tumour tissue. Using a panel of 3 hotspot mutations in *PIK3CA* and *FGFR3*, Christensen et al¹³⁴ only detected aberrations in 11% of tumour tissue samples whereas work from the same group using WES to interrogate tumour tissue identified aberrations in 100%¹³³. Patel et al¹³⁵ used a tagged amplicon sequencing (TAm-Seq) panel of 8 bladder cancer-specific genes to detect mutant DNA in tumour tissue, plasma, urinary cell pellet (UCP), and urinary supernatant (USN) in a cohort of 17 patients undergoing neoadjuvant chemotherapy for MIBC. The eight genes were *BRAF*, *CTNNB1*, *FGFR3*, *HRAS*, *KRAS*, *NFE2L2*, *PIK3CA*, and *TP53*. Aberrations were detected in tumour tissue from 12/16 (75%) patients, and on interrogating plasma, UCP and USN from this subset, mutant DNA was seen in 4/12 (33%), 5/12 (42%), and 5/12 (42%), respectively. Overall, aberrations were detected in 10/17 (59%) patients in plasma and urine samples taken prior to commencing NAC. With relatively low rates of aberration detection using focussed approaches, this data demonstrates to me that we have yet to identify the optimal panel with which to interrogate bladder cancer and this is something I will explore further in chapter 4.

Vandekerkhove et al¹³⁶ designed a 50-gene bladder cancer-specific panel based upon published literature, with target depth of 500-1000x to interrogate plasma upfront. They reported that 25/44 (56.8%) patients demonstrated a ctDNA fraction above their 2% detection threshold set. Given that ctDNA frequencies can be as low as 0.04%^{103,104}, it is perhaps not surprising that despite the broad range of targets, they detected aberrations in only 56.8% of their patients, of whom over 80% had metastatic disease. Only 1 of the 7 patients with localised disease had detectable aberrations in plasma. This illustrates the challenges previously discussed of being able to balance breadth of a panel and depth. However, Christensen et al¹³⁷ explored the feasibility of using unique identifiers (UIDs) in an effort to lower the error rate associated with NGS and so optimise ctDNA detection in plasma. They reported achieving an average detection limit of 0.4% and used a 51-gene panel to sequence 65 plasma samples from 38 patients with MIBC receiving

either neoadjuvant or palliative chemotherapy to a mean depth of 8124x. Their panel identified 24 of 38 (63.1%) aberrations previously detected with ddPCR.

The same group very recently published work using a different approach¹³⁸. Instead of using a broad panel for all patients, they used a patient-specific approach whereby tumour tissue and matched germline DNA for 68 patients with organ-confined MIBC undergoing NAC and cystectomy underwent WES to target coverage 104x. Sixteen targets were selected for each patient and multiplex PCR NGS performed to interrogate these targets in plasma at a median target coverage of 105 000 x. Samples were termed 'positive' if 2 or more targets were present. They found that the detection of ctDNA before NAC and before cystectomy had prognostic value, and that the presence of ctDNA following cystectomy predicted for relapse (100% sensitivity, 98% specificity).

In summary, research published on ctDNA in MIBC has demonstrated that higher tumour fractions are seen with higher disease burdens¹³⁶, that ctDNA is more readily detectable in urine than in plasma^{134,135}, and that the detection of ctDNA at diagnosis is prognostic¹³⁸. Furthermore, in patients undergoing cystectomy for MIBC, increased plasma ctDNA levels post-operatively have been shown to be associated with lower recurrence-free survival and overall survival¹³⁴, and in the post-operative monitoring period, detection of ctDNA was seen to precede radiological/clinical diagnosis of relapse with median lead-times of up to 101 days^{133,138}. In samples taken prior to the second cycle of NAC, median lead time over radiological diagnosis of progression was 243 days¹³⁵. These studies show great promise for ctDNA as a potential biomarker of response to neoadjuvant chemotherapy and definitive treatment, but do highlight the challenges present in determining the optimal panel for interrogating samples. While WES on tumour tissue undoubtedly gives the most information with which to then target ctDNA, scalability in the clinical setting needs to be considered. Furthermore, I suggest that the potential to interrogate plasma upfront is appealing as archived tissue may not be representative of aberrations present at the time of disease relapse, and contemporary tissue samples may not be available.

iChorCNA

In interrogating ctDNA in MIBC, most approaches to date have required *a priori* knowledge regarding specific mutations present and ctDNA levels are used as a surrogate for tumour fraction. This means that there is the potential to underestimate tumour fraction or obtain falsely negative results if the relevant aberrations have not been interrogated. However, in 2017, a new bioinformatics pipeline (iChorCNA) using copy number data derived from low-pass whole-genome sequencing (LP-WGS) to estimate tumour fraction, while considering ploidy and subclonality, was published¹⁰⁵. It was reported primarily as a tool to identify which plasma DNA samples contained sufficient ctDNA to undergo whole-exome sequencing, and was demonstrated in 520 cases of metastatic breast and prostate cancer, using data from WGS at 0.1x coverage. Seventy-three percent of cases were found to have detectable tumour fraction (>0.03). Results were also compared to data from WGS at 10-48x on plasma DNA and matched tumour biopsies (n = 7). High levels of concordance at a mega-base scale level were seen. Tumour fraction estimates from LP-WGS and WGS were concordant¹⁰⁵.

The potential to be able to assess tumour fraction without requiring knowledge of patient specific-aberrations is greatly appealing. Moreover LP-WGS is cost efficient (less than £50 per sample) and therefore has opportunities for up-scaling.

As far as I am aware, the application of this algorithm has not been reported in MIBC. However, the current knowledge base would suggest that there is no reason why it should not also be of relevance to MIBC given that copy number alterations (CNA) are common in MIBC, with 54% of patients reported to have whole-genome doubling events³¹. By exploring the TCGA bladder data further, I was able to conclude that 90.3% of patients had at least one focal CNA. The question then is whether such copy number aberrations are detectable in ctDNA?

Soave et al¹³⁹ looked at plasma DNA copy number changes with respect to 43 regions containing 37 genes using multiplex ligation-dependent probe amplification and detected CNA in 48.6% patients. This cohort included

patients with NMIBC. When considering MIBC alone, CNA were detected in 21/46 (45.7%) of patients.

Using LP-WGS, Patel et al demonstrated CNA in the plasma of 4/16 (25%) of patients commencing neoadjuvant chemotherapy (NAC). Of note, LP-WGS had detected CNA in the tumour tissue of all 16 patients. Furthermore, serial plasma samples from 5 patients undergoing NAC demonstrated evidence of tumour evolution under selective pressure of treatment.

The work to date supports my hypothesis that tumour fraction estimated from CNA from LP-WGS data may have potential as a marker of response to treatment in MIBC. However, its clinical utility may be limited in patients with early localised disease where lower levels of ctDNA may fall below the detection threshold of the iChorCNA pipeline. This will be explored in chapter 3.

1.5.3. Alternative approaches to interrogating molecular subtypes in MIBC

1.5.3.1. Evidence for a Pan-Cancer Approach

While many groups have focussed on subtypes specific to a cancer type, there is increasing interest in adopting a pan-cancer approach where tissue of origin itself is not used to arbitrarily divide cases.

Hoadley et al¹⁴⁰ explored this in a study of 3527 samples across 12 cancer types, including bladder cancer. Data from 5 genome wide platforms and one proteomic platform were integrated. They reported 11 major subtypes based on DNA copy number, DNA methylation, mRNA expression, microRNA, and protein parameters. 5 subtypes were almost identical to tissue of origin ie. members of the subgroup all had the same cancer type, but some different cancer types converged into one subtype, and some cancers diverged into multiple subtypes. Of note, bladder cancer was one of the most divergent tumour types and cases were classified into 7 of the 11 subtypes. 95.8% of bladder cases were divided into 3 main subtypes; 74/120 (61.7%) fell in to the

bladder cancer (BLCA) group (which additionally included 2 breast cancers, 1 head and neck, and 1 lung cancer), 10/120 (8.3%) were classified as LUAD-enriched (composed mainly of non-small cell lung cancers), and 31/120 (25.8%) had a squamous-like subtype which contained mostly head and neck, and lung squamous cancers. Although all the bladder cases in the squamous-like group had squamous histology, it was noted that most cases had less than 50% squamous differentiation. Higher immune signatures were also noted in the squamous-like group and on exploring outcomes of the bladder cancer cases, those in the squamous-like and LUAD-enriched groups had worse prognosis.

Additionally, one bladder case was grouped into the kidney subtype, another into glioblastoma, and 3 further cases into subtypes called 'small'- these subtypes were later discounted due to the very low number of members in each group.

On comparing bladder cancer cases in the BLCA and squamous-like subtypes, distinct differences were noted at the genomic, transcriptomic and protein levels. The squamous-like group was characterised by loss of 3p, higher levels of immune signatures and higher expression of epithelial-to-mesenchymal transition (EMT) markers. The BLCA group overall demonstrated higher levels of HER2 and Rab25 proteins.

I think it is interesting that given the comparisons drawn between the basal and luminal subtypes in breast and bladder cancer⁵³, no bladder cases fell into either of the subtypes dominated by breast cancers. This may be a reflection of the fact that this analysis integrated expression data with genomic and proteomic data.

These results show the diverse nature of BLCA and lends more support to adopting a personalised approach in MIBC. This paper also highlights the fact that a pan-cancer approach may open new avenues in the drive to identify targetable treatments, which might not otherwise be considered if limited by tissue of origin. Preliminary work looking at applying a colorectal subtype

classifier to a range of epithelial cancers had shown interesting results, particularly with regards to bladder cancer and I will discuss this now further.

1.5.3.2. Applying a colorectal cancer classifier to MIBC

In a key landmark paper, Sadanandam et al⁴⁵ applied consensus-based non-negative matrix factorisation (NMF)¹⁴¹ to publicly available gene expression data on 445 colorectal tumours. They reported the presence of 5 molecular subtypes and developed a 786-gene panel ('CRCassigner-786') to classify colorectal cancers. This was validated in further data sets with samples totalling 744. Table 1.5 below summarises the key features of each subtype. In a cohort of 197 patients for whom survival data was available, subtype did not appear to be associated with clinical outcome but on subgroup analysis of treatment-naïve patients (n=120), those with stem-like tumours had the poorest outcomes with median overall survival of less than 80 months, compared to transit-amplifying and goblet-like subtypes which were associated with the best prognosis (median overall survival not reached with median follow-up period of 122 months). The authors also noted a trend suggesting that patients with stem-like tumours gained benefit in terms of disease-free survival with adjuvant chemotherapy+/- radiotherapy, whereas adjuvant treatment in patients with TA and goblet-cell tumours appeared to have a detrimental effect. However, with relatively modest numbers in each treatment-naïve subtype (range 18-33) and overall low incidence of disease recurrence, further work in larger datasets is required to validate these findings.

Table 1.5 Colorectal subtypes identified by Sadanandam et al⁴⁵

Subtype	Features
Enterocyte	High expression of enterocyte-specific genes eg. <i>CA1</i> , <i>C12</i> , <i>KRT20</i> , <i>SLC6A3</i> , <i>AQP8</i> , <i>MS4A12</i>
Transit-amplifying	Mixed group with variable expression of stem cell and Wnt target genes; further divided into cetuximab resistant or cetuximab sensitive
Goblet-like	High expression of goblet-specific <i>MUC2</i> and <i>TTF3</i>
Stem-like	High expression of Wnt signalling targets (<i>SFRP2</i> , <i>SFRP4</i>), stem cell, myoepithelial and mesenchymal genes (<i>ZEB1</i> , <i>ZEB2</i> , <i>TWIST1</i> , <i>SNA12</i>). Low expression of differentiation markers
Inflammatory	Relatively high expression of chemokines & interferon-related genes (<i>CXCL13</i> , <i>CXCL11</i> , <i>CXCL10</i> , <i>CXCL9</i> , <i>IFIT3</i>)

Preliminary work in Dr Sadanandam's lab has suggested that bladder cancer could be subtyped using their colorectal classifier. On applying their CRCassigener-786 to TCGA bladder data (n=122), they reported that 61% of the bladder cases were allocated a subtype and 39% were considered 'mixed' i.e. features of more than one subtype¹⁴². Using the TCGA classifications available at that time (clusters I-IV)³², the CRCassigener-786 stem-like subtype appeared to overlap with cluster 2, their inflammatory subtype appeared to with the TCGA basal-like cluster 3, and the remaining CRCassigener-786 groups of TA, goblet like and enterocyte were split across the TCGA cluster 1, which was characterised by luminal features.

Given the characteristics of the colorectal subtypes, these preliminary findings appear to be broadly in keeping with what would be expected i.e. the

enterocyte and goblet-like subtypes which are associated with markers of well-differentiated cells appear to share similarities with the luminal subtypes, and the inflammatory subtype which contains chemokine markers overlaps with the basal cluster 3.

This preliminary work by Dr Sadanandam has demonstrated that MIBC can be divided into subtypes using the CRCAssigner-786. Studies have suggested that molecular subtype could be a predictive and prognostic biomarker in MIBC^{48,51-53}. As part of my work in Chapter 5, I will explore whether the CRCAssigner subtypes have clinical relevance in MIBC, with particular focus on radiotherapy response.

1.5.3.3. Nanostring platform

While RNA expression work has been mainly performed on microarray data, or more recently, on RNA-sequencing data, Dr Sadanandam's lab have been using the Nanostring platform in their work.

The Nanostring platform is essentially a digital counter. No sample amplification is required (thus eliminating inherent PCR-based errors) and the turnaround time (<48 hours from RNA extraction to data) means this is a clinically applicable platform. In brief, capture and reporter probes are hybridised to the target RNA fragments. Hybridised probes are bound to a cartridge which is then analysed by the nCounter platform which directly counts the reporter probes to give absolute results on target gene levels. In addition, the Nanostring platform is suitable for RNA samples that are degraded as is often the case with FFPE samples. While I had initially hoped to explore subtypes in MIBC using RNA-seq data generated from FFPE RNA, initial preliminary laboratory work suggested this would not be feasible due to the poor quality of FFPE RNA.

The CRCAssigner-786 has been refined to a panel of 38 genes (CRCAssigner-38). These genes form the basis of a custom Nanostring panel designed by Dr Sadanandam and his team.

Application of the Nanostring platform has been reported in various cancer types but at the time of setting up this project, I was aware of only one

publication reporting its use in MIBC. In this paper, Bellmunt et al¹⁴³ used Nanostring data as part of an integrative exploration of invasive urothelial cancers (103 cases; of whom 85 had clinical data). This work concluded that alterations in the PIK3CA/AKT/mTOR pathway had the greatest impact on survival (although this was not statistically significant), and that overexpression of CTNNB1 and PIK3CA were significantly associated with a poorer overall survival.

1.5.3.4. Clustering techniques

Clustering refers to the grouping of similar data points together so that members of one group have similar features to each other, but are distinct from members of another group. There are various techniques available but a full technical description is beyond the scope of this thesis. The majority of work in MIBC to date has used hierarchical clustering (table in appendix 5.1). A simplified description of a 'bottom-up' approach to hierarchical clustering would be that initially, each individual datapoint/sample is considered as a cluster. Points which are considered similar are grouped together, and using an iterative process, this continues until ultimately there would be one cluster containing all the points. Hierarchical clustering therefore requires subjective evaluation of the number of clusters evident¹⁴¹. There are several approaches to define how similar one datapoint is to another. One method is that of using the 'distance' between centroids. A centroid is the 'average' of a cluster and distance refers to the difference between 2 centroids.

Dr Sadanandam's clustering work has been primarily based upon NMF. This describes a robust technique which can reduce thousands of genes to a much smaller number of 'metagenes'. Patterns of expression of the metagenes can then be used to define subtypes¹⁴¹. An advantage of NMF is that, unlike with hierarchical clustering, an objective assessment of the number of groups present (k) can be inferred.

1.6. Conclusion

MIBC is a challenging disease and overall clinical outcomes are poor. There are currently no approved biomarkers in use to guide its management. A personalised medicine approach is needed to try and improve outcomes.

In this research thesis, I aim to explore MIBC at a genomic and transcriptomic level with the overarching aim of identifying potential biomarkers of clinical relevance.

At a genomic level, I will explore the potential role of plasma DNA as a prognostic marker in MIBC and assess whether the ctDNA fraction is of potential interest as an early indicator of response to chemotherapy. As part of this work, I plan to generate multiregion whole-exome sequencing data from FFPE tumour tissue to identify tumour-specific aberrations, and will further use these results alongside publicly available data to design a targeted panel with which to interrogate MIBC samples.

At a transcriptomic level, I will explore whether molecular subtype may be associated with radiotherapy response, using an existing colorectal cancer classifier but also a custom-designed gene expression panel.

To support these pilot studies, a resource of clinical material and data from patients with MIBC is essential. My first aim therefore is to establish a protocol within the Royal Marsden Hospital to allow the collection of blood, tissue and urine samples from patients with MIBC alongside clinical data. I report on establishing this protocol in Chapter 2.

1.7. References

1. Gray, P.J., *et al.* Clinical-pathologic stage discrepancy in bladder cancer patients treated with radical cystectomy: results from the national cancer data base. *Int J Radiat Oncol Biol Phys* **88**, 1048-1056 (2014).
2. Stein, J.P., *et al.* Radical cystectomy in the treatment of invasive bladder cancer: long-term results in 1,054 patients. *J Clin Oncol* **19**, 666-675 (2001).
3. CRUK. Bladder Cancer Statistics. Vol. 2016.
4. Vale, C. Neoadjuvant chemotherapy in invasive bladder cancer: a systematic review and meta-analysis. *The Lancet* **361**, 1927-1934 (2003).
5. Advanced Bladder Cancer Meta-analysis, C. Neoadjuvant chemotherapy in invasive bladder cancer: update of a systematic review and meta-analysis of individual patient data advanced bladder cancer (ABC) meta-analysis collaboration. *Eur Urol* **48**, 202-205; discussion 205-206 (2005).
6. Grossman, H.B., *et al.* Neoadjuvant chemotherapy plus cystectomy compared with cystectomy alone for locally advanced bladder cancer. *N Engl J Med* **349**, 859-866 (2003).
7. Alfred Witjes, J., *et al.* Updated 2016 EAU Guidelines on Muscle-invasive and Metastatic Bladder Cancer. *Eur Urol* **71**, 462-475 (2017).
8. Chang, S.S., *et al.* Treatment of Non-Metastatic Muscle-Invasive Bladder Cancer: AUA/ASCO/ASTRO/SUO Guideline. *J Urol* **198**, 552-559 (2017).
9. National Institute for Health & Care Excellence. Bladder Cancer: diagnosis and management. Vol. 2016 (2015).
10. Griffiths, G., *et al.* International phase III trial assessing neoadjuvant cisplatin, methotrexate, and vinblastine chemotherapy for muscle-invasive bladder cancer: long-term results of the BA06 30894 trial. *J Clin Oncol* **29**, 2171-2177 (2011).
11. Hafeez, S., *et al.* Selective organ preservation with neo-adjuvant chemotherapy for the treatment of muscle invasive transitional cell carcinoma of the bladder. *Br J Cancer* **112**, 1626-1635 (2015).
12. Lee, C.T., *et al.* Cystectomy Delay More Than 3 Months From Initial Bladder Cancer Diagnosis Results in Decreased Disease Specific and Overall Survival. *The Journal of Urology* **175**, 1262-1267 (2006).
13. Porta, N., *et al.* Hypo-Fractionation in Muscle-Invasive Bladder Cancer: An Individual Patient Data (IPD) Meta-Analysis of the BC2001 and BCON Trials. *International Journal of Radiation Oncology*Biological*Physics* **105**, S138 (2019).
14. Ploussard, G., *et al.* Critical analysis of bladder sparing with trimodal therapy in muscle-invasive bladder cancer: a systematic review. *Eur Urol* **66**, 120-137 (2014).
15. Mak, R.H., *et al.* Long-term outcomes in patients with muscle-invasive bladder cancer after selective bladder-preserving combined-modality therapy: a pooled analysis of Radiation Therapy Oncology Group protocols 8802, 8903, 9506, 9706, 9906, and 0233. *J Clin Oncol* **32**, 3801-3809 (2014).

16. James, N.D., *et al.* Radiotherapy with or without chemotherapy in muscle-invasive bladder cancer. *N Engl J Med* **366**, 1477-1488 (2012).
17. Petrelli, F., *et al.* Correlation of pathologic complete response with survival after neoadjuvant chemotherapy in bladder cancer treated with cystectomy: a meta-analysis. *Eur Urol* **65**, 350-357 (2014).
18. Choudhury, A., *et al.* MRE11 expression is predictive of cause-specific survival following radical radiotherapy for muscle-invasive bladder cancer. *Cancer Res* **70**, 7017-7026 (2010).
19. Laurberg, J.R., *et al.* Expression of TIP60 (tat-interactive protein) and MRE11 (meiotic recombination 11 homolog) predict treatment-specific outcome of localised invasive bladder cancer. *BJU Int* **110**, E1228-1236 (2012).
20. Desai, N.B., *et al.* Genomic characterization of response to chemoradiation in urothelial bladder cancer. *Cancer* **122**, 3715-3723 (2016).
21. Walker, A.K., *et al.* MRE11 as a Predictive Biomarker of Outcome After Radiation Therapy in Bladder Cancer. *Int J Radiat Oncol Biol Phys* **104**, 809-818 (2019).
22. Yafi, F.A., North, S. & Kassouf, W. First- and second-line therapy for metastatic urothelial cancer of the bladder *Curr Onc* **11**, e25-34 (2011).
23. Bellmunt, J., *et al.* Pembrolizumab as Second-Line Therapy for Advanced Urothelial Carcinoma. *N Engl J Med* **376**, 1015-1026 (2017).
24. Teo, M.Y., *et al.* DNA Damage Response and Repair Gene Alterations Are Associated with Improved Survival in Patients with Platinum-Treated Advanced Urothelial Carcinoma. *Clin Cancer Res* **23**, 3610-3618 (2017).
25. Ayers, M., *et al.* IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest* **127**, 2930-2940 (2017).
26. Rosenberg, J.E., *et al.* Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet* **387**, 1909-1920 (2016).
27. Murciano-Goroff, Y.R., Warner, A.B. & Wolchok, J.D. The future of cancer immunotherapy: microenvironment-targeting combinations. *Cell Res* **30**, 507-519 (2020).
28. Jiang, T., Chen, X., Su, C., Ren, S. & Zhou, C. Pan-cancer analysis of ARID1A Alterations as Biomarkers for Immunotherapy Outcomes. *J Cancer* **11**, 776-780 (2020).
29. Gui, Y., *et al.* Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat Genet* **43**, 875-878 (2011).
30. Kim, P.H., *et al.* Genomic predictors of survival in patients with high-grade urothelial carcinoma of the bladder. *Eur Urol* **67**, 198-201 (2015).
31. Robertson, A.G., *et al.* Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell* **171**, 540-556 e525 (2017).
32. Cancer Genome Atlas Research, N. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315-322 (2014).
33. Bellmunt, J., *et al.* Somatic Copy Number Abnormalities and Mutations in PI3K/AKT/mTOR Pathway Have Prognostic Significance for Overall

- Survival in Platinum Treated Locally Advanced or Metastatic Urothelial Tumors. *PLoS One* **10**, e0124711 (2015).
34. Cazier, J.B., *et al.* Whole-genome sequencing of bladder cancers reveals somatic CDKN1A mutations and clinicopathological associations with mutation burden. *Nat Commun* **5**, 3756 (2014).
 35. Lindgren, D., *et al.* Integrated genomic and gene expression profiling identifies two major genomic circuits in urothelial carcinoma. *PLoS One* **7**, e38863 (2012).
 36. Yin, M., *et al.* ATM/RB1 mutations predict shorter overall survival in urothelial cancer. *Oncotarget* **9**, 16891-16898 (2018).
 37. Van Allen, E.M., *et al.* Somatic ERCC2 mutations correlate with cisplatin sensitivity in muscle-invasive urothelial carcinoma. *Cancer Discov* **4**, 1140-1153 (2014).
 38. Yap, K.L., *et al.* Whole-exome sequencing of muscle-invasive bladder cancer identifies recurrent mutations of UNC5C and prognostic importance of DNA repair gene mutations on survival. *Clin Cancer Res* **20**, 6605-6617 (2014).
 39. van Oers, J.M., *et al.* FGFR3 mutations indicate better survival in invasive upper urinary tract and bladder tumours. *Eur Urol* **55**, 650-657 (2009).
 40. Wu, Q., Hoffmann, M.J., Hartmann, F.H. & Schulz, W.A. Amplification and overexpression of the ID4 gene at 6p22.3 in bladder cancer. *Mol Cancer* **4**, 16 (2005).
 41. Shen, H., *et al.* 6p22.3 amplification as a biomarker and potential therapeutic target of advanced stage bladder cancer. *Oncotarget* **4**, 2124-2134 (2013).
 42. Gallucci, M., *et al.* Status of the p53, p16, RB1, and HER-2 genes and chromosomes 3, 7, 9, and 17 in advanced bladder cancer: correlation with adjacent mucosa and pathological parameters. *J Clin Pathol* **58**, 367-371 (2005).
 43. Alizadeh, A.A., *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503-511 (2000).
 44. Perou, C.M., *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747-752 (2000).
 45. Sadanandam, A., *et al.* A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med* **19**, 619-625 (2013).
 46. Guinney, J., *et al.* The consensus molecular subtypes of colorectal cancer. *Nat Med* **21**, 1350-1356 (2015).
 47. Dyrskjot, L., *et al.* Identifying distinct classes of bladder carcinoma using microarrays. *Nat Genet* **33**, 90-96 (2003).
 48. Blaveri, E., *et al.* Bladder cancer outcome and subtype classification by gene expression. *Clin Cancer Res* **11**, 4044-4055 (2005).
 49. Sanchez-Carbayo, M., Socci, N.D., Lozano, J., Saint, F. & Cordon-Cardo, C. Defining molecular profiles of poor outcome in patients with invasive bladder cancer using oligonucleotide microarrays. *J Clin Oncol* **24**, 778-789 (2006).
 50. Lindgren, D., *et al.* Combined gene expression and genomic profiling define two intrinsic molecular subtypes of urothelial carcinoma and

- gene signatures for molecular grading and outcome. *Cancer Res* **70**, 3463-3472 (2010).
51. Sjobahl, G., *et al.* A molecular taxonomy for urothelial carcinoma. *Clin Cancer Res* **18**, 3377-3386 (2012).
 52. Choi, W., *et al.* Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer Cell* **25**, 152-165 (2014).
 53. Damrauer, J.S., *et al.* Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *Proc Natl Acad Sci U S A* **111**, 3110-3115 (2014).
 54. Seiler, R., *et al.* Impact of Molecular Subtypes in Muscle-invasive Bladder Cancer on Predicting Response and Survival after Neoadjuvant Chemotherapy. *Eur Urol* (2017).
 55. Knowles, M.A. & Hurst, C.D. Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. *Nat Rev Cancer* **15**, 25-41 (2015).
 56. Choi, W., *et al.* Intrinsic basal and luminal subtypes of muscle-invasive bladder cancer. *Nat Rev Urol* **11**, 400-410 (2014).
 57. Rebouissou, S., *et al.* EGFR as a potential therapeutic target for a subset of muscle-invasive bladder cancers presenting a basal-like phenotype. *Sci Transl Med* **6**, 244ra291 (2014).
 58. Lerner, S.P., *et al.* Bladder Cancer Molecular Taxonomy: Summary from a Consensus Meeting. *Bladder Cancer* **2**, 37-47 (2016).
 59. Paquet, E.R. & Hallett, M.T. Absolute assignment of breast cancer intrinsic molecular subtype. *J Natl Cancer Inst* **107**, 357 (2015).
 60. Cote, R.J., Esrig, D., Groshen, S., Jones, P.A. & Skinner, D.G. p53 and treatment of bladder cancer. *Nature* **385**, 123-125 (1997).
 61. Rachakonda, P.S., *et al.* TERT promoter mutations in bladder cancer affect patient survival and disease recurrence through modification by a common polymorphism. *Proc Natl Acad Sci U S A* **110**, 17426-17431 (2013).
 62. Plimack, E.R., *et al.* Defects in DNA Repair Genes Predict Response to Neoadjuvant Cisplatin-based Chemotherapy in Muscle-invasive Bladder Cancer. *Eur Urol* (2015).
 63. Mouw, K. Investigating the effect of ERCC2 mutations on DNA repair capacity and chemo-radiotherapy response in muscle-invasive bladder cancer. Vol. 2019.
 64. Efsthathiou, J.A., *et al.* Impact of Immune and Stromal Infiltration on Outcomes Following Bladder-Sparing Trimodality Therapy for Muscle-Invasive Bladder Cancer. *Eur Urol* **76**, 59-68 (2019).
 65. Yang, L., *et al.* A Gene Signature for Selecting Benefit from Hypoxia Modification of Radiotherapy for High-Risk Bladder Cancer Patients. *Clin Cancer Res* **23**, 4761-4768 (2017).
 66. Shariat, S.F., *et al.* p53, p21, pRB, and p16 expression predict clinical outcome in cystectomy with bladder cancer. *J Clin Oncol* **22**, 1014-1024 (2004).
 67. Malats, N., *et al.* P53 as a prognostic marker for bladder cancer: a meta-analysis and review. *The Lancet Oncology* **6**, 678-686 (2005).

68. Bellmunt, J., *et al.* Gene expression of ERCC1 as a novel prognostic marker in advanced bladder cancer patients receiving cisplatin-based chemotherapy. *Ann Oncol* **18**, 522-528 (2007).
69. Sakano, S., *et al.* ERCC1 and XRCC1 expression predicts survival in bladder cancer patients receiving combined trimodality therapy. *Mol Clin Oncol* **1**, 403-410 (2013).
70. Martin, R.M., *et al.* Post-transcriptional regulation of MRE11 expression in muscle-invasive bladder tumours. *Oncotarget* **5**, 993-1003 (2014).
71. Gurung, P.M., *et al.* Loss of expression of the tumour suppressor gene AIMP3 predicts survival following radiotherapy in muscle-invasive bladder cancer. *Int J Cancer* **136**, 709-720 (2015).
72. Hoskin, P.J., Rojas, A.M., Bentzen, S.M. & Saunders, M.I. Radiotherapy with concurrent carbogen and nicotinamide in bladder carcinoma. *J Clin Oncol* **28**, 4912-4918 (2010).
73. Hunter, B.A., *et al.* Expression of hypoxia-inducible factor-1alpha predicts benefit from hypoxia modification in invasive bladder cancer. *Br J Cancer* **111**, 437-443 (2014).
74. Hoskin, P.J., Sibtain, A., Daley, F.M. & Wilson, G.D. GLUT1 and CAIX as intrinsic markers of hypoxia in bladder cancer: relationship with vascularity and proliferation as predictors of outcome of ARCON. *Br J Cancer* **89**, 1290-1297 (2003).
75. Eustace, A., *et al.* Necrosis predicts benefit from hypoxia-modifying therapy in patients with high risk bladder cancer enrolled in a phase III randomised trial. *Radiother Oncol* **108**, 40-47 (2013).
76. Choudhury, A., *et al.* The predictive and prognostic value of tumour necrosis in muscle invasive bladder cancer patients receiving radiotherapy with or without chemotherapy in the BC2001 trial (CRUK/01/004). *Br J Cancer* **116**, 649-657 (2017).
77. Wouters, B. & Begg, A.C. Irradiation-induced damage and the DNA damage response. in *Basic Clinical Radiobiology* (eds Joiner, M. & van der Kogel, A.) 11-26 (Hodder Arnold, 2009).
78. Horsman, M., Wouters, B., Joiner, M. & Overgaard, J. *The oxygen effect and fractionated radiotherapy*, (Hodder Arnold, 2009).
79. Eschrich, S.A., *et al.* A gene expression model of intrinsic tumor radiosensitivity: prediction of response and prognosis after chemoradiation. *Int J Radiat Oncol Biol Phys* **75**, 489-496 (2009).
80. Eschrich, S.A., *et al.* Validation of a radiosensitivity molecular signature in breast cancer. *Clin Cancer Res* **18**, 5134-5143 (2012).
81. Ahmed, K.A., *et al.* The radiosensitivity index predicts for overall survival in glioblastoma. *Oncotarget* **6**, 34414-34422 (2015).
82. Strom, T., *et al.* Radiosensitivity index predicts for survival with adjuvant radiation in resectable pancreatic cancer. *Radiother Oncol* **117**, 159-164 (2015).
83. Do, H. & Dobrovic, A. Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. *Clin Chem* **61**, 64-71 (2015).
84. Wong, S.Q., *et al.* Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. *BMC Med Genomics* **7**, 23 (2014).

85. Bass, B.P., Engel, K.B., Greytak, S.R. & Moore, H.M. A review of preanalytical factors affecting molecular, protein, and morphological analysis of formalin-fixed, paraffin-embedded (FFPE) tissue: how well do you know your FFPE specimen? *Arch Pathol Lab Med* **138**, 1520-1530 (2014).
86. Spencer, D.H., *et al.* Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens. *J Mol Diagn* **15**, 623-633 (2013).
87. Liu, Y., *et al.* Next-generation RNA sequencing of archival formalin-fixed paraffin-embedded urothelial bladder cancer. *Eur Urol* **66**, 982-986 (2014).
88. Gerlinger, M., *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883-892 (2012).
89. Nowell, P.C. The clonal evolution of tumor cell populations. *Science* **194**, 23-28 (1976).
90. Gerlinger, M., *et al.* Intratumour heterogeneity in urologic cancers: from molecular evidence to clinical implications. *Eur Urol* **67**, 729-737 (2015).
91. Diaz-Cano, S.J., Blanes, A., Rubio, J., Matilla, A. & Wolfe, H.J. Molecular evolution and intratumor heterogeneity by topographic compartments in muscle-invasive transitional cell carcinoma of the urinary bladder. *Lab Invest* **80**, 279-289 (2000).
92. Thomsen, M.B., *et al.* Spatial and temporal clonal evolution during development of metastatic urothelial carcinoma. *Mol Oncol* **10**, 1450-1460 (2016).
93. Thomsen, M.B.H., *et al.* Comprehensive multiregional analysis of molecular heterogeneity in bladder cancer. *Sci Rep* **7**, 11702 (2017).
94. Warrick, J.I., *et al.* Intratumoral Heterogeneity of Bladder Cancer by Molecular Subtypes and Histologic Variants. *Eur Urol* (2018).
95. Peng, M., Chen, C., Hulbert, A., Brock, M.V. & Yu, F. Non-blood circulating tumor DNA detection in cancer. *Oncotarget* **8**, 69162-69173 (2017).
96. Mandel, P. & Metais, P. Les acides nucleiques du plasma sanguine chez l'homme. *Comptes rendus des seances de la Societe de biologie et de ses filiales* **142**, 241-243 (1948).
97. Schwarzenbach, H., Hoon, D.S. & Pantel, K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer* **11**, 426-437 (2011).
98. Lo, Y.M., *et al.* Presence of fetal DNA in maternal plasma and serum. *Lancet* **350**, 485-487 (1997).
99. Leon, S.A., Shapiro, B., Sklaroff, D.M. & Yaros, M.J. Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res* **37**, 646-650 (1977).
100. Stroun, M., *et al.* Neoplastic characteristics of the DNA found in the plasma of cancer patients. *Oncology* **46**, 318-322 (1989).
101. Chan, K.C., *et al.* Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. *Clin Chem* **59**, 211-224 (2013).

102. Murtaza, M., *et al.* Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* **497**, 108-112 (2013).
103. Dawson, S.J., *et al.* Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med* **368**, 1199-1209 (2013).
104. Newman, A.M., *et al.* An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med* **20**, 548-554 (2014).
105. Adalsteinsson, V.A., *et al.* Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun* **8**, 1324 (2017).
106. Todenhofer, T., Struss, W.J., Seiler, R., Wyatt, A.W. & Black, P.C. Liquid Biopsy-Analysis of Circulating Tumor DNA (ctDNA) in Bladder Cancer. *Bladder Cancer* **4**, 19-29 (2018).
107. Diehl, F., *et al.* Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci U S A* **102**, 16368-16373 (2005).
108. Conteduca, V., *et al.* Androgen receptor gene status in plasma DNA associates with worse outcome on enzalutamide or abiraterone for castration-resistant prostate cancer: a multi-institution correlative biomarker study. *Ann Oncol* **28**, 1508-1516 (2017).
109. Cancer Genome Atlas, N. Genomic Classification of Cutaneous Melanoma. *Cell* **161**, 1681-1696 (2015).
110. Trigg, R.M., Martinson, L.J., Parpart-Li, S. & Shaw, J.A. Factors that influence quality and yield of circulating-free DNA: A systematic review of the methodology literature. *Heliyon* **4**, e00699 (2018).
111. Chan, K.C., Yeung, S.W., Lui, W.B., Rainer, T.H. & Lo, Y.M. Effects of preanalytical factors on the molecular size of cell-free DNA in blood. *Clin Chem* **51**, 781-784 (2005).
112. Schwarzenbach, H., Stoecklacher, J., Pantel, K. & Goekkurt, E. Detection and monitoring of cell-free DNA in blood of patients with colorectal cancer. *Ann N Y Acad Sci* **1137**, 190-196 (2008).
113. Gormally, E., *et al.* Amount of DNA in plasma and cancer risk: a prospective study. *Int J Cancer* **111**, 746-749 (2004).
114. Mehra, N., *et al.* Plasma Cell-free DNA Concentration and Outcomes from Taxane Therapy in Metastatic Castration-resistant Prostate Cancer from Two Phase III Trials (FIRSTANA and PROSELICA). *Eur Urol* **74**, 283-291 (2018).
115. Perkins, G., *et al.* Multi-purpose utility of circulating plasma DNA testing in patients with advanced cancers. *PLoS One* **7**, e47020 (2012).
116. Gangadhar, T.C., *et al.* Feasibility of monitoring advanced melanoma patients using cell-free DNA from plasma. *Pigment Cell Melanoma Res* **31**, 73-81 (2018).
117. Ellinger, J., *et al.* Apoptotic DNA fragments in serum of patients with muscle invasive bladder cancer: a prognostic entity. *Cancer Lett* **264**, 274-280 (2008).
118. Hauser, S., *et al.* Cell-free serum DNA in patients with bladder cancer: results of a prospective multicenter study. *Anticancer Res* **32**, 3119-3124 (2012).

119. Tissot, C., *et al.* Circulating free DNA concentration is an independent prognostic biomarker in lung cancer. *Eur Respir J* **46**, 1773-1780 (2015).
120. Valpione, S., *et al.* Plasma total cell-free DNA (cfDNA) is a surrogate biomarker for tumour burden and a prognostic biomarker for survival in metastatic melanoma patients. *Eur J Cancer* **88**, 1-9 (2018).
121. Ellinger, J., *et al.* The role of cell-free circulating DNA in the diagnosis and prognosis of prostate cancer. *Urol Oncol* **29**, 124-129 (2011).
122. Cargnin, S., Canonico, P.L., Genazzani, A.A. & Terrazzino, S. Quantitative Analysis of Circulating Cell-Free DNA for Correlation with Lung Cancer Survival: A Systematic Review and Meta-Analysis. *J Thorac Oncol* **12**, 43-53 (2017).
123. Bettegowda, C., *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* **6**, 224ra224 (2014).
124. Tie, J., *et al.* Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci Transl Med* **8**, 346ra392 (2016).
125. Garcia-Murillas, I., *et al.* Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci Transl Med* **7**, 302ra133 (2015).
126. Carreira, S., *et al.* Tumor clone dynamics in lethal prostate cancer. *Sci Transl Med* **6**, 254ra125 (2014).
127. Murtaza, M., *et al.* Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat Commun* **6**, 8760 (2015).
128. Foundation, P. Developing effective ctDNA services for lung cancer. (2018).
129. Sidransky, D., *et al.* Identification of P53 Gene-Mutations in Bladder Cancers and Urine Samples. *Science* **252**, 706-709 (1991).
130. Sonpavde, G., *et al.* Circulating cell-free DNA profiling of patients with advanced urothelial carcinoma. *J Clin Oncol* **34**, abstr358 (2016).
131. McGregor, B.A., *et al.* Correlation of circulating tumor DNA (ctDNA) assessment with tissue-based comprehensive genomic profiling (CGP) in metastatic urothelial cancer (mUC). *J Clin Oncol* **36**, 453-453 (2018).
132. Birkenkamp-Demtroder, K., *et al.* Genomic Alterations in Liquid Biopsies from Patients with Bladder Cancer. *Eur Urol* **70**, 75-82 (2016).
133. Birkenkamp-Demtroder, K., *et al.* Monitoring Treatment Response and Metastatic Relapse in Advanced Bladder Cancer by Liquid Biopsy Analysis. *Eur Urol* (2017).
134. Christensen, E., *et al.* Liquid Biopsy Analysis of FGFR3 and PIK3CA Hotspot Mutations for Disease Surveillance in Bladder Cancer. *Eur Urol* **71**, 961-969 (2017).
135. Patel, K.M., *et al.* Association Of Plasma And Urinary Mutant DNA With Clinical Outcomes In Muscle Invasive Bladder Cancer. *Sci Rep* **7**, 5554 (2017).
136. Vandekerkhove, G., *et al.* Circulating Tumor DNA Reveals Clinically Actionable Somatic Genome of Metastatic Bladder Cancer. *Clin Cancer Res* **23**, 6487-6497 (2017).
137. Christensen, E., *et al.* Optimized targeted sequencing of cell-free plasma DNA from bladder cancer patients. *Sci Rep* **8**, 1917 (2018).

138. Christensen, E., *et al.* Early Detection of Metastatic Relapse and Monitoring of Therapeutic Efficacy by Ultra-Deep Sequencing of Plasma Cell-Free DNA in Patients With Urothelial Bladder Carcinoma. *J Clin Oncol* **37**, 1547-1557 (2019).
139. Soave, A., *et al.* Copy number variations of circulating, cell-free DNA in urothelial carcinoma of the bladder patients treated with radical cystectomy: a prospective study. *Oncotarget* **8**, 56398-56407 (2017).
140. Hoadley, K.A., *et al.* Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158**, 929-944 (2014).
141. Brunet, J.P., Tamayo, P., Golub, T.R. & Mesirov, J.P. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A* **101**, 4164-4169 (2004).
142. Poudel, P., *et al.* Revealing unidentified heterogeneity in different epithelial cancers using heterocellular subtype classification. *BioRxiv* (2017).
143. Bellmunt, J., *et al.* HER2 as a target in invasive urothelial carcinoma. *Cancer Med* **4**, 844-852 (2015).

Chapter 2. Establishing a Framework for Biomarker Discovery & Development in MIBC: the Collection of Clinical Material for Molecular Stratification in Muscle-Invasive Bladder Cancer (CoMB) protocol

2.1. Introduction

2.1.1. Rationale for a MIBC biomarker protocol

Translational research is dependent upon patient samples and data, and tissue biobanks are a key resource in supporting this. The Royal Marsden Hospital Generic Tissue Bank (RMH GTB) was established in 2012/13 and provides a resource for translational research projects both within the RMH/ICR and externally. All patients attending RMH are routinely invited to give their consent for blood and tissue to be collected and stored for future research by signing a 'Tissues for Research' consent form. At the time of planning initial bladder cancer translational pilots in 2015, the RMH GTB had 857 frozen tissue samples and 606 plasma samples stored under Urology, of which only 81 and 0 had 'bladder' in the topographic details respectively. Moreover, there was insufficient data available within the biobank records to determine which of the bladder tissue samples might be of use to my planned pilots. Furthermore, any samples accessed would be entirely anonymised and therefore without linked clinical data.

In order to be able to perform pilot translational studies in MIBC as part of this MD(Res), it was necessary to first establish a resource of clinical material and clinical data from patients with MIBC within the Royal Marsden Hospital. This chapter describes the protocol I established to allow the collection of human

material from MIBC patients and their use in translational work, and the preliminary laboratory results I generated to establish the blood processing protocol for plasma acquisition and tissue processing standard operating procedures.

2.1.2. Aim

My aim was to establish a protocol to allow the collection of tissue, urine, blood samples and linked clinical data from patients with MIBC to support pilot studies within my MD(Res), and future unplanned work.

2.2. Methods

2.2.1. Framework for collection of human material

The first step in setting up a resource of clinical material and data from patients with MIBC was to write a protocol detailing the eligibility criteria, sample collection schedules, clinical data collection and sample handling. The full protocol is included in Appendix 2.1 but the key points are summarised below.

2.2.1.1. Eligibility

The eligibility criteria are as follows,

- Aged 18 years or above with muscle-invasive histologically confirmed bladder cancer
- Able to give informed consent
- Able to attend appointments at RMH as per clinical follow-up schedule
- No current or prior non-bladder invasive malignancy within the last 3 years, other than non-melanoma skin cancer, clinically insignificant prostate cancer, CIN cervix

Patients are recruited to one of three groups depending on their treatment status. Group 1 includes patients who have been newly diagnosed and who have not received any treatment beyond a transurethral resection of bladder

tumour (TURBT) for their MIBC. Group 2 includes patients who have previously received radical or palliative treatment for MIBC and are under routine follow-up. Group 3 recruits patients who have relapsed following previous radical or palliative treatment and are continuing follow-up or treatment at RMH. Fig 2.1 shows the protocol schema.

In October 2018, a protocol amendment was proposed to also include patients with upper tract malignancies, and at the time of writing, the amendment is pending submission following statistician review.

2.2.1.2. Sample collection

Tissue

Archived FFPE blocks taken at diagnostic TURBT are requested from the relevant centres and stored in the RMH GTB under the CoMB protocol. Where patients have undergone further procedures e.g. cystectomy, tissue blocks are also requested where available, and the protocol also allows for storage of fresh frozen tissue.

Peripheral Blood

Up to 30mls of blood may be taken from a peripheral vein up to a maximum of 18 times per year. Samples are taken at the same time as routine clinical bloods. An amendment was proposed in October 2018 to allow research bloods to be taken separately if it is not possible to obtain samples at the same time as clinical venepuncture. Blood was initially collected in EDTA bottles and required processing by the RMH GTB within 2 hours. However, due to capacity issues within the RMH GTB that did not allow a guarantee that blood would be spun for plasma, aliquoted and frozen at -80 degrees within two hours, it was necessary to use Streck™ bottles instead (<https://www.streck.com/collection/cell-free-dna-bct/>). Streck tubes contain a preservative which prevent the release of genomic DNA. As such, samples are considered stable at room temperature for up to 14 days although, as advised by Professor Attard's team, ideally should be processed within 5 days at most.

Germline DNA

Germline samples are obtained from blood white blood cells (buffy coat of peripheral blood samples taken).

Urine samples

The protocol allows the collection of serial urine samples from patients, with a plan to store aliquots of up to 50mls. However, due to capacity issues in the RMH GTB in terms of both processing and storing the samples, it was not possible to initiate this during the timelines of my project.

Sample storage & access

Samples are stored within the RMH GTB and are allocated an anonymised unique identifier. They are catalogued using the FreezerPro system. They are released in accordance with the current standard operating procedure (SOP) ie. via a written application. On approval by the CoMB principal investigator and the RMH GTB management team, a material transfer agreement is put in place to allow access to the relevant samples. Samples used in my pilots have been catalogued on FreezerPro.

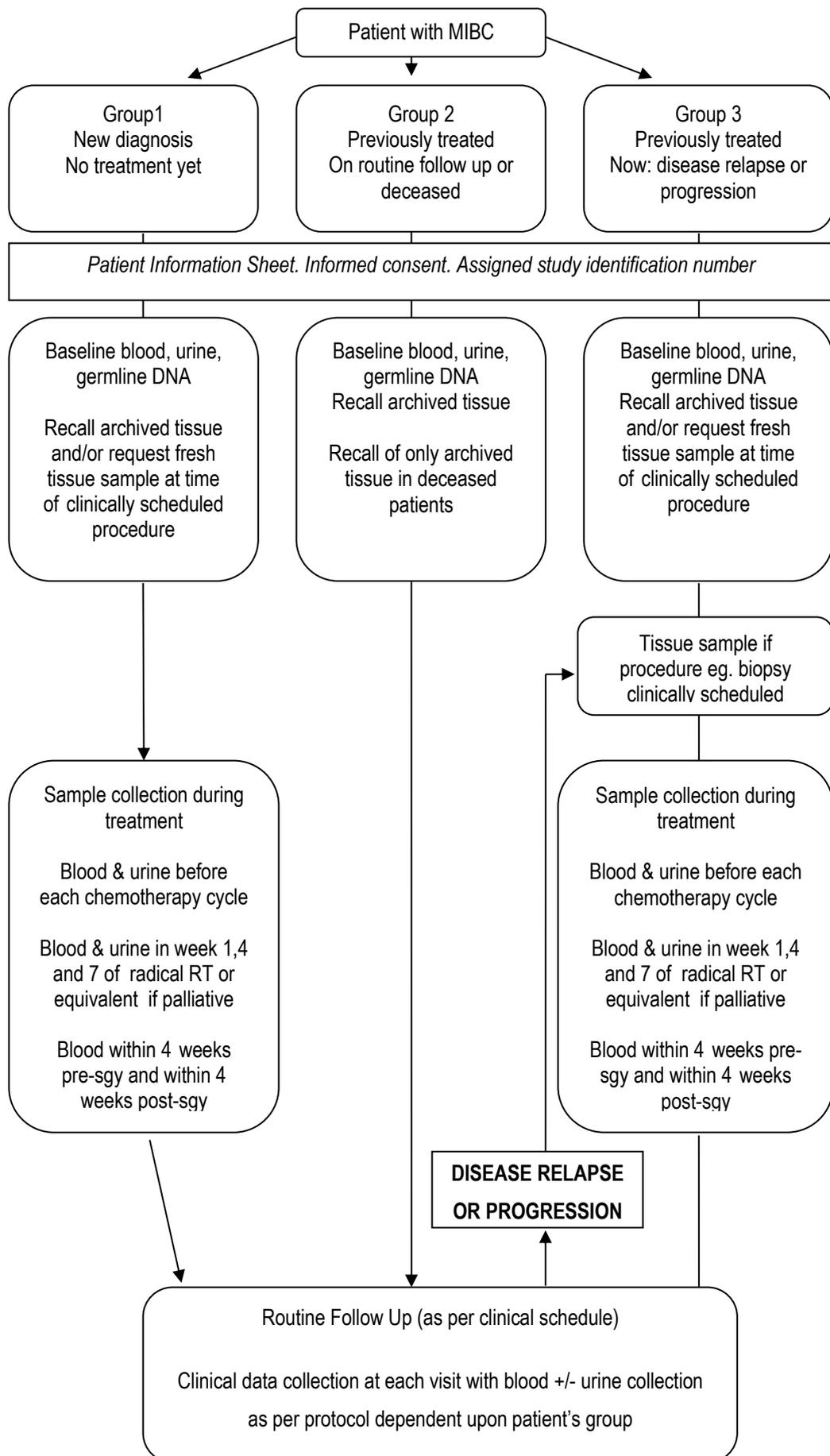


Figure 2.1 Schema of CoMB Protocol

2.2.1.3. Data Collection

Data is collected on case report forms which I created on establishing the protocol, and can be found within the protocol appendix. An amendment was made in October 2018 to reflect the introduction of immunotherapy agents.

2.2.1.4. Statistical analysis plan

It was anticipated that CoMB would support pilot studies in MIBC and potentially further expansion of those projects. It was therefore agreed to initially aim to recruit 200 patients within 3 years based upon initial statistical calculations demonstrating that with a biomarker prevalence of 33% and cohort size of 200, at least 80% power could be achieved to detect a difference in outcome of 22%. The initial suggested biomarker prevalence of 33% in the sample size calculation was based upon TCGA data¹ reporting a prevalence of statistically significant mutations of up to 49%. However, the majority of aberrations, and certainly those of subsequent interest in the pilot studies, fell short of the 33% prevalence initially stipulated. The latest protocol amendment therefore proposes instead to continue to collect samples in the context of a biobank framework rather than necessarily aiming for a defined target number, with a statistical analysis plan to be agreed for each individual pilot study.

With regards to statistical analysis, this will clearly depend upon each individual pilot being performed but the following general principles were agreed with the trial statisticians (Mansour Taghavi Azar Sharabiani and Ranga Gunapala), and were reached assuming an event rate of 40-50% and that one degree of freedom will be allowed for every 10 events:

- The minimum overall pilot study size required if survival analysis is to be performed will be 30
- If a comparison of survival outcomes is to be made between two subgroups, then the overall cohort size required will be a minimum of 40
- If comparing survival between groups subdivided according to the presence or absence of a specific aberration, the aberration should be present in at least 30% of the cohort to allow the comparison to be performed

2.3. Results

2.3.1. Key milestones

Figure 2.3 summarises the key milestones in setting up the CoMB protocol and recruitment progress up to the end of August 2018.

Of note, CoMB was very well received by a RMH patient panel in January 2015. They were very much in support of strategies pursuing a personalised approach in MIBC, and were satisfied with the proposed frequency of sample collection, even suggesting that more frequent sample collections would be acceptable.

Collection of plasma samples commenced before CoMB opened using the RMH 'Tissues for Research' consent form. Samples were processed and stored by Professor Attard's team pending formal RMH Committee for Clinical Research (CCR) approval. An application to the RMH Biomedical Research Centre (BRC) for support in collecting and processing plasma for ctDNA was successful in May 2015, ahead of CoMB formally opening in August 2015.

2.3.2. Ethical approval

I presented the CoMB protocol to the NRES Committee London- Fulham in May 2015 and approval was received for CoMB as a research tissue bank in June 2015.

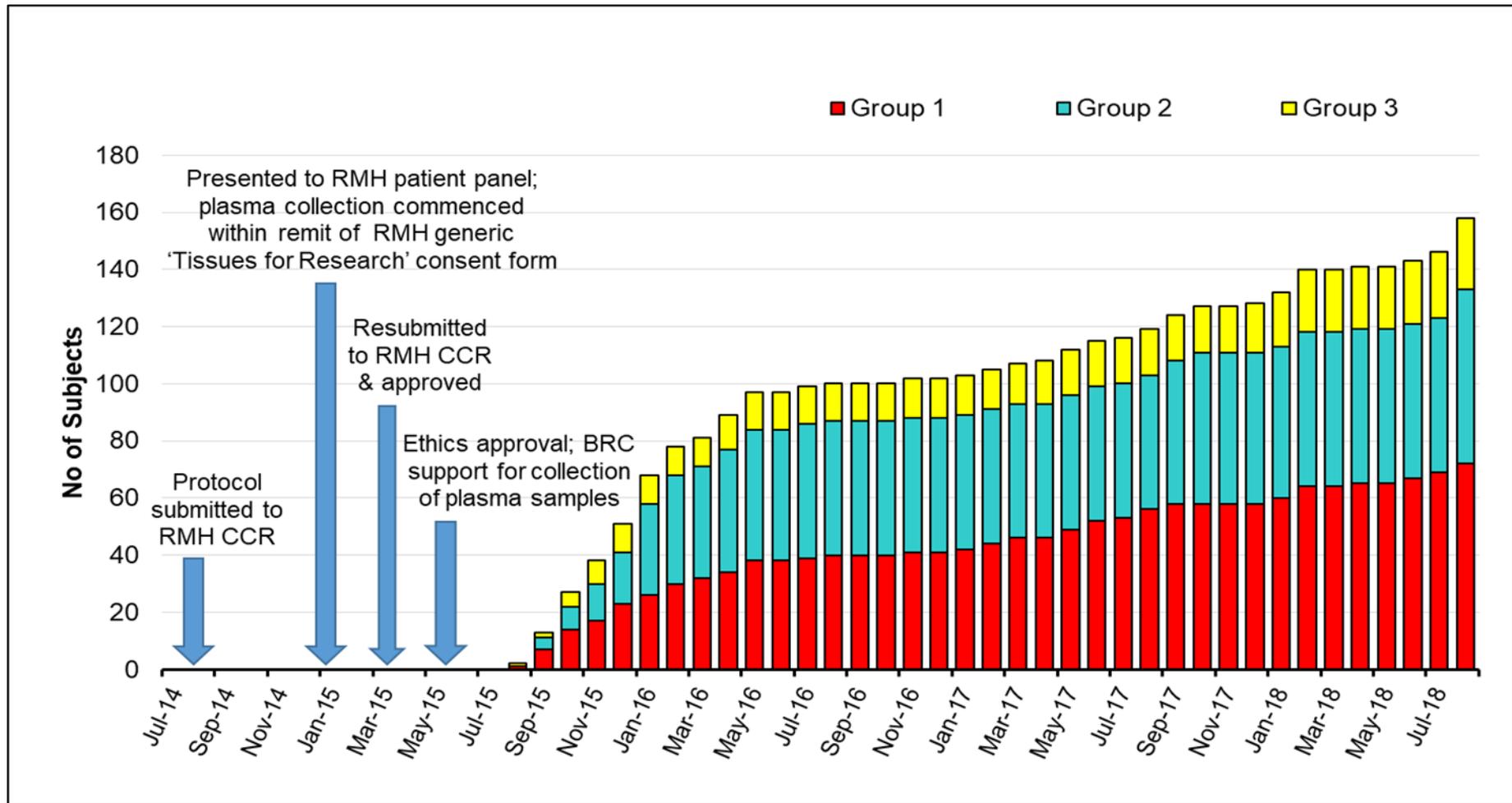


Figure 2.3 CoMB: Key Milestones and Cumulative Recruitment Over First 3 Years

2.3.3.Laboratory work undertaken to establish workflows for tissue processing standard operating protocols

2.3.3.1. Handling of FFPE blocks

Sectioning and staining

In preparation for RNA and DNA extraction from FFPE blocks, it would be necessary to prepare serial unstained sections with H&E slides at the beginning and end of the block to confirm tumour content and ensure the presence of tumour throughout.

Under the guidance of scientific officer Erica Little in Dr Sadanandam's laboratory, I therefore learnt to section FFPE blocks using the microtome and perform H&E staining on otherwise redundant animal blocks. However, in view of time considerations, my relative inexperience of sectioning and the potential relatively low cost of outsourcing, I felt that sectioning and staining of the blocks would be best performed by a professional service. After several enquiries, I made arrangements for sectioning and staining to be performed by the histopathology core facility within the CMP initially, and at a later date by the Breast Cancer Now (BCN) Histopathology Core Facility at ICR Fulham Road.

Histological assessment of FFPE slides

All samples for this MD(Res) were assessed by Dr Steve Hazell, Consultant Histopathologist at RMH Fulham Road. Areas of muscle-invasive tumour were marked out on an initial H&E slide (slide 0) **and** also the final slide to confirm the presence of tumour throughout the sections. Histology and tumour content were recorded. Areas of superficial/non-muscle invasive cancer were excluded.

RNA extraction from FFPE

I performed preliminary work in Dr Sadanandam's lab with Scientific Officers Erica Little and Chanthirika Ragulan to optimise the process of extracting RNA from FFPE.

The aim was to achieve a good yield and optimise quality where possible as determined by RNA integrity score (RIN) as assessed using the Agilent

Bioanalyzer system, with a view to potentially performing RNA-seq. Samples were also quality assessed using Nanodrop to assess quantity, 260/280 and 260/230 ratios. While multiple samples were processed from the same tissue blocks, various parameters were altered to assess the potential impact of deparaffinisation technique, slide thickness, extraction kit and elution method i.e. biological but not technical replicates were performed in this stage of the optimisation process. A table summarising my work with the team is included in appendix 2.2. Details of the final protocol can be found in Chapter 5.

In summary, RNA yield appeared to be optimised with xylene deparaffinisation and section thicknesses of 7-10um. RIN scores however were disappointing and contributed to the decision not to pursue RNA-seq. On reflection, assessment of DV200 (percentage of RNA fragments >200bp in length) would have been of use in further assessing the quality of samples extracted.

DNA extraction from FFPE

Possible approaches to extracting DNA from FFPE were discussed with members of Professor Attard's team, but also post-doc Dr Inma Spiteri Sagastume who has significant experience in the extraction of DNA from FFPE from unstained slides. The optimised protocol using xylene deparaffinisation and the Zymo Quick-DNA FFPE kit are detailed in Chapter 4.

2.3.3.2. Optimising plasma processing: Single spin vs double spin

Background

On receiving plasma samples from a patient (minimum of 2 x 10ml tubes), the RMH Biobank standard operating procedure was to spin both tubes at 1600xg for 10 minutes. Plasma from one tube was then aliquoted into a 2ml Nalgene Cryogenic vial and frozen at -80C (single spin aliquot). Plasma from the 2nd tube was aliquoted into an Eppendorf and spun again at 1600xg for 10 minutes with plasma then aliquoted into 2ml Nalgene Cryogenic vials and frozen at -80C (double spin aliquots). Under the supervision of Professor Attard and with assistance from Scientific Officer Ania Wingate, I performed a small pilot study with the aim of comparing the plasma DNA yield and characteristics from single spin and double spin aliquots, and to inform future processing of samples by the RMH biobank team, including those collected within the CoMB

protocol for use in my pilot studies. The hypothesis was that 'double spin' samples would contain less contamination from white blood cell DNA fragments.

Method

A single and double spin aliquot from a total of 22 patients were retrieved from the RMH GTB. Samples were thawed at room temperature before being transferred to 1.5ml Eppendorfs and centrifuged at 8000 xg for 10 minutes. Plasma was then transferred to a 5ml Eppendorf containing 100ul Proteinase K for every 1ml plasma, avoiding any debris at the bottom of the tube. Plasma volume transferred was recorded. The QIAamp circulating nucleic acid kit was used to extract plasma DNA as per the manufacturer's protocol (see appendix 3.1 for full protocol). Plasma DNA was quantified using the Picogreen assay. Bioanalyser analysis was also performed on all samples to assess fragment size.

Results

Using a threshold of 500bp to define low and high-size fragments, the median proportion of high-size fragments in single spin aliquots was 18.2% (IQR 8.3-34.85) compared to 8.3% in double spin aliquots (IQR 7.3-10.75). In 73% of samples, the double spin aliquots yielded a greater proportion of low-sized DNA fragments.

The results are summarised in figure 2.2 below

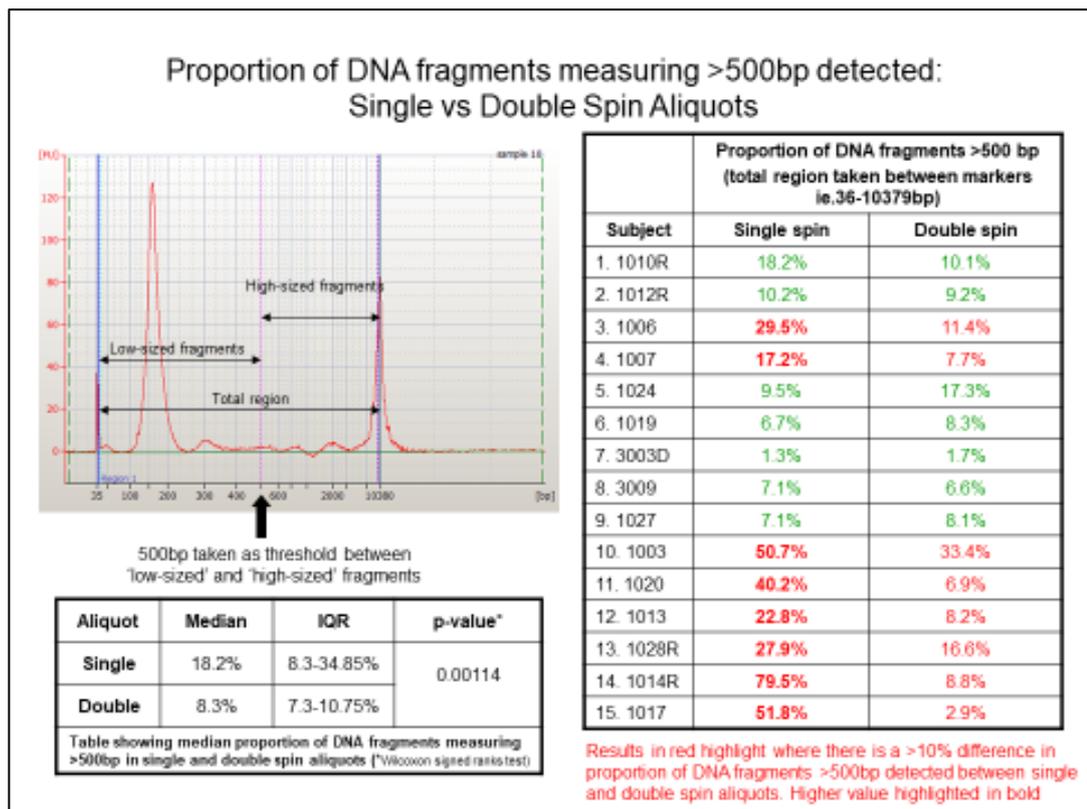


Figure 2.2 Results from single vs double spin pilot study performed to optimise plasma processing within RMH Biobank. Bioanalyser trace illustrates division of fragments into high and low-sized using a threshold of 500bp.

Conclusion

The greater proportion of high-size fragments in single spin aliquots is suggestive of contamination with DNA from white blood cells. The results suggest therefore that plasma samples should undergo a double spin before storage in order to reduce contamination with high-size fragments.

This data was presented at the 'ctDNA Management Meeting' on 22nd March 2016 and it was subsequently agreed that all blood samples received by the RMH biobank should undergo a double spin before storage at -80C.

2.3.4. Recruitment

2.3.4.1. Patients

CoMB commenced recruitment in August 2015. Patients with blood already collected under the RMH generic 'Tissues for Research' consent form were formally recruited into CoMB. As can be seen in Figure 2.3, recruitment slowed from June 2016 to May 2017, coinciding with a period with no dedicated clinical fellow. As of the end of August 2018, CoMB has a total of 158 patients (72 in Group 1, 61 in Group 2 and 25 in Group 3). Clinical data for all patients has been collected on paper case report forms and is due to be entered into an electronic database.

2.3.4.2. Tissue samples

As of 31st August 2018, a total of 301 FFPE blocks have been received for 90 patients.

2.3.4.3. Blood samples

Plasma and buffy coat samples

As of 31st August 2018, a total of 5020 vials of plasma have been stored from 140 patients across 841 time points. 1831 vials of buffy coat have been stored from 140 patients across 841 time points.

Plasma has been stored at a median of 4 time points for each patient, with a range of 1 to 29 time points.

As of end September 2018, I have extracted plasma DNA from a total of 346 vials from 139 timepoints from 55 patients. An application to RMH GTB is currently in progress to allow the transfer of plasma to Professor Attard's laboratory in University College London for ongoing plasma DNA extraction.

2.4. Use of samples in planned work

2.4.1. Overview

Within this MD(Res), I will be performing pilot studies to explore MIBC at a genomic and transcriptomic level with the overall aim of identifying potential

biomarker candidates to guide treatment strategy and develop a personalised approach to the management of MIBC.

2.4.2. Plasma DNA pilot

In this pilot, I will be assessing whether baseline plasma DNA levels may be of clinical value in patients with MIBC with regards to association with disease stage and with prognosis. Following a recent publication³, I will also be investigating whether plasma tumour fraction, estimated from copy number aberrations from low-pass whole genome sequencing, may be a potential biomarker of response to treatment in MIBC. These results will be reported in chapter 3.

2.4.3. Multiregion whole exome sequencing pilot

In chapter 4, I intend to report on the results from multiregion whole exome sequencing of tumour DNA extracted from diagnostic FFPE blocks from a pilot cohort of treatment-naïve patients commencing first-line platinum based chemotherapy, with a view to identifying tumour-specific aberrations to inform the design of a targeted next generation sequencing panel for plasma DNA, and exploring intra-tumour heterogeneity.

This work will use FFPE, plasma and buffy coat samples from CoMB in addition to clinical data. This chapter will additionally report the results from 30 patient samples profiled using Foundation Medicine's FoundationOne panel, which will additionally inform the design of the plasma DNA panel.

2.4.4. Molecular subtyping pilot

The final data chapter in my thesis focuses on molecular subtypes in MIBC, and in particular explores whether preliminary evidence suggesting MIBC can be subtyped using a colorectal cancer classifier⁴ is of clinical relevance. I will be designing a gene expression panel encompassing the existing colorectal

cancer classifier genes and genes associated with radiosensitivity to interrogate a pilot cohort of patients treated with radical radiotherapy at RMH.

This work will use FFPE samples and clinical data from CoMB. A proportion of the cohort will also have FoundationOne data available thus providing a more comprehensive view of their molecular profile.

2.5. Conclusion

I have successfully fulfilled my aim to establish a protocol to allow the collection of clinical material and data to support translational work in MIBC. Since opening in August 2015, the CoMB protocol has acquired blood, tissue and clinical data from 158 patients with MIBC, forming an invaluable resource. Samples and data from a total of 88 patients have been used within this thesis. Recruitment continues and it is anticipated that CoMB will continue to support future translational studies in MIBC.

2.6. References

1. Cancer Genome Atlas Research, N. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315-322 (2014).
2. Diehl, F., *et al.* Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci U S A* **102**, 16368-16373 (2005).
3. Adalsteinsson, V.A., *et al.* Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun* **8**, 1324 (2017).
4. Poudel, P., *et al.* Revealing unidentified heterogeneity in different epithelial cancers using heterocellular subtype classification. *BioRxiv* (2017).

2.7. Appendix

A2.1 CoMB Protocol

Collection of Clinical Material for Molecular Stratification in Patients with Muscle Invasive Bladder Cancer (CoMB)

Principal Investigator:	Professor Robert Huddart
Co-investigators:	Prof Gerhardt Attard, Dr Anguraj Sadanandam, Dr Melissa Tan, Mr Pardeep Kumar, Dr Steve Hazell, Dr Alison Tree
Statistician:	Karen Thomas

Sponsor: The Royal Marsden NHS Foundation Trust
The Institute of Cancer Research

Address: The Royal Marsden NHS Foundation Trust, Fulham Road, London SW3 6JJ
The Institute of Cancer Research, 123 Old Brompton Road, London SW7 3RP

Management: Clinical Research & Development, The Royal Marsden NHS Foundation Trust, Downs Road, Sutton, Surrey SM2 5PT

Site Address: The Royal Marsden NHS Foundation Trust, Downs Road, Sutton, Surrey SM2 5PT
The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG

Protocol Reference:	
Version Number & Date:	V3.0 October 2018
Effective Date:	
Superseded Version Number & Date (If applicable)	V2.2 Sept 2015

Contents

Investigators

Protocol Summary

1.0 Introduction

2.0 Background and rationale

3.0 Objectives

4.0 Eligibility

5.0 Methods

5.1 Patient groups

5.2 Collection of Clinical Materials

5.2.1 Tissue samples

5.2.2 Peripheral blood

5.2.3 Germline DNA samples

5.2.4 Urine samples

5.3 Planned laboratory studies

5.3.1 Plasma ctDNA in MIBC

5.3.2 Molecular subtyping in MIBC

5.4 Collection of Clinical Data

5.4.1 Baseline visit

5.4.2 Subsequent visits

5.4.3 Data Handling

5.5 Data Analysis

5.5.1 Endpoints

5.5.2 Sample size

5.5.3 Statistics

6.0 Study Organisation

6.1 Responsibilities

6.2 Start date

6.3 Completion date

7.0 Adverse events

8.0 Regulatory & Ethical Considerations

9.0 Financing, Sponsorship & Indemnity

10.0 Abbreviations

References

Appendix

Investigators

Principal Investigator

Professor Robert Huddart
Academic Radiotherapy/Division of Radiotherapy and Imaging
The Royal Marsden NHS Foundation Trust/The Institute of Cancer Research
London UK
Tel 020 8661 3457
Robert.huddart@icr.ac.uk

Co-Investigators

Dr Gerhardt Attard
Cancer Research UK Clinician- Scientist and Honorary Consultant
The Royal Marsden NHS Foundation Trust/The Institute of Cancer Research
London UK
Tel 020 8722 4137
Gerhardt.attard@icr.ac.uk
NB: Moved to UCL Jan 2018 (UCL Cancer Institute, Paul O’Gorman Building, 72 Huntley Street, WC1E 6DD; g.attard@ucl.ac.uk)

Dr Anguraj Sadanandam
Systems and Precision Cancer Medicine
Institute of Cancer Research
Cotswold Road
Surrey SM2 5NG
020 8915 6631
Anguraj.sadanandam@icr.ac.uk

Dr Melissa Tan
Clinical Research Fellow
Academic Radiotherapy/Division of Radiotherapy and Imaging
The Royal Marsden NHS Foundation Trust/The Institute of Cancer Research
London UK
020 8661 3181
Melissa.tan@icr.ac.uk

Mr Pardeep Kumar
Academic Surgery
The Royal Marsden NHS Foundation Trust
London UK
Tel 020 7808 2789
Pardeep.kumar@rmh.nhs.uk

Dr Steve Hazell
Consultant Histopathologist
The Royal Marsden NHS Foundation Trust
London UK
Tel 020 7808 2637
Steve.hazell@rmh.nhs.uk

Dr Alison Tree
Academic Radiotherapy/Division of Radiotherapy and Imaging
The Royal Marsden NHS Foundation Trust/The Institute of Cancer Research
London UK
020 8661 3624
Alison.tree@rmh.nhs.uk

Tissue Bank Manager

Janine Salter
Tissue Bank Manager/Project Lead CRUK Stratified Medicine & GEL programme
Biomedical Research Centre/Academic Services
The Royal Marsden NHS Foundation Trust/Institute of Cancer Research
Fulham Rd
Chelsea SW3 6JJ
020 8642 6011 extn 4512
Janine.salter@rmh.nhs.uk

Statisticians

Karen Thomas (Set-up – 2016)
The Royal Marsden NHS Foundation Trust
London UK
020 8661 3641
Karen.thomas@rmh.nhs.uk

Mansour Sharabiani (2016 – Oct 2018)
The Royal Marsden NHS Foundation Trust
London UK

Ranga Gunapala (Oct 2018 – present)
The Royal Marsden NHS Foundation Trust
London UK
0207 808 2186
Ranga.gunapala@rmh.nhs.uk

Clinical Trial Co-ordinator

Kelly Jones (set up – Nov 2017)

Bob Champion Unit

The Royal Marsden NHS Foundation Trust

London UK

020 8661 3070

Kelly.jones@rmh.nhs.uk

Jack Bowes (Nov 2017 – present)

Bob Champion Unit

The Royal Marsden NHS Foundation Trust

London UK

020 8661 3070

jack.bowes@rmh.nhs.uk

Protocol Summary

Title	Collection of Clinical Material for Molecular Stratification in Patients with Muscle Invasive Bladder Cancer (CoMB)
Objectives	<ol style="list-style-type: none"> 1. To acquire clinical material from patients with muscle invasive bladder cancer (MIBC) for current and future study and analysis 2. Using this material, our initial pilot studies aim to <ol style="list-style-type: none"> a. Study the spectrum of mutations at diagnosis b. Identify actionable mutations for targeted therapies c. Assess whether disease progression is associated with the emergence of subclones and if so, determine the time points at which they are detectable d. Identify potentially clinically useful prognostic biomarkers e. Identify potentially clinically useful predictive biomarkers of response to treatment f. Develop a bladder specific mutation screening panel g. Explore molecular subtypes in bladder cancer 3. To additionally acquire clinical material and data from patients with upper tract cancers (v3 Sept 2018) to support future translational work (see Appendix D)
Study population	Patients with a new or previous diagnosis of muscle invasive bladder cancer, or upper tract cancer
Eligibility:	<p>Muscle invasive histologically confirmed bladder cancer or upper tract cancer</p> <p>Able to give informed consent to participate</p> <p>Able to attend appointments at Royal Marsden Hospital (RMH) as per clinical follow-up schedule</p> <p>No prior or current non-bladder invasive malignancy within the last 3 years, other than non-melanoma skin cancer, clinically insignificant prostate cancer, CIN cervix</p> <p>Archived samples from deceased patients previously treated for biopsy proven MIBC at RMH may also be included in the study if necessary consent available eg. RMH tissues for research consent form signed</p>
Method:	<p>Patients will fall into one of three groups</p> <ol style="list-style-type: none"> 1. New diagnosis of localised or advanced MIBC or upper tract cancer ie. no treatment yet received 2. Previously treated (radically or palliatively) MIBC or upper tract cancer, and on routine follow up or deceased 3. Relapsed or progressive disease following previous treatment for MIBC or upper tract cancer <p>Patients will be asked to consent to the storage and study of tissue, blood, germline DNA samples and urine samples. Tissue may be prospectively collected and/or archival samples recalled. A maximum of 18 x 30ml samples of blood per patient may be requested each year. Clinical material will be studied using methods appropriate to achieve the objectives as set out above</p>
Study size	Aim to collect samples and data from 300 patients within the CoMB biobank as a resource for translational research

Follow up	Patients will adhere to standard clinical follow-up schedules. Data regarding investigations and subsequent treatment will be recorded at each visit. Patients will be followed up until date of death or closure of the study. Telephone or postal contact with the patient or associated healthcare professionals may be necessary to ensure complete documentation regarding outcomes
------------------	---

1.0 Introduction

Muscle invasive bladder cancer (MIBC) is a heterogeneous disease with marked variation in its behaviour and clinical outcomes. The overall prognosis is poor with 5-year overall survival of <50%¹.

The Cancer Genome Atlas Research Network report² identified a wide variety of mutations in MIBC, and there has been increasing interest in the identification of molecular markers. Such markers may be prognostic of clinical outcomes, predictive of response to treatment, or may be potential viable drug targets. There is a clinical need for such markers to help guide management decisions in MIBC, and ultimately improve patient outcomes with the delivery of more precise and personalised treatment.

We plan to collect and store clinical material from patients with MIBC to facilitate planned and future analysis, and further the understanding of this disease.

2.0 Background

Current Issues in Bladder Cancer

Bladder cancer is the 7th most common solid malignancy in the UK. It is a heterogeneous disease associated with marked variation in behaviour and clinical outcomes. Muscle invasive disease accounts for 25% of bladder cancer cases and carries the poorest prognosis with 5-year overall survival of < 50%¹.

With regards to radical treatment options, cystectomy and lymph node dissection has long been considered the gold standard. Radical radiotherapy alone is associated with a high rate (up to 50%) of incomplete response or recurrent disease, and patients with local relapse may then proceed to salvage cystectomy³. However, concomitant chemotherapy and neoadjuvant chemotherapy have been shown to improve locoregional control⁴ and overall survival⁵ respectively, and as such, trimodality treatment with transurethral resection of the bladder tumour (TURBT), neoadjuvant chemotherapy and chemoradiation is now increasingly accepted as an alternative in the radical treatment of MIBC in appropriately selected patients⁶. With this approach, salvage cystectomy rates have been reported as 10-30%^{4,6}. However, clinicians have no accurate way of identifying the subset of patients 'destined' to fail bladder preservation strategies and these patients are therefore subjected to the toxicity of both chemoradiation and surgery, with potential compromise to overall outcome with the delay to effective treatment. There is a need for a tool to identify these patients and translational work is key in achieving this.

With regards to the use of chemotherapy in bladder cancer, response rates to 1st line platinum-containing combination regimes are reported to be 40-60%⁵. It is important that those patients not responding to chemotherapy in both the neoadjuvant and palliative setting are recognised early so as to avoid unnecessary toxicity and to minimise the delay to effective treatment. A biomarker to predict or more accurately monitor response to chemotherapy would be of significant clinical use.

In the metastatic setting, 2nd line chemotherapy has a poor response rate and there are no standard treatment options beyond this. There are currently no molecularly targeted agents approved for use in bladder cancer. The identification of actionable mutations in bladder cancer may lead to the discovery of new targeted treatments.

Translational Research in Bladder Cancer

With advances in technology, there has been increasing interest in translational research in bladder cancer.

The Cancer Genome Atlas Research Network group² identified several recurrent mutations in 32 genes and potential therapeutic targets in 69% of the 131 bladder cancer samples analysed. The most common deletion seen was that of CDKN2A which is involved in cell-cycle regulation. Mutations affecting the PI3K/AKT/mTOR pathway and the RTK/RAS pathway were identified in 42% and 44% of tumours respectively. These may be potential therapeutic targets in bladder cancer. They also identified molecular subtypes using unsupervised clustering techniques and RNA sequencing, which were not dissimilar to those seen in breast cancer.

This is broadly in keeping with the work of other groups looking at molecular subtyping in bladder cancer, where subtypes have correlated with clinical outcomes⁷⁻¹⁰. Choi et al¹¹ identified a subtype they termed as 'p-53 like' which was chemoresistant. If validated and confirmed to be predictive, this would have significant clinical impact on patient treatment. Preliminary work within our group by Sadanandam et al¹² on bladder cancer cell lines has shown 3 subtypes and we would plan to further explore this.

One already highlighted clinical issue in the radical management of MIBC is the decision between cystectomy or bladder preserving strategies. Using immunohistochemical techniques, potential predictive biomarkers of outcome following radiotherapy or surgery have been identified¹³. Work is currently underway to refine the technique and prospectively validate the biomarkers as a predictive tool.

With all these techniques, there is a need for clinical tissue for analysis, and correlation with clinical outcomes is essential in order for information obtained to be clinically applicable. Compared to other tumour sites (eg. lung, pancreas), bladder cancer tissue can be obtained with relative ease via cystoscopy. The availability of serial samples and clinical outcomes will enable further analysis looking at the mechanisms underlying disease progression. Tissue sampling is however not without risk and the idea of the so called 'liquid biopsy' is hence very appealing. There has been much interest in circulating tumour cells and circulating tumour DNA (ctDNA) in cancer patients¹⁴, although little work has focused purely on bladder cancer. These may be isolated from blood or urine and would provide a non-invasive method of obtaining nucleic acid material for analysis.

Rationale

There is a clinical need for further translational work in bladder cancer to identify actionable mutations and new drug targets, and to identify biomarkers to allow more accurate personalised treatment for each individual patient.

We propose to collect clinical material from patients with MIBC for storage and study. We also propose to record clinical outcomes of those patients participating in order to facilitate translational research in bladder cancer and allow the correlation of translational results with clinical outcomes.

3.0 Objectives

- 1 To acquire clinical material from patients with MIBC (and upper tract cancers; see Appendix D) for current and future study and analysis
- 2 Using this material, our initial pilot studies aim to
 - a. Study the spectrum of mutations at diagnosis
 - b. Identify actionable mutations for targeted therapies
 - c. Assess whether disease progression is associated with the emergence of subclones and if so, determine the time points at which they are detectable
 - d. Identify potentially clinically useful prognostic biomarkers
 - e. Identify potentially clinically useful predictive biomarkers of response to treatment
 - f. Develop a bladder specific mutation screening panel
 - g. Explore molecular subtypes in bladder cancer

4.0 Eligibility

Inclusion criteria

- Patients with muscle invasive biopsy proven bladder cancer (no restriction on histological subtype or staging), or upper tract cancer (see Appendix D)
- Able to give informed consent for clinical material (tissue, blood, urine, germline DNA) to be taken and stored for analysis and correlated with the patient's clinical outcomes
- Able to attend for appointments at Royal Marsden Hospital as per clinical follow-up schedule
- No prior or current non-bladder malignancy within the last 3 years other than non-melanoma skin cancer, clinically insignificant prostate cancer, CIN of cervix
- Archived samples from deceased patients previously treated for biopsy proven MIBC at RMH may also be included within the study if necessary consent available eg. RMH tissues for research consent form signed

5.0 Method

Suitable patients will be identified in urology MDTs, urology clinics, oncology clinics and clinical trial cohorts. Bladder cancer management is focused in a single multidisciplinary clinic (Sutton Tuesday PM) and will form the centre for recruitment and follow up. We propose to initially collect clinical material from 200 patients.

Patients will fall into one of three groups depending on their stage in their treatment pathway, and the aim will be to obtain tissue, blood, urine and germline DNA samples from each patient where possible. A different patient information leaflet will be used for each setting

5.1 Patient Groups

The three patient groups are as follows:

1. Patients with a new diagnosis of localised or advanced MIBC or upper tract cancer ie. no treatment yet received

Patients will be asked to consent for the following-

- The storage and study of excess tumour +/- normal tissue taken at TURBT or diagnostic biopsy. Samples will be retrospectively recalled.
- Baseline urine and blood samples for storage and study
- Serial blood samples (up to 30mls each time, up to a maximum of 18 samples per patient per year). These will be taken in conjunction with routine clinical bloods where possible
- Serial urine samples taken at clinic visits (up to a maximum of 18 per patient per year)

Within this group, there will be a subset of patients scheduled for further procedures eg. those with a radiological diagnosis awaiting TURBT or biopsy, those due a second TURBT or patients proceeding directly to cystectomy/surgery. In these patients, consent will additionally be sought for

- A part of the fresh tissue sample taken to be donated solely for research purposes. This sample should be snap frozen or treated for long-term storage as soon as it is obtained in order to preserve its integrity for future study (see Appendix A).
- The storage and study of adjacent normal bladder tissue samples to be taken at the time of TURBT solely for research purposes.
- In those who have not yet had a TURBT, a pre-procedure urine and blood sample for storage and study will be requested. If no diagnosis of MIBC is made, any tissue already acquired will be returned to histopathology and the patient withdrawn from the study. Urine and blood samples will be disposed of in accordance with hospital policy.

2. Patients previously treated (radically or palliatively) for MIBC or upper tract cancer on routine follow up or deceased

Consent will be sought for-

- The storage and study of archival excess tumour +/- normal tissue taken at TURBT / diagnostic biopsy. Samples will be retrospectively recalled
- Baseline urine and blood samples for storage and study
- Collection of further tissue, blood and urine samples if disease recurs or progresses. At this point, patients will join the group 3 schedule below.

In deceased patients, archived tumour tissue samples may be retrieved if appropriate consent is available eg. RMH consent form for tissues for research. Samples taken from 1st September 2006 are subject to the Human Tissue Act 2004 and appropriate consent is therefore required. Samples taken before 1st September 2006 may be included without formal consent.

3. Patients with relapsed or progressive disease following previous treatment for MIBC/upper tract cancer (radical or palliative)

Consent will be sought for

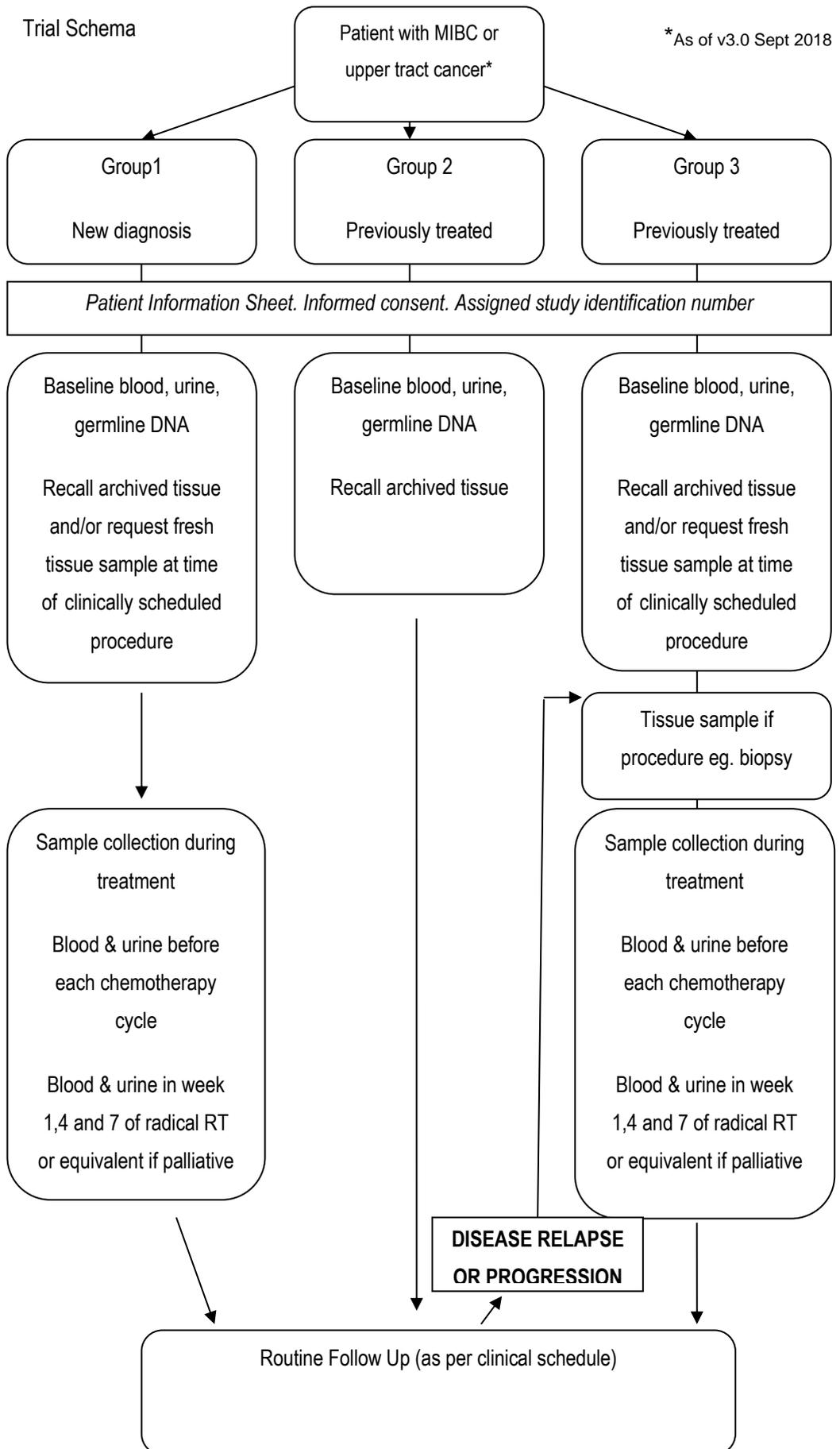
- The collection of archival tumour tissue where available for storage and study
- Baseline urine and blood samples for storage and study

- Serial blood samples (up to 30mls each time, up to a maximum of 18 samples per patient per year). These will be taken in conjunction with routine clinical bloods where possible
- Serial urine samples taken at clinic visits (up to a maximum of 18 per patient per year)
- If patients are clinically scheduled for surgery or biopsies to confirm disease relapse (from the bladder or sites of metastatic disease), they will be asked to consent to allow the storage and study of any excess tissue taken. At the time of tissue collection, one part of the tissue sample will be snap frozen or treated for long-term storage as soon as it is obtained in order to preserve its integrity for future study.

Across all groups, if patients are scheduled for repeat cystoscopy to assess treatment response or for ongoing surveillance, consent will also be requested for biopsies to be taken and tissue to be stored and studied, if it is safe to do so. If tissue is to be taken as part of the standard of care treatment, consent will be sought for excess tissue to be stored for future study. At the time of tissue collection, one part of the tissue sample will be snap frozen or treated for long-term storage as soon as it is obtained in order to preserve its integrity for future study.

If patients have previously donated samples to the RMH Biobank and further consent to participate in CoMB, they will be recruited to the most appropriate group according to their current status and the samples previously collected eg. patients who have donated pre-treatment and serial blood/urine samples will be recruited to group 1.

On entry into the study, all patients will be allocated a unique study identification number.



5.2 Collection of Clinical Materials

All clinical material collected will be assigned an identification number. For each patient, a central log will be kept of all samples requested, taken and stored. Clinical material will be stored in the RMH Biobank which is managed in accordance with RMH policy 2061 ('Generic Bio Banking and Sample Access Policy').

5.2.1 Tissue samples

Tissue samples may be prospectively collected or retrospectively retrieved for storage and study. Tissue will be securely stored in the RMH Biobank.

In patients who have consented for prospective collection of tumour samples, a part of fresh tissue will be snap frozen or treated using RNAlater solution or equivalent before freezing (see Appendix A). The remaining tissue should be processed as per local hospital protocol.

Archived formalin fixed paraffin embedded (FFPE) samples will be retrieved from the relevant centres and stored in the Biobank.

When removed from storage, tissue samples will be studied microscopically and immunohistochemically to confirm the presence of malignant bladder cells. The remainder of tissue will be studied using techniques selected by the investigators to achieve the objectives as set out in section 3.0.

5.2.2 Peripheral blood

Blood samples will be taken from a peripheral vein. The first 5mls will be discarded to minimise contamination by skin epithelial cells. 3 x 10ml samples will be collected into 3 x 10ml EDTA tubes. The samples will be spun and frozen on the same day of collection. All blood samples will be securely stored in the RMH Biobank.

Up to 30mls of blood will be taken up to a maximum of 18 times per patient per year. We anticipate that study samples will be collected at the same time as patients are scheduled for phlebotomy for clinical purposes. Where this is not possible, subjects will be asked if research bloods may be taken separately. Alternative collection tubes may be used if felt by the investigators better suited to achieving the objectives as set out in section 3.0.

As DNA is recognised to degrade in plasma over time, we aim to extract circulating DNA (cDNA) from at least 2 vials of plasma from each timepoint for every patient to optimise future work using cDNA. This will leave some plasma from each timepoint available for future work exploring other circulating nucleic acids or molecules.

5.2.3 Germline DNA samples

Non-malignant cells for germline DNA analysis will be obtained from plasma white cells extracted from EDTA tube samples. Alternative methods of collecting germline DNA such as saliva collection using an Oragene kit or buccal swabs may be considered if felt to offer significant advantages.

5.2.4 Urine samples

Selected patients will be asked to provide a voided urine sample in a sterile specimen pot at baseline. From this sample, aliquots of up to 50ml will be prepared. Samples will be securely stored in the RMH Biobank, for future study.

Serial urine samples will be collected in the same manner up to a maximum of 18 times per patient per year. It is anticipated that, where possible, samples will be requested when patients attend for their routine follow-up appointments and so extra visits should not be required.

5.3 Planned laboratory studies

The aim of these pilot studies will be to further our understanding of the molecular pathology of bladder cancer as discussed in section 3.0. These studies may include but are not limited to the following areas- gene expression analysis, microarray analysis, quantitative PCR, comparative genomic hybridisation, fluorescent in situ hybridisation, DNA sequencing, immunohistochemistry, and ctDNA analysis. .

Initial pilot work will look at plasma ctDNA and molecular subtyping in MIBC.

5.3.1 Plasma ctDNA in MIBC

We plan to perform a pilot study to isolate plasma ctDNA in patients with a diagnosis of MIBC commencing 1st line platinum based chemotherapy (palliative or neoadjuvant). While an initial study size of 30 patients for the ctDNA pilot was originally proposed in 2015 (v2.0), a surge in bladder cancer ctDNA research publications¹⁵⁻²⁰ during sample and data collection necessitated a change in study design in late 2017. Further details regarding the ctDNA pilot work can be found in Appendix D which details all current translational work using CoMB samples.

5.3.2 Molecular subtyping

We plan to perform a pilot study to assess the feasibility of extracting and processing RNA from archived MIBC tissue samples for molecular subtyping. The pilot study will also aim to refine our current molecular profiling techniques to subtype a cohort of samples to demonstrate the presence of molecular subtypes in MIBC.

5.4 Clinical Data Collection

5.4.1 Baseline data on trial entry

On entry into the study, baseline data to be collected will include patient demographics, tumour staging (TNM 7th Ed) and past medical history including performance status and smoking history.

For those patients with archived tumour material, the date and location of the diagnostic TURBT or biopsy will be noted.

For patients previously treated, treatment details and outcomes to date will be recorded. Information will be gathered from the hospital records.

5.4.2 Subsequent data collection

The clinical data to be collected at each visit for trial purposes will be the same information as would be elucidated for any clinical patient. However to ensure all required data is captured, outcomes will be recorded on a study proforma (Appendix C) at each visit. This will record the patient's current disease status with relevant investigation results and details of any subsequent treatment. Treatment outcomes and toxicity will be recorded where appropriate.

Patients will be followed up until death or closure of the study. If patients are to be discharged or care transferred to another centre, clinical outcomes may be collected by contacting the

patient or associated healthcare professionals by telephone or mail. Follow up schedules are shown in Appendix B.

5.4.3 Data Handling

Data will be collected and recorded for each patient using the study proforma (Appendix C) and identifying details will include the patient's study identification number, hospital number and initials. Data will be held on a secure database and managed in accordance with the Data Protection Act.

A central log of samples requested, taken and stored will be maintained.

The PI will permit trial-related monitoring, audits, IRB/IEC review and will provide direct access to source data or documents as required.

5.5 Data Analysis

5.5.1 Sample size

In order to establish a useful resource for translational work, we initially aimed to collect clinical samples and data from 200 patients. Initial statistical calculations had shown that with a biomarker prevalence of 33% and cohort size of 200, at least 80% power can be achieved to detect a difference in response rate/outcome of 22%. However, with increasing interest now in immunotherapy agents in MIBC and the inclusion of upper tract cancers in the CoMB protocol, we now aim to collect samples from 300 patients to additionally support potential future pilot work in these areas.

Further details regarding sample sizes for the current and planned pilots can be found in Appendix D

5.5.2 Statistics

CoMB is a biobank resource of clinical material and data for use in translational research. Statistical analysis plans will therefore depend each individual study being performed. Initial small pilot studies are anticipated to be exploratory and hypothesis generating. A statistical plan will be written and agreed for each study using CoMB samples and will be included in the Appendix. The following general principles will apply to pilot studies using CoMB samples and have been reached assuming an event rate of 40-50% and that one degree of freedom will be allowed for every 10 events.

- The minimum overall pilot study size required if survival analysis is to be performed will be 30
- If a comparison of survival outcomes is to be made between two subgroups, then the overall cohort size required will be a minimum of 40
- If comparing survival between groups subdivided according to the presence or absence of a specific aberration, the aberration should be present in at least 30% of the cohort to allow the comparison to be performed

5.5.3 Healthy volunteer data

As part of the data analysis for the initial ctDNA pilot, existing anonymised healthy volunteer data (n=10) may be used eg. to demonstrate what could be considered as a 'normal' baseline cDNA yield. This set of data relevant to the ctDNA pilot has been acquired by Prof Attard's lab

from staff volunteers and a commercial biobank. All samples were obtained with fully informed consent.

Similar healthy volunteer datasets, where available, may be used in other pilot studies. The CoMB biobank will not collect or store healthy volunteer tissue or plasma

6.0 Study Organisation

6.1 Responsibilities

The principle investigator has overall responsibility for the study and will ensure it complies with the regulatory requirements for Good Clinical Practice (GCP).

The responsibility for the day to day running of the study will be delegated to designated individuals by the Principle Investigator.

6.2 Trial Management Group

A Trial Management Group will be set up and will include the PI, co-investigators, tissue bank manager, statistician and clinical trial co-ordinator. The group will meet regularly and at least annually to review safety of participants, quality of clinical material and data collection, and review research activity related to the clinical materials acquired within this study.

6.3 Start date definition

The study will commence once sponsorship and funding has been granted.

6.4 Protocol completion

The collection of samples and data from 300 patients is expected to be completed in 60 months. Pending renewal of CoMB's ethical approval in 2020, it is anticipated that the CoMB biobank will remain within the RMH Biobank as a valuable resource of material and data for future translational work. Further expansion of the biobank may be considered in due course by the trial management team and will in part be guided by the use of CoMB samples and data to date.

7.0 Adverse events

An adverse event (AE) is any untoward medical occurrence in a participating patient administered a research procedure. The event does not necessarily have a causal relationship with participation in the trial.

A serious adverse event (SAE) is an adverse event that results in death, unplanned inpatient hospitalisation or prolongation of inpatient stay, persistent or significant disability, or is life-threatening.

As a non-interventional study, adverse events related to trial participation are anticipated to be low. Blood tests will be taken at the same time as clinical samples and so any adverse risks of venesection will not be due to study participation. No adverse events are anticipated in relation to urine collection or the retrieval of archived tissue samples.

Adverse events will be reported to the PI and managed in accordance RMH generic SOP for Adverse Events Reporting in non-CTIMPS (gSOP-33) as appropriate.

8.0 Regulatory & Ethical Considerations

8.1 Informed consent

Eligible patients will be given a patient information leaflet regarding the study and will have the opportunity to discuss the study with a member of the team. If they are happy to proceed, they will be asked to sign a consent form to confirm this. Patients are able to withdraw at any point and do not have to consent to all clinical material requests in order to participate. Samples and clinical outcome data taken before the point of study withdrawal will be retained and may be used in research projects.

8.2 Ethics

The protocol will be submitted for ethics approval and will proceed only if approved. All those involved in the study will follow practice according to International Conference on Harmonisation (ICH), Good Clinical Practice (GCP), the Declaration of Helsinki, and the Human Tissue Act.

8.3 Patient confidentiality

Patient confidentiality will be maintained in compliance with the Data Protection Act 1998.

9.0 Financing, Indemnity & Insurance

Sponsorship is being sought from RMH/ICR.

Funding for initial sample collection and ctDNA pilot will be underwritten by the RMH Academic Urology Bladder Cancer Research Fund.

NHS indemnity scheme will apply

10.0 Abbreviations

AE	Adverse event
BRC	Biomedical Research Centre
CTC	Circulating tumour cells
ctDNA	Circulating tumour DNA
GCP	Good Clinical Practice
ICR	Institute of Cancer Research
IRB	Institutional review board
IEC	Independent ethics committee
MIBC	Muscle invasive bladder cancer
NAC	Neoadjuvant chemotherapy

PI	Principle Investigator
RMH	Royal Marsden Hospital
RT	Radiotherapy
SAE	Serious adverse event
SOP	Standard operating procedure
Sgy	Surgery
TURBT	Transurethral resection of bladder tumour

11.0 References

1. Rodel C. Combined modality treatment and selective organ preservation in invasive bladder cancer: Long-term results. *J Clin Onc* 2002;20(14):3061-71
2. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder cancer. *Nature* 2014;507(7492):315-322
3. Fossa SD et al. Bladder cancer definitive radiation therapy of muscle-invasive bladder cancer. *Cancer* 1993;72(10):3036-43
4. James ND et al. Radiotherapy with or without chemotherapy in muscle-invasive bladder cancer. *NEJM* 2012;366(16):1477-88
5. Advanced Bladder Cancer (ABC) Meta-analysis Collaboration. Neoadjuvant chemotherapy in invasive bladder cancer: update of systematic review and meta-analysis of individual patient data. *Eur Urol* 2005;48(2):202-5
6. Ploussard G et al. Critical analysis of bladder sparing with trimodal therapy in muscle-invasive bladder cancer: a systematic review. *Eur Urol* 2014;66(1):120-137
7. Dyrskjot L et al. Identifying distinct classes of bladder carcinoma urine microarrays. *Nature Genetics* 2003;33:90-96
8. Blaveri E et al. Bladder cancer outcome and subtype classification by gene expression. *Clin Canc Res* 2005;11:4044-4055
9. Sjobahl G et al. A molecular taxonomy for urothelial carcinoma. *Clin Canc Res* 2012;18(12):3377-86
10. Choi W et al. Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer Cell* 2014;25:152-165
11. Damrauer JS et al. Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *PNAS* 2014;111(8):3110-3115
12. Sadanandam A et al. 2014 Unpublished preliminary data
13. Choudhury A et al. MRE11 expression is predictive of cause-specific survival following radical radiotherapy for muscle-invasive bladder cancer. *Canc Res* 2010;70:7017-7026
14. Bettgowda C et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci TRansl Med* 2014 Feb 19;6(224):224ra24. doi:10.1126/scitranslmed.3007094
15. Sonpavde, G., et al. Circulating cell-free DNA profiling of patients with advanced urothelial carcinoma. *J Clin Oncol* **34**, abstr358 (2016).

16. Birkenkamp-Demtroder, K., *et al.* Genomic Alterations in Liquid Biopsies from Patients with Bladder Cancer. *Eur Urol* **70**, 75-82 (2016).
17. Christensen, E., *et al.* Liquid Biopsy Analysis of FGFR3 and PIK3CA Hotspot Mutations for Disease Surveillance in Bladder Cancer. *Eur Urol* **71**, 961-969 (2017).
18. Birkenkamp-Demtroder, K., *et al.* Monitoring Treatment Response and Metastatic Relapse in Advanced Bladder Cancer by Liquid Biopsy Analysis. *Eur Urol* (2017).
19. Vandekerkhove, G., *et al.* Circulating Tumor DNA Reveals Clinically Actionable Somatic Genome of Metastatic Bladder Cancer. *Clin Cancer Res* **23**, 6487-6497 (2017).
20. Patel, K.M., *et al.* Association Of Plasma And Urinary Mutant DNA With Clinical Outcomes In Muscle Invasive Bladder Cancer. *Sci Rep* **7**, 5554 (2017).

Appendices

A. Handling of Tissue Samples

Tissue samples will be handled in accordance with the Human Tissue Act and the ICR RMH Joint Policy for the Removal, Storage, Use and Disposal of Human Tissue for Research. Clinical material will be stored in the RMH Biobank which is managed in accordance with RMH policy 2061 ('Generic Bio Banking and Sample Access Policy').

Samples and data may be shared with other institutions subject to a Materials Transfer Agreement.

In order to preserve the integrity of fresh tissue samples donated purely for research purposes, samples taken will be either snap frozen or immediately submerged into RNAlater solution and subsequently frozen.

RNA later

RNAlater is an aqueous solution that precipitates cell proteins and inactivates RNAses. It stabilises RNA and allows storage of tissue.

Harvested tissue should be immediately submerged into 5 volumes of RNAlater. Samples are then incubated at 4C overnight and then removed from RNAlater and stored at -80C.

B. Follow-up Schedules

Follow-up Schedule Group 1 (New diagnosis, no previous treatment): FOR RADICAL TREATMENT															
					Follow up post-radical treatment (months post completion of treatment). As per clinical need- minimum expected follow up below										
Sample	Baseline	NAC	RT	Surgery	3	6	9	12	18	24	30	36	48	60	Annually
Blood	X	X Prior to each cycle with routine bloods	X Week 1, 4, 7	X Within 4 weeks pre-sgy & 4 weeks post-sgy	x	x	x	x	x	x	x	x	x	x	x
Urine	X	X Prior to each cycle at clinical review	X Week 1, 4, 7		x	x	x	x	x	x	x	x	x	x	x
					NOT APPLICABLE POST CYSTECTOMY										
Tissue	X Archived and/or fresh sample	X Post-chemo sample		X	Consent for tissue collection and storage to be requested at each clinically scheduled cystoscopy										
ON DISEASE RELAPSE, PATIENT ENTERS SCHEDULE FOR GROUP 3															

Follow-up Schedule Group 1 (New diagnosis, no previous treatment): FOR PALLIATIVE TREATMENT				
Sample	Baseline	Chemotherapy	Radiotherapy	Follow up post treatment Schedule as per clinical need
Blood	X	X Prior to each cycle with routine bloods	X At beginning, middle and end of course	X With routine clinical bloods
Urine	X	X Prior to each cycle at clinical review	X At beginning, middle and end of course	X On attendance to clinic (minimum 6 monthly unless discharged from RMH)
Tissue	X Archived and/or fresh sample			
ON DISEASE PROGRESSION, PATIENT ENTERS SCHEDULE FOR GROUP 3				

Follow up Schedule: Group 2 Patients (Previously treated, on routine follow up)		
Sample	Baseline	Follow up as per clinical schedule
Blood	X	Annually
Urine	X	Annually
Tissue	X Archived sample	Consent for tissue collection and storage to be requested at each clinically scheduled cystoscopy
ON DISEASE PROGRESSION, PATIENT ENTERS SCHEDULE FOR GROUP 3		

Follow up Schedule: Group 3 Patients (Previously treated, relapsed)				
Sample	Baseline	Chemotherapy	RT	Follow up post treatment Schedule as per clinical need
Blood	X	X Prior to each cycle with routine bloods	X At beginning, middle and end of course	With routine clinical bloods
Urine	X	X Prior to each cycle	X At beginning, middle and end of course	On each attendance to clinic
Tissue	X If clinically indicated			On progression if clinically indicated

C. Study CRFs

CoMB: Collection of Clinical Material from Patients with Muscle Invasive Bladder Cancer
Eligibility Assessment v2.4 Oct2018

Initials: Study ID:

Date of Birth: / / Sex: Male: Female:

Eligibility

1a. Muscle invasive histologically confirmed bladder cancer Yes go to Q2 No go to Q1b

1b. If not histologically confirmed, is TURBT or biopsy planned? Yes No N/A

If answer to Q1a and 1b is no- patient is not eligible for study

2. Written informed consent to participate Yes No

3. Able to attend appointment at Royal Marsden Hospital as per clinical follow up schedule Yes No

4. NO prior or current non-bladder invasive malignancy within the last 3 years, excluding
 - non-melanoma skin Yes No
 - NCCN low risk prostate cancerclinically insignificant Patient has no prior or current history Patient has prior or current history of non-bladder malignancy and is therefore excluded from study
 prostate cancer (pT1 or pT2a Gleason ≤ 6, PSA ≤ 10 and ≤ 1cc total volume)
 - CIN cervix

If any answer to Q2-4 is 'No', patient is not eligible for study

Patient Group

Tick to indicate which group the patient is entering within CoMB study

Group 1: New diagnosis of muscle invasive bladder cancer, no treatment yet received

Group 2: Previously treated muscle invasive bladder cancer, on routine follow-up

Group 3: Relapsed or progressive disease following previous treatment for muscle invasive bladder cancer

Signature Print Name Date / /

CoMB: Collection of Clinical Material from Patients with Muscle Invasive Bladder Cancer

Groups 1-3: Baseline Visit v2.3 Aug 2018

Initials: Study ID:

Date of Birth: / / Sex: Male: Female:

Histology

1. Source of histology TURBT Biopsy Site biopsied

2. Date of diagnostic TURBT/biopsy

3. Hospital where diagnostic TURBT/biopsy performed

Royal Marsden Sutton

Royal Marsden Fulham

Kingston Hospital

Epsom Hospital

St Helier's Hospital

Croydon University Hospital

Other please specify

4. Histological subtype

Transitional cell carcinoma

Squamous cell carcinoma

Adenocarcinoma

*If mixed, please tick all that apply and mark predominant histology with an asterisk **

Neuroendocrine

- Small cell cancer

- Large cell cancer

Sarcomatoid

Other please specify

5. Grade

Grade 1

Grade 2

Grade 3

Not known

6. Was carcinoma in situ present (CIS)? Yes No

Staging Investigations at Initial Diagnosis of Muscle Invasive Bladder Cancer (MIBC)

CT chest Date Hospital

Ct abdomen Date Hospital

CT pelvis Date Hospital

Study ID:

Groups 1-3 Baseline Visit v2.3 Aug 2018

Staging Investigations at Initial Diagnosis of MIBC cont.

MRI pelvis Date Hospital

PET CT Date Hospital

Other- please specify
 Date Hospital

..... Date Hospital

..... Date Hospital

Final Staging at Initial Diagnosis of MIBC

T T1 T2 T3a T3b T4a T4b

N N0 N1 N2 N3

M M0 M1

Baseline Blood Tests at Initial Diagnosis of MIBC/Pre-treatment

Date

Hb g/l
 WBC x 10⁹/L
 Neuts x 10⁹/L
 Plt x 10⁹/L
 Creat μmol/L
 ALT U/L
 Bilirubin μmol/L
 ALP U/L
 CCa mmol/L

Baseline Study Samples

Blood samples (3 x Streck/EDTA tubes) Yes Date No

Urine sample Yes Date No

Diagnostic tissue sample received? Yes No

Signature	Print Name	Date <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
-----------	------------	--

CoMB: Collection of Clinical Material from Patients with Muscle Invasive Bladder Cancer

Group 1-3: Baseline Visit- Comorbidities v2.3 Aug 2018

Initials:

Study ID:

Date of Birth: / /

Sex: Male: Female:

Comorbidities

	Yes	No		Yes	No
Myocardial infarction	<input type="checkbox"/>	<input type="checkbox"/>			
Congestive cardiac failure	<input type="checkbox"/>	<input type="checkbox"/>			
Peripheral vascular disease	<input type="checkbox"/>	<input type="checkbox"/>			
Cerebrovascular disease	<input type="checkbox"/>	<input type="checkbox"/>			
Chronic obstructive pulmonary disease	<input type="checkbox"/>	<input type="checkbox"/>			
Connective tissue disease	<input type="checkbox"/>	<input type="checkbox"/>			
Peptic ulcer disease	<input type="checkbox"/>	<input type="checkbox"/>			
Diabetes mellitus	<input type="checkbox"/>	<input type="checkbox"/>	If yes, is it		
			Uncomplicated	<input type="checkbox"/>	<input type="checkbox"/>
			End-organ damage	<input type="checkbox"/>	<input type="checkbox"/>
Moderate to severe <u>chronic</u> kidney disease ie. GFR <60ml/min	<input type="checkbox"/>	<input type="checkbox"/>			
Hemiplegia	<input type="checkbox"/>	<input type="checkbox"/>			
Acquired immunodeficiency syndrome	<input type="checkbox"/>	<input type="checkbox"/>			
ECOG Performance Status	0	1	2	3	

Signature	Print Name	Date <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
-----------	------------	--

CoMB: Collection of Clinical Material from Patients with Muscle Invasive Bladder Cancer

Groups 1 & 3: Follow-up during current treatment: Chemotherapy v2.3 Aug 2018

Initials:

Study ID:

Date of Birth: / /

Sex: Male: Female:

Chemotherapy

1. Intent Neoadjuvant Adjuvant Palliative
2. Regime Gemcitabine / Cisplatin
- Gemcitabine / carboplatin
- Acc MVAC
- Carboplatin / Paclitaxel
- Paclitaxel
- Carboplatin / etoposide
- Atezolizumab
- Pembrolizumab
- Other

5. Date of cycle 1 / /

Date of assessment	Pre-assessment for cycle No (circle/specify)	Study samples taken? (tick if yes)		Toxicity grid completed?
		Blood	Urine	
<input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/>	1 / 2 / 3 / 4 / 5 / 6 /	<input type="checkbox"/>	<input type="checkbox"/>	Yes <input type="checkbox"/>
<input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/>	1 / 2 / 3 / 4 / 5 / 6 /	<input type="checkbox"/>	<input type="checkbox"/>	Yes <input type="checkbox"/>
<input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/>	1 / 2 / 3 / 4 / 5 / 6 /	<input type="checkbox"/>	<input type="checkbox"/>	Yes <input type="checkbox"/>
<input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/>	1 / 2 / 3 / 4 / 5 / 6 /	<input type="checkbox"/>	<input type="checkbox"/>	Yes <input type="checkbox"/>
<input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/>	1 / 2 / 3 / 4 / 5 / 6 /	<input type="checkbox"/>	<input type="checkbox"/>	Yes <input type="checkbox"/>

Continue on additional sheet if required

Assessment of response to chemotherapy (tick if none)

Imaging Investigation Performed	Date	Outcome
CT / MRI / PET	<input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/>	CR / PR / SD / PD
CT / MRI / PET	<input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/>	CR / PR / SD / PD
CT / MRI / PET	<input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/>	CR / PR / SD / PD

Study ID:

Groups 1 & 3: Follow up during current treatment- chemotherapy v2.3 Aug 2018

Cystoscopy: / / Tissue sample for study requested? Yes No

Result: No evidence of disease / Disease present

If disease present: pTa / pTis / pT1 / pT2 / pT3 / pT4

Completion of chemotherapy

Date of last cycle (day 1) <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/>	<u>Reason for stopping</u>	
	Planned treatment completed	<input type="checkbox"/>
	Toxicity	<input type="checkbox"/>
	Disease progression	<input type="checkbox"/>
	Other (please specify)	<input type="checkbox"/>
Signature	Print Name	Date <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/>

CoMB: Collection of Clinical Material from Patients with Muscle Invasive Bladder Cancer
Group 1 & 3: Follow-up during current treatment: Radiotherapy v2.3 Aug 2018

Initials: Study ID:

Date of Birth: Sex: Male: Female:

Radiotherapy

1. Intent Radical Adjuvant Palliative

2. Treatment area Bladder only
 Bladder and pelvic lymph nodes
 Bone
 Other

3. Dose Gy fractions

4. Concurrent chemotherapy? Yes 5FU/MMC
 Weekly gemcitabine
 Other

No

5. Start date

6. End date

Reason for stopping
 Planned treatment completed
 Toxicity
 Change in RT schedule
 Other (please specify):

Date of assessment	Fraction Number Due	Study samples taken? (tick if yes)		Toxicity grid completed?
		Blood	Urine	
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>				Yes <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>				Yes <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>				Yes <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>				Yes <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>				Yes <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>				Yes <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>				Yes <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>				Yes <input type="checkbox"/>

Study ID: Groups 1 & 3: Follow up during current treatment- radiotherapy v2.3 Aug 2018

Assessment of response to radiotherapy (tick if none)

Imaging Investigation performed	Date	Outcome
CT / MRI / PET	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	CR / PR / SD / PD
CT / MRI / PET	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	CR / PR / SD / PD
CT / MRI / PET	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	CR / PR / SD / PD
Cystoscopy:	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Tissue sample for study requested? Yes <input type="checkbox"/> No <input type="checkbox"/>

Result: No evidence of disease / Disease present

If disease present: pTa / pTis / pT1 / pT2 / pT3a / pT3b / pT4

Signature	Print Name	Date <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
-----------	------------	---

CoMB: Collection of Clinical Material from Patients with Muscle Invasive Bladder Cancer

Group 1 : Surgery since CoMB Recruitment v2.3 Aug 2018

Initials:

Study ID:

Date of Birth: / /

Sex: Male: Female:

Surgery

Date of procedure / /

Location Royal Marsden Hospital Fulham

Other (please specify)

Procedure performed Cystectomy

Tick all that apply Pelvic lymph node dissection

Pelvic exenteration

Ileal conduit

Neobladder

Other (please specify)

Study bloods taken? Within 4 weeks before surgery Yes Date / /

No

Within 4 weeks post surgery Yes Date / /

No

Tissue sample taken for research at time of procedure? Yes No

If yes: Snap frozen
RNA later

Final histopathological staging pT pN (/ nodes)

Signature	Print Name	Date <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> <input type="text"/>
-----------	------------	---

CoMB: Collection of Clinical Material from Patients with Muscle Invasive Bladder Cancer

Group 1-3: Routine Follow-up Since CoMB Recruitment v2.2 July 2018

Initials:

Study ID:

Date of Birth:

Sex: Male: Female:

Clinical Assessment

Date of assessment	Clinical status	Study samples taken? (tick if yes)		Toxicity grid completed?	
		Blood	Urine	Yes <input type="checkbox"/>	NA <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	No recurrence <input type="checkbox"/> Confirmed recurrence <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Yes <input type="checkbox"/>	NA <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	No recurrence <input type="checkbox"/> Confirmed recurrence <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Yes <input type="checkbox"/>	NA <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	No recurrence <input type="checkbox"/> Confirmed recurrence <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Yes <input type="checkbox"/>	NA <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	No recurrence <input type="checkbox"/> Confirmed recurrence <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Yes <input type="checkbox"/>	NA <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	No recurrence <input type="checkbox"/> Confirmed recurrence <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Yes <input type="checkbox"/>	NA <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	No recurrence <input type="checkbox"/> Confirmed recurrence <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Yes <input type="checkbox"/>	NA <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	No recurrence <input type="checkbox"/> Confirmed recurrence <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Yes <input type="checkbox"/>	NA <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	No recurrence <input type="checkbox"/> Confirmed recurrence <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Yes <input type="checkbox"/>	NA <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	No recurrence <input type="checkbox"/> Confirmed recurrence <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Yes <input type="checkbox"/>	NA <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	No recurrence <input type="checkbox"/> Confirmed recurrence <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Yes <input type="checkbox"/>	NA <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	No recurrence <input type="checkbox"/> Confirmed recurrence <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Yes <input type="checkbox"/>	NA <input type="checkbox"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	No recurrence <input type="checkbox"/> Confirmed recurrence <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Yes <input type="checkbox"/>	NA <input type="checkbox"/>

Investigations: Imaging

Imaging scan performed	Date	Outcome
	<input type="text"/>	CR / PR / SD / PD
	<input type="text"/>	CR / PR / SD / PD
	<input type="text"/>	CR / PR / SD / PD
	<input type="text"/>	CR / PR / SD / PD
	<input type="text"/>	CR / PR / SD / PD
	<input type="text"/>	CR / PR / SD / PD
	<input type="text"/>	CR / PR / SD / PD

Investigations: Cystoscopies (tick if patient had cystectomy) ()

Date	Outcome	
<input type="text"/>	No evidence of disease	<input type="checkbox"/>
	Disease present	<input type="checkbox"/>
	If disease present:	pTa /pTis / pT1 / pT2 / pT3 / pT4
	Tissue sample requested for CoMB?	Yes / No
<input type="text"/>	No evidence of disease	<input type="checkbox"/>
	Disease present	<input type="checkbox"/>
	If disease present:	pTa /pTis / pT1 / pT2 / pT3 / pT4
	Tissue sample requested for CoMB?	Yes / No
<input type="text"/>	No evidence of disease	<input type="checkbox"/>
	Disease present	<input type="checkbox"/>
	If disease present:	pTa /pTis / pT1 / pT2 / pT3 / pT4
	Tissue sample requested for CoMB?	Yes / No

Signature	Print Name	Date
		<input type="text"/>

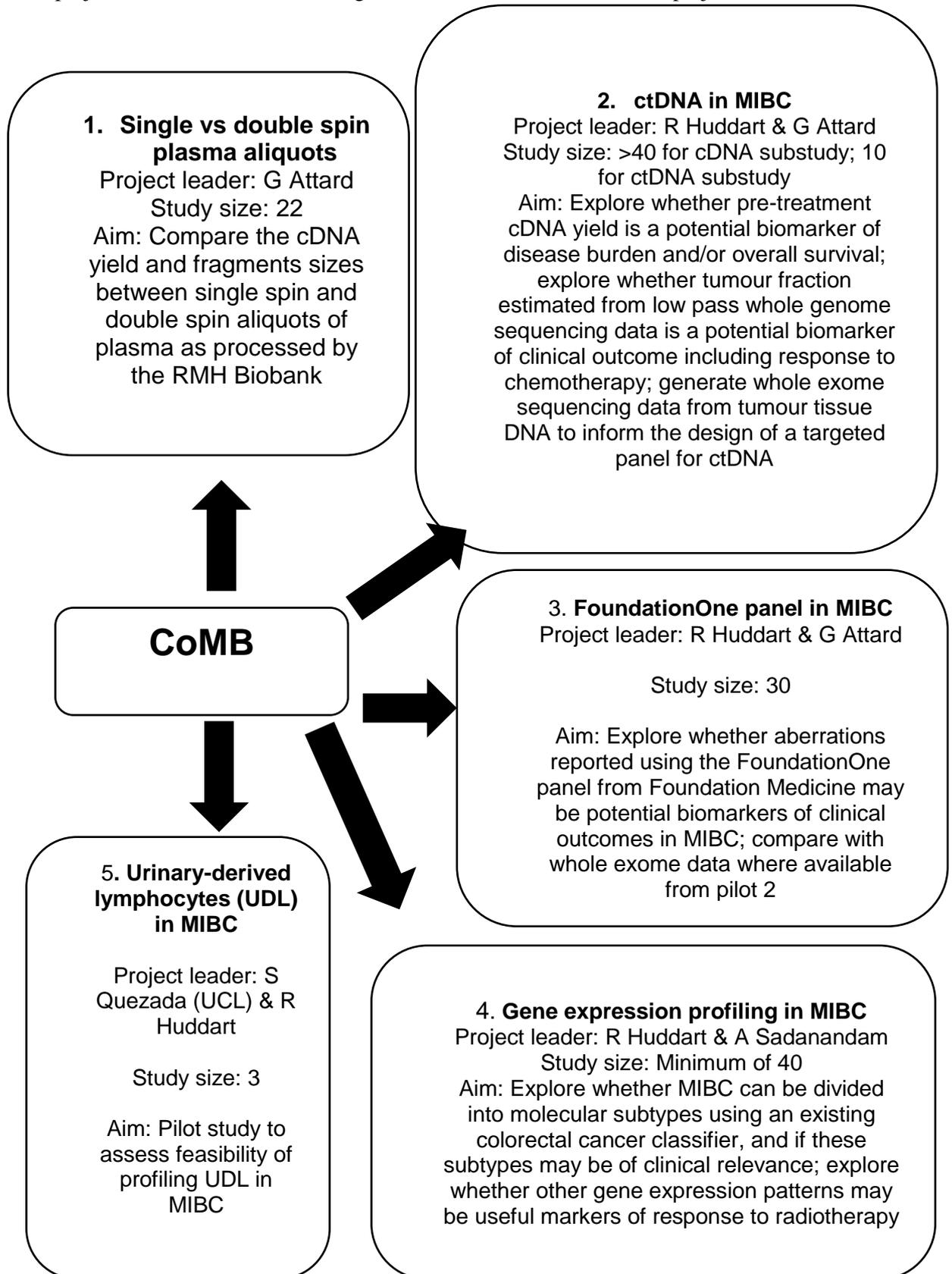
Investigations: Cystoscopies continuation sheet.

Date	Outcome
<input type="text"/>	No evidence of disease <input type="checkbox"/> Disease present <input type="checkbox"/> If disease present: pTa /pTis / pT1 / pT2 / pT3 / pT4 Tissue sample requested for CoMB? Yes / No
<input type="text"/>	No evidence of disease <input type="checkbox"/> Disease present <input type="checkbox"/> If disease present: pTa /pTis / pT1 / pT2 / pT3 / pT4 Tissue sample requested for CoMB? Yes / No
<input type="text"/>	No evidence of disease <input type="checkbox"/> Disease present <input type="checkbox"/> If disease present: pTa /pTis / pT1 / pT2 / pT3 / pT4 Tissue sample requested for CoMB? Yes / No
<input type="text"/>	No evidence of disease <input type="checkbox"/> Disease present <input type="checkbox"/> If disease present: pTa /pTis / pT1 / pT2 / pT3 / pT4 Tissue sample requested for CoMB? Yes / No
<input type="text"/>	No evidence of disease <input type="checkbox"/> Disease present <input type="checkbox"/> If disease present: pTa /pTis / pT1 / pT2 / pT3 / pT4 Tissue sample requested for CoMB? Yes / No

Signature	Print Name	Date <input type="text"/>
-----------	------------	------------------------------

D. Current pilot studies using CoMB samples (as of Sept 2018)

Since establishing CoMB, samples have been used or are planned to be used in 5 pilot projects approved by the CoMB trial management group. An updated summary of each project is included here and the figure below summarises the current projects



E. Inclusion of patients with upper tract malignancies (CoMB)

Upper tract cancer is rare. In recent years there has been increasing interest in the molecular profiling of upper tract cancers. Translational research is greatly needed to improve understanding of this rare but poor prognosis disease. To facilitate this, the eligibility criteria for CoMB will be extended to include those with upper tract cancers.

Sample collections and follow-up schedules will be as for MIBC patients in CoMB

A2.2 RNA optimisation work

Date	Processed by	Sample ID	Tissue type	Thickness	Deparaffinisation Method	Kit	Elution		Nanodrop				Bioanalyser	
									ng/ul	260/280	260/230	Total μ g	RIN	(ng/ μ l)
06.11 .14	E.L	4.1	mouse kidney A	10 x 4 μ m	MNDPS	MN	50 μ l	Single	129.7	1.98	1.67	6.485	4.3	93
	E.L	7.1	mouse kidney A	10 x 7 μ m	MNDPS	MN	50 μ l	Single	174.7	1.92	1.44	8.735	7.7	126
	E.L	10.1	mouse kidney A	10 x 10 μ m	MNDPS	MN	50 μ l	Single	99.7	1.87	1.32	4.985	3.5	76
06.11 .14	E.L	4.2	mouse kidney A	10 x 4 μ m	MNDPS	MNXS	30 μ l	Single	203.8	1.97	1.69	6.144	6.6	123
	E.L	7.2	mouse kidney A	10 x 7 μ m	MNDPS	MNXS	30 μ l	Single	275.3	1.95	1.47	8.259	NA	146
	E.L	10.2	mouse kidney A	10 x 10 μ m	MNDPS	MNXS	30 μ l	Single	181.2	1.93	1.44	5.436	6.9	107
04.12 .14	E.L	Qiagen 4	mouse kidney A	10 x 4 μ m	DPS	Qiagen	20 μ l	Double	187.9	1.96	1.8	3.758	6.6	154
	E.L	Qiagen 8	mouse kidney A	5 x 8 μ m	DPS	Qiagen	20 μ l	Double	234.8	1.95	1.73	4.696	8	206

Date	Processed by	Sample ID	Tissue type	Thickness	Deparaffinisation Method	Kit	Elution		Nanodrop				Bioanalyser	
									ng/ul	260/280	260/230	Total μ g	RIN	(ng/ μ l)
	E.L	Qiagen 10	mouse kidney A	4 x 10 μ m	DPS	Qiagen	20 μ l	Double	136	1.86	1.62	2.72	5.3	88
	E.L	Qiagen 20	mouse kidney A	2 x 20 μ m	DPS	Qiagen	20 μ l	Double	162.6	1.96	1.79	3.252	6	159
15.04 .15	E.L	Qiagen X4	Kidney B	10 x 4 μ m	Xylene	Qiagen	20 μ l	Double	766.6	1.91	1.63	15.332	2.1	914
	E.L	Qiagen X8	Kidney B	5 x 8 μ m	Xylene	Qiagen	20 μ l	Double	651.7	2.03	1.94	13.034	2.2	1011
	E.L	Qiagen X10	Kidney B	4 x 10 μ m	Xylene	Qiagen	20 μ l	Double	580.2	1.97	1.7	11.604	2.2	818
	E.L	Qiagen X20	Kidney B	2 x 20 μ m	Xylene	Qiagen	20 μ l	Double	327.3	2.03	1.94	6.546	2.2	358
15.04 .15	E.L & M.T	Qiagen D4	Kidney B	10 x 4 μ m	DPS	Qiagen	20 μ l	Double	139.6	2.02	1.84	2.792	2.3	98
	E.L & M.T	Qiagen D8	Kidney B	5 x 8 μ m	DPS	Qiagen	20 μ l	Double	239.6	1.98	1.63	4.792	2.4	213
	E.L & M.T	Qiagen D10	Kidney B	4 x 10 μ m	DPS	Qiagen	20 μ l	Double	124.5	1.93	1.62	2.49	2.3	102
	E.L & M.T	Qiagen D20	Kidney B	2 x 20 μ m	DPS	Qiagen	20 μ l	Double	142.8	1.96	1.7	2.856	2.4	135

Date	Processed by	Sample ID	Tissue type	Thickness	Deparaffinisation Method	Kit	Elution		Nanodrop				Bioanalyser		
									ng/ul	260/280	260/230	Total μ g	RIN	(ng/ μ l)	
24.04 .15	E.L & M.T	RNA A4	Kidney B	10 x 4 μ m	Xylene	Ambion	60 μ l	Single	173.8	2.13	2.04	10.248	2.5	154	
	E.L & M.T	RNA A8	Kidney B	5 x 8 μ m	Xylene	Ambion	60 μ l	Single	207.6	2.11	1.86	12.456	2.5	154	
	E.L & M.T	RNA A10	Kidney B	4 x 10 μ m	Xylene	Ambion	60 μ l	Single	165.5	2.09	1.41	9.93	2.6	124	
	E.L & M.T	RNA A20	Kidney B	2 x 20 μ m	Xylene	Ambion	60 μ l	Single	125	2.09	1.73	7.5	2.6	95	
28.05 .15	M.T & C.R	45021 Q4	Normal Kidney	10 x 4 μ m	Xylene	Qiagen	20 μ l	Double	106.9	2.08	2.15	2.138	2.7	160	
	M.T & C.R	45021 Q8	Normal Kidney	5 x 8 μ m	Xylene	Qiagen	20 μ l	Double	120.4	1.96	1.61	2.408	2.3	122	
	M.T & C.R	45021Q1 0	Normal Kidney	4 x 10 μ m	Xylene	Qiagen	20 μ l	Double	152.8	2.1	2.15	3.056	2.3	120	
	M.T & C.R	45021 Q20	Normal Kidney	2 x 20 μ m	Xylene	Qiagen	20 μ l	Double	267.3	2.1	2.15	5.346	4.7	177	
															pg/ul
01.06 .15	M.T	72486 Q4	Pancreas WT	10 x 4 μ m	DPS	Qiagen	20 μ l	Double	44.2	1.97	1.62	0.884	2.4	12583	

Date	Processed by	Sample ID	Tissue type	Thickness	Deparaffinisation Method	Kit	Elution		Nanodrop				Bioanalyser	
									ng/ul	260/280	260/230	Total μ g	RIN	(ng/ μ l)
	M.T	72486 Q8	Pancreas WT	5 x 8 μ m	DPS	Qiagen	20 μ l	Double	40.7	2.05	2.02	0.814	2.3	7030
	M.T	72486 Q10	Pancreas WT	4 x 10 μ m	DPS	Qiagen	20 μ l	Double	45.1	2.1	2.08	0.902	2.2	20110
	M.T	72486 Q20	Pancreas WT	2 x 20 μ m	DPS	Qiagen	20 μ l	Double	30.3	2.07	1.91	0.606	2.2	12300
														pg/ul
01.06 .15	M.T & C.R	72486 X4	Pancreas WT	10 x 4 μ m	Xylene	Qiagen	20 μ l	Double	2.2	2.71	1.26	0.044	2.6	1269
01.06 .15	M.T & C.R	72486 X8	Pancreas WT	5 x 8 μ m	Xylene	Qiagen	20 μ l	Double	4.9	2.23	2.01	0.098	2.5	3257
01.06 .15	M.T & C.R	72486 X10	Pancreas WT	4 x 10 μ m	Xylene	Qiagen	20 μ l	Double	2	2.44	0.82	0.040	2.1	228
01.06 .15	M.T & C.R	72486 X20	Pancreas WT	2 x 20 μ m	Xylene	Qiagen	20 μ l	Double	5	1.55	0.6	0.100	1	11
														pg/ul
17.06 .15	M.T	72487 QX8	Pancreas	5 x 8 μ m	Xylene	Qiagen	20 μ l	Double	16	2.04	1.81	0.0008	2.8	4276
17.06 .15	M.T	72487 QX10	Pancreas	4 x 10 μ m	Xylene	Qiagen	20 μ l	Double	109.8	1.39	0.72	0.0054 9	NA	NA

Date	Processed by	Sample ID	Tissue type	Thickness	Deparaffinisation Method	Kit	Elution		Nanodrop				Bioanalyser	
									ng/ul	260/280	260/230	Total μ g	RIN	(ng/ μ l)
17.06 .15	M.T	72487 AX8	Pancreas	5 x 8 μ m	Xylene	Qiagen	20 μ l	Double	23.5	1.42	0.62	0.0011 75	2.3	168
17.06 .15	M.T	72478 AX10	Pancreas	4 x 10 μ m	Xylene	Qiagen	20 μ l	Double	14.8	1.69	1.01	0.0007 4	2.8	4743

E.L.: Erica Little (Higher Scientific Officer. M.T: Melissa Tan. C.R: Chanthirika Ragulan (Scientific Officer). DPS: deparaffinisation solution. MN- Machery-Nagel.

Chapter 3. Plasma DNA and Circulating Tumour DNA in MIBC

3.1. Introduction

At the time of setting up this MD(Res), there had been very little work on plasma DNA or circulating tumour DNA (ctDNA) in MIBC beyond initial publications demonstrating that ctDNA is detectable in patients with advanced disease^{1,2}. However, as discussed in the introduction chapter, there has been a surge of publications in this area which show great promise for the role of ctDNA as a biomarker in MIBC³⁻⁹.

In this chapter, I plan to firstly explore whether baseline pre-treatment plasma DNA concentration is a potential prognostic biomarker in MIBC, and secondly explore whether ctDNA fraction, estimated using the iChorCNA algorithm¹⁰, can be tracked in sequential plasma samples and if this might be a potential biomarker of response to treatment.

3.2. Hypotheses & Aims

3.2.1. Hypotheses

1. Baseline pre-treatment plasma DNA concentration is correlated with initial disease burden in chemotherapy-naïve MIBC patients, and may be a prognostic biomarker of clinical outcome in patients with MIBC i.e. patients with a higher plasma DNA levels prior to receiving treatment have a poorer prognosis
2. Tumour fraction, estimated using copy number alteration data from low-pass whole-genome sequencing (LP-WGS) of plasma DNA, is correlated with disease burden, and is a potential biomarker of response to first line-based chemotherapy in patients with muscle-invasive bladder cancer. Furthermore, changes in tumour fraction

identified in sequential samples could be an early indicator of response or resistance to first-line platinum-based chemotherapy

3.2.2.Aims

- 1) To explore the role of baseline plasma DNA concentration as a biomarker in MIBC, I aim to,
 - a) Extract plasma DNA from baseline pre-treatment samples of chemotherapy-naïve patients in CoMB and calculate baseline plasma DNA concentration
 - b) Explore whether baseline pre-treatment plasma DNA concentration is correlated with disease burden, as defined using a) TNM staging criteria, and b) radiological RECIST v1.1 criteria
 - c) Explore whether baseline pre-treatment plasma DNA concentration is associated with overall survival
 - d) Explore, in a subset of patients, whether changes in plasma DNA concentration over time are associated with clinical outcome

- 2) To explore the role of estimated tumour fraction as a biomarker in MIBC, I aim to
 - a) Extract plasma DNA from sequential samples taken before, during and following first-line platinum chemotherapy in the neoadjuvant or palliative setting in a cohort of 10 patients
 - b) Generate LP-WGS data from sequential (pre-treatment, post-treatment and progression time points) plasma DNA samples for analysis using the iChorCNA algorithm, which estimates tumour fraction using copy number alteration results
 - c) Explore whether tumour fraction is correlated with radiological disease burden, as defined using RECIST v1.1 criteria
 - d) Evaluate whether tumour fraction estimated by the iChorCNA algorithm is a potential clinically useful biomarker of response to 1st line chemotherapy in MIBC

3.3. Method

3.3.1. Patient cohort

3.3.1.1. Plasma DNA concentration cohort

Chemotherapy-naïve patients, defined as those receiving no previous chemotherapeutic agents for MIBC other than as a radiosensitiser, within the CoMB protocol were identified. Availability of a pre-treatment plasma sample ('baseline sample') was essential for inclusion into this cohort. The absolute minimum cohort size was set at 30 in accordance with the statistical principles set out in the CoMB protocol which state that where survival analysis is to be reported, a minimum cohort size of 30 is required. All patients recruited up to 31st July 2018 were considered for inclusion in this cohort.

3.3.1.2. Tumour fraction cohort

10 subjects fulfilling the following eligibility criteria were identified from CoMB:

- Confirmed systemic treatment-naïve MIBC
- Commencing 1st line platinum-based chemotherapy in the neoadjuvant or palliative setting
- Availability of diagnostic FFPE block, baseline pre-chemotherapy plasma samples and serial plasma samples (taken before each cycle of chemotherapy, and where applicable, at disease progression)

The cohort included 5 patients with disease relapse and 5 with no disease relapse at the time of data analysis (median follow-up period of 3.1 years). Disease relapse was defined as progression or recurrence of invasive MIBC either during or after chemotherapy treatment.

All 10 patients were also included in the plasma DNA cohort above in section 1.3.1.1.

3.3.2. Plasma DNA extraction & quantification

Plasma DNA was extracted using the QIAmp Circulating Nucleic Acid Kit. The full protocol is included in Appendix 3.1. In summary, 2 double spin vials of

plasma were retrieved from the biobank for each time point of interest for each subject. Samples were thawed at room temperature before being transferred to 1.5ml DNA LoBind Eppendorfs and centrifuged at 8000XG for 10 minutes. The volume of plasma transferred was recorded.

Up to 1.5ml of supernatant was then transferred to a 5ml Eppendorfs containing 100ul Proteinase K (PK). Additional PK was added as necessary to ensure a ratio of 1ml plasma to 100ul PK. Buffers were not adjusted in volume as per the optimised protocol used by Professor Attard's team.

The final elution volume was 120ul nuclease-free water or low Tris-EDTA buffer (low TE). Nuclease-free water was used in extractions performed in 2016. However, I was later advised that low TE would be more stable for long-term storage and so extractions from 2017 onwards were eluted in low TE.

Samples were quantified using either the Picogreen QuantIt assay (2016 extractions) or the Qubit assay (2017 onwards). Extracted samples were stored at -20°C.

Baseline samples were extracted for all the plasma DNA cohort. For the 10 patients in both the plasma DNA and tumour fraction cohorts, additional samples taken during and after treatment were also extracted with the absolute minimum of pre-cycle 1, pre-cycle 3 and progression samples.

3.3.3. Measuring disease burden

I used the RECIST v1.1 criteria¹⁵ to measure radiological disease burden. These state that measurable disease is defined as lesions measuring a minimum of 10mm (or 15mm in short axis for lymph nodes). Up to 5 representative lesions were measured with no more than 2 from one organ site. Ascites, effusions and bone metastases without a soft tissue component were considered unmeasurable. Where multiple lesions were present, those most representative of all organs were selected. The largest lesions were also selected unless measurement would not be considered reproducible.

I then summed the diameters measured to record a disease burden estimate.

With regards to baseline disease burden assessment, I took measurements from imaging performed before treatment commenced and, where possible, within 28 days of when the baseline blood sample was taken. An interval of 28 days was selected in keeping with the vast majority of bladder cancer clinical trials where baseline imaging is required within 28 days prior to starting treatment. For those cases where the interval was greater than 28 days, I assessed these on a case by case basis.

Acceptable modalities for disease burden measurement were CT scans, MRI scans and for non-bladder measurements, PET scans. Where the formal report gave measurements, I checked the scan and recorded the dimension as given in the report. All lymph node measurements were visually checked to ensure they were with regards to the short axis, and where lesions were present but not measured in the report, scans were reviewed and representative lesions measured as per the criteria set out above.

In patients receiving neoadjuvant chemotherapy where staging CT chest/abdomen/pelvis (CAP) had previously been performed but there was more recent MRI bladder imaging, I used the MRI bladder scan to provide any measurements of locoregional disease.

When assessing imaging performed during the course of treatment or at progression, the same approach was applied and the imaging selected for disease burden assessment was as close to the blood draw as possible. However, in these cases, if the interval between blood draw and imaging was over 28 days, I did not exclude them as one area of interest would be to explore any potential lead-time plasma DNA analysis might give above radiological imaging. I acknowledge that in cases of rapidly progressive disease, this could potentially lead to an under- or overestimation of disease burden.

3.3.4. Whole-genome library preparation

All whole-genome libraries were prepared by me using the NEB Next Ultra II DNA Library prep kit. An optimised protocol was kindly provided by Dr Anu Jayaram (PhD student in Professor Attard's team), and the first batch of 15 samples were processed with assistance and guidance from Dr Jayaram. The full amended protocol is provided in Appendix 3.2. In brief, thawed samples were prepared into aliquots containing 25ng plasma DNA in 50ul low TE, and dispensed into a 96-well plate. End repair was followed by adaptor ligation using a 1:10 adaptor dilution in accordance with the input DNA amount. Clean-up was performed using AMPure XP beads (at 1.5x volume) followed by an 80% ethanol wash. For the first batch, NEB multiplex single index oligo primers were used. For the remaining samples, dual index primers were used. The adaptor-ligated DNA samples then underwent PCR-amplification (5 cycles). Clean up was performed using 1.5x volume of AMPure XP beads and an 80% ethanol wash. The DNA product was eluted in 33ul low TE with a final volume of 30ul transferred to a new well. 1ul of each sample was run on a Bioanalyser High Sensitivity DNA chip, and repeated at an appropriate dilution if required.

3.3.5. Low-pass whole-genome sequencing (LP-WGS)

A total of 36 plasma whole-genome libraries were sent for LP-WGS at 1x coverage to the ICR Tumour Profiling Unit (TPU).

Fastq files were generated and transferred to the Attard team bioinformatician (Dr Mariana Buongiorno Pereira) for iChorCNA analysis.

3.3.6. Estimation of tumour fraction

In brief, the LP-WGS data was divided into ~5200 bins, and a Hidden Markov Model was used to simultaneously estimate tumour content and ploidy. Solutions on whether a bin deviated from the mean level for that patient's genome (gain or loss) were allocated a likelihood ratio where the higher the

number, the more likely the solution. The solutions deemed most likely and the alternative solutions were then transferred to me for review in PDF format.

3.3.7. Analysis

3.3.7.1. Plasma DNA concentration

Median plasma DNA concentrations with IQR for patients with T2-4N0M0 disease, regional node disease (any T, N0-3, M0), and distant metastatic disease (any T, any N, M1) were reported, with a Kruskal-Wallis test performed to determine if there was any significant difference between the groups. A Mann-Whitney test was also performed to compare plasma DNA concentration between patients with or without distant metastatic disease

Further analysis explored the correlation between plasma DNA concentration and tumour burden, as measured radiologically using RECIST v1.1 criteria, using a scatter plot and Spearman Rho test.

To explore how plasma DNA concentration affects overall survival and bladder cancer-specific survival, plasma DNA concentration was treated as a continuous variable in a univariate Cox regression model. Multivariate Cox regression was also performed with the addition of M-stage as a covariate. Due to the relatively small cohort size and number of events, I was limited to a total of 2 covariates if working on the general principal that for every 10 events, one degree of freedom is permitted.

Using available healthy volunteer data (from Prof Attard's lab) on plasma DNA concentration, an arbitrary 'cut off' value of the healthy volunteer maximum was used as a threshold to define 'high' and 'low' baseline plasma DNA values. A Cox regression test was performed to compare overall and bladder cancer-specific survival for the two groups.

For the 10 patients where plasma DNA concentration from sequential samples is available, log₂ fold change in plasma DNA concentration was plotted against set time points (e.g. pre-treatment, post-treatment, progression) to explore

whether changes in plasma DNA within a patient may be of interest, and this was assessed visually.

3.3.7.2. Exploring estimated tumour fraction as a biomarker of treatment response/disease status

Median tumour fraction + IQR was calculated for samples taken at baseline for patients with M0 and M1 disease. Results were compared using a Mann-Whitney test. A Spearman correlation test was performed to explore correlation between tumour fraction and disease burden as measured radiologically using current RECIST v1.1 criteria

Log2fold change in tumour fraction against set clinical time points for each patient were calculated to explore whether changes in tumour fraction reflect a patient's pathway. All analysis was performed using Excel and SPSS.

3.4. Results

3.4.1. Patient characteristics

3.4.1.1. Cohort for plasma DNA concentration analysis

As of 31st July 2018, there were 68 and 23 patients in groups 1 and 3 respectively i.e. the groups potentially containing chemotherapy-naïve patients.

From those in group 1, 18 were excluded for the following reasons-

- No 'baseline' sample in CoMB biobank = 9
- Second malignancy post-treatment = 2
- Upper tract cancer = 2
- Not treated at RMH = 2
- Diagnosis of hepatitis B so samples discarded from biobank = 1;
- Sample in EDTA bottle not processed within 2-hour time limit = 1
- Consent form not valid (subsequently withdrawn from CoMB) = 1

Within group 3, only 2/23 patients were chemotherapy-naïve.

This resulted in a total cohort number of 52. Patient characteristics are summarised overleaf in table 3.2. The median overall survival has not yet been reached for this cohort. The current median follow-up is 2.43 years.

Table 3.2 Patient Characteristics of Plasma DNA Cohort

Characteristic	N	%
Age (median)	70.5	Range 46.5 – 89.2
Gender	Male = 41	78.8
	Female = 11	21.2
Staging		
Tx (cystectomy)	6	11.5
T2	27	57.6
T3	8	21.2
T4	11	21.2
N0	33	63.5
N1	8	15.4
N2	5	9.6
N3	6	11.5
M1 disease	11	21.2
Histology		
Predominantly TCC	52	100

3.4.1.2. Cohort for tumour fraction analysis

Table 3.3 below summarises the patient characteristics for this cohort.

Table 3.3 Patient Characteristics of Tumour Fraction Cohort

NAC = neoadjuvant chemotherapy; CRT = chemoradiation; Pall C = palliative chemotherapy; RIP BLCA = deceased, cause of death was MIBC

CoMB ID	Gender	Age at dx	Stage at dx	Treatment Received	Disease status as of 31.08.2018
1003	F	68.4	T3N0M0	NAC + CRT	No relapse
1010	M	78.9	T2N0M0	NAC + CRT	Relapse; RIP BLCA
1014	M	68.2	T3N0M0	NAC + CRT	No relapse
1016	M	72.5	T2N0M0	NAC + CRT	No relapse
1025	M	71.1	T2N0M0	NAC + CRT	No relapse
1039	M	65.6	T2N1M0	NAC + CRT	Relapse; RIP BLCA
1017	F	76.1	T2N1M0	NAC. Pall RT. Pall C	Relapse; RIP BLCA
1021	M	64.4	T3N0M1	Pall C	No relapse
1027	M	71.0	T4N2M1	Pall C	Relapse; RIP BLCA
1052	F	76.6	T4N3M1	Pall C	Relapse; on current 2 nd line

Figures summarising each patient’s treatment pathway and time points at which blood samples were taken are shown in Appendix 3.3.

The median overall survival for the ctDNA cohort has not yet been reached. The current median follow-up is 3.12 years.

3.4.2. Plasma DNA concentration

3.4.2.1. Plasma DNA concentration vs clinical staging

Plasma DNA was successfully extracted from all samples as planned and concentration results were available for all patients. Table 3.4 shows the medians for the cohort overall and by disease stage. Figure 3.1 shows the corresponding box plots.

Table 3.4 Median Plasma DNA concentrations according to disease stage. Median, interquartile range (IQR) values given. p-values calculated using Kruskal-Wallis and Mann-Whitney tests

	N	Median	IQR	p
Overall	52	11.27	6.41 – 16.30	
T2-4N0M0	30	10.89	5.74 – 13.94	0.594
Node-positive	11	11.08	8.35 – 24.19	
M1 disease	11	19.35	7.08 – 86.12	
M0 disease	30	11.08	6.00 – 13.95	0.497
M1 disease	22	19.35	7.08 – 86.12	

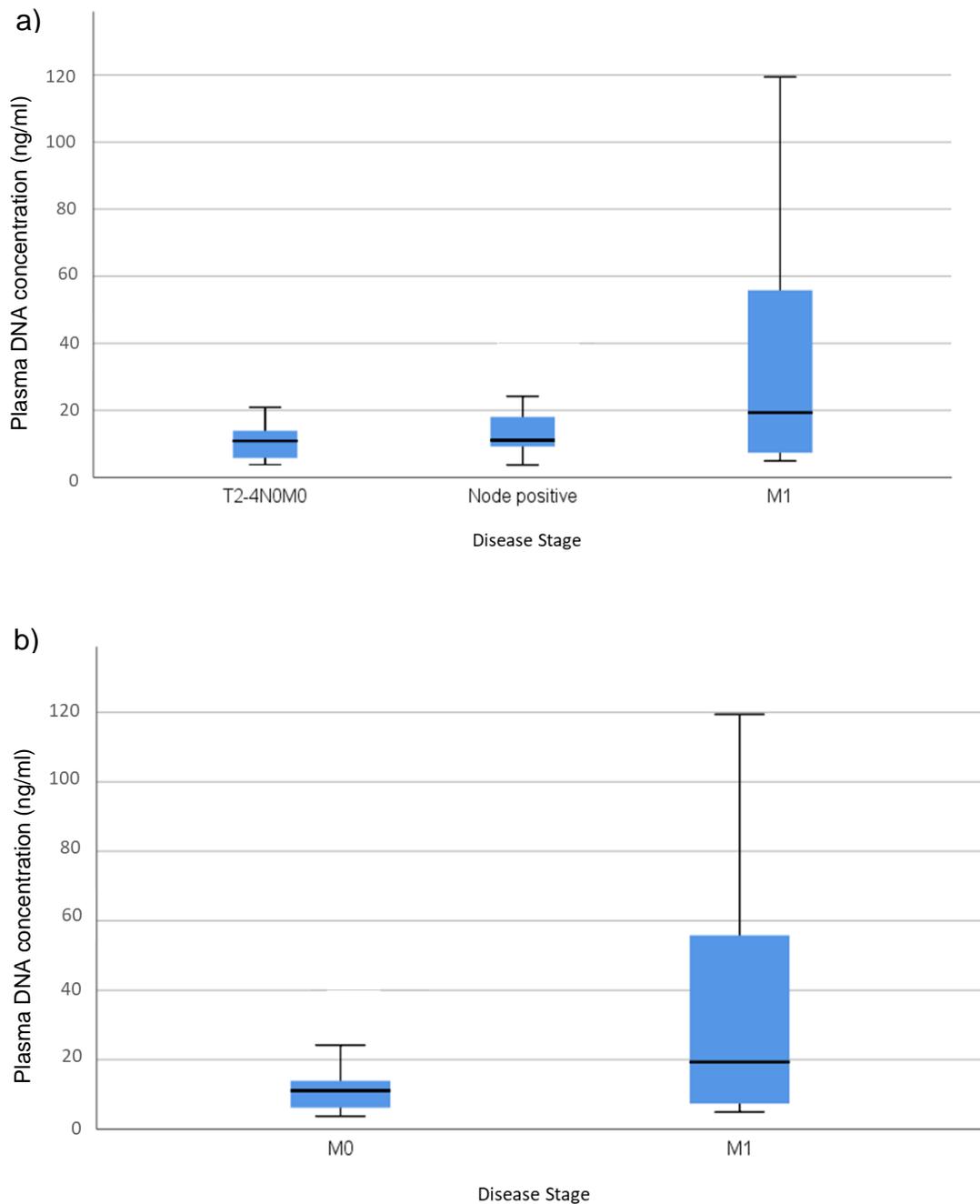


Figure 3.1 Box plots comparing plasma DNA between disease stages. a) Subjects divided into those with bladder-only disease (T2-4N0M0), involved regional lymph nodes (node positive) and distant metastatic disease (M1); b) subjects divided into those without and with distant metastatic disease (M0 and M1 respectively). Upper and lower whiskers denote 97.5th and 2.5th percentiles respectively.

3.4.2.2. Plasma DNA vs disease burden

Disease burden was measured for 41/52 (78.8%) subjects. For the 11 subjects excluded from this part of the analysis, there was no adequate post-TURBT imaging available on which to assess disease burden or interval between imaging and blood draw was significantly >28 days. Measurements taken are shown in Appendix 3.4.

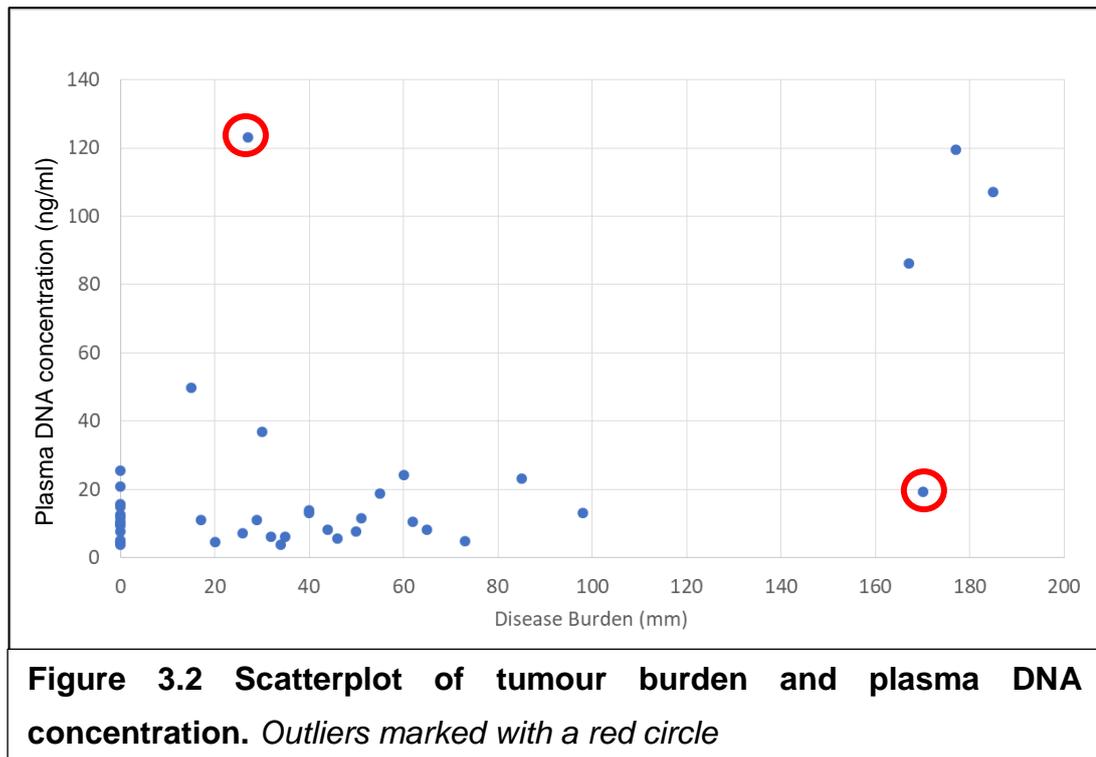
Imaging was performed at a median of 12 days (IQR 4 – 22) prior to the blood draw, and all except 3 within 28 days or less. For one of these patients (1052), the baseline staging scan was performed after TURBT but 48 days prior to blood sampling. However, the bulk of this patient's disease was in the lungs and they had a CT-guided lung biopsy performed 9 days prior to the blood draw. There had been significant disease progression in the lung metastases (the only identified site of M1 disease) and so I used this scan to provide the lung measurements. This changed the disease burden for 1052 from 99mm to 170mm. While not an ideal solution, I believe this is more representative of the disease burden at the time of the blood draw.

The second patient with an interval of over 28 days (1055) was a patient with localised disease commencing NAC. The interval was 34 days. I decided that as there was unlikely to have been any significant change in disease burden as measured on imaging in the 6 days beyond the arbitrary 28-day cut-off, it would be reasonable to include this patient within the dataset

The third patient (2059) had an interval of 32 days. This patient had M1 disease. They also had bone disease which was not measurable resulting in an underestimation of the disease burden. As this underestimation was likely to be more significant than any disease progression that may have occurred in the 4 days beyond the arbitrary 28-day cut off, I thought it would be reasonable to include this patient. Furthermore, as there were relatively few patients with a large disease burden in this cohort, and given the exploratory nature of this pilot, I felt it would be worthwhile to include this patient, accepting the potential further underestimation of disease burden.

To ensure that the inclusion of these 3 patients does not significantly alter the overall results, I performed the analysis with and without these cases.

There was no statistically significant correlation between tumour burden and plasma DNA concentration with Spearman $Rho= 0.2631$; $p=0.09651$ (Figure 3.2)



On assessing plot visually, there 2 outliers (marked with red circle). One had a disease burden of 27mm but plasma DNA concentration of 123ng/ml. Review of the medical notes showed no obvious reason for this although it was noted that the patient had a positive urine culture a week later. This raises the possibility of whether a urine infection could cause raised plasma DNA levels. Unfortunately no further plasma samples from this patient have yet been extracted. It would be of interest to see subsequent levels to try and understand any potential reasons for such a raised level.

The 2nd outlier was patient 1052 (previously discussed), with a disease burden of 170mm and plasma DNA concentration of 19.35. Given the extensive lung metastases it is perhaps surprising that the plasma DNA level was not higher.

Removing the 3 patients where complete imaging was not performed within 28 days of blood draw did not significantly alter the result (Spearman Rho = 0.203, $p = 0.221$) (Scatterplot in Appendix 3.5).

3.4.2.3. Plasma DNA concentration and survival

Cox regression with plasma DNA concentration as a continuous variable

Univariate Cox regression analysis with plasma DNA concentration as a continuous variable produced a statistically significant hazard ratio of 1.013 (95% CI 1.001 – 1.025; $p = 0.04$). For bladder cancer-specific survival, the HR was 1.021 (95% CI 1.007 – 1.034; $p = 0.003$)

Adding M-stage in as a second covariate, the HR for plasma DNA concentration with respect to overall risk of death no longer reached statistical significance (HR = 1.012, 95%CI 0.999 – 1.024, $p = 0.065$) but significance was maintained for bladder cancer-specific death (HR=1.014, 95% CI 1.001 – 1.028, $p = 0.029$).

Survival analysis using healthy volunteer data to define plasma DNA subgroups

Healthy volunteer data was available from 10 subjects. All subjects were male and under the age of 65. The median plasma DNA concentration was 3.57 ng/ml (IQR: 3.06 – 4.31). The maximum value was 6.82 ng/ml. As per the analysis plan, the maximum value was used to divide the patient cohort into those with a 'high' and 'low' plasma DNA concentrations ($n = 39$ and 13 respectively). Table 3.5 shows the patient characteristics of both groups.

Table 3.5 Patient Characteristics in the ‘High’ and ‘Low’ Plasma DNA groups

	High (Plasma DNA > 6.82ng/ml)		Low (Plasma DNA ≤ 6.82ng/ml)	
N	39		13	
Median Age	71.1yrs	IQR: 65.6 – 78.2	70.5yrs	IQR: 66.8-77.8
Gender				
<i>Males</i>	28	71.8%	13	100%
<i>Females</i>	11	28.9%	0	0%
Staging				
<i>T2-4N0M0</i>	21	53.8%	9	69.2%
<i>Node positive</i>	9	23.1%	2	15.4%
<i>M1</i>	9	23.1%	2	15.4%

Figure 3.3 shows the Kaplan-Meier curves for overall survival for patients with a ‘high’ and a ‘low’ plasma DNA baseline level as defined using healthy volunteer data. There was no statistically significant difference between the subgroups for overall survival (p=0.860) or bladder cancer-specific survival (p=0.832).

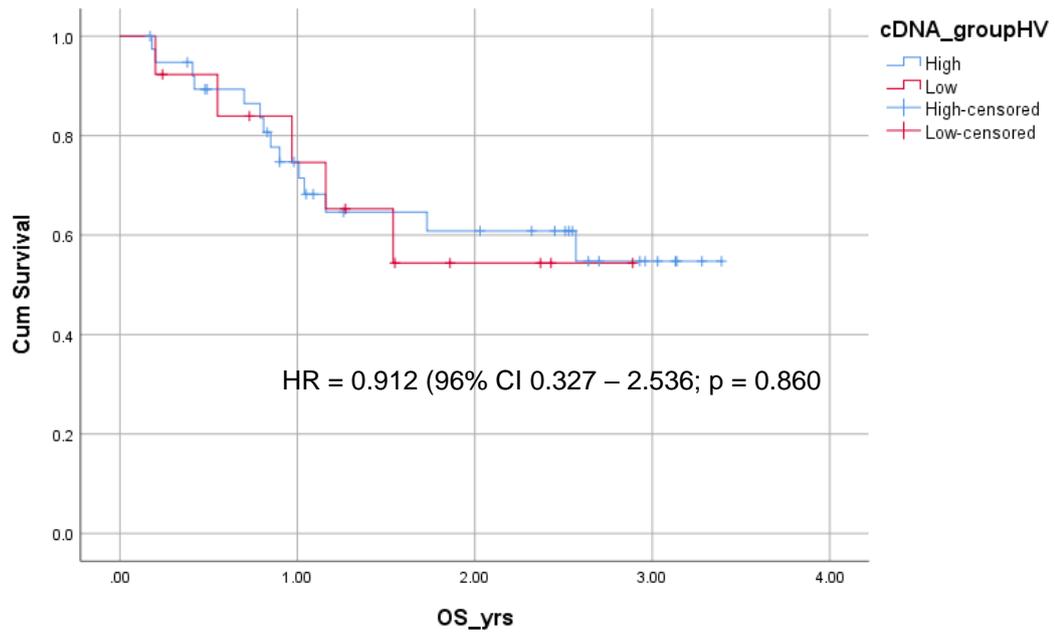


Figure 3.3 Kaplan-Meier graph showing overall survival for subjects with high and low plasma DNA concentrations

On Cox regression with the addition of M-stage as a covariate, plasma DNA group as defined by healthy volunteer data was not a statistically significant variable. The hazard ratios with 95% CI and p values are summarised in table 3.6 below.

Table 3.6 Summary of multivariate Cox regression analysis. OS = overall survival; BCSS = bladder cancer-specific survival

Endpoint	Variable	HR	95% CI	p-value
OS	Plasma DNA	0.954	0.340 – 8.256	0.928
	M-stage	3.304	1.322 – 2.678	0.011
BCSS	Plasma DNA	0.938	0.292 – 3.010	0.915
	M-stage	5.234	1.884 – 14.536	0.001

3.4.2.4. Exploration of changes in plasma DNA concentration in sequential samples

To visually assess whether changes in plasma DNA concentration from baseline might be of interest, plasma DNA concentrations from 7/10 patients receiving NAC with sequential results available were plotted using a log₂ fold change scale (figure 3.4) where a log₂ fold change of 1 is equivalent to a doubling in plasma DNA concentration from baseline, and -1 represents halving from baseline. The dashed lines represent patients with relapsed disease.

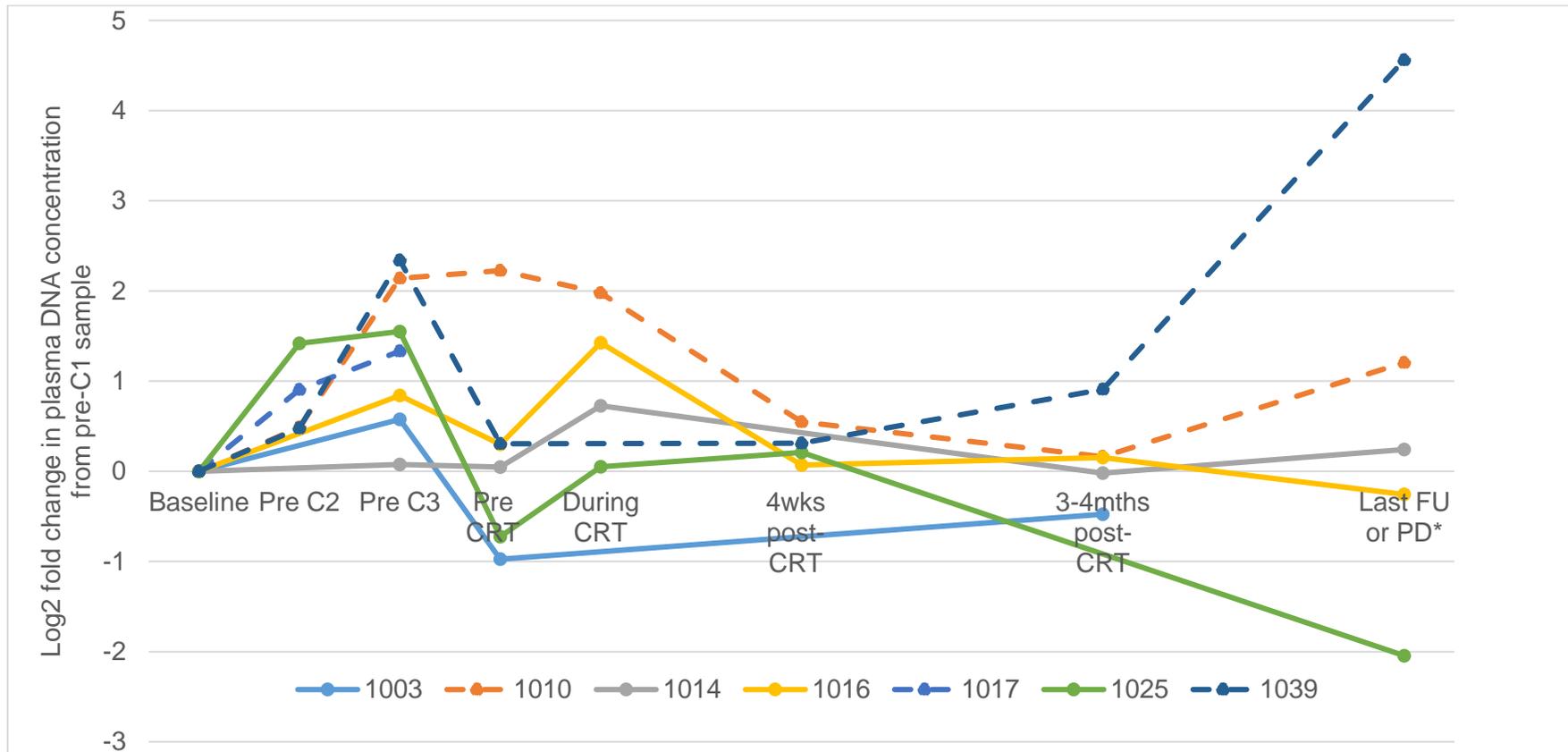


Figure 3.4. Changes in plasma DNA concentration from baseline for patients receiving neoadjuvant chemotherapy. *Dashed lines represents patients with disease progression. Solid lines represent those with no disease progression. *Time between post-CRT blood sample and 'Last FU or PD' sample ranged from 20-24 months for responders and 11months (1010) and 2 months for 1039. Abbreviations: C2 = cycle 2; C3 = cycle 3 ; CRT = chemoradiation; FU = follow-up; PD = progressive disease*

3.4.3. Tumour fraction

3.4.3.1. Estimation of tumour fraction using iChorCNA

LP-WGS data was available for 35/36 samples sent. One sample (1010 23.06.15) failed to de-multiplex despite having passed all the required quality control checks. On review, the most likely reason was a failure in index ligation.

The tumour fractions estimates and ploidy status provided as most likely by the algorithm are shown below in table 3.7. I saw no good reason to reject these results and used these figures going forward in the analysis. Of note, the ploidy status allocated for some patients changed during the course of treatment; this will be discussed further in section 3.5.2.2.

Table 3.7. Tumour fraction results for sequential samples as calculated by the iChorCNA algorithm. *Abbreviations: NAC = neoadjuvant chemotherapy; CRT = chemoradiation; PD = progressive disease, pall chemo = palliative chemotherapy*

CoMB ID	Clinical Time Point	Tumour Fraction	Ploidy
1003	-45 days C1 NAC	0	2
	-1 day C1 NAC	0.08526	2
	-1 day C3 NAC	0	2
1010	0 days C3 NAC	0	2
	98 days post CRT	0	2
	PD. 52 days before death	0.1451	2
1014	-1 day C1 NAC	0	2
	-1 day C3 NAC	0	2

Table 3.7 (continued)

CoMB ID	Clinical Time Point	Tumour Fraction	Ploidy
1016	-1 day C1 NAC	0	2
	-3 days C3 NAC	0	2
1017	-6 day C1 NAC	0	2
	-5 days C3 NAC (24 days pre-CT PD)	0	2
	6 days post CT PD	0	2
	-2 days C1 pall chemo	0.1028	2
	33 days post C2 pall chemo (78 days before death)	0.1025	2
1021	0 day C1 pall chemo	0	2
	0 days C3 pall chemo	0	2
	27 days post C6 pall chemo	0	2
	698 days from C1. No disease on scans	0	2
1025	0 days C1 NAC	0	2
	-1 day C3 NAC	0	2
1027	-44 days C1 pall chemo	0.1481	3
	-2 days C1 pall chemo	0.3158	2
	-1 day C3 pall chemo	0	2
	-15 days C6 (22 days pre CT PD)	0.1227	2

Table 3.7 (continued)

CoMB ID	Clinical Time Point	Tumour Fraction	Ploidy
1039	-7days C1 NAC	0	2
	-8 days C3 NAC	0	2
	24 days post CRT	0	2
	-1 day C1 atezolizumab	0.07027	3
	Post C2 atezolizumab (PD)	0.2166	3
1052	0 days C1 pall chemo	0.2689	4
	-1 day C3 pall chemo	0	2
	-1 day C6 pall chemo	0	2
	-2 days C1 pembrolizumab	0.1039	3
	-2 days C4 pembrolizumab	0	2

It can be seen that in the majority of cases (24/35; 68.6%), the estimated tumour fraction was 0. This included all the patients receiving neoadjuvant chemotherapy with currently no evidence of disease relapse, and one patient with metastatic lung disease who has a maintained complete response to palliative chemotherapy.

The results are further summarised in figure 3.5 below which divides the 10 patients into those with distant metastases or not, and samples into pre-treatment, pre-cycle 3 and where applicable, at progression (n=5).

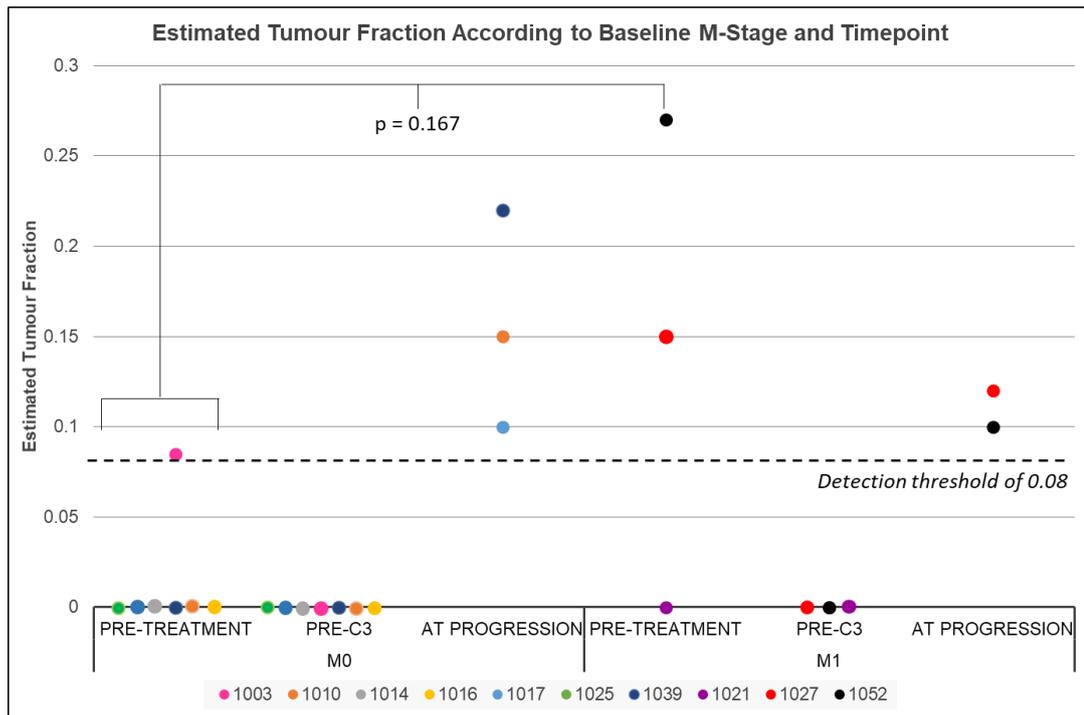


Figure 3.5 Estimated tumour fraction according to baseline M-stage and time point. 5/10 patients had disease relapse hence only 5 ‘at progression’ values available. There was no statistically significant difference in pre-treatment tumour fraction estimates between patients with and without metastatic disease

3.4.3.2. Exploring tumour fraction versus disease burden

Disease burden measurements were available for 33/35 samples. The two missing were for patient 1027 (no up-to-date baseline scan for sample taken 2 days prior to C1 PC) and patient 1003 (for blood sample taken 45 days pre-cycle 1, the CT was performed pre-TURBT whereas blood taken post TURBT). Measurements are available in Appendix 3.6.

Pre-treatment baseline imaging was performed at median of 13 days prior to blood draw (IQR = 15 days prior to blood – 7 days post blood sampling).

C3 assessment bloods done 15.5 days before imaging (IQR = 23 to 12.75 days prior to imaging).

C6 assessment bloods (n = 3) were done at 22 and 17 days before imaging, and 4 days post-imaging.

Bloods taken at disease progression were taken at a median of 5 days post-imaging (IQR = 0 – 6 days post imaging).

Fig 3.6 shows a scatter graph of disease burden vs tumour fraction with each point representing a time point i.e. each patient has several points plotted. A spearman correlation test gave a correlation coefficient of 0.600 and p value of 0.0002250.

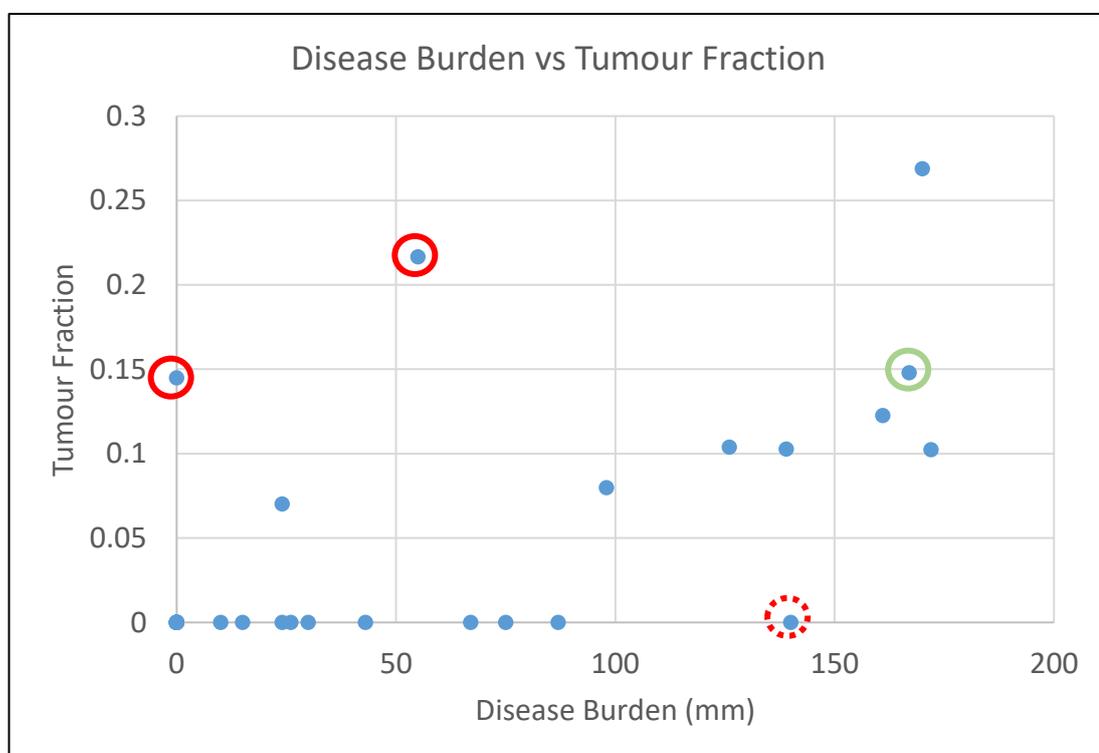


Figure 3.6 Scatter plot of disease burden and tumour fraction. *Outliers are marked with red circles. The green circle marks a baseline pre-treatment tumour fraction for patient 1027, which then fell to 0 after 3 cycles of chemotherapy (marked in dotted red circle)*

On assessing the graph visually, there are 3 outliers (marked in red circles).

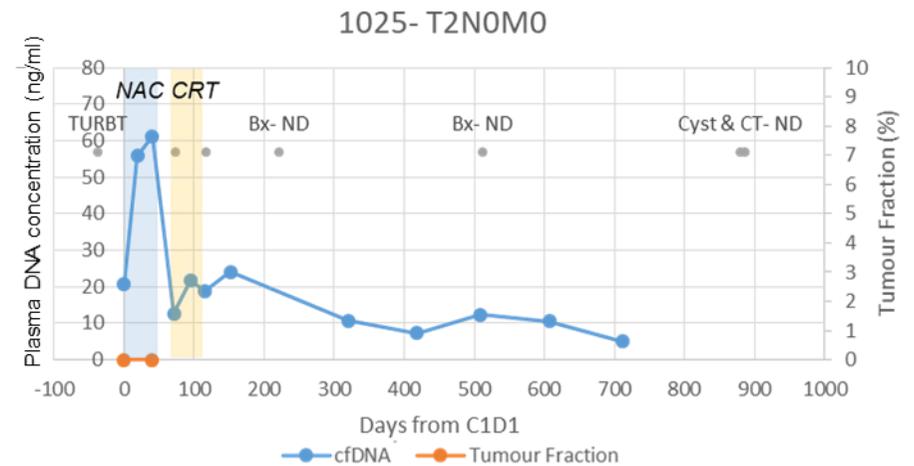
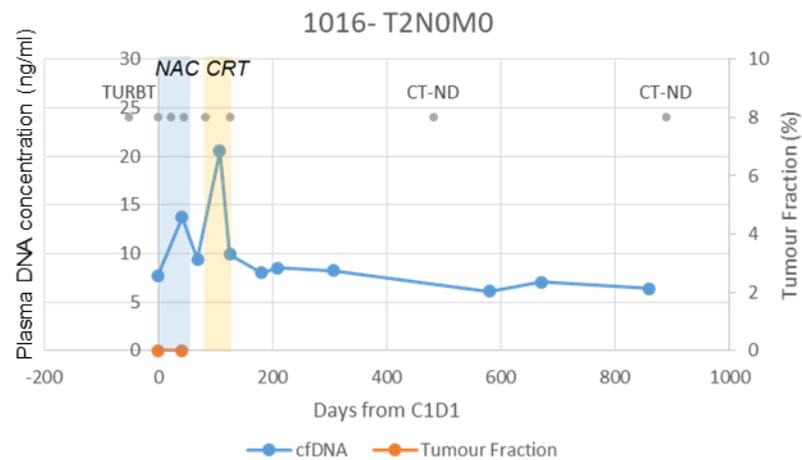
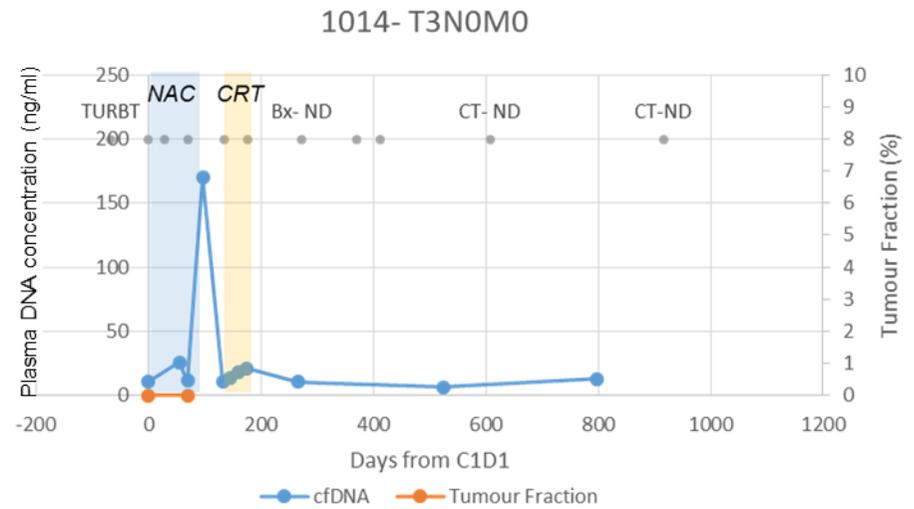
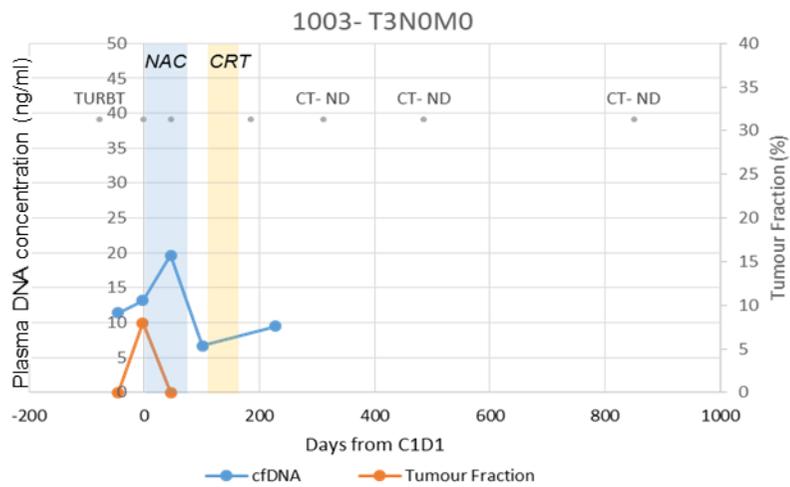
For the point at (140, 0), this relates to patient 1027 who was reported to have a partial response to 3 cycles of palliative chemotherapy. Prior to commencing treatment, patient 1027 had a disease burden of 167mm and tumour fraction of 0.1481. This point is marked in green.

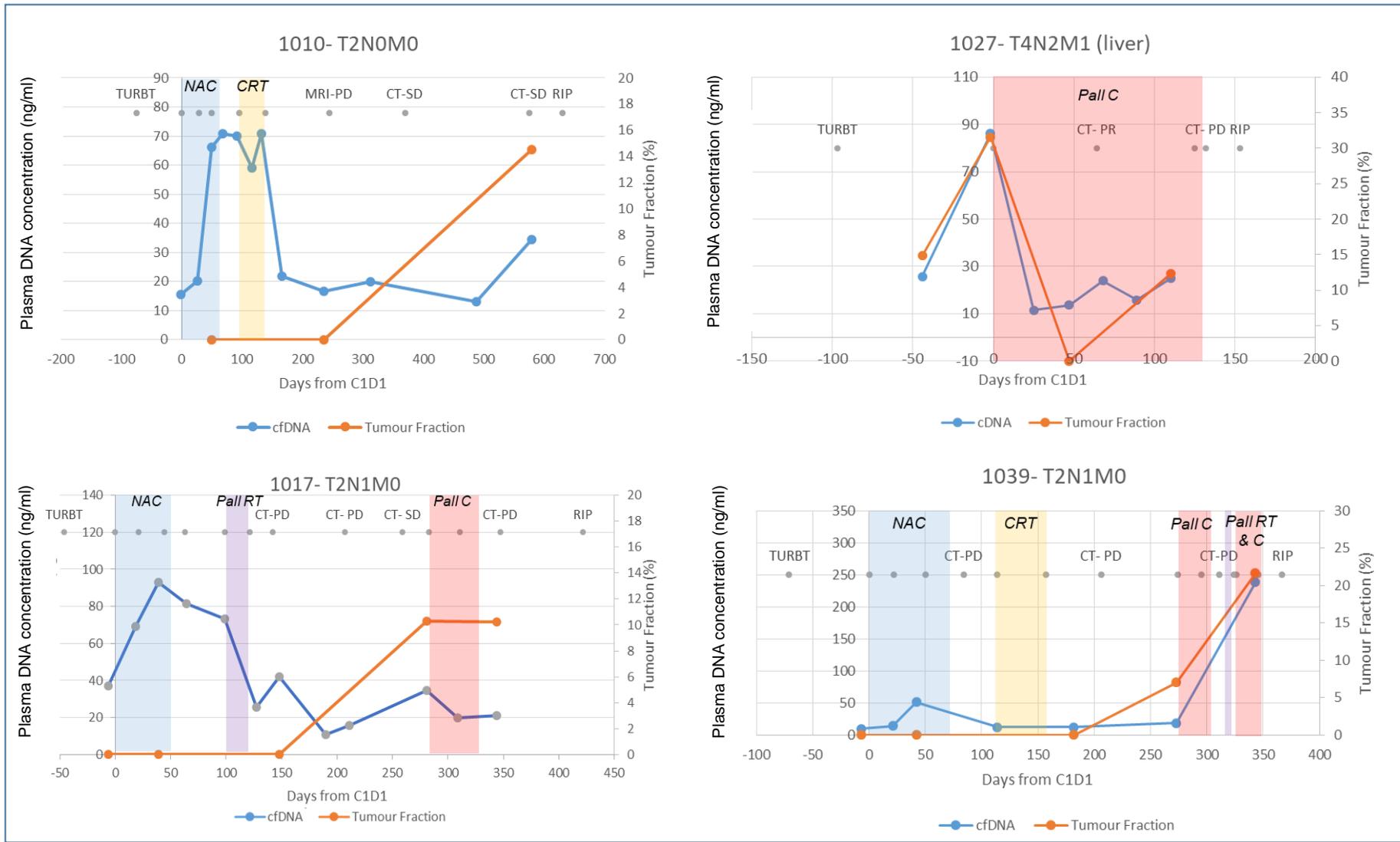
For point (0,0.145), there was no evidence of disease seen on the CT scan but the patient died 2 months later with extensive liver metastases at a different hospital (therefore no blood sample available). However, on reviewing the scans, I noted that this was a non-contrast scan and may have accounted therefore for hepatic lesions not being detected.

For the third outlier (55, 0.22), it is likely disease burden was underestimated as this patient had lytic bone metastases which were not measurable using the RECIST v1.1 criteria. Furthermore, a MRI spine done closer to time of blood sampling showed liver and pleural metastases on the localiser imaging but unfortunately it was not possible to take any measurements of disease burden on this.

3.4.3.3. Change in plasma DNA and tumour fraction during treatment

Figure 3.7 shows plasma DNA and tumour fraction levels for each patient in the context of their treatment pathway. It can be seen that in 8/10 patients, plasma DNA level rises during first line chemotherapy. In the 3 patients with detectable tumour fraction prior to commencing treatment (1003, 1027 and 1052), tumour fraction levels fell on subsequent samples and reflected response to treatment. Furthermore, in the 5 patients who progressed during or following 1st line treatment, tumour fraction was also seen to rise in all 5 cases.





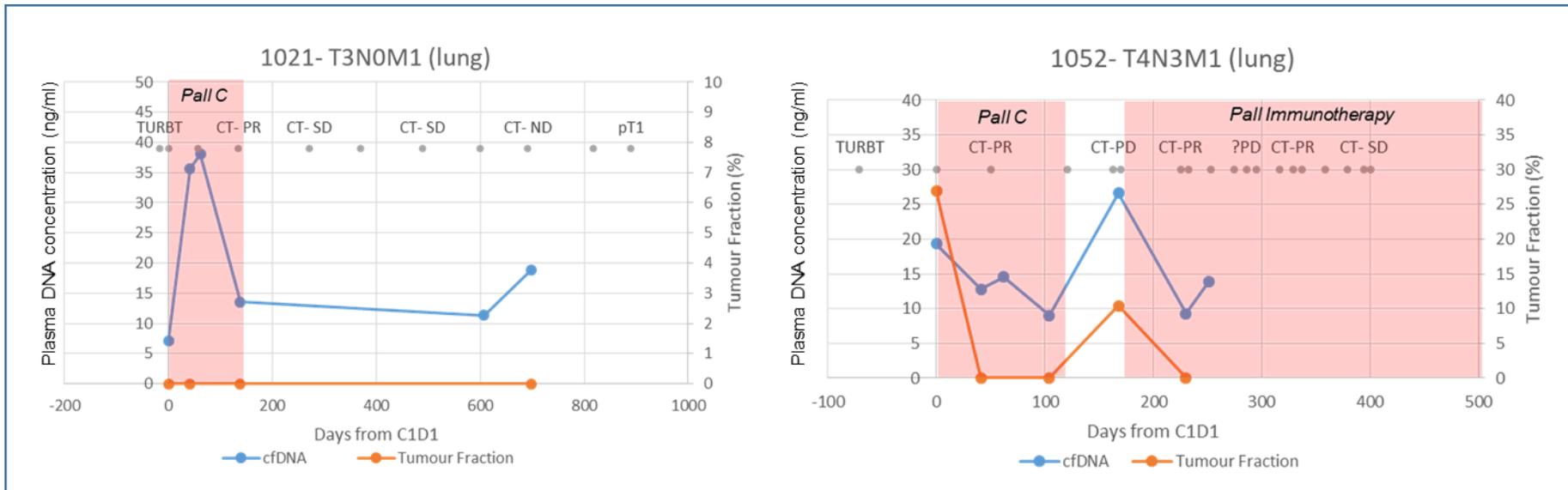


Figure 3.7. Diagrams showing plasma DNA concentration and tumour fraction results for sequential samples in context of clinical events.

Abbreviations: TURBT = transurethral resection of bladder tumour; NAC = neoadjuvant chemotherapy; CRT = chemoradiation; Pall C = palliative chemotherapy; CT- ND = nil disease seen on CT; Bx- ND = no disease on cystoscopic biopsy; CT- SD = stable disease on CT; CT- PD = disease progression on CT; RIP = deceased

3.5. Discussion

3.5.1. Plasma DNA concentration

3.5.1.1. Plasma DNA concentration and survival

The results have shown that although there is no significant correlation between baseline plasma DNA concentration and disease stage or radiological disease burden, plasma DNA concentration is associated with a small but significant hazard ratio of 1.013 and 1.021 for death and bladder cancer death respectively i.e. for every unit increase in plasma DNA concentration, there is a 1.3% and 2.1% increase in risk of death and bladder cancer death respectively.

The association between baseline plasma DNA levels and prognosis is in keeping with recent work published in prostate cancer, where baseline plasma DNA levels were reported to be an independent prognostic variable of radiological progression-free survival and overall survival in 1st and 2nd line chemotherapy settings¹⁶, in a cohort of 571 prospectively recruited patients with metastatic castration-resistant disease.

On adding the patient's M-stage (i.e. M0 or M1) to the model, the HR for risk of overall death was no longer significant but that of bladder-cancer specific death remained statistically significant at 1.014 (p=0.029). This suggests that plasma DNA concentration is an independent prognostic factor specific to bladder cancer death and that it gives further prognostic information beyond that of M-stage.

The presence specifically of visceral metastases (lung, liver, bone) is one of the two independent prognostic factors identified by Bajorin et al¹⁷ to predict survival in patients with metastatic TCC embarking on palliative platinum-based chemotherapy. In their work, they reported that patients with visceral metastases but a good performance status had a median survival of 13.3 months compared to 33 months for those with non-visceral M1 disease and a good performance status.

However, my data suggests that plasma DNA concentration may potentially offer a more sensitive way of identifying those at high risk of bladder cancer death, beyond that of simply making a distinction between M0 and M1, and of visceral or nodal metastases. Looking further at the data, 4/11 patients staged as M1 had distant nodal disease (median plasma DNA = 18.26ng/ml, range 4.93-25.53), and 7/11 patients had visceral involvement (median = 19.35ng/ml, range 6.07-119.41). The numbers are too small to draw any definitive conclusions, but it is interesting to note that while the plasma DNA medians are similar and the ranges overlap, the 3/11 M1 patients currently still alive all had visceral disease but with plasma DNA concentrations of 6.07, 7.08 and 19.35. The follow-up periods for these 3 patients are 15.1, 18.6 and 31.7 months. In contrast, the 4 patients with nodal disease who would have been quoted a median survival of 33 months died at 4.9, 10.8, 18.5 and 30.8 months post commencing chemotherapy with baseline plasma DNA concentrations of 23.09, 25.53, 4.93 and 13.43 respectively.

On the basis of this pilot data, I suggest it would be worthwhile further investigating plasma DNA in a larger cohort. The clinical implications of being able to use baseline plasma DNA as a prognostic marker would allow further risk stratification of patients at diagnosis alongside the usual factors (TNM stage, performance status) and may contribute to decision making regarding treatment strategy e.g. current international guidelines^{18,19} do not recommend staging with positron emission tomography scans (PET-CT), and UK guidelines²⁰ for MIBC recommend consideration of PET-CT only in patients with indeterminate features on CT/MRI or high risk disease (T3b+). However, it may be that plasma DNA level offers another way to identify 'occult' high risk patients and more accurate staging may then influence their treatment strategy. This pilot in plasma DNA is not without its limitations however and these are discussed further in the following sections.

3.5.1.2. Plasma DNA concentration and disease burden

On dividing the cohort according to their TNM staging, no significant difference in baseline plasma DNA concentration was seen. This result is, at first glance, not in keeping with data since published which reported that patients with

untreated metastatic MIBC had significantly higher plasma DNA levels than treatment-naïve patients with localised disease only⁶. Median plasma DNA concentrations of 20.3ng/ml and 6.4ng/ml were reported for a total of 55 plasma samples from patients with and without metastatic disease respectively, compared to my results of 19.35ng/ml and 11.08ng/ml for those with M1 and M0 disease respectively. Even though they classified patients with nodal metastases within their metastatic group, my equivalent median value of 10.08 ng/ml for patients with organ-confined disease is still higher than their value of 6.4ng/ml. The reason for this is not clear although the relatively small cohort sizes in both our studies limit any truly meaningful conclusions. Of note, the authors of this paper did not explore whether plasma DNA level was associated with clinical outcome.

The lack of significant correlation between plasma DNA concentration and disease burden in my cohort is perhaps a little at odds with my finding that plasma DNA concentration is positively associated with risk of bladder cancer-specific death, as one might hypothesise that the reason for an increased plasma DNA could be due to an increased tumour component secondary to increased tumour burden.

On the other hand, as plasma DNA levels are also affected by other benign pathological and physiological factors, it perhaps is not so surprising that no correlation was seen with disease burden. This fact, that plasma DNA rises cannot be always be attributed to malignancy alone, would also limit its use as a prognostic marker. Other potential reasons for a lack of correlation between plasma DNA concentration and radiological disease burden would include the fact that this was a small cohort, which would limit the ability to detect any small effect. However, a more significant consideration is that of the method for measuring disease burden.

Limitations of measuring disease burden

There were several limitations with regards to the measurement of disease burden.

Firstly, using the RECIST v1.1 criteria meant that disease burden would be significantly underestimated in those patients considered to have 'unmeasurable disease' such as lytic bone metastases or ascites. Furthermore, as there was a limit to the total number of lesions to be measured (maximum = 5) and the number per organ (maximum = 2), this would also have underestimated disease in a proportion of patients, particularly those with metastases mainly in one organ.

Other methods I considered included calculating the sum of the products of perpendicular measurements²¹ or a modified RECIST where all lesions measuring above 10mm are summed²². However, I decided to use the RECIST v1.1 criteria as they are widely used in reporting imaging in oncology and in bladder cancer trials. In the end, the restriction in total number of lesions and lesion per organ affected the disease measurements in 5/41 (12.2%) of this cohort (all had more than 2 measurable lesions in one organ, and 1 additionally had over 5 measurable lesions in total). One patient (1027) had 16 liver metastases documented ranging from 6 – 23 mm. The two largest were summed (17 + 23mm) with all but 2 of the remaining metastases visible to me measuring <10mm. Another patient (1052) had extensive lung disease and measurement of all lesions would have greatly changed the disease burden measurement. As this patient's lung disease measurements were taken from a limited CT-guided biopsy scan (as previously discussed in section 3.4.2.2), I did not seek to perform any formal measurements to explore this issue further although from the imaging available, it was clear to see that the disease burden would increase by at least 3-fold if all lesions were measured. Interestingly, this patient had a relatively 'low' plasma DNA baseline concentration of 19.35, given the high disease burden.

On reflection, moving forward, it might be more appropriate to sum all measurable lesions to determine disease burden. However, in cases such as 1052 where there were multiple lesions, this could prove to be very time consuming although cases such as these are likely to be the minority.

A further potential source of bias is the fact that disease measurements were taken by me and not blinded independent radiologists. I ideally would have

liked to have two independent radiologists assess anonymised scans and for inter-observer agreement to be assessed. Unfortunately this resource was not available for this pilot. I therefore had to access scans on the hospital system, meaning that they were not anonymised. A further criticism would be that measurements were taken from a variety of scan modalities, and that I used measurements from reports when available, rather than performing each and every measurement myself. However, I believe that as each scan with measurable disease was reviewed by me to ensure there was no gross error in the report, the use of measurements from the report when available was the best use of resources available to me and would also reduce potential bias given the fact that I was unblinded.

3.5.1.3. Defining 'high' and 'low' plasma DNA levels

There was no statistically significant difference in risk of death when the cohort was divided into those with a 'high' or 'low' plasma DNA level defined using the maximum healthy volunteer value of 6.82ng/ml. Unsurprisingly, on multivariable analysis, M-stage was a significant predictor of death in all analyses.

There are a few points of interest to discuss with regards to dividing the cohort according to an arbitrary plasma DNA threshold.

Firstly, the healthy volunteer data. These 10 samples were acquired by Professor Attard's lab from ICR staff volunteers (n=7) and a commercial tissue bank (n=3) over a period of 4-5 years. All samples were from males aged under 65 years, and were extracted manually using the same protocol that I used. The decision to use the maximum value as the cut-off was an arbitrary one. If the number of healthy volunteer samples had been much greater, the 97.5th percentile would have been preferable (to avoid any extreme outliers) but this was not possible on such a small cohort of 10.

With a median of 3.57 and range of 2.41 – 6.82 ng/ml, it can be seen that although the median is broadly in keeping with values reported by other groups e.g. Perkins et al²³ (median 6.5ng/ml, range 4.5-13.3ng/ml, n = 20), Adalsteinsson et al¹⁰ (median of 2.34ng/ml, range 0.55- 21.27ng/ml, n = 27),

the range (and so the maximum values) show much greater variation and this highlights the limitations of using such a small cohort of healthy volunteer data. It is reasonable to question whether healthy volunteers should have been age-matched given the fact that bladder cancer patients are typically elderly with comorbidities. At the time of designing the analysis, there was no evidence to suggest that age affects plasma DNA concentration in healthy subjects, but data has since been published that age and gender does affect plasma DNA concentration²⁴. In my cohort, the presence of comorbidities might reasonably be expected to influence levels and it is interesting to ask whether ethnicity might also have an effect on plasma DNA levels.

With regards to comorbidities, as discussed in my introduction, it is recognised that plasma DNA levels may be raised in physiological situations and in non-malignant pathological conditions such as diabetes, trauma or inflammatory conditions^{25,26}. The release of plasma DNA is thought to be likely due to cell damage although the exact mechanisms remain unclear. There is some data however to show that plasma DNA levels in patients with comorbidities such as chronic airways disease, diabetes, rheumatoid arthritis and ischaemic heart disease, are not significantly raised above levels seen in patients with no comorbidities, and that they are significantly lower than those with malignant disease²⁷.

I would therefore recommend that in future work, any healthy volunteer cohort is at least age and gender-matched. In view of the limitations of using data from a small number of healthy volunteers, post-hoc analysis of alternative thresholds for 'high' and 'low' were considered and results suggested the 75th percentile might be of relevance (see Appendix 3.7). However, given the post-hoc nature of this analysis and potential impact of multiple testing, I would suggest that this should be interpreted with care and requires evaluation in a larger cohort as a pre-planned analysis. It would be of great interest to more closely examine the relationship between plasma DNA concentration and survival, and explore whether a useful threshold e.g. using ROC analysis if considered as a continuous variable, can be defined. A larger cohort would allow a more comprehensive assessment of the relationship between plasma

DNA concentration and survival. It may be that a non-linear relationship exists between the two e.g. a relatively flat slope initially corresponding to the 'normal' or 'low' plasma DNA range, followed then by a sharp increase in risk of death, rather than the linear relationship assumed with Cox regression analysis.

On a different note, it was interesting to note all 11 female patients in the cohort fell into the 'high' plasma DNA group. Female patients with bladder cancer have been noted to present with more advanced disease and have poorer outcomes compared to their male counterparts²⁸. The reasons for this are not clearly defined but it is interesting that within this small cohort, 8/11 female patients had plasma DNA concentrations above the 75th percentile.

3.5.1.4. Serial plasma DNA measurements

Serial plasma DNA measurements were available in only a limited number of patients. I selected only those receiving NAC to illustrate on the line graph in fig 3.4 as this subgroup had more fixed time points to allow comparison across subjects. Given the fact that non-malignant causes can influence plasma DNA concentration, I was interested to see whether changes within an individual might be of interest rather than absolute figures as such. I chose a log scale as the range of plasma DNA values varied between 5.06 and 238.74, which would not have been clearly presented on a linear scale.

A limitation of the graph is that it was not possible to present the time points exactly to scale. This is less relevant for the relatively fixed time points e.g. before each chemotherapy cycle but is particularly relevant to the final time point of 'Last follow-up or disease progression'. The range of time intervals have however been provided in the footnote.

Plasma DNA concentration and tumour fraction were also plotted for each individual in figure 3.7. Of note, in 8/10 patients, plasma DNA level rises during first line chemotherapy, again consistent with previous reports of raised plasma DNA levels during systemic treatment⁶. This supports the notion that plasma DNA levels alone are unlikely to be a useful marker of response during chemotherapy given the lack of specificity for tumour dynamics²⁹. It can also

be seen that plasma DNA and tumour fraction lines do not necessarily follow each other i.e. a rise in plasma DNA does not necessarily coincide with a rise in tumour fraction.

3.5.2. Tumour fraction

3.5.2.1. Using LP-WGS and iChorCNA

Tumour fraction solutions were successfully generated from all the samples with available LP-WGS data. Unfortunately, the baseline sample for 1010 failed to demultiplex and so no result was available. However, this represents a 97.2% success rate from plasma collection to estimation of tumour fraction, and a 100% success rate for the iChorCNA pipeline in generating a solution.

Of note, the majority of samples (24/35; 68.6%) were assigned a tumour fraction of 0% and I would suggest this is likely because they fell below the detection limit of 8-9% rather than truly being 0%. This affected samples from 5/6 patients commencing neoadjuvant chemotherapy and suggests therefore that tumour fraction estimation using iChorCNA may not be of use in patients with low disease burdens. Indeed, a recent abstract reported an estimated median tumour fraction of 1.9% in metastatic urothelial cancer³⁰.

The question then is whether or not the detection limit can be improved, and initial exploration by our team bioinformatician suggests that this is not currently feasible.

3.5.2.2. Tumour fraction estimations & ploidy

On reviewing the results, I was struck by the fact that the majority of results (30/35) were with respect to a ploidy status of 2. Solutions generated for 7/10 patients were all ploidy 2 and in the remaining 3 patients, ploidy status varied. MIBC is recognised to harbour many large-scale chromosome aberrations and within the TCGA dataset, whole-genome doubling events were reported in 54%. I had therefore expected the ploidy status to be >2 in at least half of the cohort.

To further investigate this, ploidy status was inferred from the WES data using 3 algorithms (ASCAT, SEQUENZA and PureCN) by the TPU bioinformatics team. There was some variation in the results between the different pipelines but when considering the mode for each patient, 7/9 patients had a ploidy status of 3 or over. These results are available in Appendix 3.8.

The data was re-entered into the pipeline with ploidy status specified as according to the results generated from WES data. As expected, tumour fractions generated were in keeping with those generated for that specific ploidy status in the first run.

The issue of estimating ploidy status from plasma samples is complex, not least as in any tumour, the ploidy status may vary from cell to cell and the situation then becomes more complicated when considering DNA fragments in plasma. As advised by our team bioinformatician, the pipeline has an inherent bias towards a ploidy status of 2 but it is reassuring that where ploidy status is not 2, the associated tumour fraction makes biological sense. We have concluded therefore that the ploidy status generated by the pipeline should be considered more a 'means to an ends' in estimating tumour fraction rather than providing any conclusive or relevant information regarding the ploidy status of the patient's cancer at that time point.

3.5.2.3. Tumour fraction and disease burden- a potential clinical tool?

My results show that there is a statistically significant positive correlation between disease burden as measured using RECIST v1.1 criteria and tumour fraction estimated by iChorCNA from LP WGS data.

This contrasts with the results from plasma DNA and disease burden, and likely reflects the fact that tumour fraction is a more specific measurement related to the patient's cancer burden whereas plasma DNA is influenced by other non-related factors.

Disease burden was measured in the same way as for the plasma DNA analysis and so all the same limitations and potential improvements previously discussed apply here. It would be interesting to see how the improvements suggested might affect the correlation coefficient.

If a significant positive correlation was confirmed in a larger cohort and validated, tumour fraction could potentially be considered a surrogate for radiological disease burden. The advantage of tumour fraction over radiology scans would be the ability to take samples more frequently and identify response/resistance at earlier time points.^{4,7} In this very small cohort, there is some suggestion that tumour fraction could be more sensitive than radiological imaging in detecting progression e.g. patient 1010 had a tumour fraction of 14.5% despite a radiological tumour burden of 0. I accept that as the imaging performed here was suboptimal (non-contrast CT), this is not a 'fair' comparison but this example does highlight the fact that in everyday clinical practice, it is not uncommon to have renal impairment precluding the use of contrast in patients with bladder cancer, resulting in suboptimal assessment of disease state. A biomarker such as plasma tumour fraction could help assess patients more accurately.

In order to more fully assess the potential role of tumour fraction, a much larger cohort with testing of more serial samples is required. Of particular interest would be to explore at which time point tumour fraction changes. With the current data available, all 3 patients with detectable tumour fraction prior to starting treatment demonstrated a fall in tumour fraction in samples taken pre-cycle 3 and before radiological assessment of treatment response. The lead times between pre-cycle 3 tumour fraction results and radiological confirmation of disease response were 9, 17 and 23 days. However, obtaining tumour fraction results from samples taken prior to cycle 2 could reveal a considerably greater lead time, and it would be of great interest to explore this further to determine the optimal and most informative time point for tumour fraction testing. Of the 5 patients with disease progression during or following 1st line treatment, all demonstrated a rise in tumour fraction at some point but from the current available data, it is more difficult to make any meaningful statements about potential lead times and tumour fraction estimates for all serial samples available would be first required.

While lead times of up to 243 days have been reported by other groups^{4,7}, an important question to ask is whether this would have any clinical significance.

In the adjuvant setting, studies comparing immediate chemotherapy after cystectomy against deferred chemotherapy on relapse^{31,32} have shown benefit in progression-free survival with immediate treatment, but no overall survival benefit³². However, these studies were underpowered and further data is therefore needed to robustly assess potential benefit of immediate or deferred treatment at relapse. The answer here is therefore not clear but with availability of more sensitive imaging modalities e.g. PET-CT, and with increasing use of stereotactic radiotherapy for oligometastases, I believe there is an argument for more sensitive monitoring post-treatment and for pursuing early diagnosis of relapse in MIBC. Of course, patient selection here is key and not all patients would be appropriate for such an approach. A key application of tumour fraction would be to assess for persistent/residual disease after radical treatment, as this may have implications for consideration of adjuvant treatment, although clearly in this setting, the detection threshold for tumour fraction would need to be dramatically improved to make this feasible.

Since completing this analysis, work has been published looking at tumour fraction estimated by iChorCNA in 663 plasma samples from 140 patients with metastatic castrate-resistant prostate cancer³³. The authors reported that tumour fraction was correlated, amongst other parameters, with the presence and number of bone metastases, and the presence of visceral metastases. This is in keeping with my finding that tumour fraction was significantly correlated with radiological disease burden. They also reported that in patients with a baseline tumour fraction of >7% demonstrating a PSA drop of greater than 30% within 6 weeks of commencing a new treatment, tumour fraction also fell. This led the authors to suggest that tumour fraction may be a useful biomarker of initial therapeutic response. Their findings support my results and hypothesis that tumour fraction is a potentially useful biomarker, albeit limited to patients with high disease burden.

3.5.2.4. Patients with low disease burdens

My results suggest that with a tumour fraction detection threshold of 8-9%, the iChorCNA pipeline is not applicable to patients with localised disease or low disease burdens. Initial work by our bioinformatician suggests that the

detection threshold cannot be optimised. Therefore, unless the detection sensitivity can ultimately be improved, alternative strategies would therefore be required if seeking a plasma-based biomarker.

One suggestion might be to explore whether the iChorCNA pathway could be used to detect tumour fraction from urinary cell-free DNA. It has been demonstrated that cell-free DNA levels are higher in urine compared to plasma, in cohorts consisting of patients with bladder-only disease^{3,7}. It therefore follows that if urine samples are enriched in cell-free DNA compared to plasma, there may be sufficient for the iChorCNA pathway and may present a solution for those patients with localised disease in the neoadjuvant setting. This is an area I would be very interested in exploring further, and the CoMB protocol would support the collection of urine to achieve this.

3.6. Conclusion

I have achieved the aims set out within this pilot study. I have demonstrated, in a cohort of 52 patients with MIBC, that baseline plasma DNA concentration is an independent prognostic factor for risk of bladder cancer death. Using the maximum healthy volunteer plasma DNA concentration to dichotomise the cohort into those with 'high' and 'low' plasma DNA values did not demonstrate a significant difference in risk of death between the two subgroups, but further work should aim at exploring the relationship between plasma DNA concentration and survival, and indeed whether a clinically useful threshold value can be defined.

I have successfully demonstrated that tumour fraction can be estimated from LP-WGS data from plasma. While the current detection thresholds limit its use in patients with low burdens of disease, there is some evidence to suggest that tumour fraction could be a potential surrogate of radiological disease burden and may even be more sensitive in detecting disease progression.

As demonstrated by several groups, ctDNA is more readily detectable in urine including in patients with low disease burdens undergoing neoadjuvant

treatment⁷. This is likely due to direct shedding of tumour DNA fragments into urine. Future work could therefore also explore whether urinary cell-free DNA is amenable to the iChorCNA pipeline, and this may be of most relevance to patients with localised disease. An alternative approach to patients with localised disease is to utilise high-depth targeted panels/assays to detect aberrations in plasma and this is explored in the next chapter.

3.7. References

1. Bettegowda, C., *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* **6**, 224ra224 (2014).
2. Sonpavde, G., *et al.* Circulating cell-free DNA profiling of patients with advanced urothelial carcinoma. *J Clin Oncol* **34**, abstr358 (2016).
3. Birkenkamp-Demtroder, K., *et al.* Genomic Alterations in Liquid Biopsies from Patients with Bladder Cancer. *Eur Urol* **70**, 75-82 (2016).
4. Birkenkamp-Demtroder, K., *et al.* Monitoring Treatment Response and Metastatic Relapse in Advanced Bladder Cancer by Liquid Biopsy Analysis. *Eur Urol* (2017).
5. Christensen, E., *et al.* Liquid Biopsy Analysis of FGFR3 and PIK3CA Hotspot Mutations for Disease Surveillance in Bladder Cancer. *Eur Urol* **71**, 961-969 (2017).
6. Vandekerkhove, G., *et al.* Circulating Tumor DNA Reveals Clinically Actionable Somatic Genome of Metastatic Bladder Cancer. *Clin Cancer Res* **23**, 6487-6497 (2017).
7. Patel, K.M., *et al.* Association Of Plasma And Urinary Mutant DNA With Clinical Outcomes In Muscle Invasive Bladder Cancer. *Sci Rep* **7**, 5554 (2017).
8. Christensen, E., *et al.* Optimized targeted sequencing of cell-free plasma DNA from bladder cancer patients. *Sci Rep* **8**, 1917 (2018).
9. Christensen, E., *et al.* Early Detection of Metastatic Relapse and Monitoring of Therapeutic Efficacy by Ultra-Deep Sequencing of Plasma Cell-Free DNA in Patients With Urothelial Bladder Carcinoma. *J Clin Oncol* **37**, 1547-1557 (2019).
10. Adalsteinsson, V.A., *et al.* Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun* **8**, 1324 (2017).
11. Guo, G., *et al.* Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. *Nat Genet* **45**, 1459-1463 (2013).
12. Knowles, M.A. & Hurst, C.D. Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. *Nat Rev Cancer* **15**, 25-41 (2015).
13. Robertson, A.G., *et al.* Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell* **171**, 540-556 e525 (2017).
14. Soave, A., *et al.* Copy number variations of circulating, cell-free DNA in urothelial carcinoma of the bladder patients treated with radical cystectomy: a prospective study. *Oncotarget* **8**, 56398-56407 (2017).
15. Eisenhauer, E.A., *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* **45**, 228-247 (2009).
16. Mehra, N., *et al.* Plasma Cell-free DNA Concentration and Outcomes from Taxane Therapy in Metastatic Castration-resistant Prostate Cancer from Two Phase III Trials (FIRSTANA and PROSELICA). *Eur Urol* **74**, 283-291 (2018).

17. Bajorin, D.F., *et al.* Long-term survival in metastatic transitional-cell carcinoma and prognostic factors predicting outcome of therapy. *J Clin Oncol* **17**, 3173-3181 (1999).
18. Alfred Witjes, J., *et al.* Updated 2016 EAU Guidelines on Muscle-invasive and Metastatic Bladder Cancer. *Eur Urol* **71**, 462-475 (2017).
19. Chang, S.S., *et al.* Treatment of Non-Metastatic Muscle-Invasive Bladder Cancer: AUA/ASCO/ASTRO/SUO Guideline. *J Urol* **198**, 552-559 (2017).
20. National Institute for Health & Clinical Excellence. Bladder Cancer: diagnosis and management. Vol. 2016 (2015).
21. Yeh, P., *et al.* Circulating tumour DNA reflects treatment response and clonal evolution in chronic lymphocytic leukaemia. *Nat Commun* **8**, 14756 (2017).
22. Gangadhar, T.C., *et al.* Feasibility of monitoring advanced melanoma patients using cell-free DNA from plasma. *Pigment Cell Melanoma Res* **31**, 73-81 (2018).
23. Perkins, G., *et al.* Multi-purpose utility of circulating plasma DNA testing in patients with advanced cancers. *PLoS One* **7**, e47020 (2012).
24. Meddeb, R., *et al.* Quantifying circulating cell-free DNA in humans. *Sci Rep* **9**, 5220 (2019).
25. El Tarhouny, S.A., Hadhoud, K.M., Ebrahim, M.M. & Al Azizi, N.M. Assessment of cell-free DNA with microvascular complication of type II diabetes mellitus, using PCR and ELISA. *Nucleosides Nucleotides Nucleic Acids* **29**, 228-236 (2010).
26. Schwarzenbach, H., Hoon, D.S. & Pantel, K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer* **11**, 426-437 (2011).
27. Spindler, K.L., *et al.* Cell-free DNA in healthy individuals, noncancerous disease and strong prognostic value in colorectal cancer. *Int J Cancer* **135**, 2984-2991 (2014).
28. Dobruch, J., *et al.* Gender and Bladder Cancer: A Collaborative Review of Etiology, Biology, and Outcomes. *Eur Urol* **69**, 300-310 (2016).
29. Tissot, C., *et al.* Circulating free DNA concentration is an independent prognostic biomarker in lung cancer. *Eur Respir J* **46**, 1773-1780 (2015).
30. McGregor, B.A., *et al.* Correlation of circulating tumor DNA (ctDNA) assessment with tissue-based comprehensive genomic profiling (CGP) in metastatic urothelial cancer (mUC). *J Clin Oncol* **36**, 453-453 (2018).
31. Freiha, F., Reese, J. & Torti, F.M. A randomized trial of radical cystectomy versus radical cystectomy plus cisplatin, vinblastine and methotrexate chemotherapy for muscle invasive bladder cancer. *J Urol* **155**, 495-499; discussion 499-500 (1996).
32. Sternberg, C.N., *et al.* Immediate versus deferred chemotherapy after radical cystectomy in patients with pT3-pT4 or N+ M0 urothelial carcinoma of the bladder (EORTC 30994): an intergroup, open-label, randomised phase 3 trial. *Lancet Oncol* **16**, 76-86 (2015).
33. Choudhury, A.D., *et al.* Tumor fraction in cell-free DNA as a biomarker in prostate cancer. *JCI Insight* **3**(2018).

3.8. Appendix

A3.1 Plasma DNA extraction protocol

Amended from QIAamp Circulating Nucleic Acid Handbook. Protocol optimised by Prof Attard's team and specifically for up to 4ml plasma by Ania Wingate

- Thaw plasma samples at room temperature
- Set water bath to 60C
- Transfer plasma to 1.5ml DNA LoBind Eppendorfs
- Centrifuge at 8000xg for 10 minutes
- Transfer plasma to new 1.5ml DNA LoBind Eppendorfs taking care not to disturb any debris
- Add 100ul Proteinase K to 5ml DNA LoBind Eppendorfs, ensuring one 5ml tube for every 1.5ml tube
- Add up to 1.5ml plasma to a 5ml Eppendorf. If over 1ml added, add extra Proteinase K to ensure there is 100ul Proteinase K for every 1ml plasma
- Add 0.8ml Buffer to each 5ml Eppendorf. Close cap and pulse vortex for 30 seconds- place immediately in water bath at 60C for 30 minutes
- During incubation set up vacuum manifold; allow one column for each original sample. Insert appropriate number of adaptors, columns and tube extenders
- After incubation, remove tubes and add 1.8ml Buffer ACB to each 5ml Eppendorf. Close cap and pulse vortex for 30 seconds
- Incubate Eppendorfs on ice for 5 minutes
- Carefully apply the lysate-buffer ACB mixtures to the appropriate tube extender and turn on vacuum pump. Turn off pump once all the lysates have been drawn through the columns. Release the pressure to 0 mbar. Discard the tube extenders
- Add 600ul Buffer ACW1 to the columns. Leave the lid open and turn on the vacuum pump. Turn off the pump once all the buffer has been drawn through. Release the pressure to 0 mbar.

- Add 750ul Buffer ACW2 to the columns. Leave the lid open and turn on the vacuum pump. Turn off the pump once all the buffer has been drawn through. Release the pressure to 0 mbar.
- Add 750ul absolute ethanol to the columns. Leave the lid open and turn on the vacuum pump. Turn off the pump once all the buffer has been drawn through. Release the pressure to 0 mbar.
- Turn on heatblock to 56C
- Close the lids on the columns and remove them from the manifold.
- Place column in a clean 2ml collection tube and centrifuge at 20 000xg / 14 000rpm for 3 minutes
- Place the column in a new 2ml collection tube. Open the lid and put in heatblock at 56C for 10 minutes to dry the membrane completely (do not over-dry)
- Pre-heat sufficient low TE for elution at 56C
- Place column in a clean 1.5ml DNA LoBind Eppendorf. Apply 120ul low TE to the centre of the membrane and incubate at room temperature for 10 minutes
- Centrifuge at 20 000xg / 14 000rpm for 1 minute to elute the plasma DNA
- Quantify using Qubit
- Store at -20C

A3.2 Plasma Whole Genome Library Prep: Protocol

Copy produced here is an amended version of the protocol optimised by Dr Anu Jayaram

- Starting material: 25ng plasma DNA in volume of 50ul
- Transfer samples into a 96-well plate

- **End-prep**
- Add 7ul NEBNext Ultra II End Prep Reaction Buffer
- Add 3ul NEBNext Ultra II End Prep enzyme
- Seal plate and vortex to mix. Brief spin to collect all liquid from sides of wells
- Place in thermocycler with heated lid set to 75C with following parameters set- 30 minutes at 20C, 30 minutes at 65C, 4C hold

- **Adaptor Ligation**
- Add 2.5ul NEBNext Adaptor for Illumina
- Add 1ul NEBNext Ligation Enhancer
- Add 30ul NEBNext Ultra II Ligation Master Mix
- Seal plate and vortex. Spin to collect fluid from sides
- Place in thermocycler with lid temperature off and incubate at 20C for 15 minutes
- Add 3ul USER enzyme to each sample
- Seal plate. Vortex and spin to collect fluids from sides
- Place in thermocycler with heated lid set to at least 47C with the following parameters set- 15 minutes at 37C, 4C hold

- **Clean-Up of Adaptor-Ligated DNA**
- Vortex AMPure XP beads for 1 minute to ensure re-suspended
- Add 1.5x sample volume (ie. 96.5ul x 1.5 = 144.75ul) of beads to each well
- Seal, vortex and spin down gently
- Incubate at room temperature for 10 minutes

- Place plate in a magnetic stand and wait 5 minutes (or until solution clear)
- Carefully remove supernatant and discard. Do not disturb the beads
- Add 200ul freshly prepared 80% ethanol to each sample while on the magnetic stand. Incubate at room temperature for 30 seconds then remove and discard the supernatant
- Air dry the beads for up to 5 minutes (do not over-dry)
- Remove plate from magnetic stand and elute the DNA from the beads by adding 17ul low TE
- Seal plate and vortex. Spin down gently- do not allow the beads to settle.
- Incubate for 10 minutes at room temperature
- Return plate to magnetic stand and wait for solution to clear (~5 minutes)
- Transfer 15ul supernatant to a new well

- **PCR of Adaptor-Ligated DNA**
- Determine which indexes are to be used for each sample and then add the following to each well containing the adaptor-ligated DNA (15ul)
- Add 0.5ul index primer, 0.5ul universal PCR primer, 9ul low TE and 25ul NEBNext Ultra II Q5 Master Mix (NB: primers diluted 1:10 in view of low DNA input as advised by Dr Jayaram to avoid primer dimers)
- Use a multichannel pipette set at 40ul to pipette the mixture up and down at least 10 times to mix thoroughly. Seal and spin to collect fluid from sides
- Place plate in thermocycler with following PCR conditions set-

Temperature	Time	Cycles
98C	30 seconds	1
98C	10 seconds	5
65C	75 seconds	
65C	5 minutes	1
4C	Hold	NA

- **Clean-Up of PCR Amplification**

- Vortex AMPure XP beads to resuspend
- Add 1 x sample volume ie. 50ul of AMPure XP beads to each sample well. Seal plate and vortex to mix well. Spin briefly to collect fluids from sides
- Incubate at room temperature for 10 minutes
- Place plate on magnetic stand and wait for solution to clear
- Remove and discard supernatant
- Add 200ul freshly prepared 80% ethanol to each sample while on the magnetic stand. Incubate at room temperature for 30 seconds then remove and discard the supernatant
- Air dry the beads for up to 5 minutes (do not over-dry)
- Remove plate from magnetic stand and elute the DNA from the beads by adding 33ul low TE
- Seal plate and vortex. Spin down gently- do not allow the beads to settle.
- Incubate for 10 minutes at room temperature
- Return plate to magnetic stand and wait for solution to clear (~5 minutes)
- Transfer 30ul supernatant to a new well
- Store at -20

A3.3 Patient Pathways for ctDNA Cohort

Legend:

TURBT = transurethral resection of bladder tumour

CT CAP = CT chest/abdo/pelvis

PR = partial response

C1, C2 etc = cycle 1, cycle 2 chemotherapy etc

ND = no disease

CRT = chemoradiation

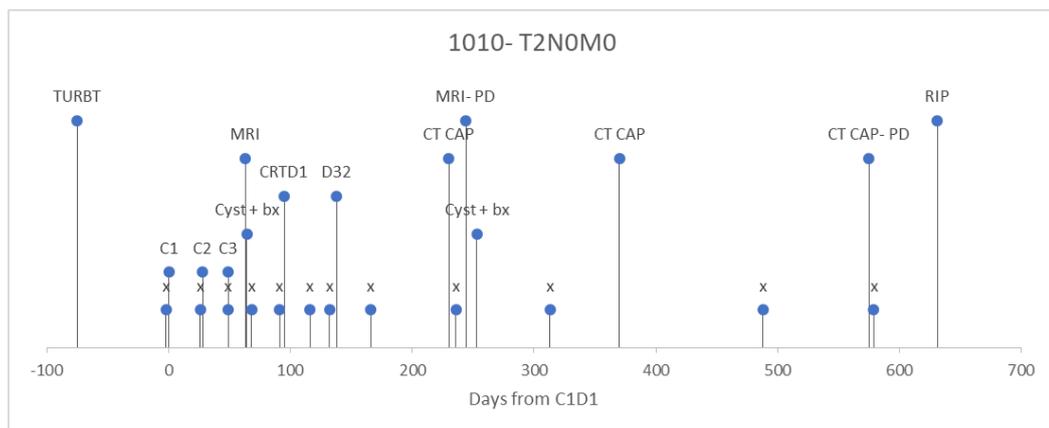
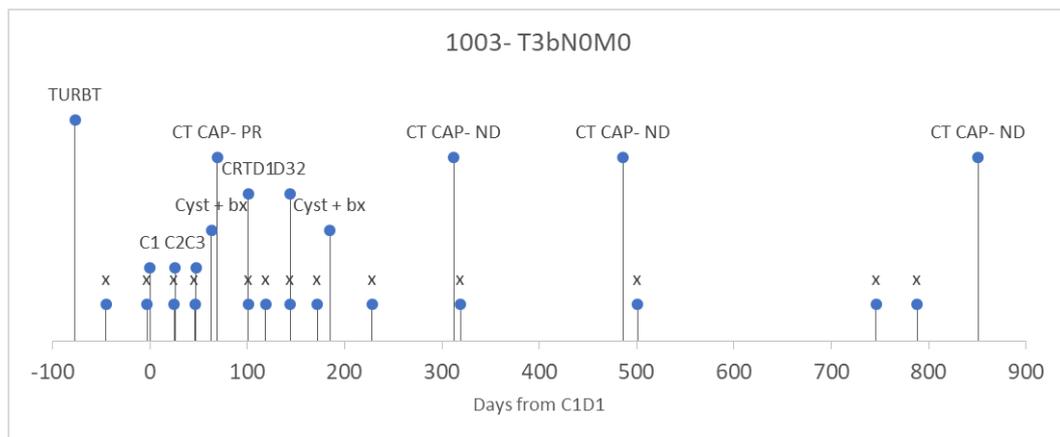
PD = progressive disease

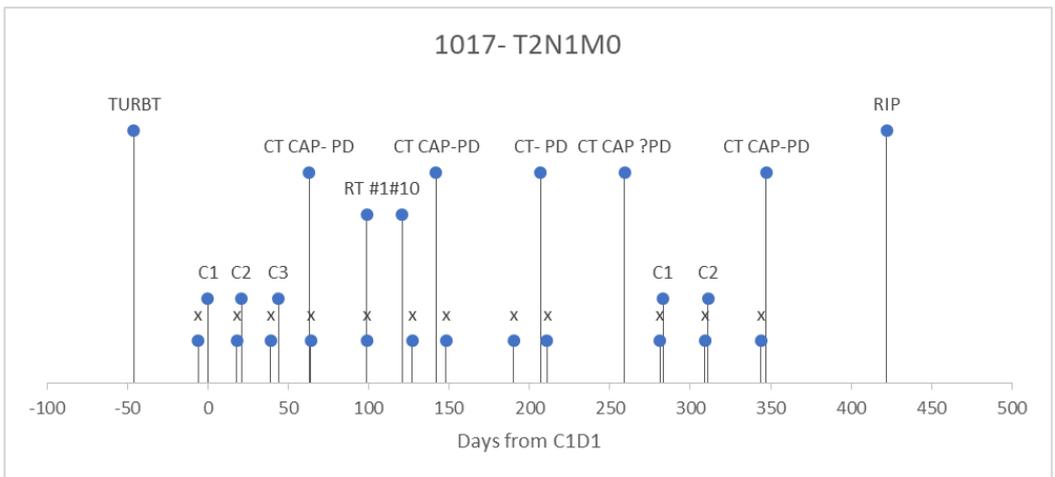
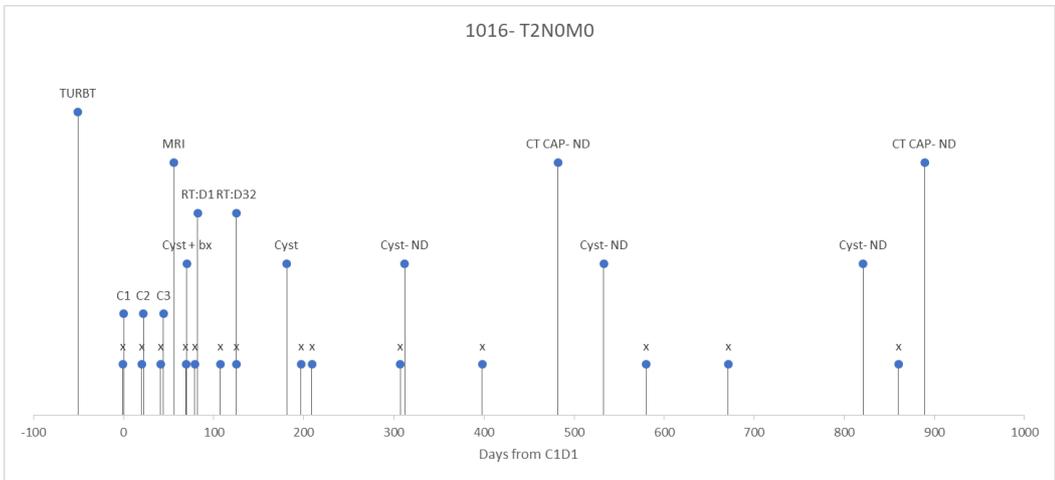
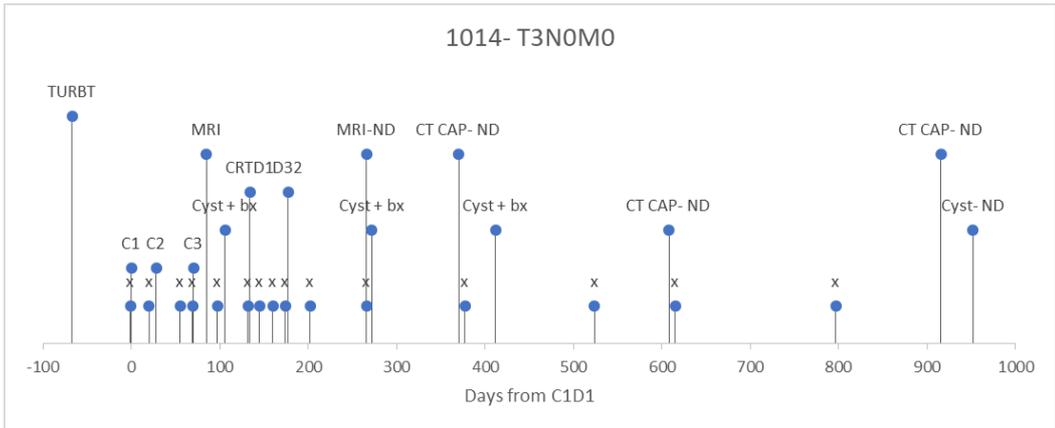
RT = radiotherapy

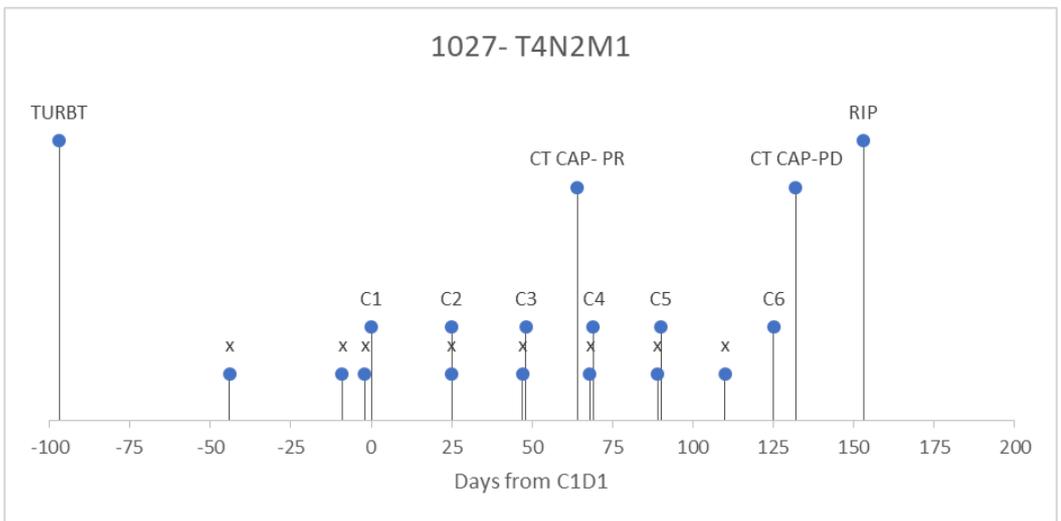
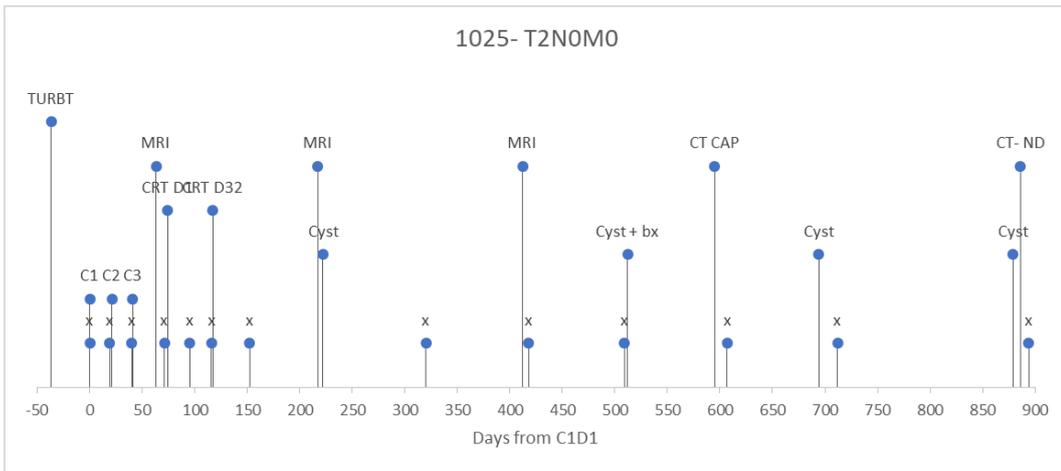
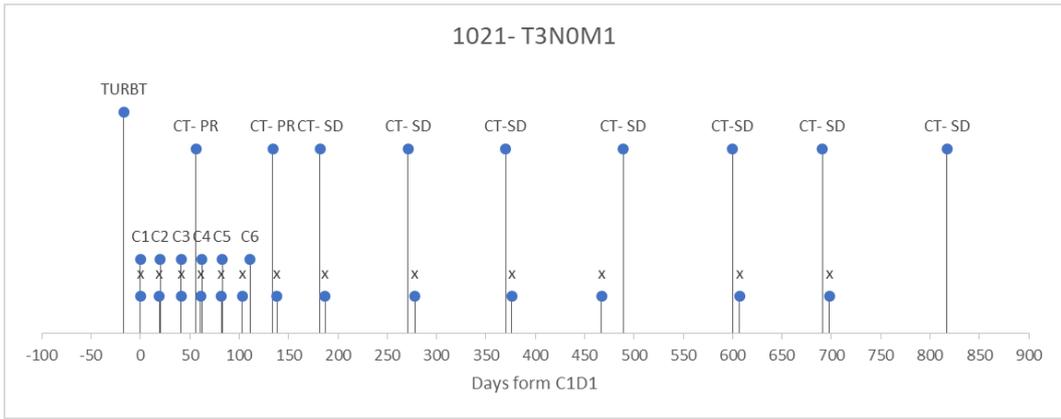
X = blood sample taken

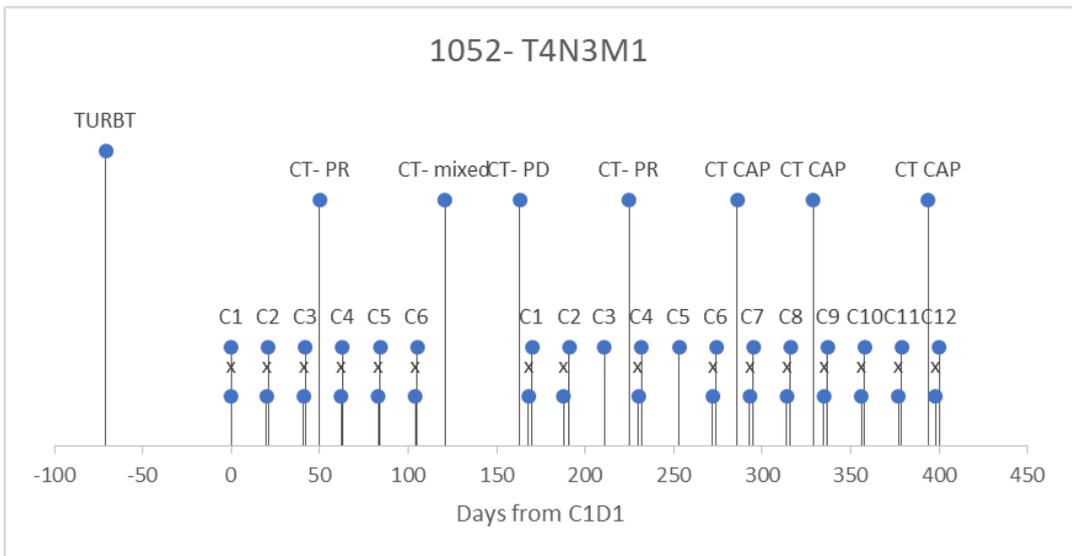
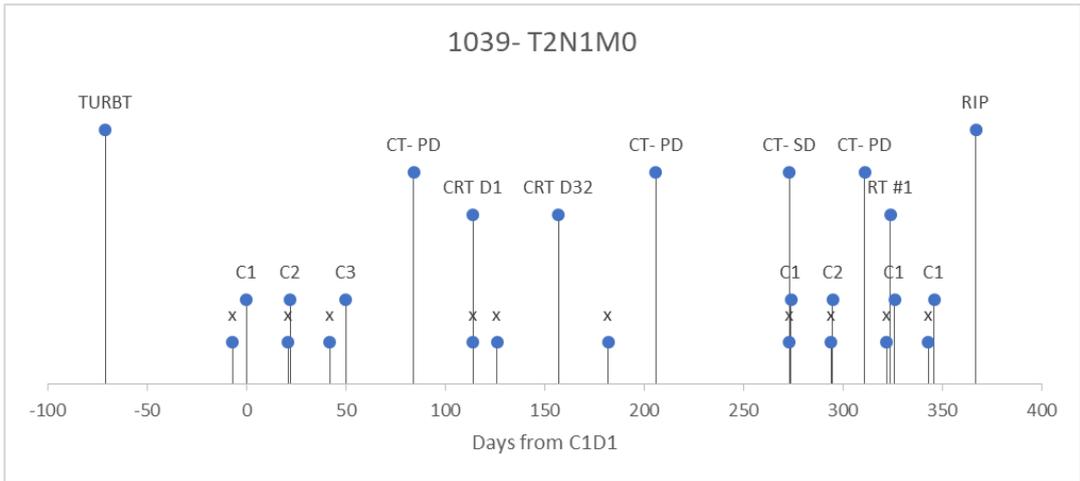
D1 / D32 = day 1 or 32 of RT

Cyst + bx = cystoscopy and biopsy









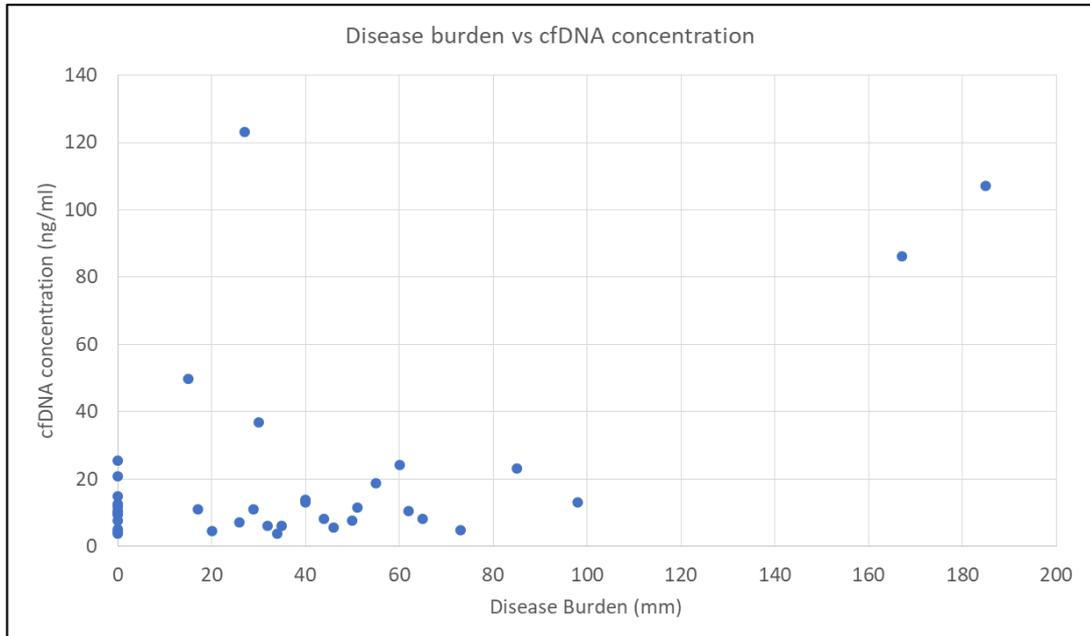
A3.4 Disease Burden Measurements for Plasma DNA concentration cohort

Abbreviations:

NM = nil measurable

ID	CT Date	Lesion 1			Lesion 2			Lesion 3			Lesion 4			Lesion 5			SUM
		Lesion 1	note	mm	Lesion 2	note	mm	Lesion 3	note	mm	Lesion 4	note	mm	Lesion 5	note	mm	
1003	22/07/2015	left bladder mass	report	98	nil			nil			nil			nil			98
1010	16/06/2015	NM		0													0
1014	28/04/2015	NM		0													0
1016	14/01/2015	NM															0
1017	29/10/2015	bladder	report	10	node	report + scan	20										30
1021	12/11/2015	Lung R lower lobe	report	26	lung	sub cm- NM											26
1025	21/04/2015	NM		0													0
1027	04/02/2016	bladder	scan	77	L pelvic side wall	report	28	R pelvic side wall	report	22			23	liver	mri report	17	167
1039	24/06/2016	PET- obturator	report	NM	bladder- NM												0
1052	27/04/2017	bladder	report	40	obturator node- 11	report	0	lung 20/06 ctpa	report	40	lung ctpa 20/06	report	19				99
1001	13/08/2015	mri bladder	report	35													35
1013	22/10/2015	mri bladder	report	40													40
1024	18/01/2016	mri NM		0													0
1028	24/06/2015	mri bladder	report	17													17
1029	21/07/2015	mri bladder	report	46													46
1007	25/09/2015	CT bladder	report	65													65
1012	03/07/2015	CT bladder	scan 121/141	40													40
1008	16/10/2015	MRI bladder	report	51													51
1020	02/12/2015	CT bladder	scan 120/140	29													29
3002	04/09/2015	PET pelvic node	report	30	liver seg met VII	report	20	nodes- NM									50
3003	23/06/2015	CT node	report	40	liver	report	110	liver 72/123	scan	35							185
1044	16/02/2017	PET bladder	report	32	bone- NM												32
1002	24/12/2014	CT bladder	report	35	PA node	report+scan	17	pretrach node	report an	21							73
1011	13/05/2015	PET bladder	scan 228/284	60													60
1018	16/11/2015	CT bladder	report	27													27
1033	11/04/2016	ct/mri difficult to define thickening		0													0
1034	31/03/2016	mri- nil		0													0
1035	07/05/2016	mri bladder	report	15													15
1040	02/08/2016	mri bladder	scan 25/40	55													55
1041	20/10/2016	pet bladder	scan nil to se	0													0
1046	24/03/2017	pet NM (small nodes)		0													0
1049	26/05/2017	mri- nil		0													0
1050	12/06/2017	pet R Cl node	report	15	bladder on pet	scan 123/196	70										85
1051	31/05/2017	ct bladder	report and sc	20													20
1053	09/06/2017	mri bladder	report	34													34
1055	09/08/2017	mri bladder	report	0													0
1056	11/08/2017	mri bladder	report	62	nodes NM												62
1057	07/09/2017	pet bladder	scan 159/196	44													44
1062	09/02/2018	mri bladder	NM	0													0
1067	29/06/2018	ct bladder	report + mri	0													0
2059	28/05/2015	ct liver	report	78	pelvic LN	report	38	LN	report	16	liver	scan 40/15	45				177

A3.5 Scatterplot of plasma DNA concentration vs disease burden with data excluded where imaging not performed within 28 days of blood sample



Scatterplot of tumour burden and plasma DNA concentration with data excluded where imaging not performed ≤ 28 days of blood sample

A3.6 Disease measurement for tumour fraction cohort

Abbreviations: NM = nil measurable

ID	Timepoint	CT Date	Blood	interval	Lesion 1			Lesion 2			Lesion 3			Lesion 4			Lesion 5			SUM
					Lesion 1	note	mm	Lesion 2	note	mm	Lesion 3	note	mm	Lesion 4	note	mm	Lesion 5	note	mm	
1003	45 days pre chemo	11/05/2015	23/06/2015	43	CT pre TURBT. Bld post TURBT															0
1003	baseline	22/07/2015	04/08/2015	13	left bladder mass	report	98	nil			nil			nil			nil			98
1003	c3 assessment	15/10/2015	22/09/2015	-23	NM															0
1010	baseline	19/05/2015	23/06/2015	35	NM															0
1010	c3 assessment- MRI	27/08/2015	13/08/2015	-14	NM															0
1010	PD 1	10/02/2016	16/02/2016	6	NM															0
1010	PD 1 MRI	24/02/2016	16/02/2016	-8	bladder thickening	report	12													0
1010	PD2 ?	20/01/2017	24/01/2017	4	NM- bone met															0
1014	Baseline	28/04/2015	05/05/2015	7	NM															0
1014	c3 assessment	30/07/2015	14/07/2015	-16	NM															0
1016	Baseline	14/01/2015	20/01/2015	6	NM															0
1016	c3 assessment	18/03/2015	03/03/2015	-15	NM															0
1017	Baseline	29/10/2015	24/11/2015	26	bladder thickening	report	10	node	report + scan	20										30
1017	c3 assessment- PD1	01/02/2016	08/01/2016	-24	right bladder wall	112/124	33	node	report + scan	34										67
1017	PD 2	20/04/2016	26/04/2016	6	right bladder wall	report	41	node	report	46										87
1017	Pre C1 pacli	15/08/2016	06/09/2016	22	right bladder wall	report	35	node	report	41	breast lesion	report	22	R paratrach	report	16	mesenteric node	report	25	139
1017	post pacli. Pre RIP	11/11/2016	08/11/2016	-3	subcut nodule	report	31	node	report + scan	40	breast lesion	report	33	R paratrach	report	21	mesenteric node	report	47	172
1021	baseline	12/11/2015	10/12/2015	28	Lung R lower lobe	report	26	lung	sub cm- NM											26
1021	c3 assessment	04/02/2016	20/01/2016	-15	Lung R lower lobe	report	15	lung	sub cm- NM											15
1021	c6 assessment	22/04/2016	26/04/2016	4	Lung R lower lobe	report	10	lung	sub cm- NM											10
1021	SD D698	31/10/2017	07/11/2017	7	NM															0
1025	baseline	21/04/2015	23/04/2015	2	NM															0
1025	c3 assessment	25/06/2015	02/06/2015	-23	NM															0
1027	baseline 44/7 pre C1	04/02/2016	02/02/2016	-2	bladder	scan	77	L pelvic side wall	report	28	R pelvic side wall	report	22			23	liver segment 4	report	17	167
1027	baseline pre C1	no scan	15/03/2016	NA				L pelvic side wall			R pelvic side wall	report								NA
1027	c3 assessment	20/05/2016	03/05/2016	-17	bladder		59	L pelvic side wall	report	20	R pelvic side wall	report	17	Liver segment 8	report	26	liver		18	140
1027	pre C6- PD	27/07/2016	05/07/2016	-22	bladder	scan	59	L pelvic side wall	report	21	R pelvic side wall	report	22	liver segment 8	report	41	liver		18	161
1039	baseline	24/06/2016	05/07/2016	11	PET- obturator	report	NM	bladder thickening-	NM											0
1039	c3 assessment	26/08/2016	23/08/2016	-3	obturator	report	NM	ext iliac node	report	NM										0
1039	Post CRT	03/02/2017	10/01/2017	-24	node	report	NM	RP node	report	24										24
1039	Pre pembro	11/04/2017	11/04/2017	0	RP node	report	24	NM												24
1039	post c2	19/05/2017	20/06/2017	32	RP node	report	24	adrenal	report	31	lytic bone- NM			MRI 28.06- liver mets, pleural mass- NM						55
1052	baseline	27/04/2017	14/06/2017	48	bladder	report	40	I obturator node	report	NM	lung bx 22/32	scan	81	lung bx 1/32	scan	49				170
1052	c3 assessment	03/08/2017	25/07/2017	-9	bladder	scan	35	I obturator node	report	NM	lung	report	25	lung	report	15				75
1052	c6 assessment	13/10/2017	26/09/2017	-17	bladder	report	0	I obturator node	report	NM	lung	report	17	lung	scan- existing measurement	26				43
1052	pre c1 pembro	24/11/2017	29/11/2017	5	bladder	scan 178/2	28	mediastinal nodes	scan 52/206	15	lung	report	46	lung	report	37	PA node - NM			126
1052	pre c4 pembro	25/01/2018	30/01/2018	5	NM															0

A3.7 Exploring alternative thresholds to define 'high' and 'low' plasma DNA subgroups

To explore alternative thresholds to dichotomise the cohort, I first divided the cohort into quartiles. Figure A3.7 shows the Kaplan-Meier curves for overall survival for the 4 subgroups.

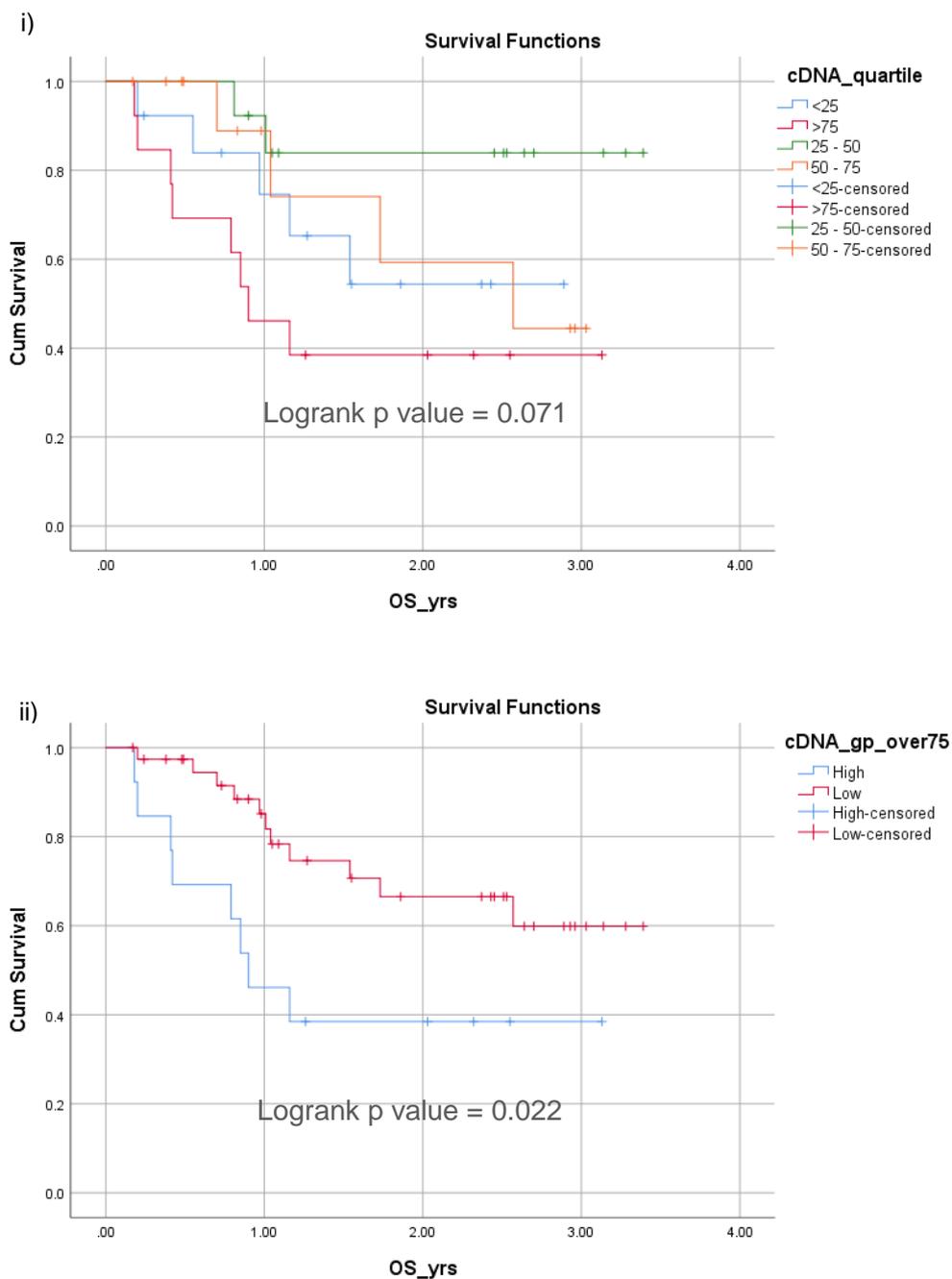


Figure A3.7 Kaplan-Meier curves demonstrating overall survival with the cohort divided according to i) plasma DNA quartile, and ii) using the 75th percentile plasma DNA value (16.3ng/ml) as the defining threshold.

There was no statistically significant difference between the groups (logrank p value = 0.071) when divided into 4 quartiles, but perhaps some suggestion that the group in the upper-most quartile had a poorer outcome compared to the others. I therefore then proceeded to use the 75th percentile value (16.3ng/ml) to divide the cohort into 'high' and 'low plasma DNA subgroups (figure A3.7ii). There was a statistically significant difference in overall survival (p = 0.022) and this was also true for bladder cancer-specific survival (p = 0.012).

Continuing to use the 75th percentile value as the cut-off to define 'high' and 'low' and so treating the plasma DNA concentration as a categorical variable, I performed univariate and multivariate Cox regression analysis.

On univariate analysis, the hazard ratio for overall risk of death was 2.784 (95% CI 1.113-6.962; p=0.029) and for bladder cancer-specific death was 3.431 (95% CI 1.233 – 9.546; p = 0.018). However, on multivariate analysis with plasma DNA subgroup and M-stage status as covariates, only the hazard ratio for bladder cancer-specific death retained statistical significance. The results are summarised in table A3.7i below

Covariate	Risk of Death			Risk of bladder cancer-specific death		
	HR	95% CI	p-value	HR	95% CI	p-value
Plasma DNA group	2.45	0.960-6.250	0.061	2.974	1.033-8.565	0.043
M-stage	2.968	1.17-7.529	0.022	4.757	1.68-13.465	0.003

Table A3.7i. Multivariate Cox regression with plasma DNA group (high/low) and M-stage as covariates

A3.8 Inferring ploidy from WES data

Sample	ASCAT		SEQUENZA			PureCN	
	Cellularity	Ploidy	Cellularity	Ploidy (Estimate)	Ploidy (mean)	Cellularity	Ploidy
1003_1A	1	2.81	0.83	5.3	3.29	0.58	3.26
1003_2C	DID NOT FIT		0.48	7	4.4	0.59	2.64
1010_1A	0.53	2.82	0.42	4.1	2.85	0.43	2.96
1010_1C	0.6	4.25	0.33	3.4	2.26	0.21	3.25
1014_1B	0.6	2.24	0.56	3.5	2.54	0.59	2.45
1014_2B	0.34	2.36	0.31	3.7	2.4	0.32	2.41
1016_1B	0.52	1.61	0.39	2.4	2.07	0.32	2.09
1016_1D (1C)	0.86	3.55	0.36	2.5	2.09	0.32	2.11
1017_2B	DID NOT FIT		0.33	3.2	2.51	0.32	2.85
1017_3C	DID NOT FIT		0.55	3.9	3.27	0.59	2.36
1021_1A	0.68	3.62	0.63	3.7	3.44	0.58	3.39
1021_2A	0.46	3.24	0.49	2.2	1.97	0.35	4.32
1025_1A	0.77	2.74	0.72	4.8	3.18	0.39	2.77
1025_1B	0.38	4.11	0.35	5.1	2.34	0.32	3.33
1039_1A	DID NOT FIT		0.82	5	3.34	0.52	2.91
1039_2B	DID NOT FIT		0.58	5.5	3.43	0.45	2.51
1016 1A	0.37	4	0.67		2	0.32	2
1016 1C	DID NOT FIT		0.46		2	0.21	2
1027 1A	0.5	2.5/3.9	0.86		3.5	0.85	3.5
1027 2A	DID NOT FIT		0.85		3.5	0.82	3.5

Chapter 4. Exploring the Use of Multiregion Whole Exome-Sequencing of FFPE Tumour Tissue and Publicly Available Data to Identify Plasma ctDNA Targets in MIBC

4.1. Introduction

4.1.1. Overview

Circulating tumour DNA (ctDNA) is a valuable tool in cancer research and is of great interest in MIBC. A key challenge remains that of identifying and quantifying the ctDNA component in plasma.

In chapter 3, I demonstrated that in patients with a higher disease burden, tumour fraction may be estimated from LP-WGS data using the iChorCNA pipeline. However, this is not sensitive enough to be applicable to patients with lower burdens of disease, including those with organ-confined disease only.

In this chapter, I will explore an alternative approach to interrogating plasma ctDNA which should be applicable to those patients with low burdens of disease. Firstly, I aim to assess the feasibility of performing high depth multiregion whole exome sequencing on DNA extracted from formalin-fixed paraffin-embedded (FFPE) tumour tissue with a view to identifying both dominant and sub-clonal aberrations to track in plasma. This data will additionally provide some insight into intra-tumour heterogeneity in MIBC. The second approach will use the whole exome data in combination with data from a commercially available panel and publicly available data to design a targeted panel for use in MIBC to interrogate plasma ctDNA samples. The targeted panel will aim to interrogate clinically relevant and potentially actionable

targets in plasma, and to capture as large a proportion of MIBC patients as possible by the inclusion of targets relevant to MIBC.

4.1.2. Using a commercially available panel

In September 2017, the opportunity arose to send 30 samples to Foundation Medicine for genomic profiling using their FoundationOne panel, courtesy of Roche.

The FoundationOne panel consists of 236 cancer-related genes and 47 introns. Results from a cohort of 295 patients with stage III/IV MIBC¹ showed that 294/295 had at least one clinically relevant genomic alteration (defined as aberrations linked to drugs on the market or in trials). The most common aberrations detected were *CDKN2A* (34%), *FGFR3* (21% including fusions), *PIK3CA* (20%), *ERBB2* (17%; 50:50 amplification and substitution).

An appealing feature of the FoundationOne panel was the inclusion of the *TERT* promoter, a region that was not readily encompassable within the whole exome sequencing planned. The results from sequencing using this commercially available panel will also be reported in this chapter, and will additionally support the design of a targeted panel.

4.2. Hypotheses and Aims

4.2.1. Hypotheses

Whole-exome sequencing at 500x coverage is achievable on FFPE-derived tumour DNA from patients with MIBC, and will identify a broad spectrum of aberrations, including subclonal mutations at low abundance

Comparison of whole-exome data from separate regions from each patient will give some insight into intra-tumour heterogeneity

Aberrations identified using a commercial targeted sequencing panel may be associated with clinical outcomes

Aberrations identified via whole-exome sequencing and targeted sequencing can be used in combination with publicly available data to design a targeted NGS panel to sequence plasma ctDNA.

4.3. Aims

To macrodissect and extract DNA from multiple regions of tumour from diagnostic FFPE blocks for a total of 10 patients with MIBC

To prepare libraries for WES from a minimum of 2 regions per patient, and achieve a target depth of 500x

To analyse the WES data and describe the range of aberrations identified, with focus on examples of intra-tumour heterogeneity, and compare against publicly available data

To analyse aberrations reported by a commercially available targeted sequencing panel with regards to clinical outcomes, and compare results with those from WES data

To use data generated and publicly available data to design a targeted NGS panel to interrogate genes of interest in MIBC, which may be applied to plasma DNA

4.4. Materials and Methods

4.4.1. Cohort selection

4.4.1.1. WES cohort

The cohort of 10 patients used in the tumour fraction pilot (Chapter 3; see 3.4.1.2) were used in this study. To summarise, the following criteria was used-

- Group 1 patients ie. treatment-naïve
- Received neoadjuvant or palliative platinum-based chemotherapy

- Availability of FFPE block from TURBT- presence of at least 2 regions of tumour suitable for macrodissection on histological assessment
- Availability of plasma samples- pre-treatment, before each cycle, post-completion of chemotherapy, and if applicable, at disease progression.

The cohort consisted of 5 patients with disease relapse and 5 with no relapse at the time of data analysis (median follow-up 3.12 years). Disease relapse was defined as the recurrence or progression of invasive disease during or after chemotherapy treatment.

4.4.1.2. FoundationOne cohort

FFPE blocks were sent for testing using the FoundationOne panel from a total of 30 patients with MIBC. Ten of the thirty were from the WES cohort mentioned above. The remaining 20 patients were selected from the RNA molecular subtyping cohort (see chapter 5), with a focus on those with disease relapse.

4.4.2. Processing of samples

4.4.2.1. Histological assessment

In order to inform the best way to proceed with regards to extracting DNA from FFPE, an initial batch of 26 FFPE blocks from 23 patients had a single 2-3um H&E slide prepared by the Histopathology team at the RMH Centre for Molecular Pathology (CMP) in July 2016. These slides were assessed by Dr Steve Hazell to confirm tumour content and the number of tumour regions present. Those with tumour content of at least 60% and at least 2 tumour regions were then returned to CMP for serial sectioning of 15 slides of 10um thickness, and a final H&E slide. The final H&E slides were again assessed by Dr Hazell to ensure presence of tumour throughout the sections and no significant change in morphology of the marked tumour regions. The results are shown in Appendix A4.1.

Further FFPE blocks were sent to the Breast Cancer Now Histopathology Core Facility at the ICR Chelsea in 2017/18 on my return from maternity leave, as

the RMH CMP no longer had the capacity to support this work. Following successful test DNA extractions (see Appendix A4.2), 10 serial sections of 10um thickness were requested with H&E slides at the beginning and end i.e. slide 0 and slide 11. All slides were assessed for presence of muscle-invasive tumour by Dr Steve Hazell, with tumour areas marked on the H&E slides.

Figure 4.1 illustrates the work flow for multiregion WES.

4.4.2.2. DNA extraction

The full protocol is provided in Appendix A4.3. In summary, slides were deparaffinised by submerging them in xylene for up to 1 minute or until the paraffin was visibly dissolved. Slides were then submerged in 100% alcohol for 1 minute and ddH₂O for a minimum of 1 minute.

Using the H&E slides as a guide, regions were macrodissected separately using a 16G needle into 1.5ml LoBind Eppendorfs containing Zymo digestion mix. All work was performed in a fume cupboard due to the use of xylene.

On completion, the Eppendorfs containing tissue and digestion mix were incubated overnight at 55C as per the protocol for 12-16 hours. DNA extraction was performed the following morning. A key amendment advised by Dr Inma Spiteri was to retain the initial filtrate and pass this through the same column once the protocol had been completed. This was to ensure any DNA in the filtrate was also captured. The test extractions confirmed that there was additional DNA to be captured (as documented in Appendix A4.2), and so this amendment was performed for all samples. Samples were quantified using Qubit analyser and stored at -20C.

4.4.2.3. Germline DNA extraction

Germline DNA was extracted from patient-matched buffy coat samples. The full protocol is provided in the Appendix A4.4. The Qiagen Mini kit was used and the manufacturer's protocol followed.

4.4.2.4. Whole exome library generation

Whole exome libraries were generated from DNA extracted from 2 tumour regions and germline DNA for each patient. Each patient (n=10) therefore had a total of 3 libraries prepared.

The Agilent Sureselect XT HS kit was selected for use but due to a 2-month supply delay, an initial trial using samples from 2 patients with the Sureselect XT2 kit (pooled capture) was performed with Dr Spiteri. Exome capture was unsuccessful, most likely due to uneven pooling of samples (see A4.5 for protocol overview and results).

The Sureselect XT HS protocol is provided in Appendix A4.6. Libraries from tumour and germline DNA were prepared according to the manufacturer's instructions.

The input DNA used for all samples (tumour and germline) was 210ng in 51ul. Any volume lacking was made up with low TE and volumes were concentrated as required. While the maximum DNA input for the XT HS protocol is 200ng in 50ul, the extra 10ng and 1ul was added to allow for loss of material and volume during sonication and sample transfer.

A preliminary batch of 2 tumour and 1 germline DNA sample from 2 patients i.e. total 6 samples were prepared using the XT HS protocol. Libraries were prepared by me with supervision from Dr Anu Jayaram and Ania Wingate. The bioanalyser traces for the whole exome libraries are shown in Appendix A4.7. Samples were sent to TPU for WES in early January 2018 and results received mid-end February 2018.

As initial results were satisfactory, the decision was made to proceed with the remaining samples. However, it was agreed to use a FFPE repair kit and fewer amplification cycles (given high gDNA library yields achieved) to try and reduce the PCR duplication rate to 25%. The target PCR duplication rate of 25% was determined following discussions with Prof Attard's team and Dr Spiteri who has significant experience in working with FFPE DNA.

Whole exome libraries from the second batch of 8 patients were prepared independently by me. There was insufficient gDNA library to proceed to capture for patient 1052. Exploration suggested that this may be due to high levels of degradation of the DNA, perhaps due to diathermy effect which had been noted on histological review. Attempts to repeat this library were felt unlikely to be successful. Additional libraries from patient 1016 were therefore prepared to complete the flowcell and allow the second batch of samples to be sequenced. Therefore a total of 16 tumour and 7 germline samples from 8 patients were sent in the second batch to TPU.

All sequencing was performed by TPU. Libraries were loaded on to Nova S2 flowcells (v1) and sequenced using PE 100 cycles on the Illumina Novaseq platform.

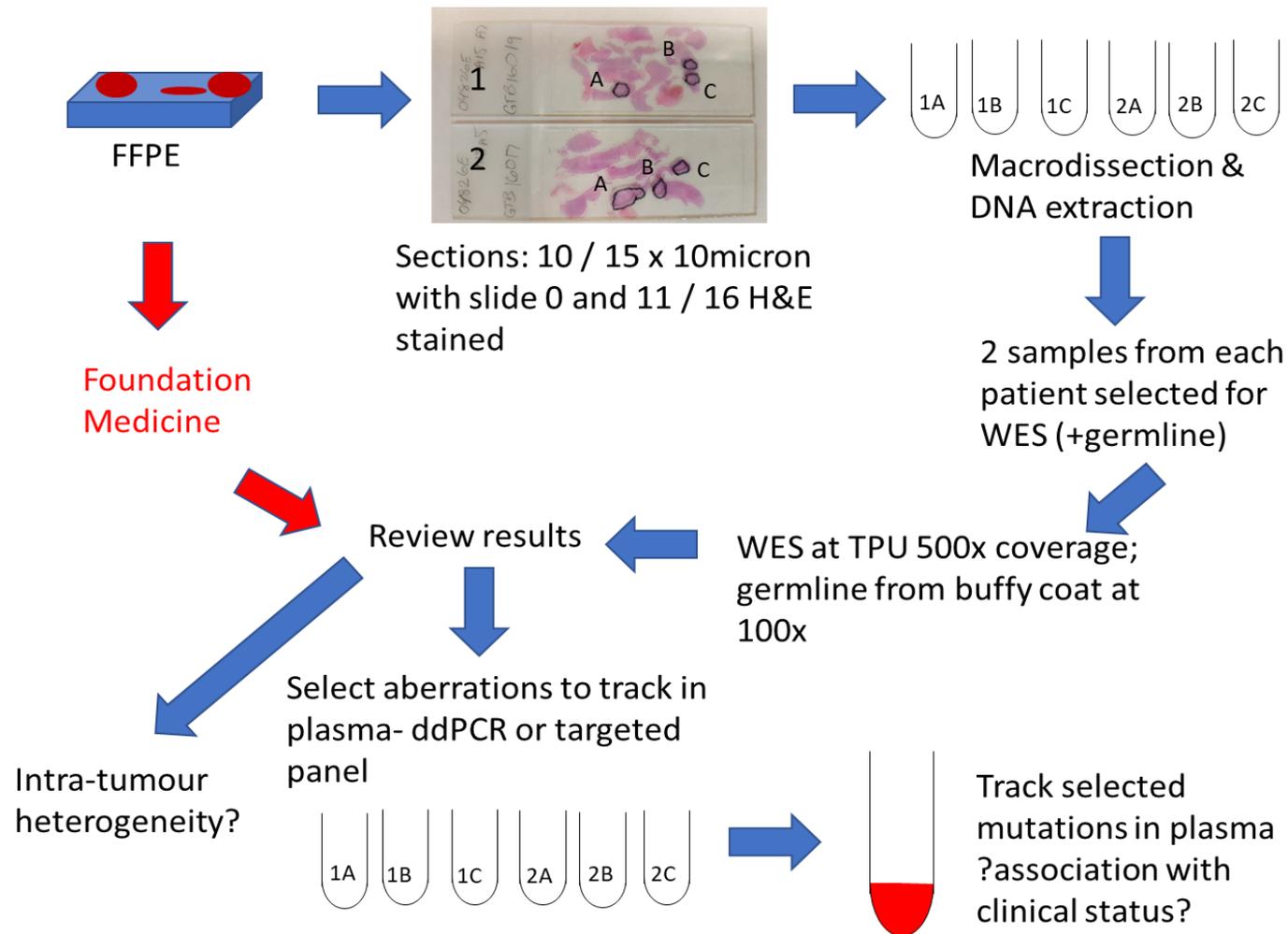


Figure 4.1 Anticipated workflow for multiregion whole exome sequencing

4.4.3. Foundation Medicine FoundationOne Panel

FFPE blocks from 30 patients were sent for testing. One block failed and a second block from that same patient was sent and successfully tested. On histological assessment, one patient was reported to have carcinosarcoma and not TCC as recorded in clinical documents. This patient was therefore excluded from data analysis meaning the final cohort was 29. The FoundationOne reports listed aberrations considered clinically significant or 'reportable' in the context of cancer with respect to data in the current literature, and reports also listed variants of unknown significance. On requesting the criteria for 'reportable' aberrations, Foundation Medicine have responded that

“databases such as COSMIC as used to determine reportability of variants as well a periodic review of the published medical literature by the Foundation Medicine team. Variants that appear on commercial clinical reports have sufficient evidence supporting oncogenicity based on internal guidelines defined in FMI’s variant reporting pipeline”.

(Email correspondence from Roche Medical Information)

No further precise criteria were made available. A data transfer agreement for the VCF files was drawn up and forwarded to the ICR legal team for evaluation. Unfortunately this did not come through in the timescale of this project and so analysis has been performed on the reports provided, and email correspondence with the Foundation Medicine bioinformatic team for any queries.

The full FoundationOne gene list is provided in Appendix A4.8.

4.4.4.Data analysis

4.4.4.1. Bioinformatics

Initial bioinformatics analysis for the whole exome data was performed by Ritika Chauhan, Alistair Rust and Pradeep Ramagiri in the TPU Bioinformatics Team. The reports generated stated that BWA (version 0.7.5a) was used to align reads to the GRCh37 human reference genome. PCR duplicates were removed before further processing using PicardTools. Variant calling was performed using MuTect (version 1.1.4) and the Genome Analysis Tool Kit (GATK; version 2.7-2). Excel spreadsheets and a word document report were provided detailing the aberrations detected, genomic location, predicted functional impact (as assessed using ANNOVAR), coverage and number of alternative reads (variant allele frequency). The raw data files were also transferred to me. I used IGV (integrative genome viewer) to review aberrations of interest.

The GATK-called aberrations were filtered according to TPU standard criteria which selected those where tumour VAF was >40% and <60% with the normal VAF of <20% or >80%, and using the same cut offs but with tumour and normal switched.

Copy number alterations were reported in an excel spreadsheet as regions of gain, loss, amplification or deletion. The TPU team also provided Circos plots summarising the copy number data.

Any genes of particular interest or aberrations where there was a discrepancy between that reported by TPU or the FoundationOne panel were explored by me using IGV.

Limited additional bioinformatician support meant that I was unable to use the MutSig 2CV and GISTIC pipelines as anticipated to identify statistically significant aberrations and further assess copy number changes within the timeframe of this project. Therefore, to filter out artefact and identify likely significant aberrations, I compiled a list of 64 genes comprising the 58 significantly mutated genes reported by the TCGA and additional DNA damage repair genes of interest highlighted in the literature^{2,3} (Appendix A4.9).

I focussed my analysis on this subset of core genes. Synonymous variants were excluded. Copy number analysis was based on the reports provided by TPU.

4.4.4.2. Statistical analysis

Whole-exome sequencing cohort (n = 9)

The most commonly altered genes and specific aberrations were described in terms of frequency and also those affecting the greatest proportion of patients. The cohort was too small to perform any formal statistical analysis of whether the presence of an aberration was associated with an outcome.

FoundationOne cohort (n=29)

The individual reports for each patient were collated together for analysis, and the range of aberrations described for all those reported. Analysis was performed to include all reported aberrations and just those of clinical significance (as labelled by the FoundationOne report).

Where an altered gene or aberration was present in 30% or more of the FoundationOne cohort, statistical comparison was made using a Fisher's exact test and/or Kaplan Meier analysis with logrank assessment to explore whether the aberration/gene status was associated with clinical outcome.

Clinical endpoints for this cohort included:

- Overall survival- defined as time from 1st treatment to death or censored at date last known alive
- Bladder cancer-specific survival- defined as time from 1st treatment to death from bladder cancer; data censored at death from any non-bladder cancer cause or date last known alive for surviving patients
- Progression-free survival- defined as time from 1st treatment to disease progression or death; data censored at date last known alive for those surviving patients with no evidence of disease relapse
- Locoregional relapse free survival- defined as time from 1st treatment to locoregional relapse (recurrence of disease in the bladder or regional lymph nodes, including superficial bladder cancer); data censored at

death from non-bladder cancer cause, metastatic disease preceding locoregional relapse by 3 months or more or date last known alive in those with no locoregional relapse.

Locoregional relapse free survival was explored in the subset of patients that received radical radiotherapy (n = 25).

The cohort was too small to perform any formal comparison of survival outcomes between subgroups but Kaplan-Meier analysis was performed to report median OS with 95% CI and 2-year survival rates for subgroups.

To explore the potential association of DNA damage repair (DDR) genes with outcome, the cohort was additionally divided into those with or without an alteration in *ATM*, *FANCD2*, *PALB2*, *BRCA1* and *BRCA2*. These genes were selected based on published work² which suggested MIBC patients with aberrations in these genes had improved clinical outcomes. Of note, their panel also included *ERCC2* which was not targeted on the FoundationOne panel.

4.5. Results

4.5.1. FoundationOne Panel

4.5.1.1. FoundationOne cohort

Table 4.1 overleaf summarises the patient characteristics for the cohort sent for profiling using Foundation Medicine's FoundationOne panel. As previously mentioned, one patient was excluded as they did not have TCC on histological review leaving a final cohort of 29 patients. Median follow-up was 3.0 years (95% CI 2.9 – 3.0). Minimum follow up was 1.09 years.

With regards to relapse patterns, the cohort included 18 patients with relapsed disease. A total of 9 patients had locoregional recurrence (+/- M1 disease), and of these patients, only 6 had invasive locoregional recurrence. The median progression free survival was 1.02 years (95% CI 0.68 – 1.36).

The median overall survival for the group overall was 2.8 years (95% CI 1.2 – 4.4). Of note, all deaths were due to bladder cancer.

Table 4.1. Patient Characteristics of FoundationOne cohort.
Abbreviations: NAC = neoadjuvant chemotherapy; CRT = chemoradiation; NMIBC = non muscle-invasive bladder cancer

		No relapse N = 11	Relapse N = 18	Total N = 29
Age	<i>Median (range)</i>	72.7 (46.1-81.5)	76.42 (61.3-90.9)	
Gender	<i>Male</i>	9	13	23
	<i>Female</i>	2	5	7
Staging	<i>T2N0M0</i>	4	5	9
	<i>T3N0M0</i>	3	6	10
	<i>T4N0M0</i>	0	0	0
	<i>Any T, N1-3, M0</i>	2	4	6
	<i>Any T, any N, M1</i>	2	3	5
Treatment	<i>NAC</i>	10	10	20
	<i>CRT</i>	10	15	25
	<i>Palliative chemo</i>	1	2	3

Table 4.1 (continued)

		No relapse N = 11	Relapse N = 18	Total N = 29
Pattern of relapse	<i>NMIBC only</i>	0	1	1
	<i>Locoregional recurrence (+/- NMIBC)</i>	0	4	4
	<i>LRR + M1</i>	0	4	4
	<i>M1 only</i>	0	9	9
	<i>Alive at date of analysis</i>	11 (range 1.1 – 4.0 yrs; median 2.96)	2	13

4.5.1.2. Overview of aberrations reported by FoundationOne panel

Within the cohort of 29 patients, a total of 708 aberrations were reported across 228 genes. Of these, 264 aberrations across 80 genes were considered to be of clinical significance, with the remaining 444 labelled as variants of uncertain significance (VUS). All patients had at least 4 clinically significant aberrations reported, with an average of 7 and range of 4 – 17.

When considering all 708 aberrations reported, those affecting the most patients involved the *TERT* promoter and *TP53*. These 2 genes were each altered in 69% of the cohort. Aberrations in *ATM*, *ERBB2* and *ERBB3* were seen in 35% of patients. These results are further summarised in Table 4.2.

Of note, the range of median exon coverage for samples sent was 227 – 975x with a median of 594 (IQR 405 – 689).

Table 4.2 Most prevalent aberrant genes affecting the cohort as reported by FoundationOne panel (clinically significant and VUS)

Aberrant Gene	No. of patients with relapse (n=18)	No. of patients with no relapse (n=11)	Total no. of patients affected (%)
<i>TERT</i> promoter	13 (72%)	7 (64%)	20 (69%)
<i>TP53</i>	15 (83%)	5 (45%)	20 (69%)
<i>ATM</i>	5 (28%)	5 (45%)	10 (35%)
<i>ERBB3</i>	5 (28%)	5 (45%)	10 (35%)
<i>ERBB2</i>	7 (39%)	3 (27%)	10 (35%)
<i>MLL2</i>	5 (28%)	4 (36%)	9 (31%)
<i>RB1</i>	5 (28%)	4 (36%)	9 (31%)
<i>KDM6A</i>	8 (44%)	1 (9%)	9 (31%)
<i>FGFR3</i>	5 (28%)	3 (27%)	8 (28%)

When focussing on only those aberrations considered clinically significant, *TP53* and the *TERT* promoter alterations were present again in 69%. This time however, the next most commonly altered genes were *KDM6A* in 31% and *RB1* in 27.6%. Table 4.3 below summarises further.

Table 4.3 Most prevalent aberrant genes affecting the cohort as reported by the FoundationOne panel (clinically significant only)

Aberrant Gene	No. of patients with relapse (n=18)	No. of patients with no relapse (n=11)	Total no. of patients affected (%)
<i>TP53</i>	15 (83%)	5 (45%)	20 (69%)
<i>TERT</i> promoter	13 (72%)	7 (64%)	20 (69%)
<i>KDM6A</i>	8 (44%)	1 (9%)	9 (31%)
<i>RB1</i>	4 (22%)	4 (26%)	8 (28%)
<i>ERBB2</i>	5 (28%)	2 (18%)	7 (24%)
<i>ARID1A</i>	5 (28%)	2 (18%)	7 (24%)
<i>PIK3CA</i>	3 (17%)	3 (27%)	6 (21%)
<i>RAF1</i>	3 (17%)	2 (18%)	5 (17%)
<i>CDKN2B</i>	4 (22%)	1 (9%)	5 (17%)
<i>MLL2</i>	2 (11%)	3 (27%)	5 (17%)
<i>CDKN2A</i>	4 (22%)	1 (9%)	5 (17%)

When considering specific mutations, the most common were 124C>T in the *TERT* promoter (17/29), loss of *CDKN2A/B* (5/29), *ERBB2* amplification (5/29), *RAF1* amplification (5/29), and 146C>T in the *TERT* promoter 146 C>T (4/29; one of whom also had the 124C>T alteration). Amplifications were also seen in *SDHC* (5/29) and *DDR2* (4/29) but these were reported as of uncertain significance. The *TP53* Q331* aberration was reported in 3/29.

4.5.1.4. Exploring association of aberration with clinical outcome

Figure 4.2 below shows the aberrations considered clinically significant reported in at least 3 patients within the cohort (>10%).

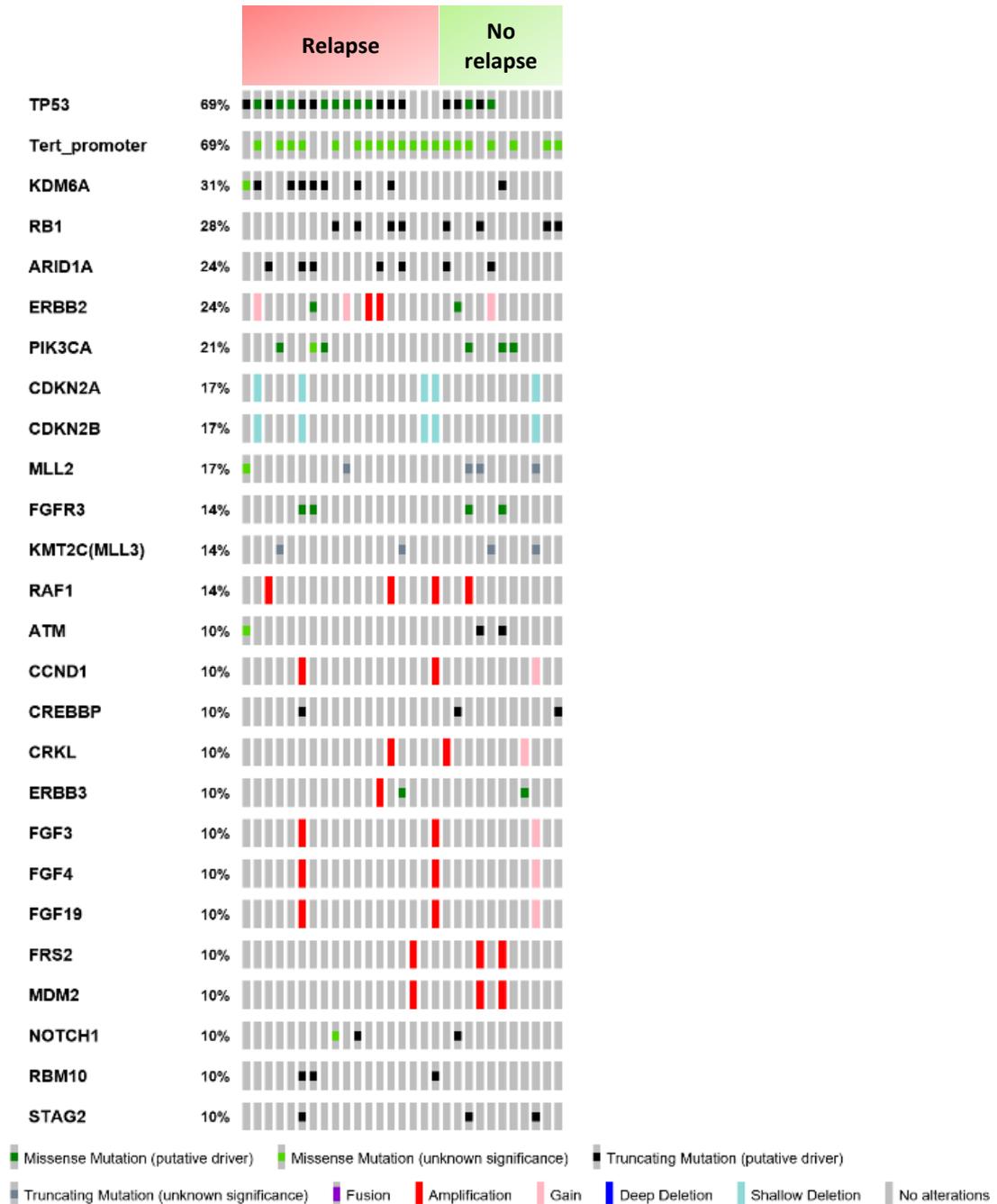


Figure 4.2 Oncoprint figure illustrating aberrations of clinical significance identified by FoundationOne panel in at least 3/29 patients (produced using online tool at cBioportal^{4,5})

Survival outcomes

TP53, *TERT* promoter and *KDM6A* aberrations were present in at least 30% of the cohort. Exploration of mutation status and outcomes for each of these genes was therefore performed. There was a suggestion that patients with mutant *TP53* or *KDM6A* have poorer outcomes. Figure 4.3 shows the Kaplan-Meier curves for progression-free survival and overall survival according to those with or without a a) *TP53* aberration, b) *TERT* promoter mutation, and c) *KDM6A* mutation. Table 4.4 below shows the median and 2-year overall survivals and progression-free survivals for the same patient groupings.

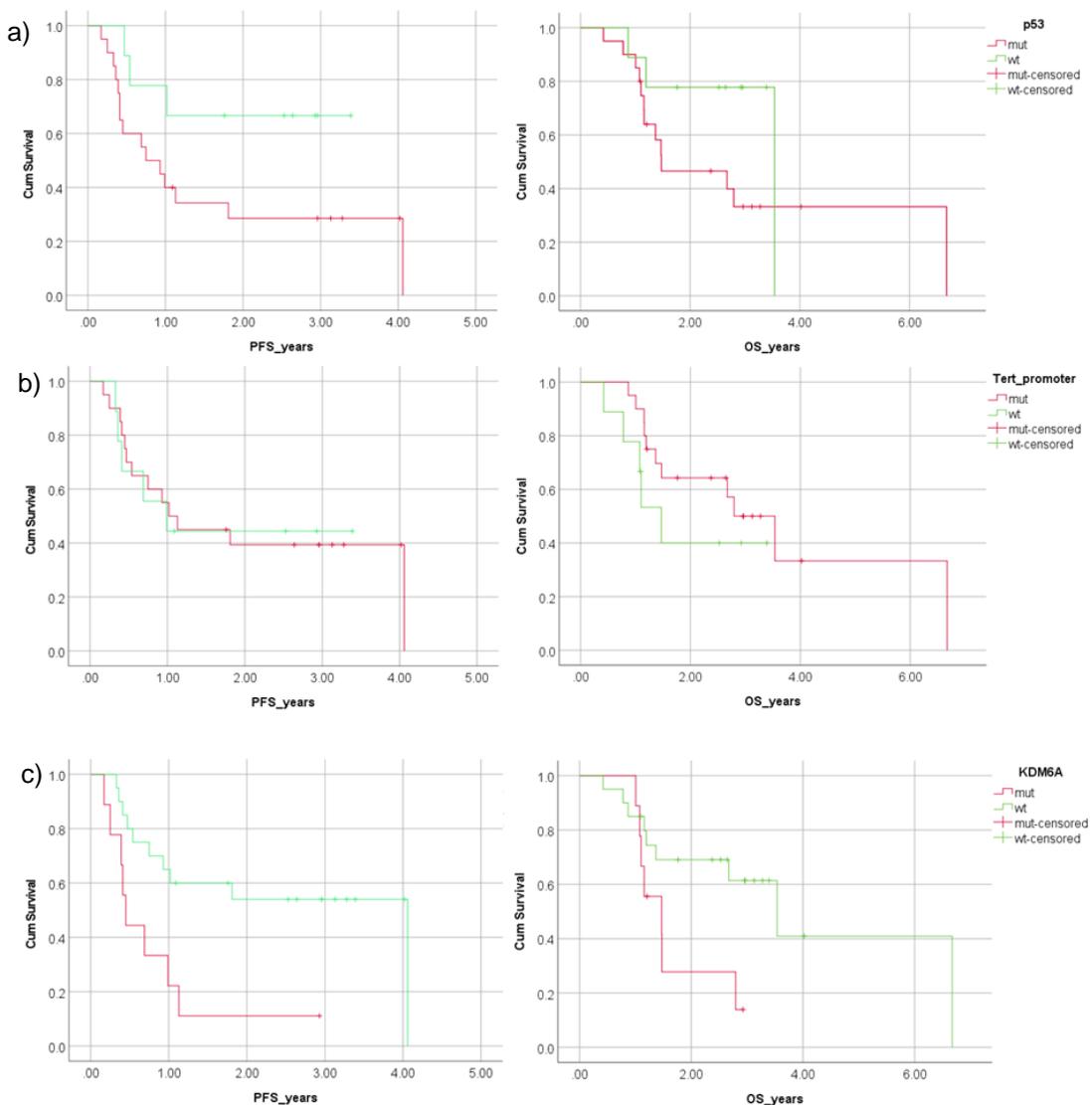


Figure 4.3 Kaplan-Meier curves for progression-free survival (PFS) and overall survival (OS) according to mutation status of a) *TP53*, b) *TERT* promoter, and c) *KDM6A*

Table 4.4 Median and 2-year probabilities of progression-free survival (PFS) and overall survival (OS) for patients according to *TP53*, *TERT* promoter and *KDM6A* status. Subgroup numbers were considered too small for formal statistical comparison.

Abbreviations: WT = wildtype, MUT = mutant, NR = not reached

Gene	Frequency of MUT	Median PFS (yrs; 95% CI)		2-year PFS		Median OS (yrs; 95% CI)		2-year OS (yrs)	
		WT	MUT	WT	MUT	WT	MUT	WT	MUT
<i>TP53</i>	20/29 (69%)	NR	0.75 (0.12–1.28)	67%	29%	NR	1.47 (0-3.11)	78%	47%
<i>TERT</i> promoter	20/29 (69%)	0.99 (0.11– 1.87)	0.1.02 (0.58– 1.46)	44%	39%	1.47 (0.97-1.97)	2.79 (1.91-3.67)	40%	64%
<i>KDM6A</i>	9/29 (31%)	4.06 -	0.45 (0.33 – 0.57)	54%	12%	3.54 (2.00-5.07)	1.47 (0.71-2.23)	69%	28%

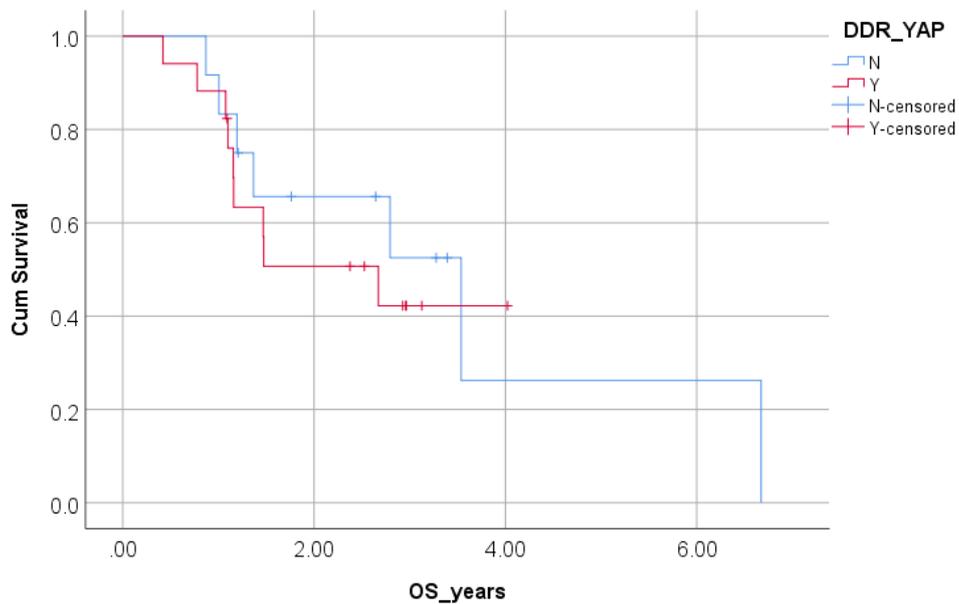
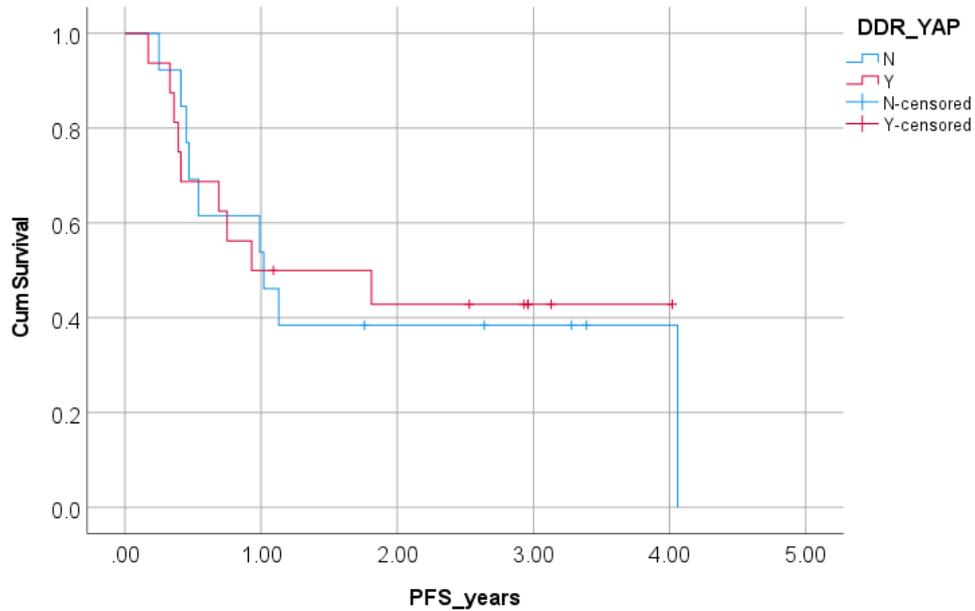
Response to chemotherapy

Twenty patients received NAC and 3 had palliative chemotherapy. Of these, only 4 had progression on treatment. It was therefore not feasible to perform any analysis to compare aberration profiles between patients with or without response to chemotherapy.

DNA damage repair genes

Fifty-nine percent (17/29) patients had at least one aberration in a DNA damage repair (DDR) gene as defined by Yap et al² i.e. *ATM*, *FANCD2*, *PALB2*, *BRCA1* and *BRCA2*. *ERCC2* was not on the FoundationOne panel and therefore could not be included for this cohort. 12/17 had only one aberration. *ATM* was the most commonly aberrant gene from this group affecting 10/29 patients, followed by *BRCA1* (n=5), *FANCD2* (n=4), *BRCA2* (n=3) and *PALB2* (n=1).

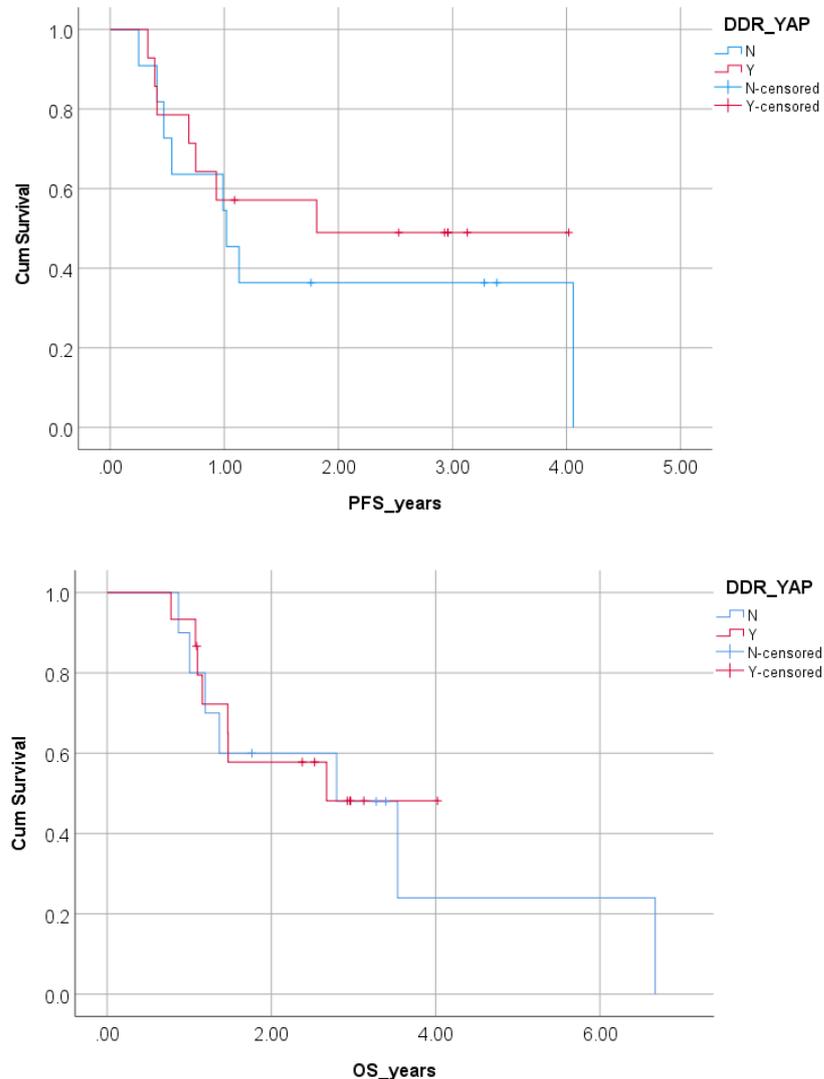
Kaplan-Meier analysis of progression-free survival and overall survival for the cohort according to the presence or absence of at least one of the 5 DDR genes listed did not suggest a significant difference in outcomes. While the subgroup numbers are too small for formal statistical comparison, the Kaplan-Meier curves for progression-free survival and overall survival appear similar for patients with or without an alteration in one of the 5 DDR genes listed (Figure 4.4).



	<i>DDR Alteration?</i>	
	<i>No</i>	<i>Yes</i>
<i>Median PFS (yrs)</i>	1.02 (95% CI 0.33-1.71)	0.99 (95% CI = 0.00-2.87)
<i>2-year PFS</i>	38%	43%
<i>Median OS (yrs)</i>	3.54 (95% CI = 1.35-5.73)	2.67 (95%CI = 0.61-4.74)
<i>2-year OS</i>	67%	50%

Figure 4.4 Kaplan-Meier curves for progression-free survival (PFS) and overall survival (OS) according to presence or absence of a DNA damage repair gene aberration (red and blue lines respectively) as defined by Yap et al². Subgroup numbers were considered too small for formal statistical comparison.

When considering only those treated with radical chemoradiation (n=25), Kaplan-Meier curves for PFS and OS again did not suggest any significant difference between those with or without a DDR alteration (figure 4.5).



	<i>DDR Alteration?</i>	
	<i>No (n = 10)</i>	<i>Yes (n = 15)</i>
<i>Median PFS (yrs)</i>	1.02 (95% CI = 0.38-1.66)	1.81 (95% CI = not generated)
<i>2-year PFS</i>	36%	49%
<i>Median OS (yrs)</i>	2.80 (95% CI = 0.86-4.73)	2.67 (95% CI = not generated)
<i>2-year OS</i>	60%	58%

Figure 4.5 Kaplan-Meier curves for progression-free survival (PFS) & overall survival (OS) according to presence or absence of a DDR gene aberration as defined by Yap et al² for patients treated with chemoradiation. Subgroup numbers were considered too small for formal statistical comparison.

With regards to locoregional relapse in the subset of patients treated with radical radiotherapy, figure 4.6 shows the Kaplan-Meier curves for locoregional relapse-free survival (LR_RFS) for patients with or without a DDR alteration. The median LR_RFS in patients with no DDR alterations was 1.13 years (no 95% CI generated) with 2-year LR_RFS of 48%. The median LR_RFS has not been reached for patients with a DDR alteration; 2 year LR_RFS was 73% in this subgroup. Again the numbers are too small to draw any conclusions but looking at the curves, there is perhaps a suggestion of those with a DDR aberration having better LR_RFS than those without a DDR alteration.

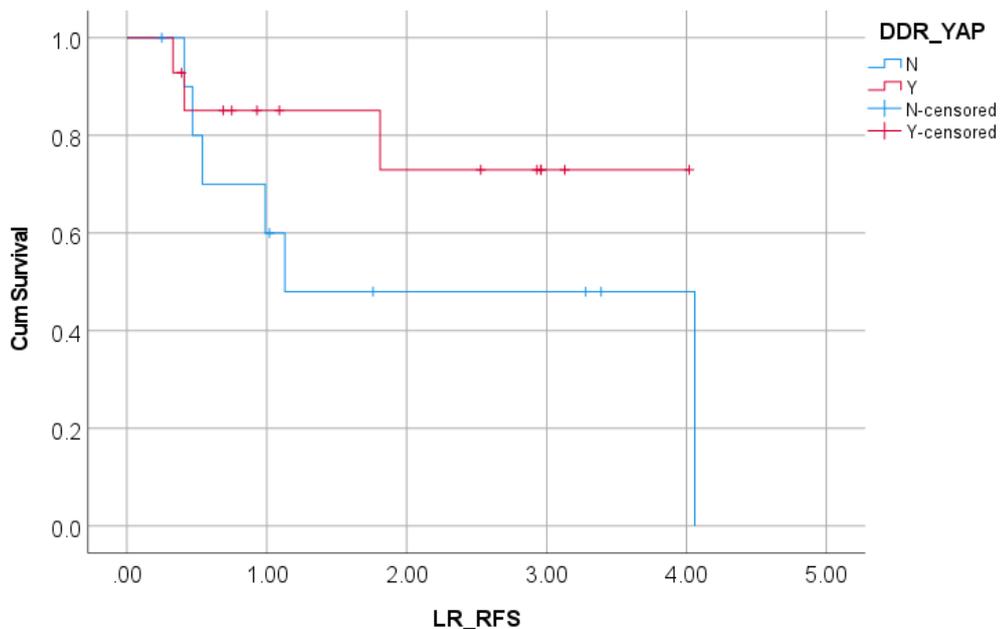


Figure 4.6 Kaplan-Meier curve showing locoregional relapse-free survival (LR_RFS) in the 25 patients treated with radical radiotherapy +/- chemotherapy, according to the presence or absence of a DDR alteration. Subgroup numbers were considered too small for formal statistical comparison.

4.5.3. Whole-exome sequencing

4.5.3.1. Assessment of tumour content and histology for WES

Table 4.5 summarises the number of tumour regions initially identified for each patient and the histology.

Table 4.5 FFPE samples assessed for each patient in WES cohort

Patient	No. FFPE blocks sectioned	Total No. Regions	Histology	Tumour Content
1003	2	6	TCC	40% (x3) 60% (x3)
1010	1	3	TCC	70%
1014	2	6	TCC; sarcomatoid and giant cell differentiation	100%
1016	1	4	TCC	70%
1017	4	10	TCC	50% x2 90% x 3 100% x 5
1021	2	5	TCC; significant desmoplasia	50%
1025	1	3	TCC	60%
1027	2	2	TCC	100%
1039	2	6	TCC	90% x3 80% x3
1052	2	4	TCC; squamous differentiation	50-60%

4.5.3.2. DNA extraction

Total DNA yields from each region with photos of the H&E stained slides are documented in Appendix A4.10. An example is shown below in figure 4.7.

Of note, for patients 1039 and 1052, there was insufficient DNA to proceed with whole exome library preparation after extraction from 10 x 10um sections. A further 10 sections were taken from the same blocks, with first and last H&E slides (Appendix A4.10) assessed to confirm presence of tumour throughout.

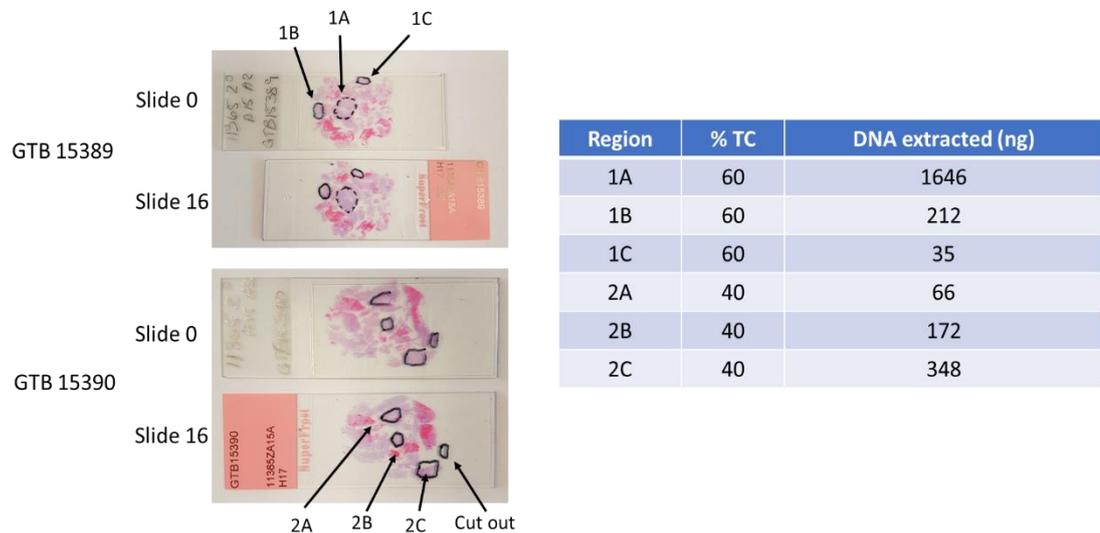


Figure 4.7 Example of H&E slides marked for macrodissection with table showing DNA yield from each region

DNA was extracted from a total of 54 FFPE tumour regions. The mean yield from each region was 960ng with range from 13 – 11048ng.

4.5.3.3. Whole-exome sequencing cohort

The same 10 patients in the plasma ctDNA cohort in Chapter 3 were used for the multiregion whole-exome sequencing cohort. Their patient characteristics can be found in Table 3.3 and individual treatment pathways in Appendix A3.3.

The median overall survival for the cohort has not yet been reached. The median follow-up period is 3.12 years.

4.5.3.5. Overview of results

Quality control parameters

Table 4.6 below shows the quality control parameters of samples sent for WES.

For the first batch, the median depth of coverage ranged from 447-567 with reads on target (83.0-83.8%) as expected for FFPE DNA. The PCR duplicate rate ranged between 39.5 – 47.5% for tumour samples.

For the second batch, the median depth achieved on tumour samples ranged from 199 – 790. The cause for the huge variation is likely related to the PCR duplicate rate which ranged from 43.3-80.6%. Potential reasons for this are discussed in section 4.5.1.1.

Table 4.6 Quality control parameters for a) Batch 1, and b) Batch 2, of samples subjected to WES

a) Batch 1

Sample Name	Total Unique Reads	% Reads Mapped	% Reads on Target	% of Duplicates	Median Depth	Mean Coverage
1027_1A	450,613,917	96.59%	83.25%	41.6%	567	722
1027_2A	356,503,248	96.31%	82.35%	47.5%	447	558
1027_G	89,540,105	97.39%	83.63%	18.5%	120	149
1016_1A	408,520,712	96.28%	82.07%	41.4%	552	641
1016_1C	399,946,383	97.07%	83.00%	39.5%	539	628
1016_G	91,944,805	97.40%	83.80%	16.1%	127	154

b) Batch 2

Sample Name	Total Unique Reads	% Reads Mapped	% Reads on Target	% of Duplicates	Median Depth	Mean Coverage
1016_G	91,944,816	97.40%	83.8%	16.1%	127	154
1003_1A	254,884,287	94.59%	77.5%	76.6%	251	288
1003_2C	305,789,965	95.71%	79.7%	71.4%	317	356
1003_G	142,369,293	98.59%	89.7%	23.2%	149	188
1010_1A	539,797,648	97.40%	82.9%	47.6%	566	653
1010_1C	516,386,936	97.32%	83.0%	50.0%	558	625
1010_G	138,919,571	98.45%	88.5%	21.3%	142	181
1014_1B	472,506,527	96.99%	81.7%	53.5%	495	564
1014_2B	463,484,364	97.05%	81.9%	49.3%	493	554
1014_G	176,695,090	98.52%	88.9%	22.1%	177	232
1017_2B	506,855,472	96.97%	82.4%	53.2%	539	609
1017_3C	492,994,698	96.84%	82.3%	57.4%	526	591
1017_G	160,389,773	98.20%	84.3%	35.2%	169	195
1021_1A	754,181,212	97.43%	83.8%	43.3%	790	922
1021_2A	551,635,731	97.09%	82.8%	50.9%	591	665
1021_G	141,324,788	98.26%	83.8%	33.9%	146	171
1025_1A	296,650,745	95.53%	78.3%	70.1%	296	338
1025_1B	259,501,105	93.00%	73.1%	75.9%	245	274
1025_G	169,401,995	98.26%	87.4%	23.5%	176	217
1039_1A	457,116,100	96.97%	82.4%	54.4%	431	552
1039_2B	358,167,621	96.29%	79.7%	63.9%	343	416
1039_G	164,401,473	98.30%	86.8%	23.8%	173	208
1016_1B	210,359,542	93.42%	77.6%	78.4%	213	237
1016_1D	192,622,272	93.43%	76.0%	80.6%	199	212

Somatic mutations

In total, MuTECT analysis identified 21 584 single nucleotide variations across 7244 genes from WES data of 20 samples from 9 patients. 975 were labelled as low severity, 12884 of moderate severity and 7725 of high severity. GATK analysis additionally revealed a total of 731 602 aberrations. 505 333 of low severity, 212 760 of moderate severity and 13 509 of high severity. Appendix A4.11 shows the most frequently aberrant genes using the TPU-generated filtered reports.

Cross-referencing the TPU-filtered results against the 64-gene list as previously described identified 48 altered genes of interest. However, I noted that WES reported *TP53* aberrations in 8/9 patients compared to 6/9 on FoundationOne. Closer inspection of the data revealed that the TPU GATK-filtered results provided had also included likely single nucleotide polymorphisms (SNPs) e.g. *TP53* ProArg72. I therefore manually curated the results for those genes affecting 2 or more patients, to reflect somatic non-synonymous mutations only. Table 4.7 below shows the genes identified affecting 2 or more patients within this cohort. The full table (pre- and post-SNP removal) is shown in Appendix A4.12.

The aberrant genes affecting the most patients were *TP53*, *ERBB3* and *KMT2C*.

Table 4.7 Aberrant genes affecting the greatest proportion of patients from analysis of a subset of 64 core genes from whole-exome sequencing data. For each gene, the number of high, moderate or low predicted functional impact aberrations is shown.

Gene	No. aberrations				No. samples	No. patients
	HIGH	MODERATE	LOW	TOTAL		
<i>TP53</i>	7	14	12	33	11	6
<i>ERBB3</i>	0	10	4	14	12	6
<i>KMT2C</i>	1	4	4	9	9	6
<i>ERCC2</i>	0	6	2	8	8	4
<i>RHOA</i>	0	6	2	8	8	4
<i>STAG2</i>	2	4	3	9	7	4
<i>NFE2L2</i>	3	4	0	7	5	4
<i>ASXL1</i>	0	6	2	8	8	3
<i>FAT1</i>	2	0	7	9	6	3
<i>ATM</i>	2	4	2	8	6	3
<i>KDM6A</i>	2	0	3	5	5	3
<i>BRCA1</i>	2	1	7	10	5	3
<i>SPTAN1</i>	0	0	6	6	4	3
<i>FANCD2</i>	0	0	6	6	4	3
<i>KMT2A</i>	2	1	2	5	4	3
<i>ARID1A</i>	4	4	2	10	6	2
<i>METTL3</i>	0	6	0	6	6	2
<i>ELF3</i>	1	4	0	5	4	2
<i>KMT2D</i>	4	2	2	8	4	2
<i>PIK3CA</i>	2	4	0	6	4	2
<i>ASXL2</i>	2	0	3	5	4	2
<i>FGFR3</i>	0	4	0	4	4	2
<i>ERBB2</i>	0	5	7	12	3	2
<i>CDKN1A</i>	3	0	0	5	3	2

Specific mutations

Focussing on the 64-gene core subset of interest, no single mutation was observed in more than one patient.

Copy number alterations

There was marked variability in the number of copy number alterations reported between patients and even between an individual's samples (figure

4.8). When considering the number of regions with copy number changes, the range was from 28 regions in patient 1014 to 925 regions in patient 1027.

A total of 1357 regions of copy number alterations were reported with more losses/deletions (n = 1263) compared to gains/amplifications (n = 688). Circos plots are shown in Appendix A4.13.

Four of the nine patients had regions of amplification. The most common regions were 6p22.3 and 7p21.1, which were each amplified in 2 patients. Candidate genes associated with 6p22.3 include *E2F3*, *ID4*, *MBOAT1* and *CDKAL1*. Genes associated with 7p21.1 include *TWIST1*.

With regards to regions of deletion, 4p16.1 was deleted in 7/9 patients. This region is associated with a group of *USP17L* genes which play a de-ubiquinating role.

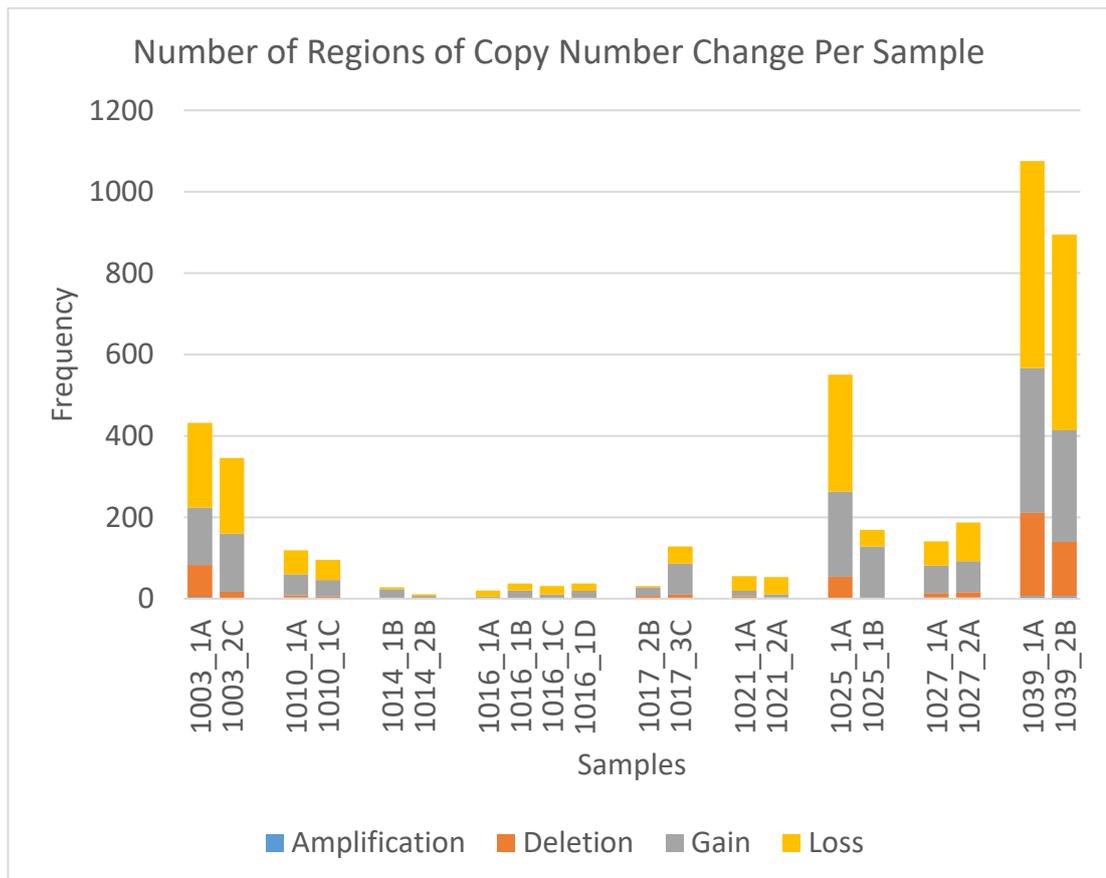


Figure 4.8 Stacked bar graph illustrating the number of regions with copy number change for each sample

Further copy number analysis was challenging without further bioinformatician support. For this cohort of 9 patients, results from the FoundationOne panel had reported that the most common copy number changes were amplification of *RAF1* (3p25.2), *Bcl2l1* (20q11.21) and *ASXL1* (20q11.21); each were reported in 2/9 of the whole-exome cohort. Genes associated with 6p22.3, 7p21.1 and 4p21.1 do not appear to be on the FoundationOne panel.

4.5.3.6. Comparing WES and FoundationOne results

Comparing the results from WES and the FoundationOne panel, several discrepancies were noted with aberrations reported in WES data but not by the FoundationOne panel, and vice versa.

For the 9 patients with both WES and FoundationOne results, FoundationOne testing reported a total of 61 clinically significant aberrations in 45 genes. Cross referencing the MuTECT/filtered GaTK WES data showed that 7 aberrations in 4 genes affecting 5 patients were reported by FoundationOne panel but not in WES data.

On further review, it became apparent that 5/7 aberrations reported exclusively by FoundationOne panel were in fact detected in the WES data but had also been seen at similar frequency in the germline data, thereby suggesting that these were not somatic mutations. For the remaining 2 aberrations reported by FoundationOne (both frame-shift mutations in *KDM6A*), these again were found to be present in the WES data when I reviewed them in IGV. On discussion with TPU, it became clear that these aberrations had been filtered out during their pipeline to report only aberrations on target. A table detailing the discrepancies is in Appendix A4.14.

Eight aberrations in genes tested by the FoundationOne panel were reported by WES but not FoundationOne. On further discussion with Foundation Medicine, in 5/8 cases, the aberration was seen in the FoundationOne data but not reported as it fell below their frequency threshold. The remaining 3 aberrations were all a pTrp33* aberration in *NFE2L2*. However, these aberrations were seen at a coverage level in WES of <50x in all but one of the

samples. The low depth here means it is not possible to confidently call this as a true aberration.

18 copy number alterations of clinical significance were reported across 6/9 patients by the FoundationOne panel. 17 correlated with TPU results. For patient 1027, FoundationOne reported a *MYC1* amplification. While this was not reported in the WES data, the genomic location for *MYC1* was in a region of gain.

4.5.3.7. ERCC2

A total of 8 samples from 4 patients demonstrated an aberration in *ERCC2*. 2/8 aberrations from 2 patients were of predicted low impact. The remaining 6 aberrations across 2 patients were considered to have moderate impact.

No intra-tumour heterogeneity was seen with regards to *ERCC2* with all samples from each of the 2 patients demonstrating the same aberration; all four samples from patient 1016 demonstrated a Thr484ala mutation, and both samples from patient 1021 had a Glu86gln substitution. Table 4.8 below shows the aberrations detected and the variant allele frequency. Although the variant allelic frequency (VAF) ranges from 9 – 22.9% in patient 1016, the tumour content also varies and may explain the different VAFs seen.

Table 4.8 *ERCC2* aberrations detected in multiregion WES

Patient	Sample	TC	Amino Acid	VAF Tumour	Depth Tumour	Alt Normal	VAF Normal	Depth Normal
1016	1A	30	p.Thr484Ala	0.09	435	0	0	158
1016	1C	80	p.Thr484Ala	0.229	437	0	0	158
1016	1B	80	p.Thr484Ala	0.176	102	0	0	158
1016	1D	50	p.Thr484Ala	0.184	98	0	0	158
1021	1A	50	p.Glu86Gln	0.27	847	0	0	223
1021	2A	50	p.Glu86Gln	0.213	780	0	0	223

Of note, patients 1016 and 1021 had received neoadjuvant and palliative cisplatin chemotherapy respectively, with no evidence of disease progression or invasive relapse at the time of analysis with follow-up periods from starting chemotherapy of 3.39 and 2.64 years. This is interesting to note in the context of published work suggesting that *ERCC2* aberrations are associated with cisplatin sensitivity in MIBC⁶.

4.5.3.8. Exploring intra-tumour heterogeneity

SNVs

Comparison of moderate and high predicted impact aberrations reported by TPU affecting the subset of 64 core genes demonstrated differences between samples from individual patients. These are summarised in Appendix A4.15.

A striking apparent finding was the presence of a *TP53* aberration in only one of the two samples from patient 1014. However, on reviewing the sequencing data, it became clear that the same aberration was present in the second sample but had been filtered out from the TPU report as its frequency fell below the arbitrary threshold set.

Patient 1016 had 4 tumour samples sent. One had a moderate impact *KMT2A* aberration but only at a frequency of 5.4% (4/74). This aberration was not seen at all in the other 3 tumour samples. Further reassessment of the tumour content of these 4 samples showed that the sample with the *KMT2A* aberration had a higher tumour content than the other samples, suggesting that if the aberration is considered 'real', one explanation for the difference seen between samples might be due to tumour content rather than true intra-tumour heterogeneity.

Two patients however had aberration patterns that would not be explainable by differing tumour contents:

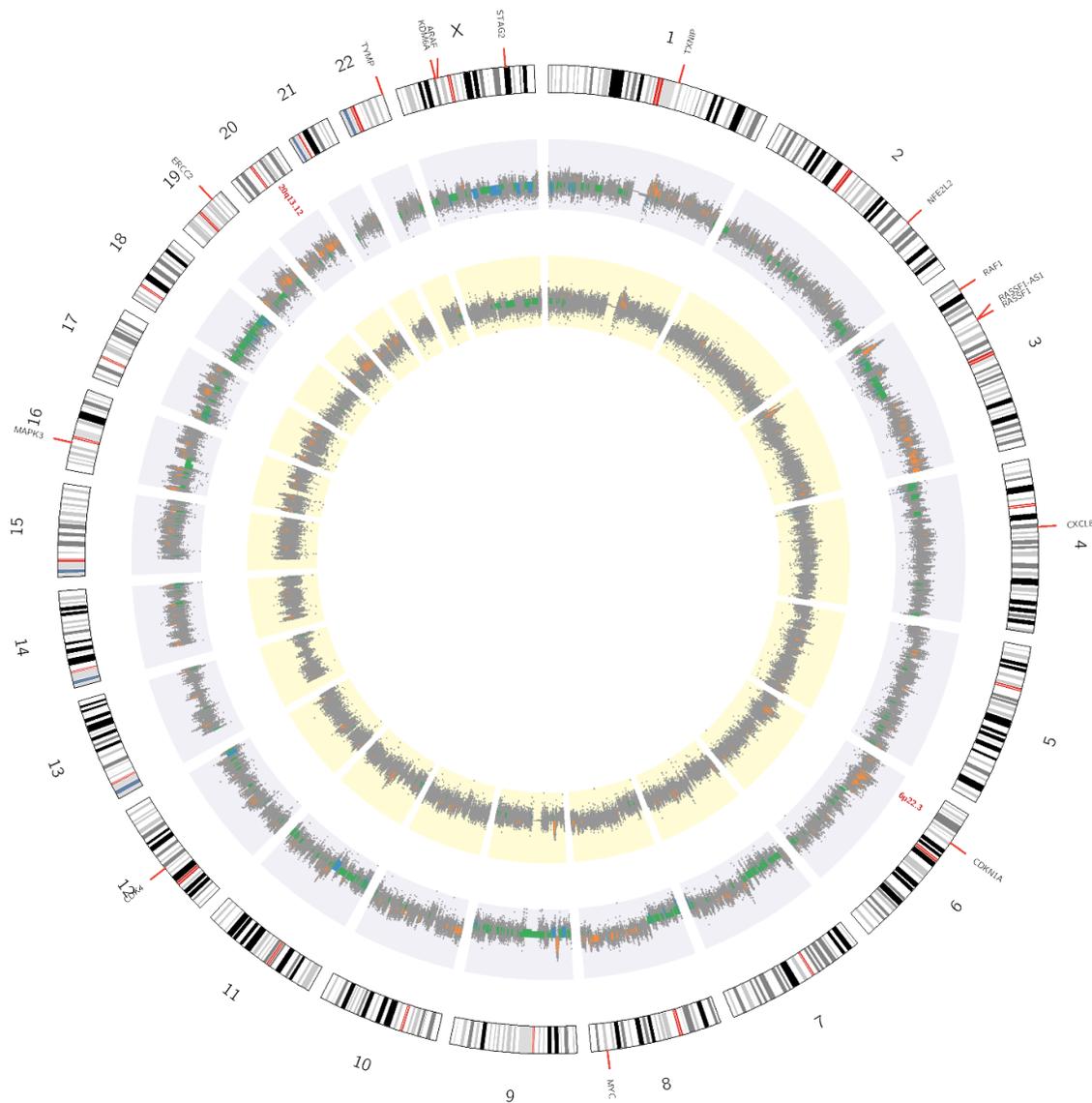
1) Patient 1003 had 2 samples sequenced; sample 2C (40% tumour content) had an *ERBB3* aberration detected at a frequency of 13.4% (68/506) but this aberration was not seen at all in sample 1A despite good coverage of this region of over 500x and tumour content of 60%. Sample 1A had a high impact

aberration in *TMCO4* at a frequency of 37.4% but this alteration was not detected at all in sample 2C (at a coverage of 111x).

2) Patient 1027 had a *CDKN2A* mutation at 10.2% frequency (32/313) in sample 2A (tumour content 100%) but not in 1A which also had tumour content of 100%. An *ERBB3* aberration was also seen at 5.2% (24/459) in 1A but not in 2A.

Copy number alterations

Comparing copy number changes between samples also identified many differences, mainly involving regions of amplification which were seen only as regions of gain in the paired sample. One patient, 1025, however had quite marked differences and this is best illustrated by the circos plot (see figure 4.9)



Outer Ring : Chromosome position
 Middle Ring: : 1025_1A
 Inner Ring: : 1025_1B

- : High gain (amplification)
- : Shallow gain
- : One copy loss
- : Homozygous deletion

Figure 4.9 Circos plot for 1025

Note the differences between sample 1025_1A (middle ring) and 1025_1B (inner ring) e.g copy loss seen in chromosome 18, gain in chromosome 3

4.6. Discussion

4.6.1. Whole-exome sequencing in MIBC

4.6.1.1. Feasibility of performing multiregion high-depth WES

Coverage and PCR duplicates

In this pilot work, I have shown that it is feasible to perform multiregion high depth whole-exome sequencing on DNA extracted from FFPE MIBC tumour tissue. I have set up a workflow for the preparation of FFPE blocks, macrodissection, DNA extraction and library preparation. It was pleasing to see that target coverage was achieved in most cases but there was significantly more variability with the second batch of samples sent in terms of coverage and PCR duplicate rate.

This was disappointing as I had hoped that by adding a FFPE repair step before library construction and reducing PCR cycles for the second batch, the PCR duplicate rate would have fallen. On reviewing the process, I suspect that the poorer PCR duplication rate and resulting lower coverage was due to a lower DNA input at the commencement of library preparation after FFPE repair i.e. whilst I had accounted and compensated for DNA loss during sonication by adding an additional 5% at the start, I did not make any additional adjustments for loss during FFPE repair. Furthermore, although I performed fewer PCR cycles for the second batch (9 compared to 11), the better quality 'repaired' DNA amplified more readily resulting in overamplification despite the reduction in cycles.

In future work, I would therefore suggest that in addition to adding an extra 5% to account for DNA loss in the sonication process as previously done for all samples where possible, a further amount, perhaps 10%, should be added to account for loss during the FFPE repair step in order to ensure maximum possible input for library generation. Ensuring as close to 200ng input DNA for library construction is the best way to reduce PCR duplicates and optimise sequencing capacity.

Ideally, I would have liked to perform a small trial using the same batch 1 samples but with the addition of the FFPE repair kit and fewer amplification cycles first before proceeding with the remaining samples. This would have allowed me to assess the effect of the modifications and would have no doubt alerted me to the further modifications required. Unfortunately this was simply not possible within the timescale available.

Despite these limitations, I feel however that I have achieved my aim set out to demonstrate the feasibility of performing high depth multiregion sequencing on FFPE-derived DNA, and my work has set up a good foundation upon which to build and optimise the process further.

With regards to PCR duplicates, one consideration at the time was whether to use the unique molecular identifiers (UMIs) to identify duplicates rather than identification of identical reads. This however would require an entirely different pipeline and high levels of sequencing to ensure robust aberration calls. At the time, as pipelines using UMIs were still in development, it was agreed not to use them currently but future work could include to re-analyse the data using the UMIs and assess the impact of doing so.

Data Analysis

This work has generated an enormous amount of data and work is currently underway by one of Prof Attard's bioinformaticians to more fully analyse it.

However, for the purposes of this MD(Res), I wanted to gain an overview of the results and any key interesting features. My first approach was to use the reports generated by TPU, which contained a total of 8625 calls for 9 patients from MuTect and GATK pipelines. However, during my analysis, and particularly when comparing data with results from the FoundationOne panel, it became apparent to me that the filtering criteria used by TPU had resulted in the omission of some key aberrations, including a *TP53* frameshift mutation. I therefore made the decision to use the full list of MuTect and filtered GATK calls which included 29609 aberrations in an attempt to avoid missing any key results. This proved more challenging than I had initially anticipated given the amount of data generated, and I also discovered that using the filtered GATK

list would still result in the omission of a *TP53* frameshift mutation as the frequency fell just short of the 40% arbitrary threshold used by TPU. Using the full list of GATK calls with MuTect aberrations however would mean reviewing a total of 753187 aberrations, with copy number alteration data yet to be analysed! During my analysis it also became apparent that high-impact frameshift mutations in *KDM6A* in 2 patients had not been reported by TPU at all as they were filtered out when selecting 'on-target' reads. This highlights to me that in order to best explore the data generated, it will be necessary to start with the raw data and this is the approach being taken by my bioinformatician colleague currently.

Given the enormous scale of data generated, I therefore adopted a pragmatic approach and focussed on a subset of 64-genes which have either been reported in the literature as significantly mutated in MIBC or are of potential interest in the context of patients receiving NAC +/- CRT i.e. DNA damage repair genes. I also focussed on non-synonymous somatic aberrations. Of course, using a subset of genes previously identified as being significantly mutated in MIBC greatly limits the analysis and likely neglects potentially interesting results. However, in the face of a huge amount of data, and after several unsuccessful attempts by me to analyse the raw data, I feel this was an entirely reasonable and logical starting point.

Using this approach, I found that the most frequent aberrant genes in my data were *TP53*, *ERBB3* and *KMT2C* which were each altered in 6/9 patients. In the updated TCGA data (accessed via online data portal November 2019), these aberrations were seen in 48%, 72% and 21.6% of subjects respectively. It is not possible to make any comparisons in aberration frequency here given my very small cohort and also the differences in sequencing parameters between the datasets. My data was generated from WES with a coverage range of 312-922x (mean 520x) on FFPE DNA while the TCGA data has been generated from WES on fresh-frozen tissue with mean coverage of 85x. It is therefore likely that by sequencing at a greater depth, my data includes subclonal aberrations that would not have been detected by sequencing at 85x.

However, my data undoubtedly also contains FFPE artefact and noise given the higher depths of sequencing, and the challenge is in determining what is real and what is not. Pipelines such as MutSig 2CV are valuable in determining aberrations which are statistically significant⁷, and tools to more readily identify FFPE artefact and aid analysis⁸ have been developed. While it has not been possible to run either MutSig 2CV nor GISTIC (for copy number alterations) within the timescale of this project, this is something I have discussed with the bioinformatician team for future work and it will be interesting to see these results.

4.6.1.2. Intra-tumour heterogeneity

One of my aims was to assess the feasibility of multiregion sequencing in MIBC and explore heterogeneity in MIBC.

In my limited results, possible intra-tumour heterogeneity was demonstrated by two of my cohort who each had an *ERBB3* aberration present in only one of two of their samples. In both patients, there was good sequencing coverage and the aberrations were both of predicted moderate functional impact. Furthermore, although no *ERBB3* mutations were documented in the FoundationOne patient reports, after liaising with the Roche Medical Information team, I was reassured to discover that the specific *ERBB3* aberrations were seen for both patients in their data but had been filtered out due to low frequencies (4% for patient 1003 and 1% for patient 1027). This confirmation of the presence of *ERBB3* mutations, albeit at low frequencies, serves to validate my corresponding WES results and the absence of the aberration from the accompanying sample is in keeping with a demonstration of intra-tumour heterogeneity.

One of the challenges in exploring intra-tumour heterogeneity in MIBC is in determining true heterogeneity from artefact and noise. In an interesting study by Shi et al¹², they explored the use of WES in assessing intra-tumour heterogeneity. Using technical replicates, different target coverages of 90, 160 and 184x, and high depth amplicon sequencing for validation (median 605x), they concluded that '34-80% of somatic variants contributing to genetic

heterogeneity are technical noise'. They also concluded that performing WES at standard depths is likely inadequate to assess intratumour heterogeneity.

In this respect, although a small pilot, I think that my approach of using high depth multiregion sequencing forms a good base from which to further explore heterogeneity in MIBC. I think there is more to be gained from the existing data and feel that it is worthy of further investigation. Gaining some understanding in to heterogeneity will be key in understanding treatment resistance and disease progression, and to form more successful treatment paradigms.

4.6.2. Targeted sequencing in MIBC: A commercial panel

4.6.2.1. Limitations of using a commercial panel

The FoundationOne reports did not provide any information with regards to whether aberrations in tumour suppressor genes were mono-allelic or bi-allelic. Unfortunately, it was not possible to obtain the raw data, which would have potentially allowed analysis of this, within the timescale of this project. This is therefore a limitation of using this platform in a research context as without this information, it is not possible to comment on the potential functional impact of the aberration. The lack of access to the raw data was not ideal as it then became necessary to contact Foundation Medicine with any specific queries regarding results, and it was not possible to explore the data myself. A further limitation was that their precise criteria for listing an aberration as reportable was not made available.

4.6.2.2. TP53 and KDM6A status associated with outcomes

Although a small cohort of patients, it was interesting to see the divergence of the Kaplan-Meier curves for those with mutant vs wildtype *TP53* and *KDM6A*, with patients with mutant status appearing to have worse PFS and OS. With such a small cohort, formal statistical comparison was not performed. However, with regards to *TP53*, my findings are in keeping with previous work demonstrating that patients with mutant *TP53* have poorer outcomes following radical cystectomy¹³⁻¹⁵. To the best of my knowledge, there has not been any published work reporting an association of *TP53* mutation status with outcome

following radical radiotherapy +/- chemotherapy. As my pilot results in a radiotherapy-treated cohort are in keeping with results from cystectomy cohorts, this would be in keeping with *TP53* mutation status being a prognostic biomarker rather than predictive biomarker of response to treatment.

It has been previously reported that *TP53* aberrations are more frequently seen in patients with extra-vesical disease or regional lymph node involvement¹⁴. These clinicopathological features are recognised to be adverse prognostic factors¹⁶ so the association between *TP53* mutations and poorer outcomes fits with this. Interestingly, in my cohort, 8 patients were staged as T2N0M0 (i.e. no extravescical spread or locoregional lymph nodes) but all 8 had a *TP53* aberration. Conversely, of the 21 patients with extra-vesical or regional lymph node disease, only 12 (57%) had a *TP53* aberration. It may be that some of the 8 patients staged as T2N0M0 were under-staged; upstaging at the time of cystectomy is recognised to be common in MIBC¹⁷ and of course, in this cohort of radiotherapy-treated patients, there was no surgical staging.

KDM6A (also known as *UTX*) codes for a histone demethylase and is involved in chromatin modification. Pan-cancer studies have reported that bladder cancer has the highest frequency of *KDM6A* aberrations¹⁸, and its tumour suppressor activity is often lost in MIBC due to inactivating mutations. Recent work has shown that in mice with a *KDM6A* deficiency and heterozygous p53 status, there was increased activation of cytokine and chemokine pathways, increased cancer stem cells and increased bladder cancer development, suggesting a role for *KDM6A* in the carcinogenesis of MIBC and raising its profile as a potential target¹⁹. Another recent interesting report²⁰ looked at whether *KDM6A* (located on the X-chromosome) may play a role behind the increased incidence of bladder cancer in males. As part of their analysis, they accessed the TCGA dataset and reported that in females, mutant *KDM6A* was associated with poorer disease-free survival but this association was not seen in males²⁰. My cohort of 29 included only 7 females, of whom 4 had a *KDM6A* aberration. It would therefore not be feasible to assess for any gender differences in my data. However, the poorer clinical outcomes seen in the

TCGA cystectomy dataset is in line with the pilot results from my radiotherapy cohort. It would be interesting to explore this further, particularly as there is no other published data I am aware of looking at *KDM6A* status in MIBC patients treated with radiotherapy.

In contrast to the *TP53* and *KDM6A* results, there appeared to be no clear separation of the curves for progression-free survival with regards to *TERT* promoter status. However, while 2-year overall survival probabilities of 40% for wildtype and 64% for mutant *TERT* promoter status were demonstrated, it is difficult in this small cohort to know whether this apparent difference is true or significant. The suggestion that *TERT* promoter aberrations may be associated with better outcomes certainly sits at odds with previously published data which has demonstrated that *TERT* promoter alterations in urothelial cancer are associated with poorer clinical outcomes^{18,21,22}. These cohorts have consisted of patients with metastatic bladder cancer (n >300)¹⁸, those with 'urothelial cancer' (mix of NMIBC, MIBC, upper tract cancers and metastatic disease; n = 389)²² and those with NMIBC or MIBC (n = 327)²¹. Somewhat frustratingly, the latter two studies did not analyse those with MIBC separately so it is difficult to draw any direct comparisons. Furthermore, details on staging and treatment received were not provided within these publications.

4.6.2.3. DNA damage repair genes

When exploring whether DDR gene status was associated with outcome in this small cohort, the Kaplan-Meier curves for PFS, OS and LR_RFS appeared fairly similar for patients with or without an alteration in one of the 5 DDR genes selected. There was perhaps a suggestion of the curves separating when considering LR_RFS in those treated with chemoradiation with 2-year LR_RFS probabilities of 73% and 48% seen, but the cohort is unfortunately too small to draw any conclusion or perform any meaningful statistical analysis. The hypothesis would be that patients with an altered DNA damage pathway may be more radiosensitive as they may not be able to repair radiation-induced DNA damage as readily as those with an intact pathway, and has been explored by other groups³.

The small range of genes chosen for my analysis were taken from a paper published in 2014 which reported that patients with an alteration in at least one of *ATM*, *ERCC2*, *FANCD2*, *PALB1*, *BRCA1* or *BRCA2* had improved relapse-free survival post-cystectomy. Of these 6 genes, *ERCC2* was omitted as this gene was not on the FoundationOne panel. This was frustrating as there was much interest in the potential role of *ERCC2* as a biomarker of response to platinum-based chemotherapy⁶ in MIBC, and it was the most common DDR alteration in the chemoradiation study mentioned above³. Interestingly, the FoundationOne panel has since been updated but still does not include *ERCC2*. This reflects the fact that despite much excitement about *ERCC2*, its role as a biomarker in the clinical setting has yet to be demonstrated in a large prospective trial.

More recent work interrogated 34 DDR genes in metastatic bladder cancer, of which 20 were on the FoundationOne panel so it would be interesting, in the first instance, to revisit my data and expand my analysis of DDR genes to include these.

4.6.2.4. The *TP53* Pro72Arg polymorphism

One interesting observation when comparing the FoundationOne and whole-exome results for the 9 patients with both available was the number of *TP53* aberrations. The FoundationOne results identified 6/9 patients with a clinically significant *TP53* aberration whereas the WES reports from TPU reported a *TP53* aberration in 8/9 patients. On reviewing the WES data, it became apparent that the two additional patients reported by TPU to have a *TP53* aberration had a Pro72Arg *TP53* alteration. However, I was able to see that this aberration was present at a higher frequency in each patient's germline sample, and that this particular aberration is one of the most common *TP53* polymorphisms (rs1042522), with frequency noted to increase in a North to South distribution within the Northern hemisphere from 17% to 63% (Sweden to Nigeria)²³.

Although not a somatic mutation, this polymorphism is of potential relevance in this setting. Dumont et al²⁴ demonstrated that the Pro72 (P72) variant was less efficient at inducing apoptosis than the Arg72 (R72) variant. More recent

work has also suggested that the polymorphism, when present in mutant p53 results in increased cell migration and invasion in osteosarcoma, prostate and lung cancers²⁵. Furthermore, Basu et al²⁶ reported that breast cancer patients with mutant p53 and the R72 variant had a significantly poorer prognosis than those without the polymorphism.

Interestingly, on reviewing the unfiltered data, all 9 patients had the P72 variant. In 2 patients, this was the only aberration in *TP53*, seen also in the germline data. The remaining 7 patients had other somatic aberrations in *TP53*. It would be interesting to explore this further in a larger cohort and see whether, as reported in breast cancer, the presence of this polymorphism in aberrant *TP53* confers a poorer prognosis²⁶. Previous work²⁷ has suggested that the codon 72 polymorphism may be associated with risk of urothelial cancer in the context of smokers, and further work has explored whether this polymorphism and MDM2 variants may influence outcomes following chemoradiation for bladder cancer²⁸.

It would be interesting to explore the role of the Pro72Arg polymorphism within an expanded CoMB cohort to see whether previous published results suggesting possible susceptibility to developing disease or outcomes from treatment are reproduced, and if any further insight can be gained from our patient cohort.

4.6.3. Designing a targeted panel for plasma

4.6.3.1. Using data generated and publicly available data

In this chapter, I have generated high-depth whole-exome data for 9 patients with MIBC, and additionally received high-depth targeted sequencing data for a further 20 patients. I had planned to use this data with publicly available data, namely the TCGA database, to design a targeted sequencing panel for plasma in MIBC. However, since commencing this project, several groups published work in a similar vein, each using different approaches²⁹⁻³¹, as discussed in my introduction.

It is interesting to note that if considering the ddPCR and Tam-Seq panel targets used in the published literature^{31,32}, the range interrogated would have potentially encompassed a total of 23/29 patients from my FoundationOne cohort, although it should be noted that the vast majority (20/29) would have been due to the inclusion of all *TP53* exons (Appendix A4.16). However, one of the key challenges when working with plasma is the detection of aberrations at low frequencies. It would be interesting to include *TERT* promoter aberrations in a targeted panel; these were present in 69% of my cohort and this is consistent with published data¹⁸. Targeting these alongside *TP53* exons should therefore encompass a significant proportion of MIBC patients and would have the potential to capture 26/29 (89.7%) of my patient cohort. This is an exciting prospect but in order to be a useful tool in monitoring patients over time, the panel would need to include a broader range of targets i.e. one would not be able to detect other aberrations which may become more prominent with evolving treatment resistance or disease progression if they are not on the panel.

Considering the published work to date, some interesting questions with relevance to the direction of my work are raised.

Firstly, what is the best approach to sequence plasma ctDNA in MIBC? In this project, I have started to explore the feasibility of both a patient-specific approach, by performing high depth multiregion WES on tumour tissue, with a view to selecting targets to track in plasma, and also considered the use of a targeted panel to sequence plasma upfront using data generated and publicly available data.

The data published so far suggests that a patient specific approach appears to be the most successful in terms of maximising chances of ctDNA detection rather than upfront plasma sequencing using predefined assays or panels. Although whole-exome sequencing of tumour tissue would give an overview of any aberrations present that could be tracked in plasma, a targeted panel has the advantages of requiring less DNA input and achieving much higher coverage for less cost. I would therefore support the use of a bladder cancer-specific targeted sequencing panel to identify aberrations present in tumour

tissue, followed by interrogation of plasma for selected targets using ultra-sensitive methods such as ddPCR or Tam-Seq, or highly targeted NGS as used by Christensen et al³³. The questions that then arise are what targets should be on the panel, what sequencing depth of tumour tissue is necessary, and is it sufficient to sequence only one tumour tissue sample when there is evidence of intra-tumour heterogeneity in MIBC?

With regards to panel design, groups have used publicly available data to identify frequently mutated genes in MIBC for inclusion on their panels^{33,34}. On reviewing their lists, there are 28 targets in common between their panels of 50 and 51 genes. Selecting targets based upon their frequency is logical but I would additionally add targets of interest such as DNA damage repair genes and in the context of our patient cohort, genes potentially associated with radiotherapy response e.g. *NBN*³⁵, and also immunotherapy e.g. *POLD1/POLE*³⁶. At this point, I am not able to finalise a panel as this will also depend upon the further analysis of my WES data, and discussions regarding the feasible number of targets and depth. I have however included a list of candidates of interest in Appendix A4.17. Of note, sequencing capabilities have dramatically increased over the course of my MD(Res) and it might be that, depending on the final target panel, upfront sequencing of plasma is soon feasible at ultra-high coverage. Currently however, given that there is still a lot to discover on the dynamics of MIBC progression and treatment resistance, it makes sense to use broader sequencing techniques on tissue first to identify potential plasma targets.

Regarding depth of sequencing required for tumour tissue and intra-tumour heterogeneity, I note that Christensen et al³³ had a mean depth of 104x when performing WES on tumour tissue and were able to select 16 aberrations to track for each individual patient with very promising results. So what extra benefit would my high-depth data bring? And is there a place for multi-region sequencing here? It is interesting that in the majority MIBC studies reported to date, sequencing of tumour tissue has been performed on a single diagnostic sample. Jamal-Hanjani et al³⁷ performed multiregion WES on lung cancer tissue samples and reported that had only one tumour sample been analysed,

76% of mutations considered subclonal could have been labelled as clonal. This might account for why some aberrations detected in tissue are then not detected in plasma, although I accept that the more likely explanation is that of low frequencies. In my preliminary exploration of heterogeneity, if only a single sample had been analysed from each of my subjects, an *ERBB3* aberration of predicted moderate functional impact would have been missed in 2 patients and would not have been selected as targets to track in plasma.

With regards to sequencing depth, some might suggest however that aberrations present at such low frequency to require high-depth sequencing to detect are unlikely to be of any clinical significance, and therefore such analysis is unnecessary. I would say that while they may not be of direct relevance at the commencement of a treatment pathway, expansion of such subclones may be drivers of treatment resistance or disease progression and are thus of interest when considering the dynamics of MIBC.

Saying this, when paired tissue samples taken at diagnosis and local recurrence for 5 MIBC patients treated with chemoradiation were compared³, targeted sequencing results showed that the vast majority (median 92%; range 74-100%) of aberrations were shared between the samples, and alterations unique to the recurrence were uncommon with a mean of 4% (range 0-16%). The authors suggested that treatment failure in these cases were due to primary resistance rather than emerging subclones, and might suggest that performing high depth sequencing for subclones may be less relevant, at least in the setting of investigating radioresistance in MIBC. This however of course is a tiny cohort and further investigation is warranted. The CoMB cohort represents an excellent resource in which to investigate this further with the collection of post-NAC and post-radiotherapy biopsy samples.

Whether such high depth sequencing up to 500x in FFPE-derived DNA is helpful however is a reasonable question. Through this work, I have had first-hand experience in the challenges of working with FFPE material and the subsequent data analysis. In the context of selecting plasma targets, which will then be even more challenging to detect at even lower frequencies, the optimal target coverage perhaps has yet to be elucidated. What is clear

however is that multiregion high depth whole exome sequencing has its place in exploring heterogeneity and clonal evolution, as demonstrated in non-small cell lung cancer³⁷ by the TracerX group. While admittedly this work used fresh frozen samples and therefore bypassed the inherent issues of working with FFPE, the application of such a paradigm would be of huge interest in MIBC. Furthermore, as FFPE material has been and remains the primary method of preserving tissue samples in clinical practice, the optimisation of pipelines using FFPE-derived material is essential if we are to make best use of this resource.

4.6.3.2. Using a commercial panel

It could be suggested that the commercially available panels, which have detected ctDNA in the plasma of up to 90% of patients with metastatic urothelial cancer³⁸⁻⁴⁰, could be of use in this setting, and preclude the need to custom-design a panel. However, these panels omitted many of the commonly mutated genes in bladder cancer, including those which may be potentially actionable e.g. chromatin-modifying genes, and other limitations with regards to data interpretation have already been discussed in section 4.6.2.1. While early panels included only 7 of the top 20 significantly mutated genes in MIBC³⁸, later panels have expanded the targets and the FoundationOne panel used in my work included 20 of the 23 most significantly mutated genes from the TCGA report. The omissions were *ERCC2*, *SPTAN1* and *ELF3*. I note that panels specifically for plasma have also been commercially developed and it would be interesting to see how these perform in MIBC.

4.6.4. Future work

This pilot work has identified several areas for future work. Firstly, the FoundationOne results suggested a prognostic role for *TP53* and *KDM6A* status and it would be worthwhile expanding the cohort to investigate this further. Although these findings are not novel and have already been explored in larger cohorts, previous work has focussed on patients managed with cystectomy and I think it would be worthwhile performing this analysis in a large cohort of patient treated with chemoradiation. Of note, although my

FoundationOne cohort did include patients with metastatic disease, the majority (86%) received radical radiotherapy +/- chemotherapy.

The same applies for *TERT* promoter status as although my data here was less convincing for any association with clinical outcome, other groups have suggested some significance to *TERT* promoter status with relation to outcomes.

It would also be interesting to explore DDR genes further in the context of outcomes following radiotherapy. Pan-cancer work has demonstrated that MIBC has a greater frequency of DDR alterations than other solid cancers⁴¹. Although there has been some published work looking at DDR alterations in MIBC, these studies have focussed on response to neoadjuvant chemotherapy and not radiotherapy^{6,42}.

With regards to the WES data, I await results from further analysis by the bioinformatician team with great interest. The results will hopefully help guide further work in plasma, both in guiding selection of aberrations to target in plasma in the current ctDNA cohort and also in further defining a targeted panel. What will be of most interest is exploring whether sequencing FFPE-derived tumour DNA at a target 500x gives a practical advantage in selecting plasma targets of relevance or if it is best placed for exploring heterogeneity and phylogenetics. I have already discussed ways in which my workflow could be optimised e.g. reduction of PCR duplicates etc., and it might be useful to look again at whether UMIs may be helpful. I would suggest that regardless of sequencing depth, multiregion sequencing has a place in selecting plasma targets to give confidence in whether clonal or subclonal aberrations are being selected.

The surge of publications in this area during my MD(Res) has been both helpful and frustrating at times given the similarities in work published and that which I have been trying to achieve. However, the CoMB cohort are unique in that they represent patients treated with chemoradiation and not cystectomy as in the other published work. Further work now should focus on assessing plasma targets in sequential samples as discussed above, perhaps with a

focus on radiotherapy response. One potential project of interest would be to compare aberration profiles from samples pre- and post-radiotherapy to explore mechanisms behind primary radioresistance or disease recurrence. Several studies have shown that ctDNA is more readily detectable in urine and so this is also an area for future work.

4.7. Conclusion

In this chapter, I have successfully set up a workflow for high depth multiregion whole-exome sequencing from FFPE-derived DNA in MIBC. Analysis of the data generated is ongoing but initial findings have demonstrated the feasibility of this approach. The final results will inform future work and aid design of a targeted panel. While I had initially planned to design a targeted sequencing panel for plasma in MIBC, my results and data from work published since would suggest that a patient-specific approach is most appropriate.

This has been a fast-moving field with promising advances reported with regards to the role of ctDNA in MIBC. Although I have not been able to achieve all my aims within the timescale of this project, I think my work here has set a good foundation for future work and the CoMB collection provides a valuable and ready resource for such projects.

4.8. References

1. Ross, J.S., *et al.* Comprehensive genomic profiling of 295 cases of clinically advanced urothelial carcinoma of the urinary bladder reveals a high frequency of clinically relevant genomic alterations. *Cancer* **122**, 702-711 (2016).
2. Yap, K.L., *et al.* Whole-exome sequencing of muscle-invasive bladder cancer identifies recurrent mutations of UNC5C and prognostic importance of DNA repair gene mutations on survival. *Clin Cancer Res* **20**, 6605-6617 (2014).
3. Desai, N.B., *et al.* Genomic characterization of response to chemoradiation in urothelial bladder cancer. *Cancer* **122**, 3715-3723 (2016).
4. Cerami, E., *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-404 (2012).
5. Gao, J., *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, p11 (2013).
6. Van Allen, E.M., *et al.* Somatic ERCC2 mutations correlate with cisplatin sensitivity in muscle-invasive urothelial carcinoma. *Cancer Discov* **4**, 1140-1153 (2014).
7. Lawrence, M.S., *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214-218 (2013).
8. Kato, M., *et al.* A computational tool to detect DNA alterations tailored to formalin-fixed paraffin-embedded samples in cancer clinical sequencing. *Genome Med* **10**, 44 (2018).
9. Thomsen, M.B., *et al.* Spatial and temporal clonal evolution during development of metastatic urothelial carcinoma. *Mol Oncol* **10**, 1450-1460 (2016).
10. Thomsen, M.B.H., *et al.* Comprehensive multiregional analysis of molecular heterogeneity in bladder cancer. *Sci Rep* **7**, 11702 (2017).
11. Knowles, M.A. & Hurst, C.D. Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. *Nat Rev Cancer* **15**, 25-41 (2015).
12. Shi, W., *et al.* Reliability of Whole-Exome Sequencing for Assessing Intratumor Genetic Heterogeneity. *Cell Rep* **25**, 1446-1457 (2018).
13. George, B., *et al.* p53 gene and protein status: the role of p53 alterations in predicting outcome in patients with bladder cancer. *J Clin Oncol* **25**, 5352-5358 (2007).
14. Kim, P.H., *et al.* Genomic predictors of survival in patients with high-grade urothelial carcinoma of the bladder. *Eur Urol* **67**, 198-201 (2015).
15. Malats, N., *et al.* P53 as a prognostic marker for bladder cancer: a meta-analysis and review. *The Lancet Oncology* **6**, 678-686 (2005).
16. Stein, J.P., *et al.* Radical cystectomy in the treatment of invasive bladder cancer: long-term results in 1,054 patients. *J Clin Oncol* **19**, 666-675 (2001).
17. Gray, P.J., *et al.* Clinical-pathologic stage discrepancy in bladder cancer patients treated with radical cystectomy: results from the national cancer data base. *Int J Radiat Oncol Biol Phys* **88**, 1048-1056 (2014).

18. Zehir, A., *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* (2017).
19. Kobatake, K., *et al.* Kdm6a Deficiency Activates Inflammatory Pathways, Promotes M2 Macrophage Polarization, and Causes Bladder Cancer in Cooperation with p53 Dysfunction. *Clin Cancer Res* **26**, 2065-2079 (2020).
20. Kaneko, S. & Li, X. X chromosome protects against bladder cancer in females via a KDM6A-dependent epigenetic mechanism. *Sci Adv* **4**, eaar5598 (2018).
21. Rachakonda, P.S., *et al.* TERT promoter mutations in bladder cancer affect patient survival and disease recurrence through modification by a common polymorphism. *Proc Natl Acad Sci U S A* **110**, 17426-17431 (2013).
22. Isharwal, S., *et al.* Prognostic Value of TERT Alterations, Mutational and Copy Number Alterations Burden in Urothelial Carcinoma. *Eur Urol Focus* **5**, 201-204 (2019).
23. Beckman, G., *et al.* Is p53 polymorphism maintained by natural selection? *Hum Hered* **44**, 266-270 (1994).
24. Dumont, P., Leu, J.I., Della Pietra, A.C., 3rd, George, D.L. & Murphy, M. The codon 72 polymorphic variants of p53 have markedly different apoptotic potential. *Nat Genet* **33**, 357-365 (2003).
25. Barnoud, T., Parris, J.L.D. & Murphy, M.E. Common genetic variants in the TP53 pathway and their impact on cancer. *J Mol Cell Biol* **11**, 578-585 (2019).
26. Basu, S., *et al.* Mutant p53 controls tumor metabolism and metastasis by regulating PGC-1alpha. *Genes Dev* **32**, 230-243 (2018).
27. Kurodo, Y., Tsukino, H., Nakao, H., Imai, H. & Kato, T. p53 codon 72 polymorphism and urothelial cancer risk. *Cancer Letters* **189**, 77-83 (2003).
28. Shinohara, A., *et al.* Association of TP53 and MDM2 polymorphisms with survival in bladder cancer patients treated with chemoradiotherapy. *Cancer Sci* **100**, 2376-2382 (2009).
29. Birkenkamp-Demtroder, K., *et al.* Genomic Alterations in Liquid Biopsies from Patients with Bladder Cancer. *Eur Urol* **70**, 75-82 (2016).
30. Christensen, E., *et al.* Liquid Biopsy Analysis of FGFR3 and PIK3CA Hotspot Mutations for Disease Surveillance in Bladder Cancer. *Eur Urol* **71**, 961-969 (2017).
31. Patel, K.M., *et al.* Association Of Plasma And Urinary Mutant DNA With Clinical Outcomes In Muscle Invasive Bladder Cancer. *Sci Rep* **7**, 5554 (2017).
32. Birkenkamp-Demtroder, K., *et al.* Monitoring Treatment Response and Metastatic Relapse in Advanced Bladder Cancer by Liquid Biopsy Analysis. *Eur Urol* (2017).
33. Christensen, E., *et al.* Early Detection of Metastatic Relapse and Monitoring of Therapeutic Efficacy by Ultra-Deep Sequencing of Plasma Cell-Free DNA in Patients With Urothelial Bladder Carcinoma. *J Clin Oncol* **37**, 1547-1557 (2019).
34. Vandekerkhove, G., *et al.* Circulating Tumor DNA Reveals Clinically Actionable Somatic Genome of Metastatic Bladder Cancer. *Clin Cancer Res* **23**, 6487-6497 (2017).

35. Berlin, A., *et al.* NBN gain is predictive for adverse outcome following image-guided radiotherapy for localized prostate cancer. *Oncotarget* **5**, 11081-11090 (2014).
36. Wang, F., *et al.* Evaluation of POLE and POLD1 mutations as biomarkers for immunotherapy outcomes across multiple cancer types. *JAMA Oncol* **5**, 1504 - 1506 (2019).
37. Jamal-Hanjani, M., *et al.* Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl J Med* (2017).
38. Sonpavde, G., *et al.* Circulating cell-free DNA profiling of patients with advanced urothelial carcinoma. *J Clin Oncol* **34**, abstr358 (2016).
39. Agarwal, N., *et al.* Characterization of metastatic urothelial carcinoma via comprehensive genomic profiling of circulating tumor DNA. *Cancer* **124**, 2115-2124 (2018).
40. McGregor, B.A., *et al.* Correlation of circulating tumor DNA (ctDNA) assessment with tissue-based comprehensive genomic profiling (CGP) in metastatic urothelial cancer (mUC). *J Clin Oncol* **36**, 453-453 (2018).
41. Abbosh, P.H. & Plimack, E.R. Molecular and Clinical Insights into the Role and Significance of Mutated DNA Repair Genes in Bladder Cancer. *Bladder Cancer* **4**, 9-18 (2018).
42. Plimack, E.R., *et al.* Defects in DNA Repair Genes Predict Response to Neoadjuvant Cisplatin-based Chemotherapy in Muscle-invasive Bladder Cancer. *Eur Urol* (2015).

4.9. Appendix

A4.1 Histological review of slides

CoMB	GTB Number	Assessment of slide 0		Assessment of final slide
		Histopath	Tumour content %	
1001	15373	solid TCC	70	same
1002	15367	solid TCC	80	same
1002	15371	solid TCC	70	same
1003	15390	solid TCC	60	same but one region lost
1003	15389	solid TCC	40	same but one region lost
1006	15385	micropapillary TCC & solid TCC	80	same but only one region
1007	15398	solid TCC / ?squamous	70	same but one cutting out
1007	15386	solid TCC / ?squamous	70	same
1008	15430	Solid TCC. Desmoplasia ++	50	same but one cutting out
1010	15428	solid TCC	70	same but one cutting out
1013	15459	solid TCC	80	same
1013	15460	solid TCC	60	2 regions cutting out
1014	16017	sarcomatoid TCC	100	same
1014	16019	giant cell TCC	100	same
1016	15442	solid TCC	70	same but one cutting out
1017	15450	solid TCC	50- but crushed++	badly cutting out
1020	15432	solid TCC	80	badly cutting out
1021	15426	Solid TCC. Desmoplasia ++	50	same
1021	15427	Solid TCC. Desmoplasia ++	50	same
1025	15418	solid TCC	80	same
1033	16036	solid/glandular TCC	80	same but cutting out
1033	16037	solid TCC	60	same but cutting out
1035	16087 = 16215	solid TCC	70	very different

A4.2 Test DNA extraction from FFPE

DNA extracted from patients 1033 and 1013 in a preliminary test to optimise procedure if necessary, assess likely yields and inform future processing of samples. DNA extraction protocol as per Appendix A4.3



1033 GTB 16036



1033 GTB 16037



1013 GTB 15459



1013 15460

CoMB	GTB	Region	Histology	Tumour content	DNA yield (ng/ul)	Total (ng)
1033	16036	1A	TCC	80%	1A1 = 47.8 1A2 = 14.0	2719
1033	16037	2A	TCC	60%	2A1 = 4,64 2A2 = 1.15	259

CoMB	GTB	Region	Histology	Tumour content	DNA yield (ng/ul)	Total (ng)
1013	15459	1A	TCC	80%	1A1 = 11.6 1A2 = 1.92	2090
1013	15459	1B	TCC	80%	1B1 = 4.6 1B2 = 0.956	
1013	15459		TCC	80%	1C1 = 23.6 1C2 = 3.79	
1013	15460		TCC	60%	2A1 = 82 2A2 = 18.1	5126
1013	15460		TCC	60%	2B1 = 8.36 2B2 = 1.52	
1013	15460		TCC	60%	2C1 = 4.56 2C2 = 1.24	

A4.3 Protocol: Deparaffinisation and DNA extraction from FFPE

Deparaffinisation protocol adapted from Dr Inma Spiteri (Post-Doc in Dr Sottirivo's team) Zymo Quick-DNA FFPE Kit- amendments as advised by Dr Inma Spiteri. Main adaption was performing a second elution using the initial filtrate to ensure maximum yield of DNA. As test extractions confirmed there was additional DNA to capture using a second elution, this was performed for all samples.

- Fill a coplin jar with xylene
- Dip the FFPE slide into the xylene until paraffin removed

- Dip the FFPE slides into 100% ethanol in a 2nd coplin jar
- Place the slides into a coplin jar filled with 2 x de-ionised water
- Macrodissect samples directly into 1.5ml DNA-LoBind Eppendorfs containing digestion mix (45ul nuclease-free water, 45ul 2x digestion buffer, 10ul Proteinase K)
- Following macrodissection, incubate samples for 12-16 hours at 55C
- Transfer samples to heatblock set to 94C and incubate for 20 minutes
- Remove samples and add 5ul RNase A. Pipette up and down to mix
- Incubate for 5 minutes at room temperature
- Add 350ul Genomic Lysis Buffer to each sample and vortex to mix
- Add 135ul isopropanol to each sample and mix thoroughly. Centrifuge at >12 000 xg for 1 minute
- Transfer supernatant to a Zymo-Spin IIC Column in a collection tube. Centrifuge at 10 000 xg for 1 minute
- Keep filtrate but transfer spin column to a new collection tube
- Add 400ul of Genomic DNA Wash 1 to the spin column and centrifuge at 10 000 x g for 1 minute. Discard the flow-through
- Add 700ul of Genomic DNA Wash 2 to the spin column and centrifuge at 12 000 x g for 1 minute. Discard the flow-through
- Add 200 ul of Genomic DNA Wash 2 to the spin column and centrifuge at 12 000 x g for 1 minute
- Transfer spin column to a 1.5ml DNA LoBind Eppendorf. Add 50ul pre-warmed (55C) DNA elution buffer to each sample to elute. Incubate for 5 minutes at room temperature
- Centrifuge at top speed for 30 seconds to elute the DNA
- Transfer filtrate previously kept aside to the same spin column. Centrifuge at 10 000 xg for 1 minute
- Add 400ul of Genomic DNA Wash 1 to the spin column and centrifuge at 10 000 x g for 1 minute. Discard the flow-through
- Add 700ul of Genomic DNA Wash 2 to the spin column and centrifuge at 12 000 x g for 1 minute. Discard the flow-through
- Add 200 ul of Genomic DNA Wash 2 to the spin column and centrifuge at 12 000 x g for 1 minute

- Transfer spin column to a new 1.5ml DNA LoBind Eppendorf. Add 50ul pre-warmed (55C) DNA elution buffer to each sample to elute. Incubate for 5 minutes at room temperature
- Centrifuge at top speed for 30 seconds to elute the DNA
- Quantify all samples using Qubit fluorometer
- Store at -20C

NB: 1st filtrate was labelled as CoMB ID, region number then 1. 2nd filtrate was labelled as CoMB ID, region number then 2

A4.4 Protocol: Germline DNA extraction from buffy coat

Amended from manufacturer's protocol- Qiagen DNA Mini Blood Kit

- Thaw samples at room temperature
- Pipette 20ul proteinase K into a 1.5ml LoBind Eppendorf for each sample
- Add 200ul buffy coat to the Eppendorf
- Add 200ul Buffer AL to the Eppendorf; mix by pulse-vortexing for 15 seconds
- Incubate at 56C for 10 minutes
- Centrifuge briefly to collect solution
- Add 200ul 100% ETOH and pulse-vortex for 15s
- Apply mixture to a QIAamp Minispin column in a 2ml collection tube. Close cap and spin at 6000 xg for 1 minute. Put column in clean collection tube. Discard the filtrate
- Add 500ul Buffer AW1. Close cap and spin at 6000 xg for 1 minute. Put column in clean collection tube. Discard the filtrate
- Add 500ul Buffer AW2. Close cap and spin at full speed for 3 minutes
- Place column in new collection tube and spin at full speed for 1 minute
- Place column in 1.5ml LoBind Eppendorf. Add 200ul distilled water and incubate at room temperature for 5 minutes.
- Spin at 6000xg to elute the DNA
- Store at -20C

A4.5 Protocol: Sureselect XT2 (pooled capture) for whole exome libraries

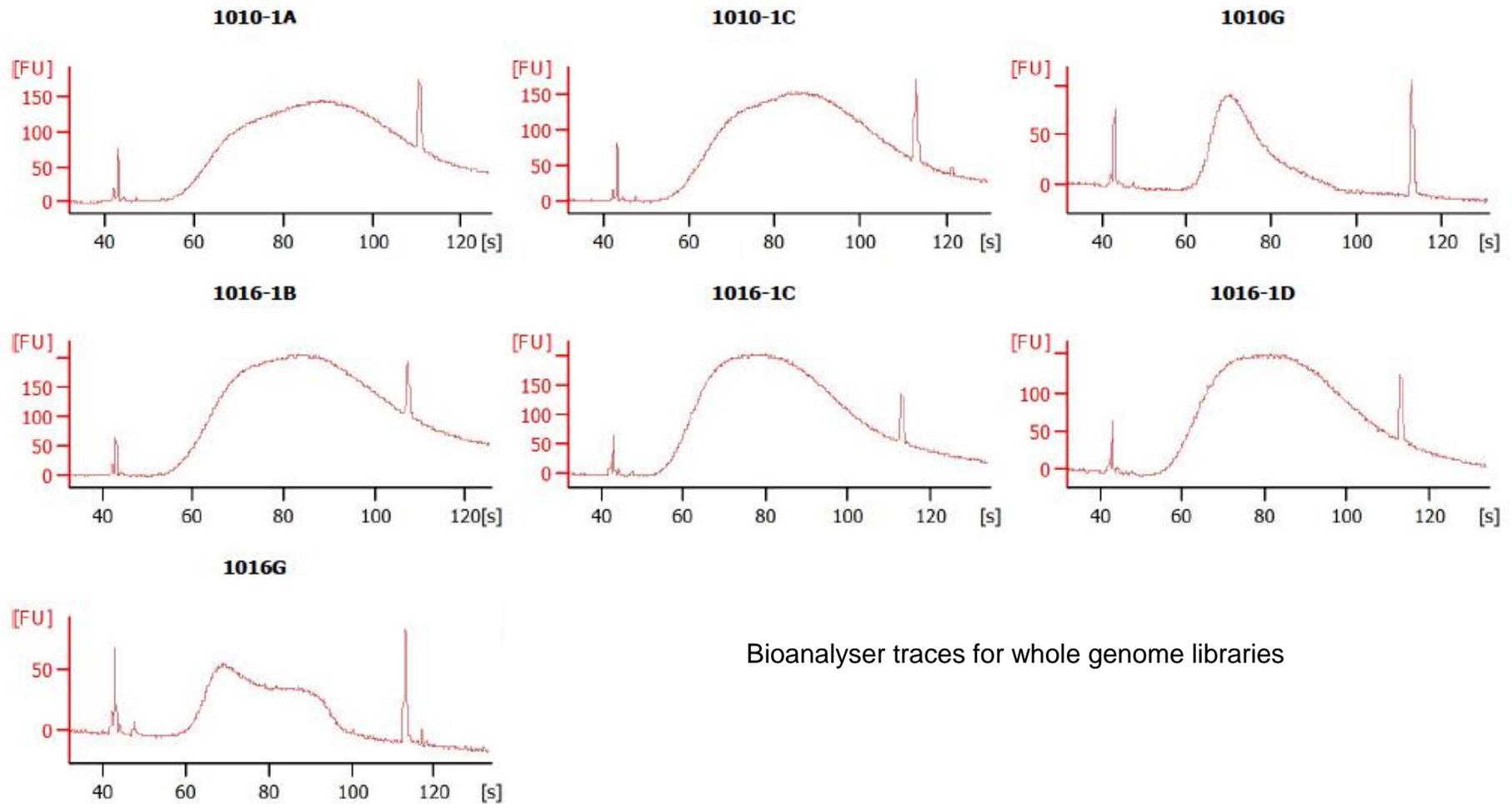
A total of 7 samples from patients 1010 and 1016 were selected as high yields of DNA had been obtained.

Patient	Region	Total DNA (ng)
1010	1A	1583
	1B	520
	1C	1324
1016	1A	838
	1B	1020
	1C	1388
	1D	1153

Protocol performed in accordance with manufacturer's instructions with a few amendments. In brief,

- 500ng input for each sample
- Samples sonicated using 2 x 5 minute cycles on Diagenode for tumour DNA and 2 x 12 minute cycles for germline DNA
- Ends repaired using 50ul of SureSelect End Repair Reaction mix; mixed by pipetting up and down
- Incubate at 20C for 30 minutes
- Sample purified using AMPure XP beads
- 2ul of SureSelect dA-tailing master mix applied to each sample; mixed by pipetting up and down
- Plate placed in thermal cycler: 37C for 30 minutes then 60C for 10 minutes
- Pre-capture indexing adaptor ligated
- Indexed DNA purified using AMPure XP beads
- Index library amplified; 7 cycles performed in accordance with optimisation protocol performed by Dr Spiteri
- Library amplified using AMPure XP beads; assessed with Bioanalyser

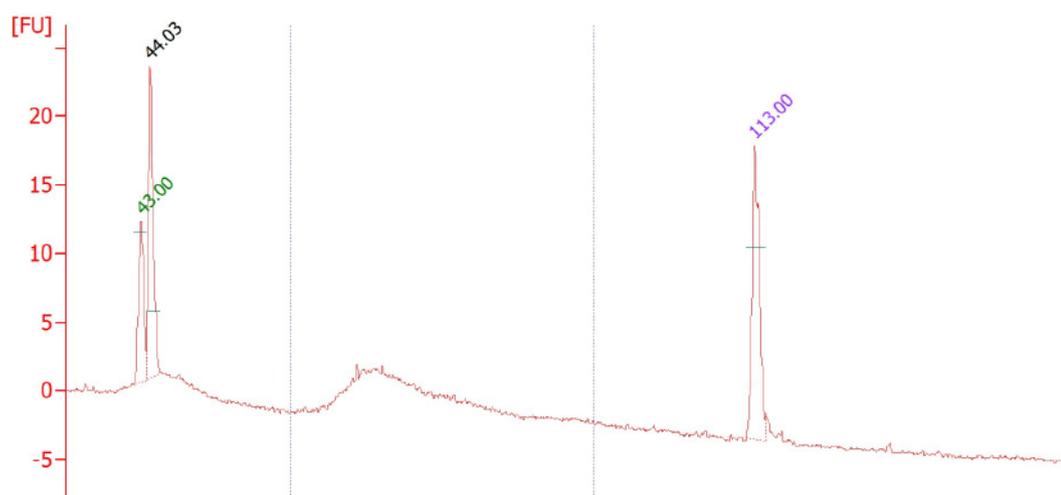
- gDNA library pools hybridised to capture (uneven amounts given my ai to achieve different levels of coverage for germline and tumour DNA)
- Hybridised DNA captured using streptavidin beads
- Captured libraries amplified (8 cycles)
- Amplified libraries purified using AMPure XP beads
- Bioanalyser to assess libraries



Bioanalyser traces for whole genome libraries

Patient	Region	Index	Planned gDNA input (ng)	Concentration (ng/ul)	Amount available (ng)
1010	1A	B02	187.5	9.16	219.84
	1C	C02	375	10.3	247.2
	Germline	D02	93.75	4.20	100.8
1016	1B	E02	187.5	11.4	273.6
	1C	F02	187.5	13.7	328.8
	1D	G02	375	10.7	256.8
	Germline	H02	93.75	3.38	81.12

Due to gDNA library yields, it was not possible to pool the libraries as planned. As the total amount obtained was the minimum total gDNA input of 1500ng, I proceeded with exome capture using the inputs available.



Bioanalyser trace for pooled library. 1.5nM in 20ul

Insufficient for the half Nova flowcell allocated; need minimum 2.0nM in 75ul.
Dr Spiteri's libraries were sufficient therefore issue with my samples.

A4.6 Protocol: Agilent XT HS whole exome library kit (and FFPE repair)

Amended version of manufacturer's protocol + Covaris shearing instructions with tables from official Agilent protocol (vA1 July 2017).

Step 3 (FFPE repair) was added for batch 2; details amended from NEB protocol

DAY 1

- 1) Prepare samples
 - for tumour and germline samples, prepare 210ng in 51ul (excess for sonication/transfer loss). Volume made up with low TE

- 2) Shear the DNA
 - Fill Covaris with water to 5L mark
 - Turn machine on and de-gas- may take up to an hour
 - Transfer all of the sample ie.51ul into microtubes- push pipette through white lid. LABEL MICROTUBES. MAX 8 microtubes in one run therefore if >8, need multiple runs
 - Put microtubes in red rack and screw down to secure
 - Press load position when Covaris ready. Sit red rack on arm
 - Select method and edit. Check sample position, check parameters
 - o For FFPE DNA
 - Duty factor 10%
 - Peak incident power (PIP) 175
 - Cycles per burst 200
 - Treatment time = 240 seconds
 - Bath temp = 2-8deg
 - o For germline (buffy coat therefore high-quality)
 - As above as discussed with team at lab meeting 02.05.18
 - Press save then run
 - Once finished- press load position to remove samples.
 - Press service position to empty water. Press 'yes' to home?
 - Degas on- gets rid of extra water. Degas off. Service position. Use syringe in left hand drawer to get rid of water on top of black tray.
 - Exit. Shut down laptop. Turn off machine

- 3) FFPE DNA repair

- NEB FFPE DNA repair kit. TAKE OUT AMPURE BEADS BEFORE!!!
- Transfer sample to well (make up volume to 53.5ul)
- Add 6.5ul FFPE DNA Repair buffer and 2ul FFPE DNA repair mix
- Mix by pipetting up and down, spin to collect liquid
- Incubate at 20C for 15 mins
- Make up 80% ETOH
- Vortex AMPure beads for 1 min
- Add 186ul beads to sample and vortex
- Incubate for 5 mins at room temp
- Magnetic stand (5mins)
- Remove and discard cleared solution
- Add 200ul 80% ETOH to each sample while in magnetic stand, incubate for 30s then remove alcohol. Repeat and ensure all ETOH removed
- Air dry for 5 mins on stand (do not overdry)
- Remove from stand
- Elute in 50ul (volume needed for XTHS library prep) low TE
- Seal, vortex and incubate for 2 mins at room temp
- Magnetic stand
- Transfer entire volume to new well

4) Repair and dA-tail the DNA ends

- Thaw on ice
 - o End-repair A Tailing buffer (yellow cap)
 - o Ligation buffer (purple cap)
- Make ligation master mix- needs to be at room temp for 30-45 mins before use in step 5. When ready to make mix, take out T4 DNA ligase (blue cap) and put on ice
- Vortex thawed ligation buffer (purple cap) for 5s at high speed (NB- very viscous so vortex well before use and mix well using pipette set to 80% total volume to pipette up and down 15-20 times!)
- For ligation mix

Reagent	Volume for 1 reaction	Volume for 8 reactions (includes excess)
Ligation Buffer (purple cap)	23 µl	207 µl
T4 DNA Ligase (blue cap)	2 µl	18 µl
Total	25 µl	225 µl

- Pre-program thermal cycler with heated lid on with following-

Step	Temperature	Time
Step 1	20°C	15 minutes
Step 2	72°C	15 minutes
Step 3	4°C	Hold

* When setting up the thermal cycling program, use a reaction volume setting of 70 μ L.

- Press start then pause to allow lid to heat up
- Vortex end repair A tailing buffer (yellow cap) for 5s at high speed. Make sure all solids dissolved otherwise keep vortexing
- Prepare End repair/dA-tailing master mix- mix by pipetting up and down 15-20 times. Spin briefly and KEEP ON ICE

Reagent	Volume for 1 reaction	Volume for 8 reactions (includes excess)
End Repair-A Tailing Buffer (yellow cap or bottle)	16 μ l	144 μ l
End Repair-A Tailing Enzyme Mix (orange cap)	4 μ l	36 μ l
Total	20 μl	180 μl

- Transfer repaired sample to 96 well plate (Label plate and note positions!). If <50ul, make up volume using low TE
- Add 20ul of end repair/dA-tailing master mix to each sample well. Pipette up and down 15-20 times using pipette set at 60ul
- Seal plate, spin samples then place in thermal cycler
- Press play!
- Get ice in cool box ready for plate!
- Thaw the adaptor oligo mix (white cap)
- Take AMPure beads out of fridge

5) Ligate the molecular-barcoded adaptor

- Once at 4deg hold, transfer samples to ice
- Preprogram thermal cycler with heated lid on as below. Press start then pause to allow lid to heat up

Step	Temperature	Time
Step 1	20°C	30 minutes
Step 2	4°C	Hold

* When setting up the thermal cycling program, use a reaction volume setting of 100 μ L.

- Add 25ul of the ligation master mix to each sample. Mix by pipetting up and down 15-20 times using a pipette set at 85ul.

- Add 5ul Adaptor oligo mix (white cap) to each sample. Mix by pipetting up and down 15-20 times using a pipette set at 85ul
 - Seal plate. Spin samples down and place in thermal cycler. Press play
 - Prepare 70% ethanol for use in step 9
 - Thaw pre-capture PCR mix reagents and indexes (in -20 pre PCR pack) on ice
 - o Herculase II Fusion DNA Polymerase (red cap)
 - o 5xHerculase II reaction buffer (clear cap)
 - o 100 nM dNTP Mix (green cap)
 - o Forward primer (brown cap)
 - o Indexes to be used (black caps)
- 6) Purify sample using AMPure XP beads
- Vortex AMPure beads for 1 min
 - Add 80ul AMPure beads to each sample. Pipette up and down to mix
 - Incubate for 5 mins at room temp
 - Put in magnetic rack and wait 5-10 mins for solution to clear
 - Remove and discard cleared solution from each well. DO NOT DISTURB THE BEADS
 - Leave plate in magnetic rack and add 200ul of 70% ETOH to each sample
 - Wait 1 min then remove the ethanol. Repeat once ie. Add 200ul ETOH to each sample, wait 1 min then remove. Ensure all ETOH removed- use P20 pipette if needed.
 - Air dry the samples (typically around 4 mins in PCR hood). DO NOT OVERDRY BEADS
 - Add 35ul nuclease-free water to each sample well
 - Seal plate and vortex. Spin briefly.
 - Incubate for 2 mins at room temp
 - Take seal off and put in magnetic rack. Wait 5 mins for solution to clear
 - Transfer all supernatant (approx. 34.5ul) to new well- do two rounds using pipette set at 17.25ul.
- 7) Amplify adaptor-ligated library
- Take out POST PCR beads from fridge
 - Preprogram thermal cycler with heated lid on as below- NB- POST PCR

Table 16 Pre-Capture PCR Thermal Cycler Program*

Segment	Number of Cycles	Temperature	Time
1	1	98°C	2 minutes
2	8 to 14, based on input DNA quality and quantity (see Table 17)	98°C	30 seconds
		60°C	30 seconds
		72°C	1 minute
3	1	72°C	5 minutes
4	1	4°C	Hold

* When setting up the thermal cycling program, use a reaction volume setting of 50 µL.

- For FFPE DNA with 200ng input, use 9 cycles
- For germline DNA with 200ng input, use 8 cycles
- Make pre-capture PCR reaction mix ON ICE

Table 18 Preparation of Pre-Capture PCR Reaction Mix

Reagent	Volume for 1 reaction	Volume for 8 reactions (includes excess)
5× Herculase II Reaction Buffer (clear cap)	10 µl	90 µl
100 mM dNTP Mix (green cap)	0.5 µl	4.5 µl
Forward Primer (brown cap)	2 µl	18 µl
Herculase II Fusion DNA Polymerase (red cap)	1 µl	9 µl
Total	13.5 µl	121.5 µl

- Add 13.5ul of the PCR reaction mix to each sample
- Add 2ul of the appropriate index to each sample
- Cap the wells then vortex at high speed for 5s. Spin briefly
- Take plate and 70% ETOH down to post PCR
- Press play on thermal cycler and once thermal block reaches 98deg, put samples in and close lid
- Take bioanalyser reagents out if planning to do on Day 1

- 8) Purify amplified library
 - Vortex beads for 1 min
 - Add 50ul of beads to each sample well. Pipette up and down 15-20 times
 - Incubate for 5 mins at room temp
 - Put in magnetic rack and wait for solution to clear (5 mins)
 - Remove and discard cleared solution. DO NOT DISTURB BEADS
 - Add 200ul 70% ETOH to each well
 - Wait 1 min then remove ETOH
 - Repeat ie. Add 200ul ETOH then remove as much as poss (use P20)
 - Air dry samples. DO NOT OVERDRY
 - Add 15ul nuclease free water to each sample well
 - Seal plate. Vortex to mix and spin down briefly
 - Incubate for 2 mins at room temp
 - Remove seal. Put in magnetic rack and wait until solution clear (2-3 mins)
 - Transfer supernatant to new well (15ul)

- 9) BIOANALYSER / store at -20
 - Need 500-1000ng for hybridisation in a volume of 12ul (make up with nuclease free water)

DAY 2

- 1) Hybridise to capture library
 - Take out streptavidin beads from fridge
 - Get ice
 - Thaw reagents as below

Table 19 Reagents for Hybridization

Kit Component	Storage Location	Thawing Conditions	Where Used
SureSelect XT HS and XT Low Input Blocker Mix (blue cap)	SureSelect XT HS Target Enrichment Kit ILM Hyb Module, Box 2 (Post PCR),* -20°C	Thaw on ice	page 45
SureSelect RNase Block (purple cap)	SureSelect XT HS Target Enrichment Kit ILM Hyb Module, Box 2 (Post PCR),* -20°C	Thaw on ice	page 46
SureSelect Fast Hybridization Buffer (bottle)	SureSelect XT HS Target Enrichment Kit ILM Hyb Module, Box 2 (Post PCR),* -20°C	Thaw and keep at Room Temperature	page 47
Capture Library	-80°C	Thaw on ice	page 47

* May also be labeled as *SureSelect XT HS and XT Low Input Target Enrichment Kit ILM Hyb Module, Box 2 (Post PCR)*.

- Preprogram thermal cycler with heated lid on as below-

Table 20 Pre-programmed thermal cycler program for Hybridization*

Segment Number	Number of Cycles	Temperature	Time
1	1	95°C	5 minutes
2	1	65°C	10 minutes
3	1	65°C	1 minute
4	60	65°C [†]	1 minute
		37°C	3 seconds
5	1	65°C [†]	Hold

* When setting up the thermal cycling program, use a reaction volume setting of 30 µl (final volume of hybridization reactions during cycling in Segment 4).

† Reducing the hybridization temperature to 60°C (Segments 4 and 5) may improve coverage for AT-rich regions of some libraries.

- Start the program then pause immediately to allow lid to heat up
- Put 500-1000ng DNA in new plate and make up to 12ul using nuclease-free water
- Add 5ul Sureselect XTMS and XT Low input blocker mix (blue cap) to each sample
- Cap, vortex and spin down
- Transfer sealed samples to thermal cycler and press play
- NB- need to pause at step segment 3 to add reagents!
- Have 15 minutes to prepare RNase block and hybridisation mix
 - o 25% solution of Sureselect RNase block (1 part RNase block:3 parts water). Mix well and keep on ice

Table 22 Preparation of RNase Block solution

Reagent	Volume for 1 reaction	Volume for 8 reactions (includes excess)
SureSelect RNase Block	0.5 µl	4.5 µl
Nuclease-free water	1.5 µl	13.5 µl
Total	2 µl	18 µl

- o Hybridisation mix as below for whole exome (>3Mb)- combine at room temp. mix well by vortexing for 5s then spin down briefly. Use immediately!! DO NOT KEEP AT ROOM TEMP FOR EXTENDED PERIODS

Table 23 Preparation of Capture Library Hybridization Mix for Capture Libraries ≥ 3 Mb

Reagent	Volume for 1 reaction	Volume for 8 reactions (includes excess)
25% RNase Block solution (from step 5)	2 μ l	18 μ l
Capture Library ≥ 3 Mb	5 μ l	45 μ l
SureSelect Fast Hybridization Buffer	6 μ l	54 μ l
Total	13 μl	117 μl

- PRESS PAUSE ON THERMAL CYCLER ONCE REACHES SEGMENT 3 ie. After 15 mins!

Table 21 Thermal cycler program for Hybridization with required pause

Segment Number	Number of Cycles	Temperature	Time
1	1	95°C	5 minutes
2	1	65°C	10 minutes
3	1	65°C	1 minute (PAUSE cyclor here)
4	60	65°C	1 minute
		37°C	3 seconds
5	1	65°C	Hold*

* Begin the capture steps on page 49 when the thermal cycler starts the 65°C Hold segment.

- Open lid and keeping plate in thermal cycler, add 13ul of hybridisation mix to each sample well. Mix well by pipetting up and down 8-10 times (total now 30ul)
- Seal wells with fresh CAPS! NOT STICKY SHEET. MAKE SURE PROPERLY SEALED otherwise evaporation loss
- Vortex briefly and spin briefly to remove bubbles. Put back in thermal cycler
- Close lid and press play
- Set up other thermal cycler ready for prewarming buffer- 70deg with lid 80deg

2) Preparing streptavidin-coated magnetic beads- start about 45 mins after commencing hybridisation

- Vortex beads
- BATCH WASH
 - o For each sample, add 50ul of the beads to an Eppendorf
 - o Add 200ul of Sure select binding buffer (from room temp post PCR box) for each sample (ie. N x 200ul) to the same Eppendorf

- Pipette up and down 15-20 times
 - Put in a magnetic stand and wait for solution to clear (>5 mins)
 - Remove and discard the supernatant
 - Repeat the wash twice more (from adding binding buffer) so total 3 washes
 - Resuspend the beads in n x 200ul binding buffer
 - Put 200ul of suspended beads in wells of a new plate, ready to transfer sample over
- 3) Capture hybridised DNA using streptavidin-coated beads
- Remove samples from thermal cycler once reached hold and immediately transfer entire volume (30ul each) to each of the wells now containing beads USING A MULTICHANNEL PIPETTE
 - Pipette up and down 5-8 times then seal with fresh caps
 - Incubate on a mixer (1400-1800rpm) for 30 mins at room temp
 - Reset cycler to 70 deg
 - While mixing- pre-warm the buffer
 - Need 6 lots of 200ul Wash Buffer 2 per sample
 - Taking into account evaporation, aliquot required number of 250ul volumes of Wash Buffer 2 into a new plate in x number of vertical rows, and cap them in vertical rows
 - Put in thermal cycler to warm
 - Once 30-min incubation period over, briefly spin samples.
 - Place in magnetic rack and wait for solution to clear
 - Remove and discard supernatant
 - Resuspend beads in 200ul in Wash Buffer 1 (NOT WARMED). Pipette up and down x 15-20.
 - Put in magnetic stand. Remove and discard cleared solution
 - Remove from magnetic stand
 - Remove one strip of caps from pre-warmed buffer and use 100ul multichannel pipette to resuspend beads in 200ul prewarmed Wash Buffer 2
 - Pipette up and down 15-20 times
 - Seal wells with fresh caps and vortex at high speed for 8s
 - Spin briefly- do not pellet the beads- make sure still in suspension
 - Incubate for 5 mins at 70deg with heated lid on
 - Remove and place on magnetic stand
 - Wait 1 min for solution to clear then remove and discard supernatant
 - Repeat for total of 6 washes ie. Remove caps from next row of prewarmed buffer and add 200ul of prewarmed buffer to resuspend beads etc.....
- AFTER LAST WASH, MAKE SURE ALL BUFFER REMOVED
 - Put plate on ice
 - Add 25ul nuclease free water to each sample well. Pipette up and down 8 times to resuspend beads

- 4) Amplify captured libraries
 - Take out AMPure beads from fridge
 - Thaw reagents below and keep on ice

Table 26 Reagents for pre-capture PCR amplification

Component	Storage Location	Where Used
Herculase II Fusion DNA Polymerase (red cap)	SureSelectXT HS Target Enrichment Kit, ILM Hyb Module Box 2 (Post PCR), * -20°C	page 56
5× Herculase II Reaction Buffer (clear cap)	SureSelectXT HS Target Enrichment Kit, ILM Hyb Module Box 2 (Post PCR), -20°C	page 56
100 mM dNTP Mix (green cap)	SureSelectXT HS Target Enrichment Kit, ILM Hyb Module Box 2 (Post PCR), -20°C	page 56
SureSelect Post-Capture Primer Mix (clear cap)	SureSelectXT HS Target Enrichment Kit, ILM Hyb Module Box 2 (Post PCR), -20°C	page 56

* May also be labeled as *SureSelectXT HS and XT Low Input Target Enrichment Kit ILM Hyb Module Box 2 (Post PCR)*.

- Preprogram thermal cycler as below. 9 cycles for WES

Table 27 Post-capture PCR Thermal Cycler Program

Segment	Number of Cycles	Temperature	Time
1	1	98°C	2 minutes
2	9 to 14 See Table 28 for recommendations based on Capture Library size	98°C	30 seconds
		60°C	30 seconds
		72°C	1 minute
3	1	72°C	5 minutes
4	1	4°C	Hold

Table 28 Post-capture PCR cycle number recommendations

Capture Library Size/Description	Cycles
Libraries <0.2 Mb	14 cycles
Libraries 0.2–3 Mb (includes SSeI XT HS and XT Low Input ClearSeq Comp Cancer)	12 cycles
Libraries 3–5 Mb	10 cycles
Libraries >5 Mb (includes SSeI XT HS and XT Low Input Human All Exon V6 and Clinical Research Exome V2 libraries)	9 cycles

- Start the program then press pause to allow lid to heat up
- Make PCR mix as below ON ICE. Vortex to mix

Table 29 Preparation of post-capture PCR Reaction mix

Reagent	Volume for 1 reaction	Volume for 8 reactions (includes excess)
Nuclease-free water	12.5 µl	112.5 µl
5× Herculase II Reaction Buffer (clear cap)	10 µl	90 µl
Herculase II Fusion DNA Polymerase (red cap)	1 µl	9 µl
100 mM dNTP Mix (green cap)	0.5 µl	4.5 µl
SureSelect Post-Capture Primer Mix (clear cap)	1 µl	9 µl
Total	25 µl	225 µl

- Add 25ul of PCR mix to each sample (on ice)
- Mix by gently pipetting up and down- avoid splashing samples on wall. DO NOT SPIN

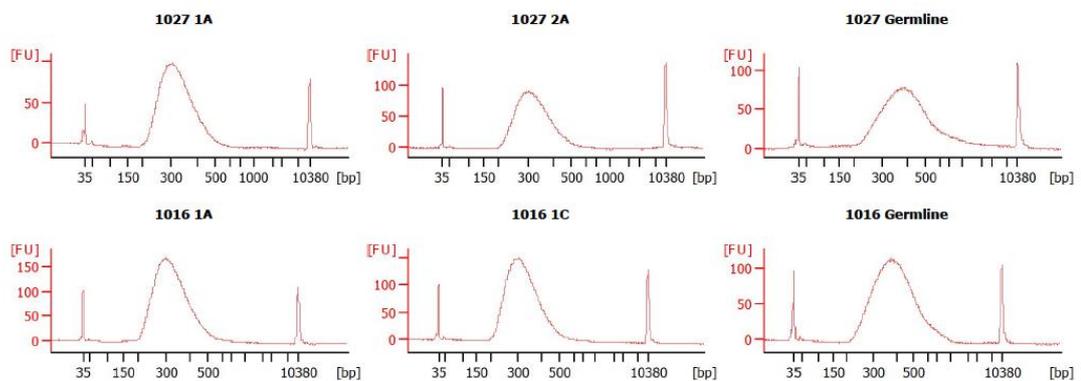
- Seal and put in thermal cycler. Press play
- If doing bioanalyser on Day 2 then take out reagents
- Prepare 70% ETOH (need 400ul per sample plus excess)
- Once complete, spin briefly then place in magnetic stand
- Wait 2 mins for solution to clear
- Transfer supernatant to new well (approx. 50ul) (use P20x2 and P10?)

5) Purify amplified captured libraries using AMPure beads

- Vortex beads
- Add 50ul beads to each sample. Pipette up and down x 15-20
- Incubate at room temp for 5 mins
- Put on magnetic stand and wait for solution to clear (3-5 mins)
- Remove and discard cleared solution
- Add 200ul 70% ETOH
- Wait 1 min then remove ETOH
- Repeat ETOH wash- ensure all ethanol removed after 2nd wash
- Air dry samples
- Add 25ul nuclease free water to each sample well
- Seal plate, vortex and briefly spin
- Incubate for 2 mins at room temp
- Magnetic stand for 2 mins until solution clear
- Transfer 25ul supernatant to new well!!

6) Bioanalyser

A4.7 Bioanalyser traces for whole exome library results



A4.8 FoundationOne panel of genes

ABL1	CCND3	ERG	GNAQ	MAGI2	PDGFRA	SLIT2
ABL2	CCNE1	ERRF1	GNAS	MAP2K1	PDGFRB	SMAD2
ACVR1B	CD274	ESR1	GPR124	MAP2K2	PKD1	SMAD3
AKT1	CD79A	EZH2	GRIN2A	MAP2K4	PIK3C2B	SMAD4
AKT2	CD79B	FAM46C	GRM3	MAP3K1	PIK3CA	SMARCA4
AKT3	CDC73	FANCA	GSK3B	MCL1	PIK3CB	SMARCB1
ALK	CDH1	FANCC	H3F3A	MDM2	PIK3CG	SMO
AMER1	CDK12	FANCD2	HGF	MDM4	PIK3R1	SNCAIP
APC	CDK4	FANCE	HNF1A	MED12	PIK3R2	SOCS1
AR	CDK6	FANCF	HRAS	MEF2B	PLCG2	SOX10
ARAF	CDK8	FANCG	HSD3B1	MEN1	PMS2	SOX2
ARFRP1	CDKN1A	FANCL	HSP90AA1	MET	POLD1	SOX9
ARID1A	CDKN1B	FAS	IDH1	MITF	POLE	SPEN
ARID1B	CDNK2A	FAT1	IDH2	MLH1	PPP2R1A	SPOP
ARID2	CDKN2B	FBXW7	IGF1R	MPL	PRDM1	SPTA1
ASXL1	CDKN2C	FGF10	IGF2	MRE11A	PREX2	SRC
ATM	CEBPA	FGF14	IKBKE	MSH2	PRKAR1A	STAG2
ATR	CHD2	FGF19	IKZF1	MSH6	PRKCI	STAT3
ATRX	CHD4	FGF23	IL7R	MTOR	PRKDC	STAT4
AURKA	CHEK1	FGF3	INHBA	MUTYH	PRSS8	STK11
AURKB	CHEK2	FGF4	INPP4B	MYC	PTCH1	SUFU
AXIN1	CIC	FGF6	IRF2	MYCL	PTEN	SYK
AXL	CREBBP	FGFR1	IRF4	MYCN	PTPN11	TAF1
BAP1	CRKL	FGFR2	IRS2	MYD88	QKI	TBX3
BARD1	CRLF2	FGFR3	JAK1	NF1	RAC1	TERC
BCL2	CSF1R	FGFR4	JAK2	NF2	RAD50	Tert Promoter
BCL2L1	CTCF	FH	JAK3	NFE2L2	RAD51	TET2
BCL2L2	CTNNA1	FLCN	JUN	NFKBIA	RAF1	TGFBR2
BCL6	CTNNB1	FLT1	KAT6A	NKX2-1	RANBP2	TNFAIP3
BCOR	CUL3	FLT3	KDM5A	NOTCH1	RARA	TNFRSF14
BCORL1	CYLD	FLT4	KDM5C	NOTCH2	RB1	TOP1
BLM	DAXX	FOXL2	KDM6A	NOTCH3	RBM10	TOP2A
BRAF	DDR2	FOXP1	KDR	NPM1	RET	TP53
BRCA1	DICER1	FRS2	KEAP1	NRAS	RICTOR	TSC1
BRCA2	DNMT3A	FUBP1	KEL	NSD1	RNF43	TSC2
BRD4	DOT1L	GABRA6	KIT	NTRK1	ROS1	TSHR
BRIP1	EGFR	GATA1	KLHL6	NTRK2	RPTOR	U2AF1
BTG1	EP300	GATA2	KMT2A	NTRK3	RUNX1	VEGFA
BTK	EPHA3	GATA3	KMT2C	NUP93	RUNX1T1	VHL
C11orf30	EPHA5	GATA4	KMT2D	PAK3	SDHA	WISP3
CARD11	EPHA7	GATA6	KRAS	PALB2	SDHB	WT1
CBFB	EPHB1	GID4	LMO1	PARK2	SDHC	XPO1
CBL	ERBB2	GLI1	LRP1B	PAX5	SDHD	ZBTB2
CCND1	ERBB3	GNA11	LYN	PBRM1	SETD2	ZNF217
CCND2	ERBB4	GNA13	LZTR1	PDCD1LG2	SF3B1	ZNF703

Panel includes select rearrangements of BCR, ETV1, ETV4-6, Tmprss2, microsatellite instability and tumour mutational burden

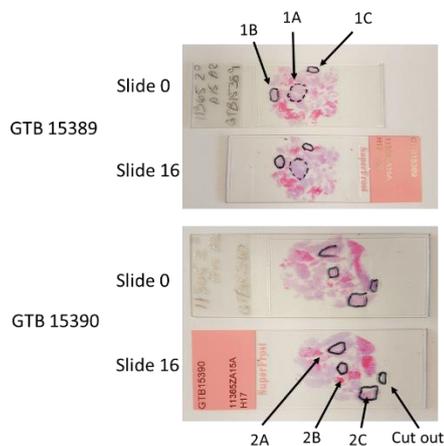
A4.9 Subset of 64 genes

Genes derived from TCGA significantly mutated gene list and DDR genes from literature

Genes			
ACTB	ERCC2	NFE2L2	STAG2
ARID1A	FANCD2	NRAS	TAF11
ASXL1	FAT1	NUP93	TMCO4
ASXL2	FBXW7	PALB2	TP53
ATM	FGFR3	PARD3	TSC1
BRCA1	GNA13	PIK3CA	USP28
BRCA2	HES1	PSIP1	ZBTB7B
BRIP1	HIST1H3B	PTEN	ZFP36L1
C3orf70	HRAS	RAD50	
CASP8	KANSL1	RB1	
CDKN1A	KDM6A	RBM10	
CDKN2A	KLF5	RHOA	
CREBBP	KMT2A	RHOB	
CUL1	KMT2C	RXRA	
ELF3	KMT2D	SF1	
EP300	KRAS	SF3B1	
ERBB2	MB21D2	SPN	
ERBB3	MBD1	SPTAN1	
	METTL3	SSH3	

A4.10 Histological assessment of FFPE blocks for WES & DNA yields

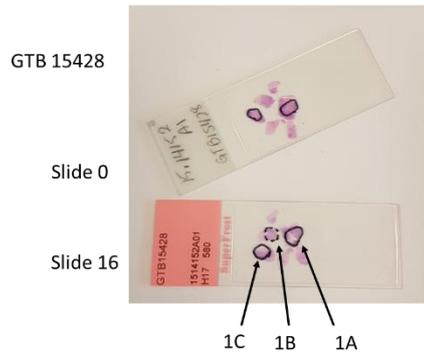
1003



Region	% TC	DNA extracted (ng)
1A	60	1646
1B	60	212
1C	60	35
2A	40	66
2B	40	172
2C	40	348
G	NA	

Notes:

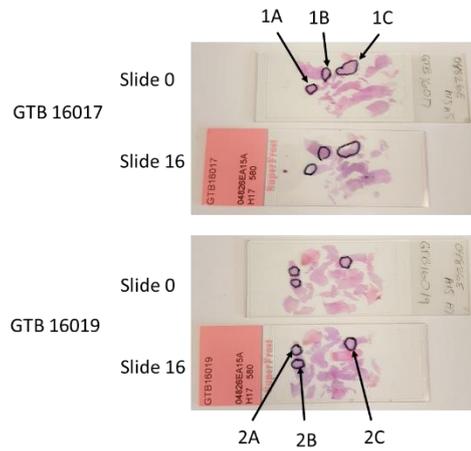
1010



Region	% TC	DNA extracted (ng)
1A	70	1583
1B	70	520
1C	70	1324
G	NA	

Notes:

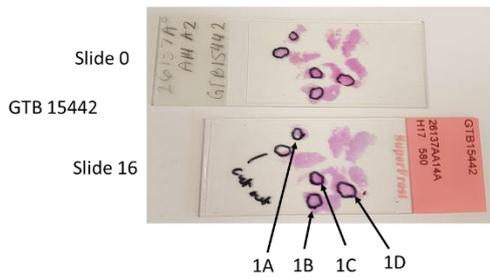
1014



Region	% TC	DNA extracted (ng)
1A	100	1179
1B	100	1993
1C	100	1685
2A	100	380
2B	100	2224
2C	100	1022
G	NA	

Notes:

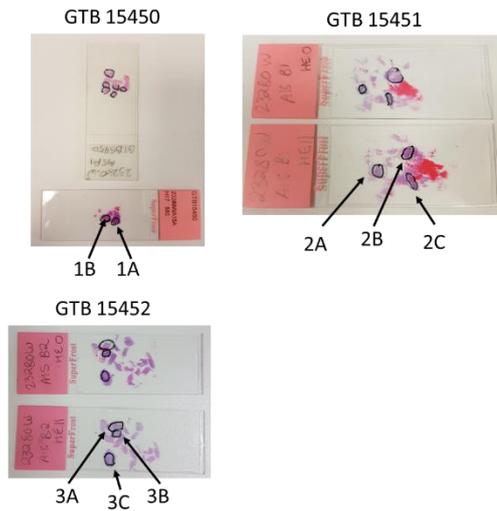
1016



Region	% TC	DNA extracted (ng)
1A	70	838
1B	70	1020
1C	70	1388
1D	70	1153
G	NA	

Notes:

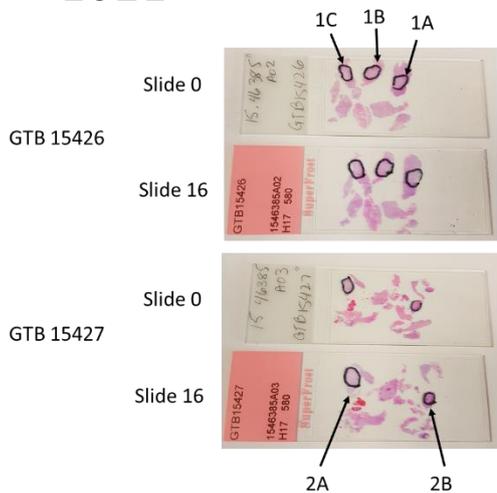
1017



Region	% TC	DNA extracted (ng)
1A	50	81.6
1B	50	81.3
2A	90	483.36
2B	90	2697
2C	90	385.92
3A	100	984
3B	100	483.36
3C	100	1883.04
4A	100	204
4B	100	354
G		

Notes: 1A/B- crushed ++

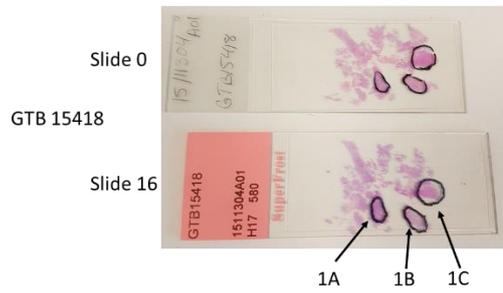
1021



Region	% TC	DNA extracted (ng)
1A	50	1244
1B	50	945
1C	50	995
2A	50	1279
2B	50	844
G	NA	

Notes:

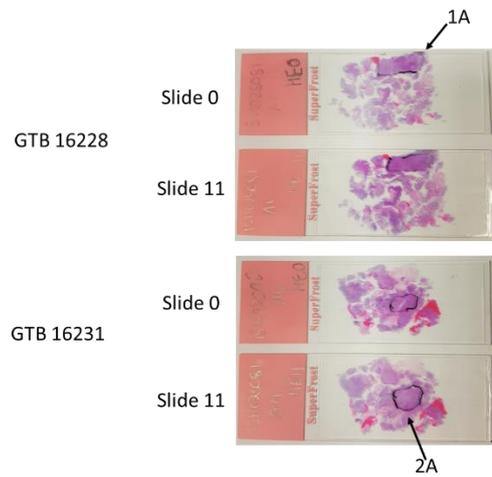
1025



Region	% TC	DNA extracted (ng)
1A	60	581
1B	60	210
1C	60	290
G	NA	

Notes:

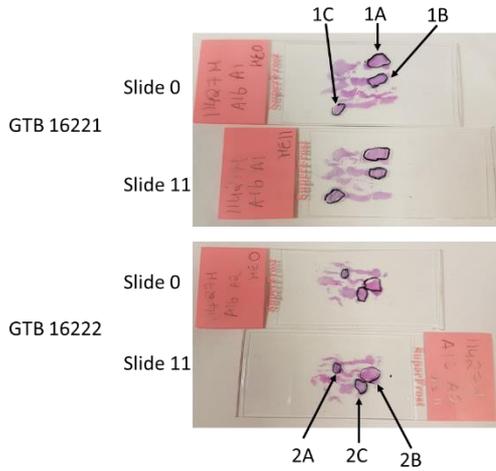
1027



Region	% TC	DNA extracted (ng)
1A	100	11048
2A	100	5735
G	NA	

Notes: XT HS WE library prep. 1A repeated

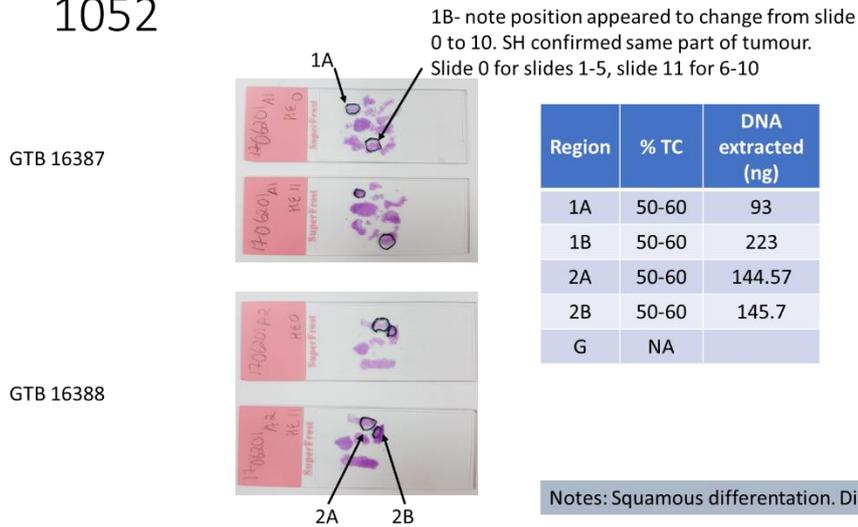
1039



Region	% TC	DNA extracted (ng)
1A	90	243
1B	90	50
1C	90	34
2A	80	44
2B	80	145
2C	80	39
G	NA	

Notes: SH had originally only marked one region per block (1A and 2B) as viable so not surprising v low yield from others

1052

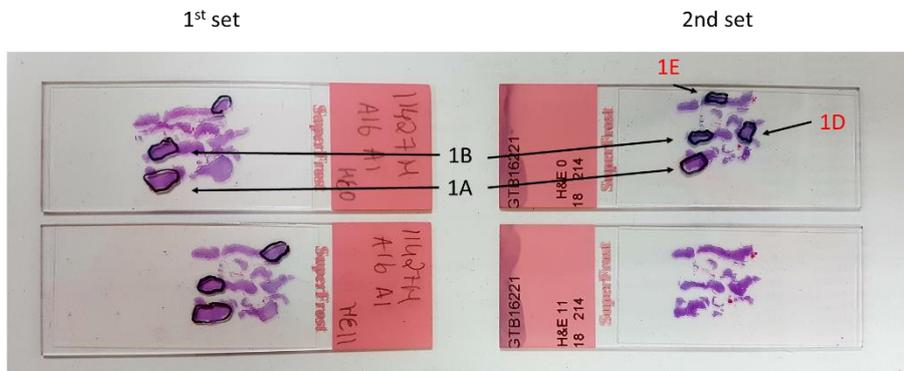


1B- note position appeared to change from slide 0 to 10. SH confirmed same part of tumour.
Slide 0 for slides 1-5, slide 11 for 6-10

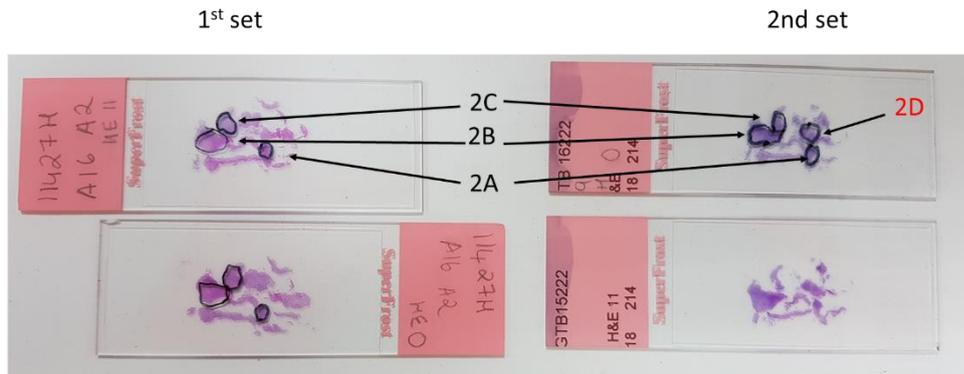
Region	% TC	DNA extracted (ng)
1A	50-60	93
1B	50-60	223
2A	50-60	144.57
2B	50-60	145.7
G	NA	

Notes: Squamous differentiation. Diathermy.

1039- resectioned (GTB 16221)



1039 resectioned- GTB 16222

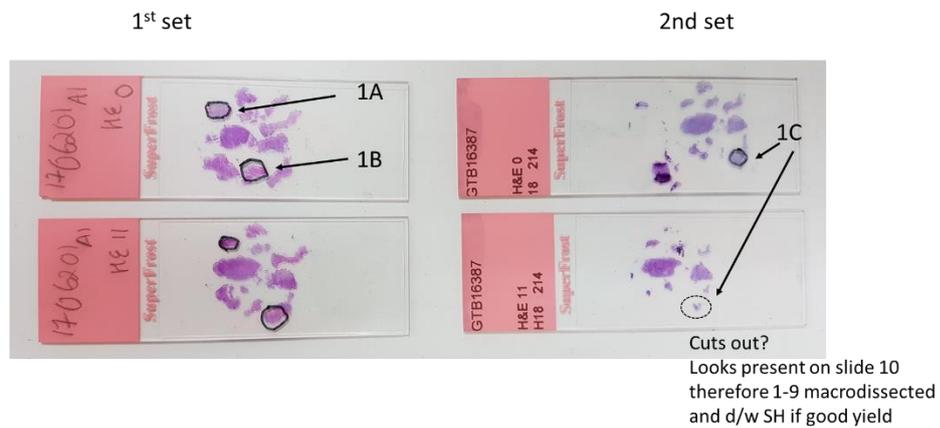


1039- resectioned

Region	Tumour Content	DNA extracted March 2018 (ng)	DNA extracted 04.04	Total
1A	90	243	35	278
1B	90	50	5	55
1C	90	34	NA	34
1D		NA	24	24
1E		NA	13	13
2A	80	44	17	61
2B	80	145	112	257
2C	80	39	10	49
2D		NA	38	38

Notes: SH had originally only marked one region per block (1A and 2B) as viable so not surprising v low yield from others. 2nd batch- diathermy effect ++

1052 resectioned- GTB 16387



1052 resectioned- GTB16388



1052- resectioned

Region	Tumour Content	DNA extracted March 2018	DNA extracted 04.04.18	Total
1A	50-60	93	NA	93
1B	50-60	223	NA	223
1C		NA	33	33
2A	50-60	144	NA	144
2B	50-60	145	NA	145
2C	10-20	NA	946	946

Notes: 2nd batch- diathermy effect ++

A4.11 Most aberrant genes from all WES data

Combining the final TPU reports for each sample (i.e. moderately and high impact MuTECT aberrations and filtered GATK results as detailed above), analysis revealed that of the 8624 genes, the most frequently aberrant gene was TTN with a total of 26 aberrations in 10 samples from 6 patients. The second most mutated gene was TP53 with 24 aberrations in 14 samples from 8 patients. MUC12 was third with 20 aberrations in 11 samples from 7 patients. A table showing the aberrant genes affecting at least 5/9 of the cohort is shown overleaf.

Gene	No of aberrations	No of samples affected	No of patients affected
TP53	24	14	8
MUC12	20	11	7
TTN	26	10	6
USP6	20	9	6
PCM1	11	9	6
RP11-467N20.5	8	8	6
WRN	7	7	5
SYNE1	8	8	5
ZNF208	7	6	5
PHF21A	6	5	5
RANBP2	11	8	5
MUC16	12	9	5
NBPF10	16	8	5
GOLGA6D	10	8	5
HYDIN	12	9	5
ERBB3	10	10	5
ANAPC1	10	8	5

A4.12 Table of results using subset of 64 genes

Gene	No. aberrations				No. samples	No. patients
	HIGH	MODERATE	LOW	TOTAL		
TP53	7	17	12	36	14	8
KMT2C	1	4	8	13	10	7
ERBB3	0	10	4	14	12	6
ERCC2	0	6	2	8	8	4
RHOA	0	6	2	8	8	4
FAT1	2	3	7	12	7	4
STAG2	2	4	3	9	7	4
ATM	2	5	2	9	7	4
KDM6A	2	2	3	7	7	4
BRCA1	2	4	7	13	6	4
NFE2L2	3	4	0	7	5	4
ASXL1	0	6	2	8	8	3
ELF3	1	5	0	6	5	3
ERBB2	0	6	7	13	4	3
SPTAN1	0	0	6	6	4	3
FANCD2	0	0	6	6	4	3
KMT2A	2	1	2	5	4	3
ARID1A	4	4	2	10	6	2
METTL3	0	6	0	6	6	2

Gene	No. aberrations				No. samples	No. patients
	HIGH	MODERATE	LOW	TOTAL		
KMT2D	4	2	4	10	4	2
PIK3CA	2	4	0	6	4	2
ASXL2	2	0	3	5	4	2
FGFR3	0	4	0	4	4	2
CDKN1A	3	2	0	5	3	2
KANSL1	0	3	2	5	3	2
CDKN2A	0	3	0	3	2	2
MBD1	0	2	0	2	2	2
SF1	0	8	0	8	4	1
CREBBP	2	2	1	5	2	1
FBXW7	4	0	0	4	2	1
TSC1	0	2	0	2	2	1
RB1	0	0	2	2	2	1
RHOB	0	2	0	2	2	1
HRAS	0	2	0	2	2	1
PTEN	2	0	0	2	2	1
SF3B1	0	2	0	2	2	1
GNA13	0	2	0	2	2	1
USP28	0	0	2	2	2	1
TAF11	0	2	0	0	2	1
RXRA	0	0	2	2	1	1
EP300	0	0	1	1	1	1
BRCA2	0	0	1	1	1	1
KLF5	0	0	1	1	1	1
MB21D2	0	1	0	1	1	1
SSH3	0	1	0	1	1	1
TMCO4	1	0	0	1	1	1
BRIP1	0	0	0	0	0	0
PALB2	0	0	0	0	0	0
RAD50	0	0	0	0	0	0
ZFP36L1	0	0	0	0	0	0
KRAS	0	0	0	0	0	0
PSIP1	0	0	0	0	0	0
C3orf70	0	0	0	0	0	0
ZBTB7B	0	0	0	0	0	0
PARD3	0	0	0	0	0	0
CUL1	0	0	0	0	0	0
NRAS	0	0	0	0	0	0
RBM10	0	0	0	0	0	0
ACTB	0	0	0	0	0	0
CASP8	0	0	0	0	0	0
HIST1H3B	0	0	0	0	0	0

Gene	No. aberrations				No. samples	No. patients
	HIGH	MODERATE	LOW	TOTAL		
ERBB2	0	0	0	0	0	0
NUP93	0	0	0	0	0	0
SPN	0	0	0	0	0	0
HES1	0	0	0	0	0	0
Grand Total	48	137	94	277		

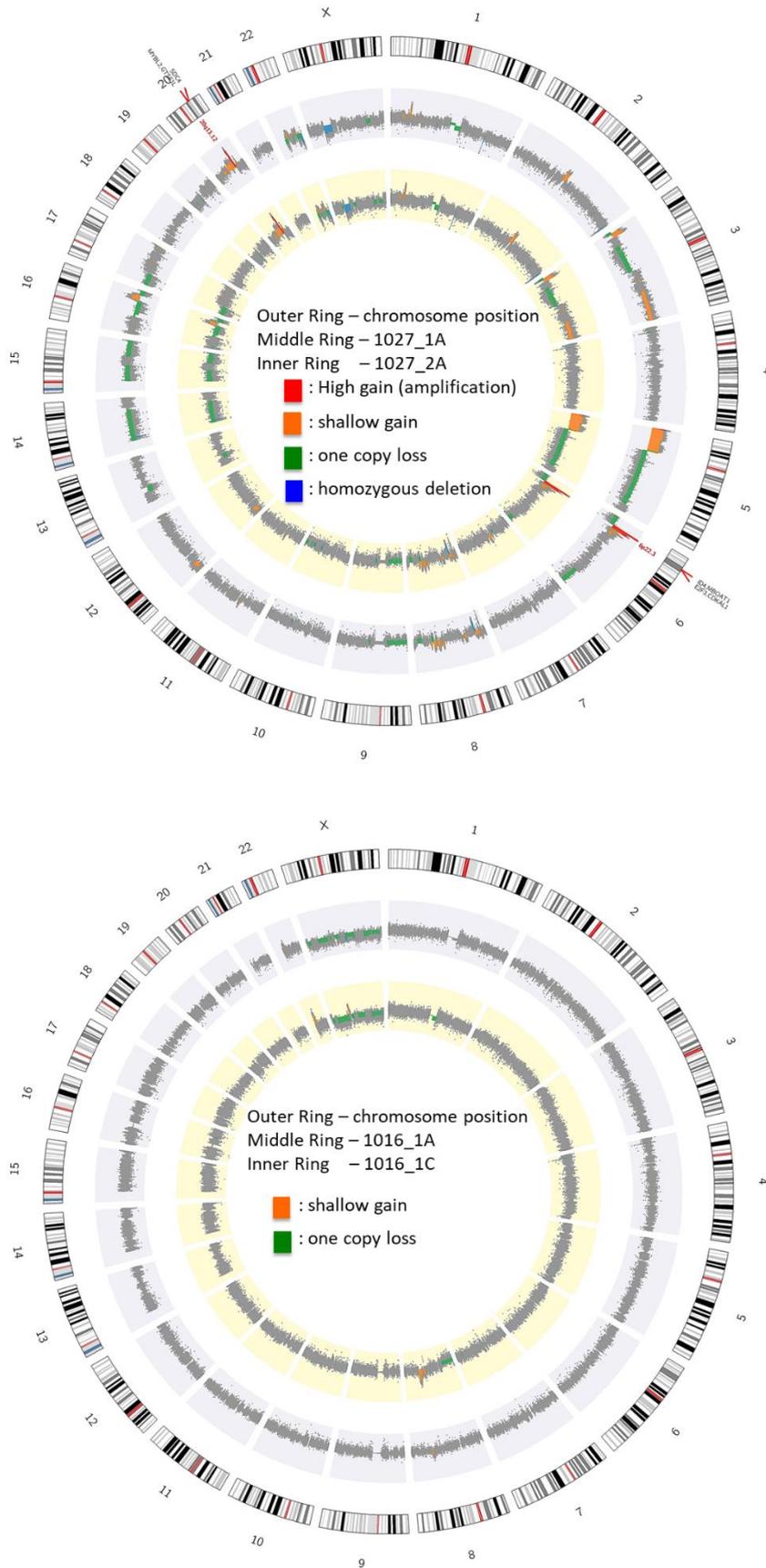
Table replicated but with likely polymorphisms removed

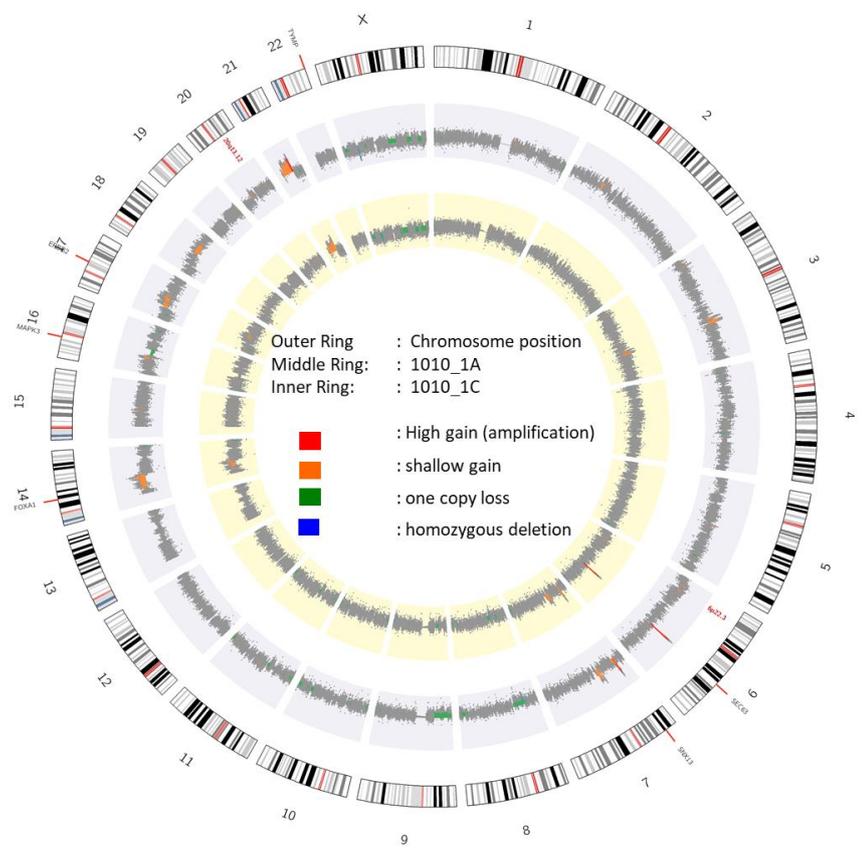
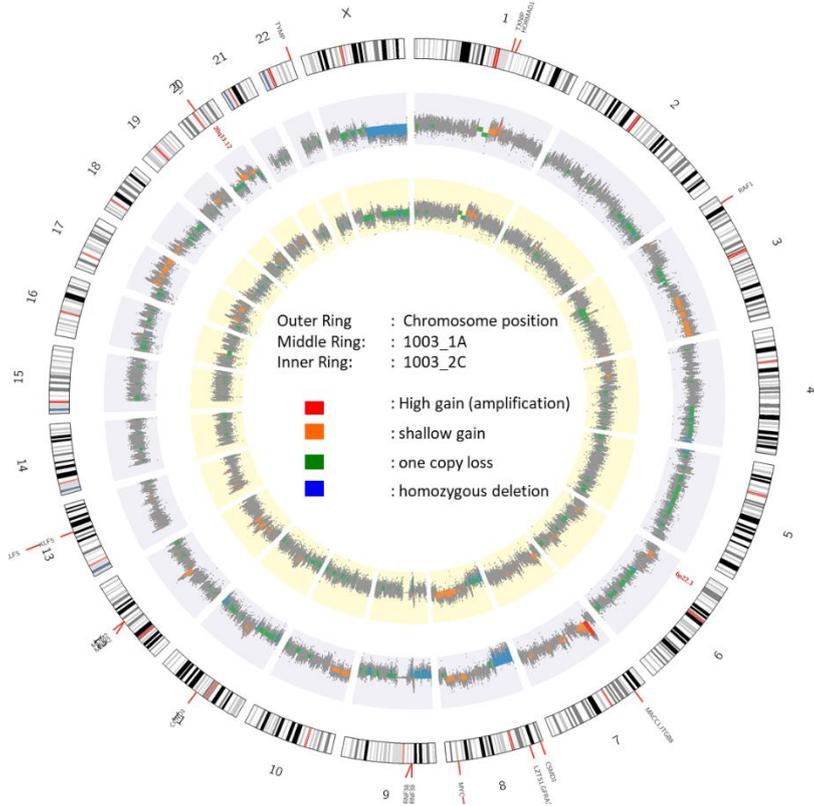
Gene	No. aberrations				No. samples	No. patients
	HIGH	MODERATE	LOW	TOTAL		
TP53	7	14	12	33	11	6
KMT2C	1	4	8	13	10	7
ERBB3	0	10	4	14	12	6
ERCC2	0	6	2	8	8	4
RHOA	0	6	2	8	8	4
FAT1	2	0	7	9	6	3
STAG2	2	4	3	9	7	4
ATM	2	4	2	8	6	3
KDM6A	2	0	3	5	5	3
BRCA1	2	1	7	10	5	3
NFE2L2	3	4	0	7	5	4
ASXL1	0	6	2	8	8	3
ELF3	1	4	0	5	4	2
ERBB2	0	5	7	12	3	2
SPTAN1	0	0	6	6	4	3
FANCD2	0	0	6	6	4	3
KMT2A	2	1	2	5	4	3
ARID1A	4	4	2	10	6	2
METTL3	0	6	0	6	6	2
KMT2D	4	2	4	10	4	2
PIK3CA	2	4	0	6	4	2
ASXL2	2	0	3	5	4	2
FGFR3	0	4	0	4	4	2
CDKN1A	3	0	0	5	3	2
KANSL1	0	0	2	2	2	1
CDKN2A	0	1	0	1	1	1
MBD1	0	0	0	0	0	0

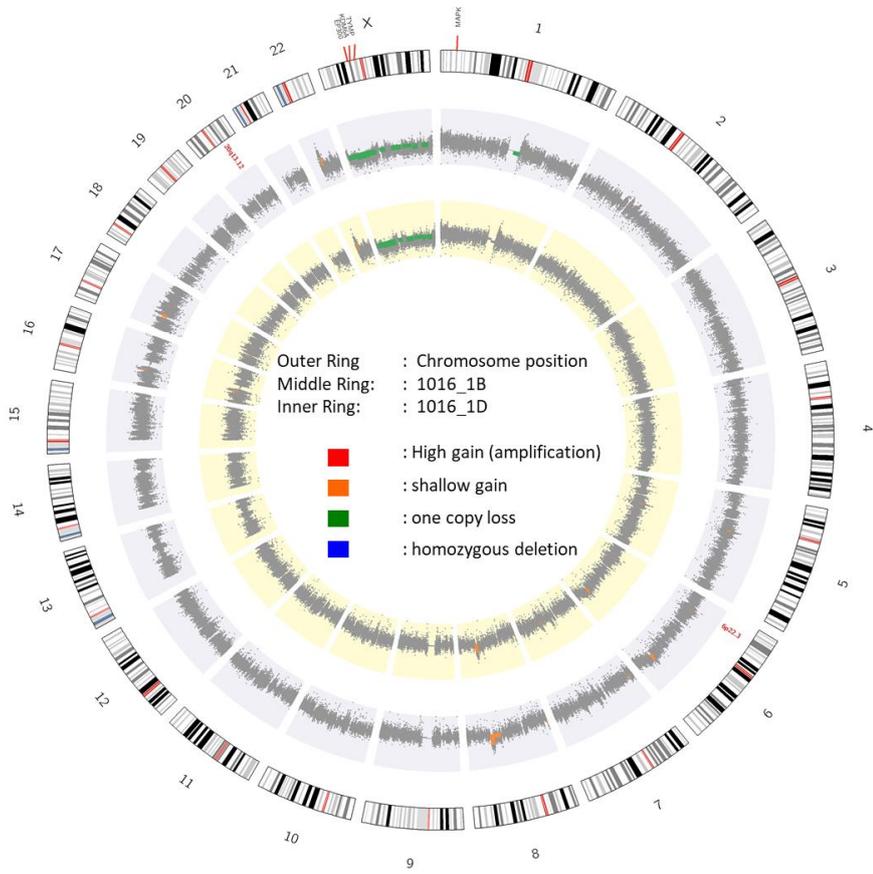
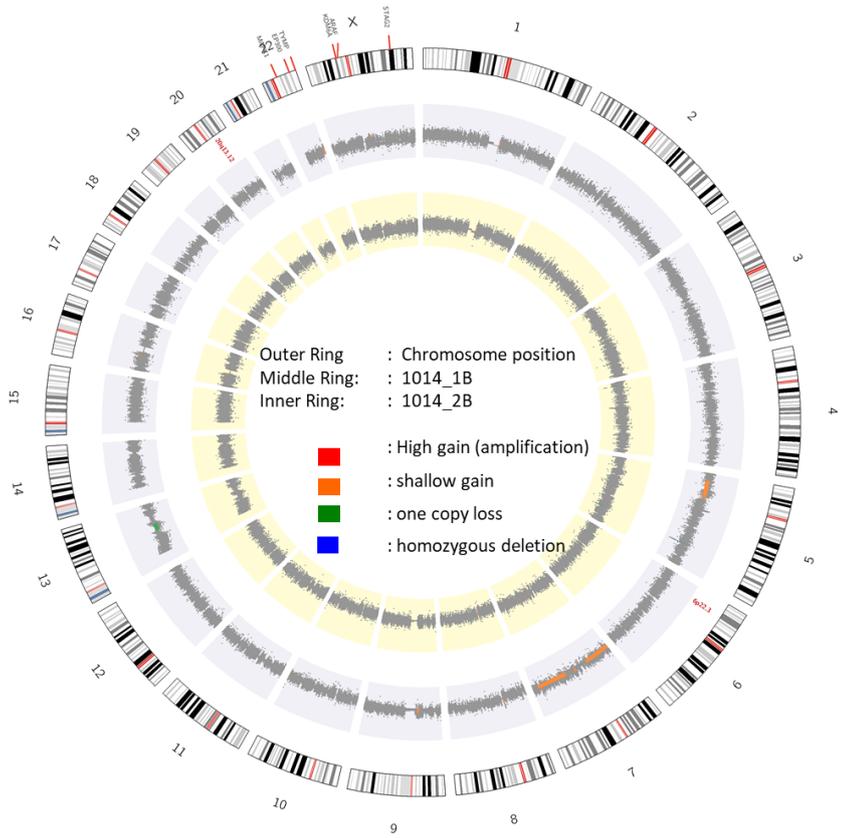
Above table replicated but with synonymous variants removed

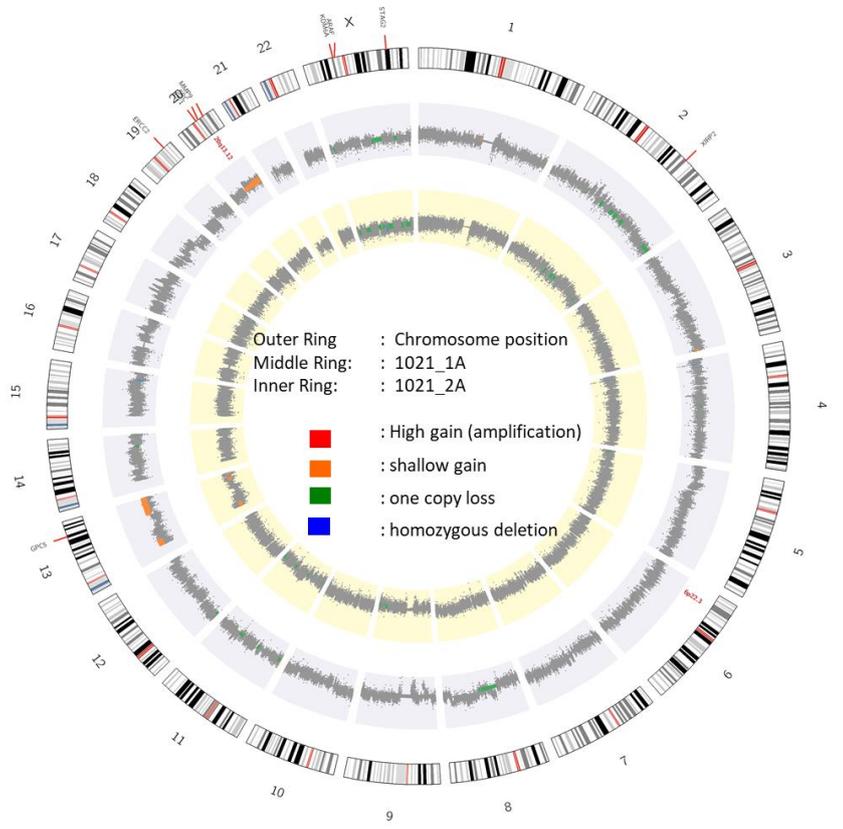
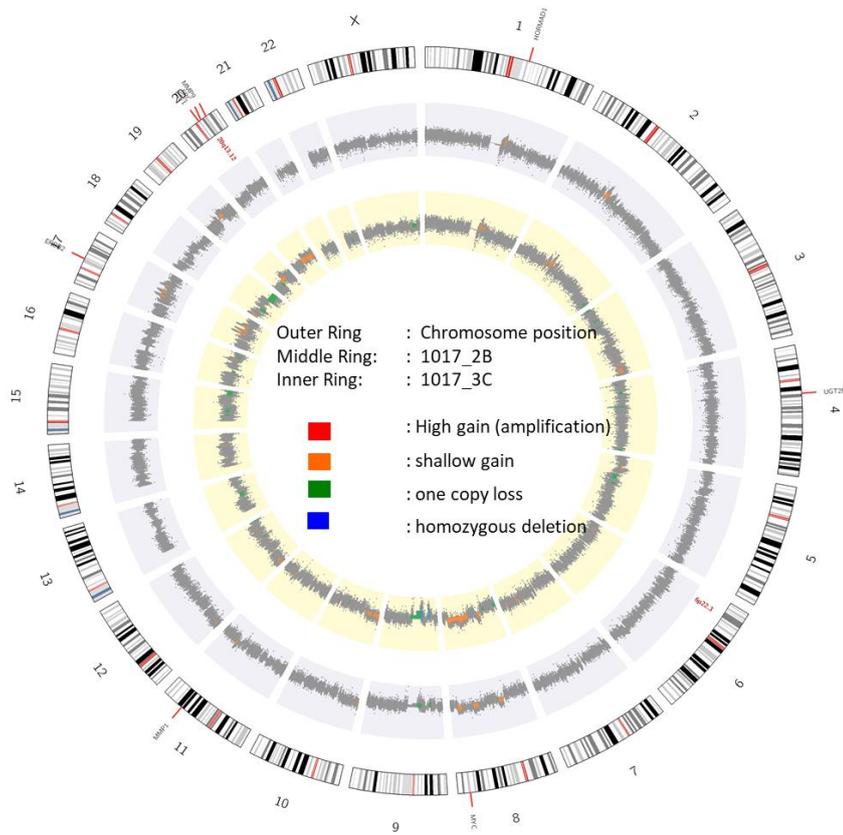
Gene	No. aberrations				No. samples	No. patients
	HIGH	MODERATE	LOW	TOTAL		
TP53	7	14	12	33	11	6
KMT2C	1	4	4	9	9	6
ERBB3	0	10	4	14	12	6
ERCC2	0	6	2	8	8	4
RHOA	0	6	2	8	8	4
STAG2	2	4	3	9	7	4
NFE2L2	3	4	0	7	5	4
FAT1	2	0	7	9	6	3
ATM	2	4	2	8	6	3
KDM6A	2	0	3	5	5	3
BRCA1	2	1	7	10	5	3
SPTAN1	0	0	6	6	4	3
FANCD2	0	0	6	6	4	3
KMT2A	2	1	2	5	4	3
ASXL1	0	6	2	8	8	3
ELF3	1	4	0	5	4	2
ERBB2	0	5	7	12	3	2
ARID1A	4	4	2	10	6	2
METTL3	0	6	0	6	6	2
KMT2D	4	2	2	8	4	2
PIK3CA	2	4	0	6	4	2
ASXL2	2	0	3	5	4	2
FGFR3	0	4	0	4	4	2
CDKN1A	3	0	0	5	3	2
KANSL1	0	0	0	0	0	0
CDKN2A	0	1	0	1	1	1
MBD1	0	0	0	0	0	0

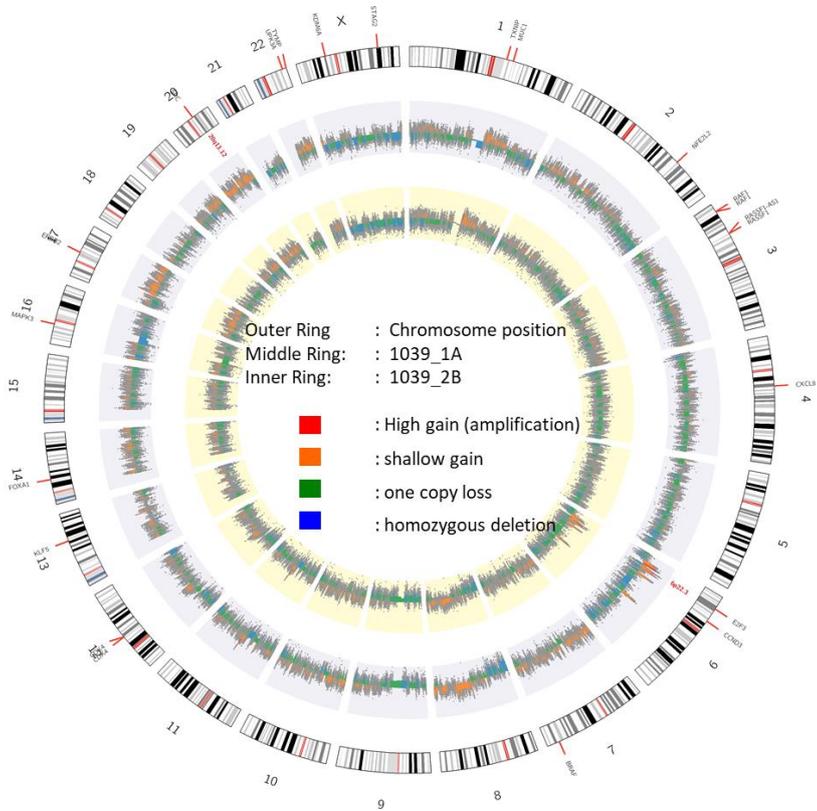
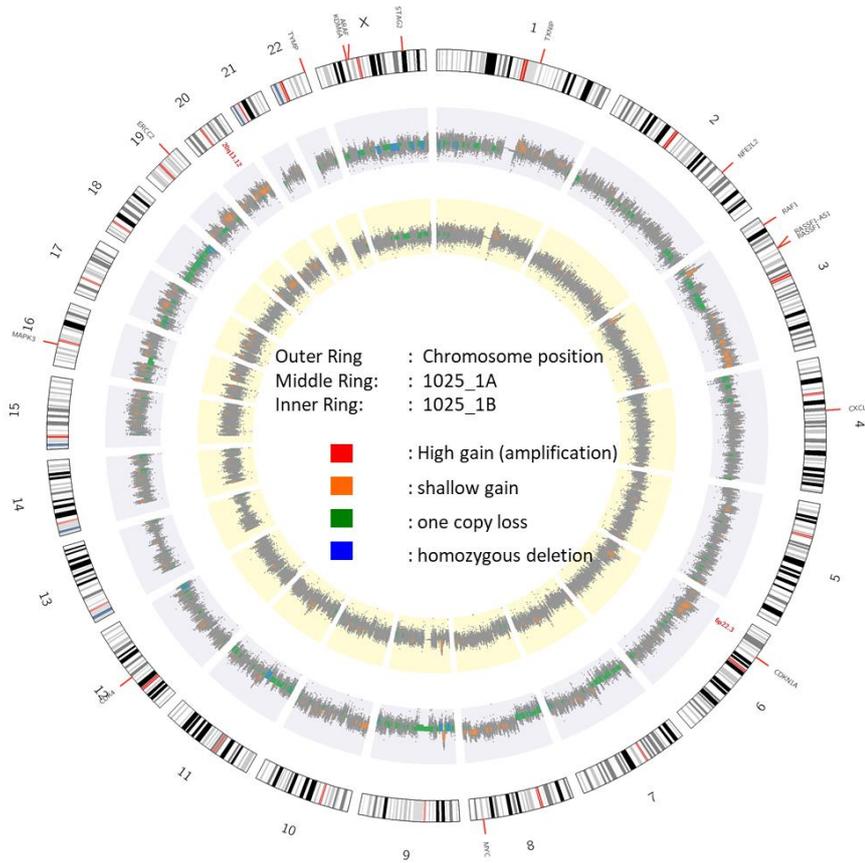
A4.13 Circos plots for copy number alterations











A4.14 FoundationOne and WES discrepancies

Table of aberrations reported by FoundationOne but not by TPU WES

Patient	Gene	Aberration	Outcome
1014	<i>FAT1</i>	T2997M	Present in both tumour samples at frequencies of 53.6% and 55.8% but also seen in germline sample at 39.5%; not a recognised SNP but given frequency germline frequency, had been discounted as somatic mutation.
1039	<i>KDM6A</i>	Q1006fs*5	Present on reviewing raw data but removed during TPU analysis as target regions only reported
1027	<i>FAT1</i>	R885L	Present in both tumour samples at 67% and 59.8%, but also in germline at 44.9%. Not a recognised SNP but discounted as a somatic mutation
1025	<i>ERBB3</i>	N369S	Present in both tumour samples at 59% and 55.4%, and in germline at 42%. Not a recognised SNP but discounted as a somatic mutation
1017	<i>FAT1</i>	Y1672C	Present in both tumour samples at 65.9% and 79.9%, and germline at 46.1%. Not a recognised SNP.
	<i>KDM6A</i>	R1054fs*29	Present on reviewing raw data on IGV. Confirmed with TPU that aberration had been filtered out of the results provided as not in target region.

Patient	Gene	Aberration	Outcome
	<i>BRCA1</i>	M1361I	Present in both tumour samples at 63.9% and 71%, germline 38.8%. Recognised SNP in dbSNP.

Table showing aberrations reported by TPU but not FoundationOne

Sub	Sample	Gene	Aberration	VAF on WES	Outcome
1003	2C	<i>ERBB3</i>	p.met658val	13% (68/506) Only 1/518 reads in sample 1A	4% frequency therefore filtered out
1003	1A	<i>NFE2L2</i>	p.Trp33*	23.5% (but only 18 depth)	Not seen on FM
1014	2B	<i>KMT2C</i>	p.Val919Leu	8.6% (11/128) Not seen in other sample or in germline but germline depth = 20	5% MAF on FM but filtered out; plus snp exists
1021	1A	<i>KMT2C</i>	p.Arg904*		6% but filtered out
1021	1A	<i>NFE2L2</i>	p.Trp33*	4.3%	Not seen

Sub	Sample	Gene	Aberration	VAF on WES	Outcome
1025	1B	<i>NFE2L2</i>	p.Trp33*	16.1% (but 36 depth)	Not seen
1027	1A	<i>ERBB3</i>	p.Gln847Glu	5%	1% filtered out
1039	2B	<i>KMT2C</i>	p.Arg909Lys		18% VAF but filtered out

A4.15 Exploring intra-tumour heterogeneity

Patient	Gene	Aberration (impact)	Sample	VAF (alteration/coverage)
1003	<i>ERBB3</i>	pMet658V (moderate)	1A	0% (0/500)
			2C	13.4% (68/506)
	<i>TMCO4</i>	p.Gln13* (high)	1A	37.4% (34/91)
			2C	0% (0/111)
1014	<i>ERBB2</i>	6 x aberrations (moderate)	1B	Seen on IGV for all but very low frequency (0.05% (6/1023) to 1.88% (10/533))
			2B	Range 5.1% (28/554) to 9.9% (19/191)
	<i>KMT2C</i>	p.Val919Leu (moderate)	1B	On IGV 16/179
			2B	8.6% (11/128)
	<i>TP53</i>	p.Pro153fs	1B	47.3% (depth 793)
			2B	Seen on IGV but had been filtered out by TPU falling short of arbitrary thresholds.....
1016	<i>KMT2A</i>	p.Arg1350Cys (moderate)	1A	0
			1B	5.4% (4/74)
			1C	0
			1D	0

Patient	Gene	Aberration (impact)	Sample	VAF (alteration/coverage)
1017	SSH3	p/Thr65Lys (moderate)		13/487
			3C	9.5% (35/269)
1021	ELF3	p.Ser163fs	1A	51.6% (depth 931)
				Seen on IGV 241/1007;
	KMT2C	p.Arg904*	1A	7.9% (7/89)
				8/99
1025	BRCA1	p.Ser1634Gly	1B	5.3% (9/171)
			1A	6/351
	CDKN1A	p.Gly125fs	1A	46% (759)
			1B	20.2% on IGV
1027	CDKN2A	p.Arg53Trp (moderate)	1A	0
			2A	10.2% (32/313)
	ERBB3	p.Gln847Glu	1A	5.2% (24/459)
			2A	0
	MB21D2	p.Gln311Glu	1A	3.2% (19/593)
			2A	Igv 19/911
1039	KMT2C	p.Arg909Lys	2B	5.7% (4/70)
				0/47

A4.16 Specific assays used by other groups to interrogate plasma DNA

Aberration	Reference	Prevalence in FoundationOne cohort
PIK3CA E545K	Birkenkamp-Demtroder et al; Patel et al	2/29
PIK3CA pG545	Patel et al	0
PIK3CA pHC1047	Patel et al	1
FGFR3 S249C	Birkenkamp-Demtroder et al; Patel et al	2/29
FGFR3 Y373C p.y375	Birkenkamp-Demtroder et al	1/29
FGFR3 p.K652	Patel et al	0
BRAF pV600	Patel et al	0
CTNNB1 pT41	Patel et al	0
CTNNb1 pS45	Patel et al	0
HRAS pG12	Patel et al	0
KRAS pG12	Patel et al	2
KRAS pG13	Patel et al	0
KRAS pQ61	Patel et al	0
NFE2L2 pG31	Patel et al	0
TP53- all exons	Patel et al	20/29

A4.17 Targeted panel candidates

Candidates should include the genes most commonly mutated in MIBC. Vandekerkhove et al³⁴ and Christensen et al³⁷ used public data to determine the most common aberrant genes when designing their panels of 50 and 51 genes respectively. There were 28 genes in common when comparing the lists and I think it is reasonable to include these on future panels. The most common aberrant genes from my FoundationOne and WES data have also been added, in addition to DDR genes that have been reported to potentially be associated with outcomes^{2,3}.

Gene	Notes
AKT1	Targets present on both panels [1,2]
ARID1A	
ATM	
BRAF	
CDKN1A	
CDKN2A	
CREBBP	
CTNNB1	
EP300	
ERBB2	
ERBB3	
ERCC2	
FBXW7	
FGFR3 (including introns)	
HRAS	
KDM6A	
KMT2D	
KRAS	
NF1	
NRAS	
PIK3CA	
PTEN	
RB1	
RXRA	
STAG2	
TERT promoter	
TP53	
TSC1	
ASXL1	FoundationOne data= aberrant in 24%
NOTCH1	FoundationOne data= aberrant in 24%
KMT2C	FoundationOne data= aberrant in 67%
RHOA	WES data- aberrant in 44%
NFE2L2	WES data- aberrant in 44%
BRIP1	DDR genes proposed to be associated with trend towards better outcomes in CRT cohort [3]
FANCD2	
RAD50	
BRCA1	
PALB2	
CCND1	Aberrant in 16% of CRT cohort [3]
MDM2	Aberrant in 12% of CRT cohort [3]; mutually exclusive with TP53

Chapter 5. Exploring Molecular Subtype as a Biomarker of Radiosensitivity in Muscle-Invasive Bladder Cancer

5.1. Introduction

5.1.1. Molecular subtyping in muscle-invasive bladder cancer

There is great interest in the use of molecular subtyping to more accurately define a cancer beyond its histological features. The potential for subtype to be a predictive biomarker has interested many groups and some data, albeit based on small numbers, has suggested that subtype may be able to predict chemoresistance in muscle-invasive bladder cancer (MIBC)¹.

The potential of whether molecular subtyping could be a predictor of response to radiotherapy in MIBC has been little explored, and will be the focus of this chapter. Based on data from pan-cancer studies² and preliminary work within Dr Sadanandam's laboratory³ suggesting similarities between the proposed subtypes in colorectal and bladder cancer, I set out to explore whether a colorectal subtype classifier, the CRCAssigner-38, could be applied to MIBC. I expect to confirm that MIBC can be classified into colorectal subgroups, as data has shown that MIBC can be broadly divided into those with demonstrating markers of differentiation (luminal) and those with basal features, with subtypes also demonstrating stem-like and inflammatory features. These molecular features were also seen in the CRCAssigner-38 subtypes, as discussed in my introduction chapter. The critical question to be addressed, however, in this pilot study is whether the colorectal subtypes are of clinical relevance in MIBC.

In addition to using publicly available expression data, I have generated my own gene expression data from a retrospective cohort of MIBC patients treated with radical radiotherapy. This pilot study will also explore whether the Radiosensitivity Index (RSI)⁴ is of potential relevance in MIBC.

5.2. Hypotheses and Aims

5.2.1. Hypotheses:

MIBC can be subdivided into 5 molecular subtypes based upon gene expression profiles by using an existing colorectal cancer classifier (CRCAssigner-38)

CRCAssigner-38 subtypes identified in MIBC may be of clinical relevance with regards to clinical outcomes and may be of interest as a potential biomarker

Molecular subtype may be a biomarker of response to radiotherapy, and comparison of gene expression profiles of patients with MIBC treated with radical radiotherapy who have achieved locoregional control and those who have not, will identify potential candidate biomarkers of radiotherapy response

Gene expression profiling of multiple regions of tumour from one patient will demonstrate heterogeneity with respect to subtype allocated

5.2.2. Aims

To apply CRCAssigner-38 to publicly available gene expression data and explore whether colorectal cancer subtype is associated with response to neoadjuvant chemotherapy

To assess the feasibility of extracting RNA from diagnostic FFPE blocks obtained from a cohort of patients who have completed radical daily radiotherapy for MIBC, for interrogation on the Nanostring panel; RNA will be extracted from multiple regions with a view to exploring intratumour heterogeneity in the context of molecular subtype allocation

To design a Nanostring panel encompassing the colorectal subtyping genes, TCGA bladder subtyping genes and potential radiosensitivity genes, and apply this to the cohort of radiotherapy-treated MIBC patients

To divide the cohort into 5 colorectal subtypes using CRCAssigner-38 and assess whether subtype may have clinical relevance in the context of MIBC patients treated with radical radiotherapy +/- chemotherapy

To explore subtypes present using the custom-designed Nanostring gene panel and assess whether subtype allocated may have clinical relevance in the context of MIBC patients treated with radical radiotherapy +/- chemotherapy

To compare gene expression profiles in patients considered responders or non-responders to radical radiotherapy and identify those which are differentially expressed.

To explore whether the Radiosensitivity Index (RSI) described by Torres-Roca's group⁴ is of relevance in MIBC using the Nanostring platform

5.3. Methods

5.3.1. Application of CRCAssigner-38 to Publicly Available Data

Normalised gene expression data and clinical data from the GSC dataset⁵ was downloaded from GEOmnibus (GSE87304). This dataset consisted of 305 MIBC patients who received neoadjuvant chemotherapy. An R-script provided by Dr Sadanandam's bioinformatician team was applied to subtype cases using their current 38-gene panel (see Appendix A5.1 for script). The method involves correlation of gene expression data (from the above dataset) to a published centroid^{7,8} (summary of gene expression for each gene in each subtype).

It should be noted that the subtyping algorithm assigns a primary subtype, where possible, to each sample according to pre-defined thresholds. However, where values do not quite meet the thresholds (correlation coefficient less than 0.15 or the difference between highest and second highest correlation

coefficient is less than 0.06), samples may be labelled as mixed or undetermined.

Each result was reviewed and correlated with the NAC response data available within the downloaded dataset. The contact for the GSE87304 dataset was emailed to request survival data, which was not readily available on the file downloaded from the GEO Omnibus website.

MIBC expression data from the TCGA dataset was downloaded using the Broad Institute Firehose resource.

5.3.2. Patient cohort

5.3.2.1. Nanostring cohort

Patients who had completed radical daily radiotherapy +/- chemotherapy for MIBC at the Royal Marsden Hospital (RMH) were identified from the CoMB database. However, in order to achieve a sufficient cohort size within the timeframe of this project, it was also necessary to search the RMH database for potential additional patients. This was also necessary in order to achieve a balance of patients with and without locoregional recurrence post-radiotherapy. A list of patients treated with radical radiotherapy with a diagnosis of bladder cancer was therefore obtained from the RMH Data Department in early 2015. By cross-referencing with the RMH computerised notes system (EPR), I was able to identify a further group of potential patients. These patients were invited to participate in CoMB on attendance at their next routine follow-up appointment. If the patient was deceased but had a signed copy of the RMH Tissues for Research form on EPR, they were recruited into CoMB as per the protocol. Where no consent form was found on the system, I reviewed their paper notes and scanned any signed RMH Tissues for Research forms found into the EPR system. The patient could then be recruited into CoMB group 2 as per the protocol.

5.3.2.2. Foundation Medicine cohort

As discussed in chapter 4, I secured the opportunity to send up to 30 FFPE blocks to Foundation Medicine for testing with their FoundationOne panel, courtesy of Roche. 10 patients from the circulating tumour DNA pilot had blocks sent, of whom 3 were also included in the Nanostring cohort. A further 20 blocks were sent from patients within the Nanostring cohort, primarily from those with relapsed disease.

5.3.3. Processing of FFPE blocks

5.3.3.1. Sectioning and staining

Diagnostic FFPE blocks were requested from the CoMB biobank or from the relevant peripheral hospital. Sectioning and staining was performed by the team at the Breast Cancer Now Histopathology Core Facility based in ICR Chelsea. An initial 2-3um section underwent H&E staining then 10 x 10um unstained serial sectioned slides (numbered) with a final 2-3um section for H&E staining were taken. Stained slides were reviewed by Dr Steve Hazell (consultant histopathologist at RMH). Areas of viable muscle-invasive tumour were marked out and assessment of histology and tumour content made.

An initial batch of 31 blocks were sectioned and assessed for extraction, with a 2nd batch of 24 blocks sent thereafter.

Batch 1 contained 31 patients. On pathological review, one had carcinosarcoma only and no TCC despite clinical documents stating a diagnosis of TCC. This patient was therefore excluded and did not undergo macrodissection.

Batch 2 contained 24 patients (including 2 from batch 1 where there was insufficient RNA on initial extraction). On pathological review, one patient only had NMIBC present and so was excluded.

Therefore, blocks were sectioned from a total of 53 patients with 51 going forward to macrodissection. Appendix A5.2 details the histological assessments made.

5.3.3.2. Macrodissection

Sections were processed in batches of up to a maximum of 80 sections at a time. Xylene deparaffinisation was performed using Dr Sadanandam's lab's protocol (see Appendix A5.3).

Macrodissection was performed using a 16G needle. Macrodissected tissue was collected into a labelled 1.5ml RNA LoBind Eppendorf containing 200ul 100% ethanol. Once all samples in that batch had been macrodissected, samples were centrifuged at 13 000 rpm for 5 minutes. The ethanol was then removed without disturbing the tissue pellet. Samples were placed (with lid open) in a thermoblock at 55C for approximately 5 minutes (or until dry). Samples were stored at -20C.

For the first batch, areas with different tumour content or different histologies were macrodissected separately- this resulted in a total of 46 tumour regions from 30 patient blocks being macrodissected.

However, due to insufficient concentrations of RNA extracted from several patients in batch 1, I decided to macrodissect multiple regions of tumour (unless of differing histology) into one Eppendorf to ensure an adequate concentration to proceed with Nanostring testing. A total of 25 regions were therefore macrodissected from 23 patient blocks.

Samples were all deparaffinised and macrodissected by me.

5.3.3.3. Dual RNA and DNA extraction

Dual extraction was performed using the Ambion Recoverall kit. The protocol (optimised by Dr Sadanandam's team) is provided in Appendix A5.4. In brief, macrodissected tissue samples were thawed at room temperature. Digestion buffer and protease was added to each sample. Samples were incubated overnight in a thermoblock for 16 hours at 50C.

Samples were checked at 15 hours to ensure adequate digestion. If there was significant undigested tissue remaining, an additional 1-2ul protease was added and the sample vortexed. Additional incubation time was given beyond 16 hours if required to ensure adequate digestion of tissue.

Samples were then transferred to 80C for 15 minutes, before the addition of isolation additive and transfer to a filter cartridge in a new collection tube. Samples were centrifuged. The filter cartridge was transferred to a new collection tube and stored at 4C for DNA extraction later. The RNA extraction protocol was then completed on the filtrate as per the manufacturer's instructions. Following buffer washes and treatment with a DNase mix, RNA was eluted in a volume of 20ul pre-warmed nuclease-free water and a double elution was performed i.e. eluate was re-applied to the filter column. Samples were then kept on ice pending the DNA extractions, and all samples were quantified using Nanodrop.

A total of 70 dual extractions were performed on samples from 51 patients. A table of yields obtained and Nanodrop parameters can be found in Appendix A5.5 and A5.6. There was adequate RNA to proceed with Nanostring testing in 25/30 patients from batch 1, and 19/21 patients from batch 2.

I performed dual extraction from the first 4 samples under the guidance of Dr Elisa Fontana (PhD student in Dr Sadanandam's team), and all other extractions were performed independently by me.

5.3.4. Nanostring Protocol

5.3.4.1. Panel design and construction

I designed a 144-gene panel which included the 60 CRCAssigner-38 genes with 10 housekeeping genes, TCGA MIBC subtype genes⁶, DNA damage repair (DDR) genes and genes potentially associated with radiosensitivity, including those in the Radiosensitivity Index⁴. The full list is provided in Appendix A5.7

The oligonucleotide probes were ordered from Nanostring. Master stocks and working pools were constructed with Dr Elisa Fontana as per the manufacturer's instructions (provided in Appendix A5.8)

5.3.4.2. Nanostring runs

The Nanostring platform ideally requires an input of 100ng RNA in a volume of 7ul. However, Dr Sadanandam's team have achieved satisfactory results with 100ng RNA in volumes up to 10ul. Using these parameters, a total of 48 samples from 44 patients were analysed using my custom-designed panel.

The 4 Nanostring runs were performed with Dr Elisa Fontana. I prepared the necessary dilutions, and observed Dr Fontana prepare the samples for the first run. I then prepared the samples for the remaining 3 runs and Dr Fontana set the samples up on my behalf in the Nanostring Prep Station. Transcripts were counted using the automated Nanostring nCounter system.

The protocol and samples include in each runs are shown in Appendix A5.9 and A5.10 respectively.

5.3.5. Statistics and analysis

5.3.5.1. Sample size

Sample size estimation by the team's bioinformatician (Dr Gift Nyamundanda) showed that a minimum of 35 subjects would be required in order to demonstrate the 5 colorectal subtypes with 80% power (see Appendix A5.11).

The CoMB protocol states that a minimum cohort size of 40 is required if survival is to be compared between two subgroups (in this case- patients with or without locoregional relapse). Dr Sadanandam's team advised that in their experience, up to a third of RNA extractions are inadequate to proceed with Nanostring and therefore I planned to extract RNA from a minimum of 53 patients to ensure a minimum cohort size of 40 would be achieved.

5.3.5.2. Analysis of publicly available dataset

All analysis was performed using R v3.4.1 and v3.5.1, Excel and SPSS.

CRCAssigner-38 subtypes were allocated to samples from the Genomic Single Classifier (GSC) and TCGA datasets using R. In the GSC dataset, a Chi-square test was performed to compare the response rates to neoadjuvant

chemotherapy (defined by the authors as <pT2) between the subtypes allocated.

Subtype allocated by the CRCAssigner-38 and both the GSC and TCGA subtypes were compared to explore any overlap between the classifiers. The GSC labels samples as either basal, luminal, luminal infiltrated or claudin-low. The TCGA identified 5 subtypes following non-negative matrix factorisation (NMF) consensus clustering of RNA-sequencing data from 408 chemotherapy-naïve MIBC samples. They were basal squamous, luminal infiltrated, luminal papillary, luminal and neuronal)⁶.

Kaplan-Meier analysis and the log-rank tests were planned to compare survival outcomes between the CRCAssigner-38 subtypes in the GSC and TCGA datasets. However, survival data for the GSC dataset did not become available so analysis was not possible.

5.3.5.3. Radiotherapy cohort clinical endpoints

The main clinical endpoints of interest were as follows-

- Locoregional relapse-free survival
 - Defined as time free of disease recurrence in the regional nodes and/or superficial or invasive disease in the bladder; measured from start of radiotherapy with data censored at any preceding distant metastases (if over 30 days before locoregional failure), second primary, death from non-bladder cause, or date last known alive.
- Invasive locoregional relapse-free survival
 - As above but excluding NMIBC as an event; data censored as above and at NMIBC recurrence
- Overall survival
 - Defined as time from start of radiotherapy to date of death, with data censored at date last known alive in those not deceased.

5.3.5.4. Analysis of Nanostring data

Data normalisation

Count data was normalised by Dr Elisa Fontana using the nSolver Analysis Software (v3.0). Low quality samples flagged by the software were removed. 2 samples were removed at this point leaving data available for 46 of the 48 samples tested, from 43/44 patients.

Subtype allocation

Expression data was subjected to non-negative matrix factorisation (NMF) clustering using established R pipelines used by Dr Sadanandam's team. Subtypes were identified according to the following classifications:

- CRCAssigner-38
- TCGA

As the TCGA classification system was not publicly available, it was necessary to re-create it from publicly available data on a subset of 234 TCGA subjects. This was performed by bioinformatician Dr Gift Nyamundanda. The recreated centroids were used to assign TCGA subtypes to the GSC and radiotherapy cohorts

- Using the custom Nanostring panel

NMF was performed by Dr Gift Nyamundanda to establish whether any other novel subtypes were present. It was pre-planned that should the CRCAssigner-38 subtypes not appear to be of clinical relevance in bladder cancer, then NMF analysis using the Nanostring gene panel would be repeated with the CRCAssigner-38 specific genes excluded i.e. with 91 genes

Descriptive statistics were used to summarise the patient/sample characteristics in each subtype identified. Cohort numbers within each subtype for the radiotherapy cohort were anticipated to be too low to perform statistical comparison of survival, and so median overall survival with 95% CI were reported for each subtype.

Exploring differentially expressed genes

The cohort was divided into radiotherapy responders and non-responders based upon the presence or absence firstly of locoregional recurrence, and then of invasive locoregional recurrence. A Shapiro-Wilk test on the log2 normalised data confirmed a non-normal distribution and so Mann-Whitney tests were used to explore for differentially expressed genes.

To minimise the effects of multiple testing, a subset of 36 DNA damage repair and candidate radiosensitivity genes were tested and the Benjamini-Hochberg correction was applied using a false discovery rate of 0.05.

The Radiosensitivity Index (RSI)

I planned to use the rank-based algorithm discussed in section 1.4.4.2 to calculate the RSI for each sample tested. Unfortunately this analysis could not be performed as it became apparent during data analysis that one of the RSI genes had not been correctly targeted.

5.4. Results:

5.4.1. Applying CRCAssigner-38 to GSC dataset

The CRCAssigner-38 successfully allocated 291/305 cases to a colorectal subtype (or mix of subtypes). 14 cases were labelled undetermined.

Table 5.1 below shows the distribution across the CRC subtypes and compares them against the subtypes assigned by the GSC. Figure 5.1 illustrates the subtype allocation for each sample using both classifications systems alongside NAC response.

Table 5.1 Subtypes assigned to 305 cases MIBC using CRCAssigner-38 (CRC38) vs GSC subtypes. Undeter = undetermined. Greatest overlaps have been highlighted in orange.

GSC \ CRC38	GSC					
	Basal	Claudin-low	Infiltrated luminal	Luminal	Undeter	Total
Enterocyte	3	1	5	18	16	43
Goblet-like	5	0	1	23	10	39
Inflammatory	21	20	6	7	25	79
Stem-like	16	13	22	7	14	72
TA	7	4	0	4	4	19
Mixed	7	4	2	18	8	39
Undeter	0	3	2	4	5	14
Total	59	45	38	81	82	305

Overall, the GSC basal and claudin-low appeared to overlap with the CRCAssigner-38 inflammatory and stem-like groups, and the GSC infiltrated luminal appeared to overlap with CRCAssigner-38 stem-like tumours. GSC luminal tumours were divided between the CRCAssigner-38 enterocyte and goblet-like subtypes, likely reflecting the shared characteristic of expressing markers of differentiation. Of note, over a quarter of cases were labelled as 'undetermined' by the GSC classifier compared to <5% by the CRCAssigner-38.

A Chi square test showed no statistically significant difference in NAC major response rate (defined as <pT2 at cystectomy) between the CRCAssigner-38

subtypes ($p = 0.407$; table 5.2 below). This is in keeping with the original GSC analysis which concluded that GSC subtype was not predictive of NAC response but did have some prognostic relevance in terms of overall outcomes. I was not able to correlate CRC subtype with initial pre-treatment stage as or survival as this information was not available.

Table 5.2 Data showing NAC major responses (defined as $pT2$) for the GSC cohort when subtyped using the CRCAssigner-38 classifier

Subtype	Major response	No response	Not assessable	p-value
Enterocyte	16/43 (37.2%)	25/43 (58.1%)	2/43 (4.7%)	0.407
Goblet like	19/39 (48.7%)	18/39 (46.2%)	2/39 (5.1%)	
Inflammatory	34/79 (43.0%)	41/79 (51.9%)	4/79 (5.1%)	
Stem like	23/72 (31.9%)	43/72 (59.7%)	6/72 (8.4%)	
Transit amplifying	10/19 (52.6%)	9/19 (47.7%)	0/19 (0%)	
Undetermined	6/14 (42.9%)	8/14 (57.1%)	0/14 (0%)	
Mixed	10/39 (25.6%)	27/39 (69.2%)	2/39 (5.1%)	

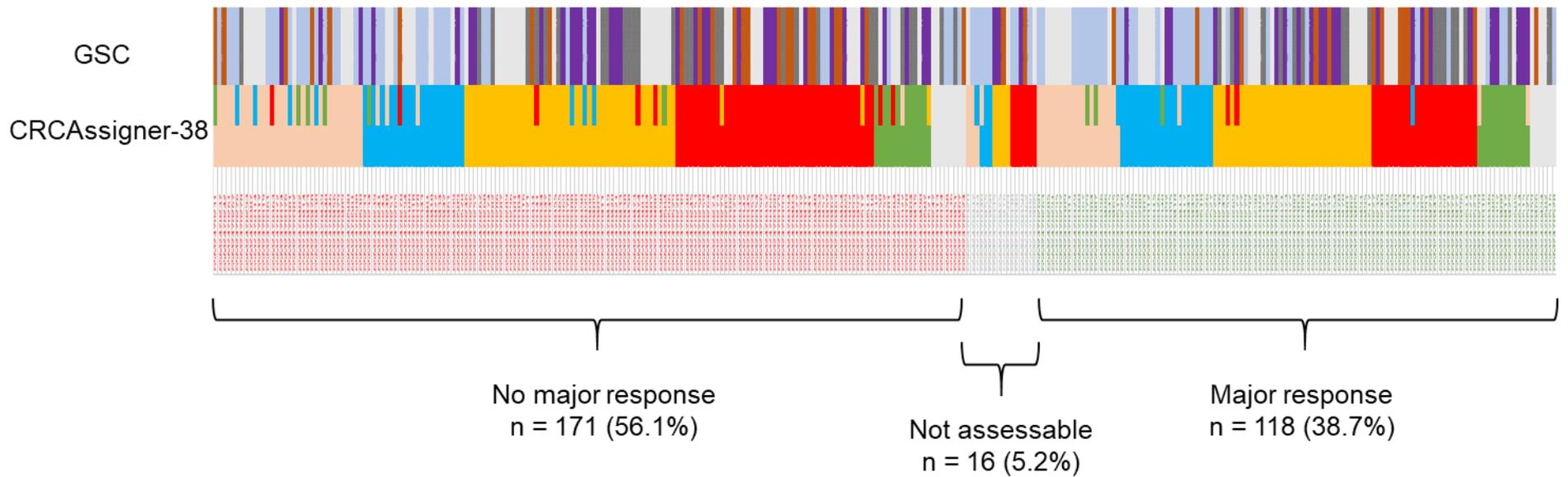


Figure 5.1 Diagram summarising subtype allocated to each sample according to GSC and CRCAssigner-38 classifications. Each vertical line represents a patient. Where a mixed subtype was reported by the CRCAssigner-38, this has been represented with a split in the colour bar

GSC		CRCAssigner-38	
	Basal		Enterocyte
	Claudin-low		Goblet-like
	Infiltrated luminal		Inflammatory
	Luminal		Stem-like
	Not assessable		Transit-amplifying
			Undetermined

5.4.2. Radiotherapy patient cohort

5.4.2.1. Patient characteristics

An overview of the patient characteristics is shown below in table 5.3. The full table is in Appendix A5.12.

Of note, a significant proportion of patients had high risk disease, including 3 with para-aortic nodal involvement, which would be treated palliatively in many centres rather than with radical chemoradiation. Unlike many series, 75% received neoadjuvant chemotherapy and all but one patient had concurrent chemotherapy. Just over half this cohort were treated within a radiotherapy dose escalation trial, and hence received more than the standard 64Gy in 32 fractions.

26 patients had a post-radiotherapy biopsy result available, Seventy-seven percent (20/26) had a complete pathological response i.e. pT0. Three patients (12%) had residual pT2 disease and the remaining 3 (12%) had CIS. An additional 13 patients underwent cystoscopy alone which showed no evidence of residual disease. The remaining 4 patients had imaging follow-up at 3-6 months with no evidence of local disease recurrence.

Table 5.3 Characteristics of radiotherapy cohort

**Some patients had taken part in a dose escalation study hence >64Gy*

	N	%
Age	Median: 71.85 yrs	Range: 46.1 – 90.9
Male	34	79.1
Female	9	20.9
Disease stage		
<i>T2-4 N0 M0</i>	29	67.4
<i>Any T N1-3 M0</i>	11	25.6
<i>Any T Any N M1</i>	3 (PA nodes)	7.0
Histology		
<i>TCC</i>	37	86.0
<i>Small cell/neuroendocrine</i>	3	7.0
<i>Other</i>	3	7.0
Neoadjuvant chemotherapy	32	74.4
Concurrent chemotherapy	42	97.7
RT Dose		
<i>64Gy in 32#</i>	21	48.8
<i>>64 Gy in 32#*</i>	22	52.2
Cystoscopy 3-4 months post RT	39	90.7
Biopsy result available	26/39	66.7

5.4.2.2. Clinical Outcomes

At a median follow-up period of 3.80 years, 22/43 (51.2%) patients had experienced disease relapse.

The median progression-free survival was 3.80 years (95% CI 1.519-6.081).

The patterns of disease recurrence at first relapse were:

- 10/43 (23.3%) had NMIBC
 - 6/10 (60.0%) patients had concurrent NMIBC present at the time of diagnosis of MIBC.
 - 3/10 subsequently went on to develop metastases (2.5, 8 and 11 months later);
 - 1/10 developed locoregional nodal disease 5 months later
 - 1/10 developed invasive bladder disease at same site of original disease 26 months later
 - Remaining 5/10 had NMIBC relapse only
- 7/43 (16.3%) had M1 disease (including 1 with local node recurrence, and one with invasive bladder recurrence)
- 5/43 (11.6%) had invasive locoregional relapse
 - 4/5 relapsed within the bladder only
 - 2 of these patients developed distant metastases 8.5 months and 15 months later
 - 1/5 had only a pelvic nodal relapse

A total of 12/43 (27.9%) developed distant metastatic disease at some point, and 17/43 (39.5%) had locoregional recurrence (LRR; of which 9/17 were muscle-invasive disease). The median LRR disease-free survival was 3.82 years (95% CI 2.44 – 5.21).

4 patients had salvage cystectomy (2 for pT3/4 disease and 2 patients with NMIBC only). A total of 17/43 (39.5%) patients have died and the median overall survival for the group is 6.41 years (95%CI 0.95 – 11.78). The median bladder cancer-specific survival has not yet been reached.

The 2-year LRR disease-free survival was 66.3%.

In terms of subsequent data analysis, the groups of interest are defined as follows-

a) Patients with locoregional recurrence (n = 17)

Patients with bladder and/or pelvic nodal relapse, including those with concurrent or subsequent distant metastatic disease or NMIBC, and

b) Patients with invasive locoregional recurrence (n = 9)

- Patients with bladder and/pelvic nodal relapse, including those with concurrent or subsequent distant metastatic disease, but NOT NMIBC

5.4.2.3. Applying CRCAssigner-38 to the MIBC Radiotherapy Cohort

CRCAssigner-38 allocated 36/46 (78.3%) samples to a subtype. 6/46 samples were deemed to be a mix of subtypes, and 4/46 samples were labelled as undetermined. Appendix A5.13 and A5.14 show further details regarding the samples labelled as mixed or undetermined.

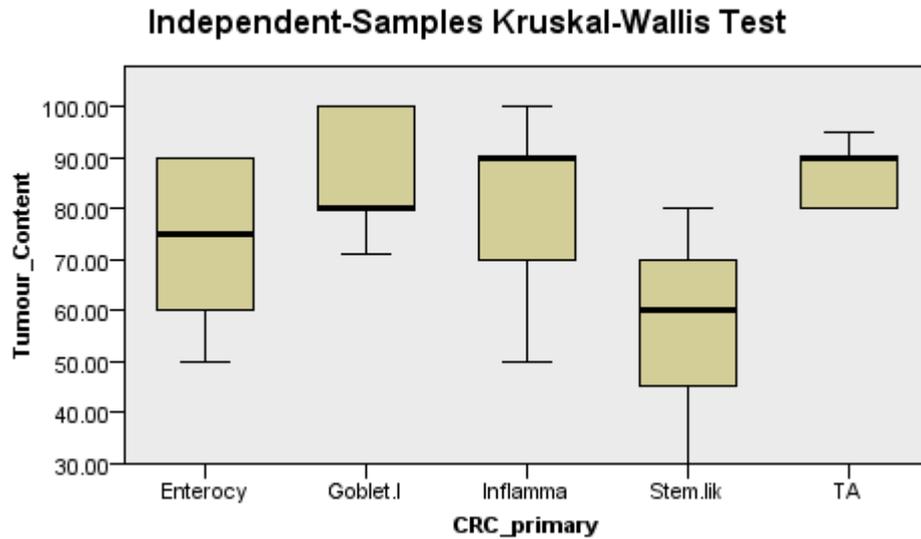
When considering the primary subtype allocated, 15/43 (34.9%) patients were allocated to the stem-like group, followed by 9/43 (20.9%) labelled as enterocyte and 8/43 (18.6%) as inflammatory. 6/43 (14.0%) and 5/43 (11.6%) were allocated to goblet-like and transit-amplifying subtypes respectively (Appendix A5.15).

Clinicopathological Features According to CRCAssigner-38 Subtype

Using primary subtype allocated, I explored whether subtype was associated with clinicopathological features. There was no significant difference in T-stage ($p=0.94$), N-stage ($p=0.95$) or M-stage ($p=0.37$) (tables in Appendix A5.16).

However, there was a statistically significant difference in tumour content between the CRC subtypes allocated ($p=0.0005$); figure 5.2. Post-hoc Dunn-Bonferroni tests showed significant differences between the stem-like subtype

and goblet, TA and inflammatory subtypes (Appendix A5.17), with stem-like tumours having a lower tumour content (mean = 57%).



	Mean tumour content (%)	Std dev
Enterocyte	74.5	14.62
Goblet-like	85.1	12.00
Inflammatory	80.0	15.81
Stem-like	57.0	14.54
TA	87.0	6.71

Figure 5.2 Box plot showing the mean tumour content for each CRC subtype; mean values with standard deviation shown in table

Relapse patterns according to CRCAssigner-38 subtypes

Table 5.4 below shows the distribution of patients into the CRCAssigner-38 subtypes and the pattern of relapse at a median follow-up of 3.80 years. Where multiple samples have been tested from one patient, the patient has only been included once, under the most representative subtype (as determined by the subtype of the majority of samples). Patients returned as 'undetermined' or 'mixed' were labelled according to the primary subtype.

Table 5.4 Distribution of CRCAssigner-38 subtypes within radiotherapy cohort (n=43) and pattern of relapse at median follow-up period of 3.80 years. NB: 2007 considered as stem-like. 2009 considered as enterocyte.

Inv = invasive

	N	Any relapse	Inv LRR +/- M1	Non-inv LRR +/- M1	M1 only	No relapse at last follow up or death from other cause
Enterocyte	9 (21%)	5 (56%)	1 (11%)	4 (44%)	0 (0%)	4 (44%)
Goblet-like	6 (14%)	3 (50%)	1 (17%)	0 (0%)	2 (33%)	3 (0%)
Inflammatory	8 (18%)	3 (38%)	2 (25%)	1 (13%)	0 (0%)	5 (63%)
Stem-like	15 (35%)	9 (60%)	4 (27%)	2 (13%)	3 (33%)	6 (40%)
Transit-amplifying	5 (12%)	2 (40%)	1 (20%)	1 (20%)	0 (0%)	3 (60%)

Clinical outcomes according to CRCAssigner-38 subtypes

Results from Kaplan-Meier analysis are shown in figure 5.3 and table 5.5 overleaf. There is perhaps a suggestion that stem-like tumours have poorer outcomes but the subtype numbers are too small to make any formal statistical comparison..

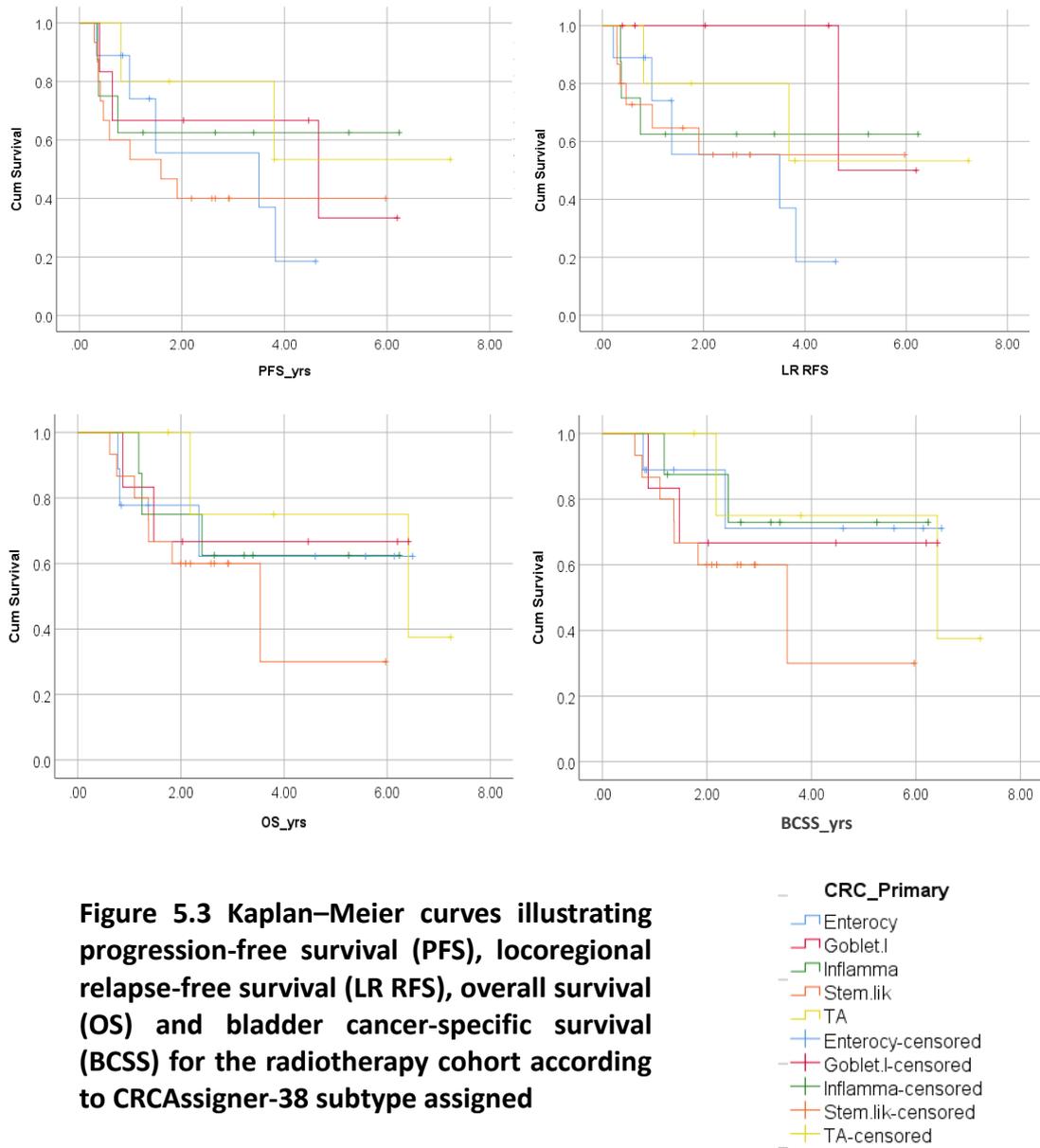


Figure 5.3 Kaplan–Meier curves illustrating progression-free survival (PFS), locoregional relapse-free survival (LR RFS), overall survival (OS) and bladder cancer-specific survival (BCSS) for the radiotherapy cohort according to CRCAssigner-38 subtype assigned

Table 5.5 Table showing median and 2 year progression-free survival (PFS), locoregional relapse-free survival (LR RFS), overall survival (OS) and bladder cancer-specific survival (BCSS) according to CRCAssigner-38 subtype. *NR = not reached.*

Primary CRCAssigner-38 subtype	N	Median PFS (yrs) (95%CI)	2-yr PFS	Median LR RFS (yrs) (95%CI)	2-yr LR RFS	Median OS (yrs) (95% CI)	2-yr OS	Median BCSS (yrs) (95% CI)	2-yr BCSS
Enterocyte	9	2.19 (0.0 – 7.80)	56%	3.50 (0 – 8.05)	55%	NR	78%	NR	89%
Goblet	6	4.66 (0.0 – 10.68)	67%	4.66	100%	NR	68%	NR	68%
Inflammatory	8	NR	63%	NR	64%	NR	75%	NR	88%
Stem-like	15	1.59 (0 – 3.26)	40%	NR	55%	3.54 (1.07-6.01)	60%	3.54 (1.07 – 6.01)	60%
Transit-amplifying (TA)	5	NR	80%	NR	80%	6.41 (0.07 – 12.75)	100%	6.41 (0.07 – 12.75)	100%

5.4.2.4. Applying TCGA subtypes to the MIBC Radiotherapy Cohort

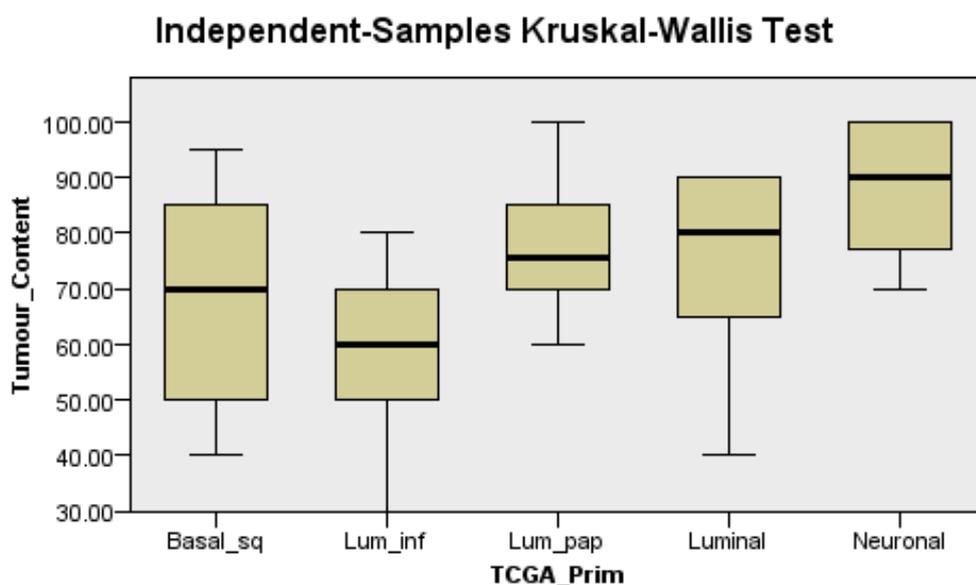
Using the recreated TCGA centroids, 38/46 (82.6%) samples were assigned to a subtype. 4/46 samples were deemed to be a mix of subtypes and 2/46 were labelled undetermined. Appendix A5.18 provides further details regarding the samples labelled mixed or undetermined.

When considering the primary subtype allocated, 15/43 (34.9%) patients were allocated to basal squamous, 10/43 (23.3%) to luminal papillary with the remaining 18 patients divided equally between luminal, luminal infiltrated and neuronal subtypes (14.0% each). Of note, the 3 cases with small cell/neuroendocrine differentiation were assigned to the neuronal subtype.

Clinicopathological features according to TCGA subtype

Using primary subtype allocated, there was no significant difference in T-stage ($p = 0.9212$), N-stage ($p = 0.7594$), or M-stage ($p = 0.6414$) between the allocated TCGA subtypes (Appendix A5.19).

Exploring primary TCGA subtype allocated and tumour content showed a trend towards luminal-infiltrated having lower tumour content ($p=0.051$) with mean 58.6% (figure 5.4).



TCGA subtype	Tumour content (%)	STDEV
Basal squamous	68.75	18.11
Luminal infiltrated	58.57	17.72
Luminal papillary	76.6	12.87
Luminal	74.29	19.02
Neuronal	87.83	12.17

Figure 5.4 Box plot showing the mean tumour content for each TCGA subtype; mean values with standard deviation (STDEV) shown in table

Relapse patterns according to TCGA subtype

Table 5.6 below shows the distribution of the radiotherapy cohort into the TCGA subtypes, using the recreated centroids, and the patterns of relapse. For the two patients with more than one sample sent, the more prevalent subtype was selected for this analysis. This was not possible for patient 2009 where 2 samples were tested with differing subtype allocations, and so the same sample as used for the CRCAssigner-38 analysis was selected for consistency.

Table 5.6 Relapse patterns at a median follow-up of 3.80 years according to subtype allocated using recreated TCGA centroids. Patients where more than one sample sent: 2007 considered as luminal infiltrated. 2009 as neuronal.

	Total	Inv LRR +/- M1	Non-inv LRR +/- M1	M1 relapse only	No relapse at last follow up or death from other cause
Luminal	6 (14%)	0 (0%)	1 (17%)	1 (17%)	4 (67%)
Luminal papillary	10 (23%)	1 (10%)	2 (20%)	2 (20%)	5 (50%)
Luminal infiltrated	6 (14%)	3 (50%)	1 (17%)	0 (0%)	2 (33%)
Basal Squamou s	15 (35%)	3 (20%)	2 (13%)	1 (7%)	9 (60%)
Neuronal	6 (14%)	2 (33%)	2 (33%)	1 (17%)	1 (17%)

Formal statistical comparison was not performed due to small subcohort numbers. However, when dividing the cohort into those with basal or luminal subtypes, there was no statistically significant difference in locoregional disease-free survival ($p = 0.826$) or overall survival ($p=0.549$) (Appendix A5.20).

Clinical outcomes according to TCGA subtype

Figure 5.5 and table 5.7 overleaf show the results from Kaplan-Meier analysis of the radiotherapy cohort according to primary TCGA subtype using the recreated centroid classifier. Luminal infiltrated tumours appear to have poorer

outcomes but again, the cohort is too small to make any formal statistical comparisons.

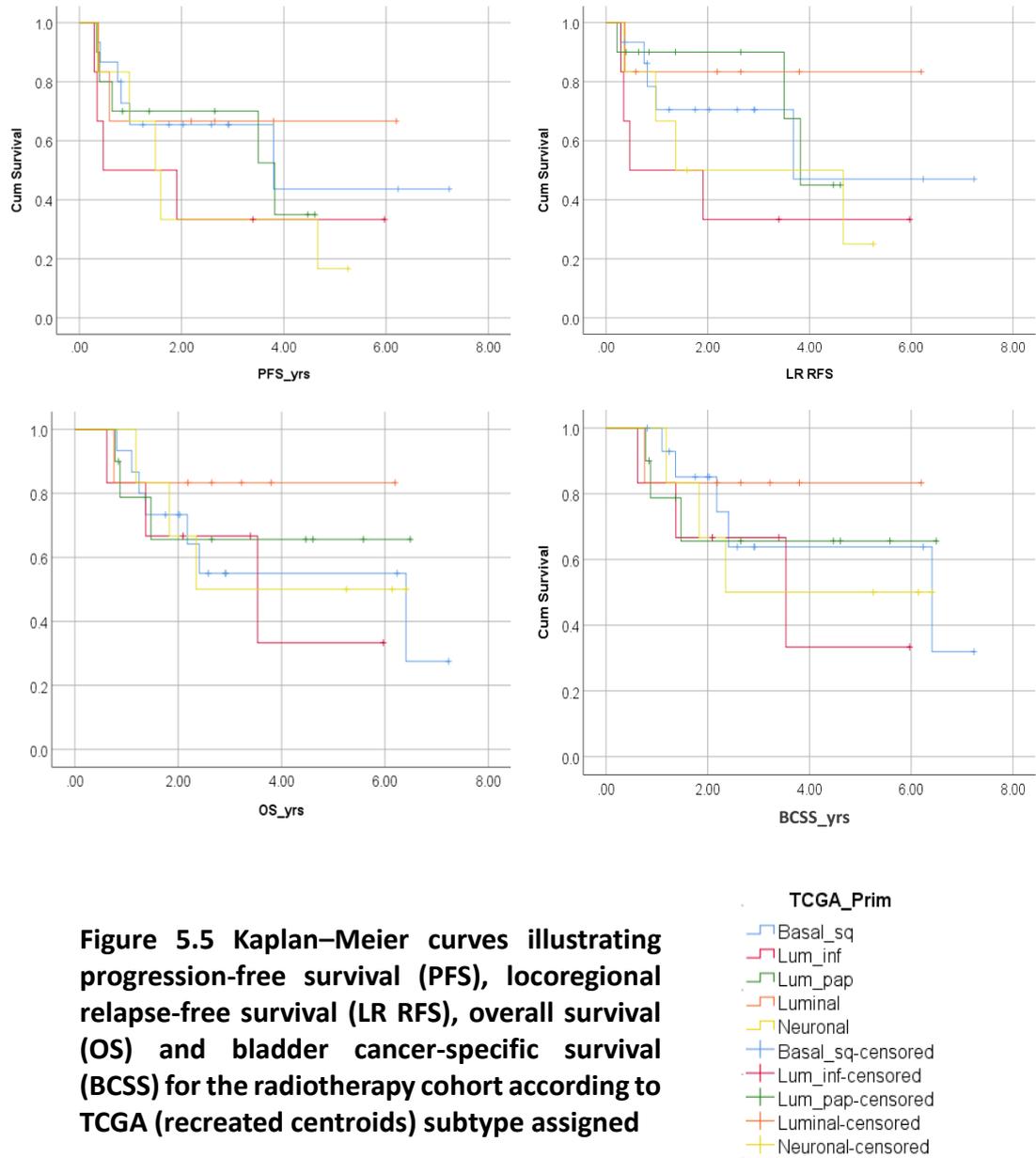


Figure 5.5 Kaplan–Meier curves illustrating progression-free survival (PFS), locoregional relapse-free survival (LR RFS), overall survival (OS) and bladder cancer-specific survival (BCSS) for the radiotherapy cohort according to TCGA (recreated centroids) subtype assigned

Table 5.7 Table showing median and 2-year rates for progression-free survival (PFS), locoregional relapse-free survival (LR RFS), overall survival (OS) and bladder cancer-specific survival (BCSS) according to primary TCGA subtype (recreated centroids) in the radiotherapy cohort. *NR = not reached*

TCGA (recreated centroids)	N	Median PFS (yrs) (95% CI)	2-yr PFS	Median LR RFS (yrs) (95% CI)	2-yr LR RFS	Median OS (yrs) (95% CI)	2-yr OS	Median BCSS (yrs) (95% CI)	2-yr BCSS
Basal squamous	15	3.80 (0.00 – 8.78)	67%	3.68	71%	6.41 (0.51-12.31)	74%	6.41 (0.57 – 12.25)	85%
Luminal infiltrated	6	0.47 (0.00 – 2.34)	33%	0.47 (0 – 2.34)	34%	3.54 (0.30 – 6.78)	67%	3.54 (0.30 – 6.78)	67%
Luminal papillary	10	3.82 (0.45 – 7.20)	70%	3.82 (3.17 – 4.47)	50%	NR	66%	NR	66%
Luminal	6	NR	68%	NR	84%	NR	84%	NR	84%
Neuronal	6	3.80 (1.52 – 6.08)	33%	1.37 (0-4.91)	50%	2.35 (no values generated)	67%	2.35 (no values generated)	67%

5.4.3. A new subtyping classifier with clinical relevance in MIBC?

5.4.3.1. 3 subtypes identified using 134-gene panel

Patterns of relapse

NMF clustering using expression data for the 134 genes included on the Nanostring panel revealed 3 subtypes (Appendix A5.21). Table 5.8 below shows the patterns of relapse according to subtype allocated.

Table 5.8 Relapse patterns for the 3 subtypes identified using the 134-gene panel at median follow-up of 3.80 years. NB: 2007 considered as subtype 3 (all 3 samples subtype 3); 2009 as subtype 1 (both subtype 1)

NMF134	Total	Inv LRR +/- M1	Non-inv LRR +/- M1	M1 relapse only	No relapse at last follow up or death from other cause
1	18 (42%)	2 (11.1%)	6 (33.3%)	2 (11.1%)	8 (44%)
2	11 (26%)	3 (27.2%)	1 (9.1%)	1 (9.1%)	6 (55%)
3	14 (32%)	4 (28.6%)	1 (7.1%)	2 (14.3%)	7 (50%)

Formal statistical comparison was not possible given the small subgroup numbers but on visually inspecting the data, there does not appear to be an association with NMF134 subtype and clinical outcome.

Figure 5.6 and table 5.9 overleaf show the Kaplan-Meier analysis for the 3 subtypes identified.

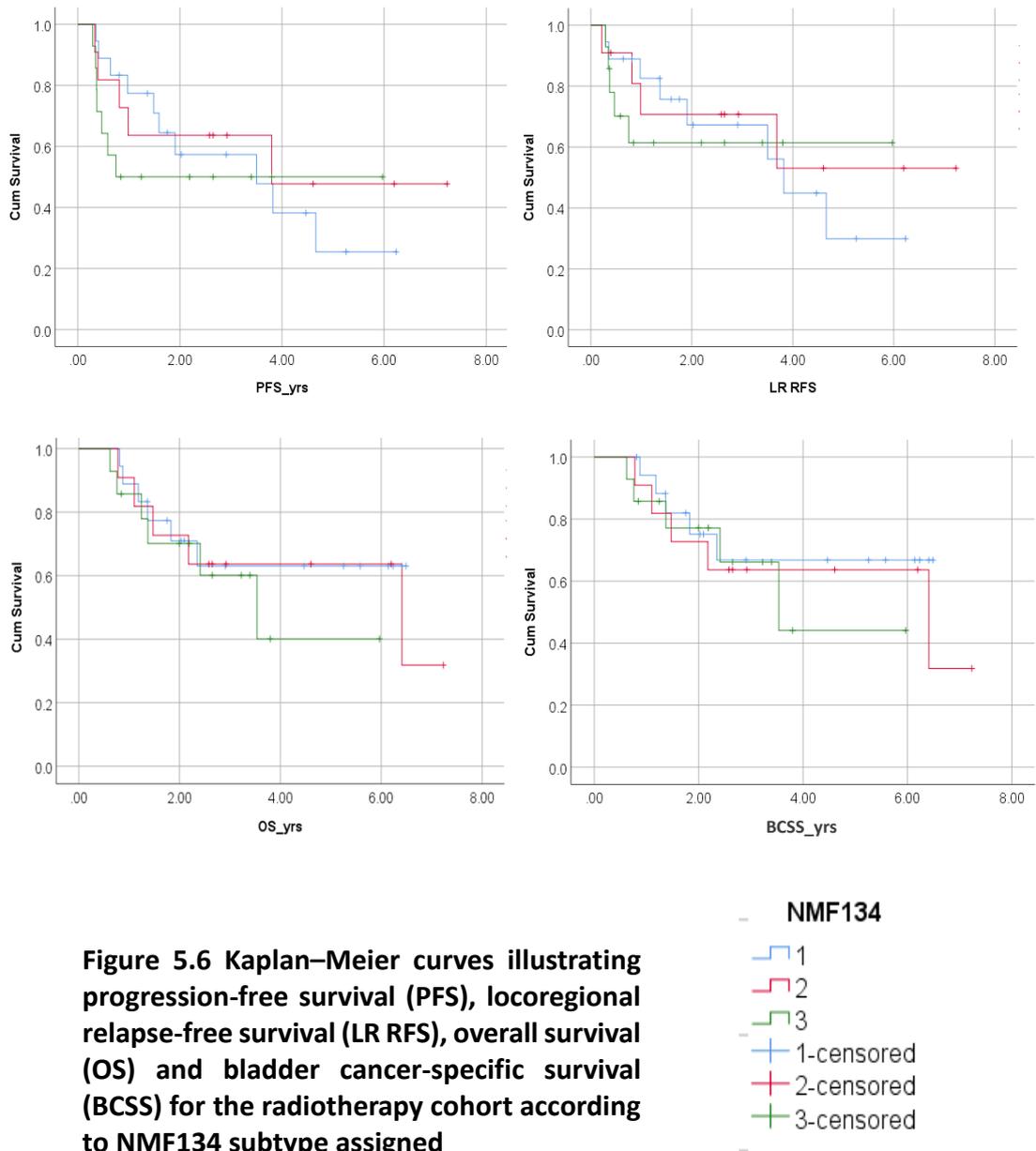


Figure 5.6 Kaplan–Meier curves illustrating progression-free survival (PFS), locoregional relapse-free survival (LR RFS), overall survival (OS) and bladder cancer-specific survival (BCSS) for the radiotherapy cohort according to NMF134 subtype assigned

Table 5.9 Median and 2-year rates for progression free survival (PFS), locoregional relapse-free survival (LR RFS), overall survival (OS) and bladder cancer-specific survival according to NMF134 subtype.

NMF134	N	Median PFS (yrs) (95%CI)	2-yr PFS	Median LR RFS (yrs) (95% CI)	2-yr LR RFS	Median OS (yrs) (95% CI)	2-yr OS	Median BCSS (yrs) (95% CI)	2-yr BCSS
1	18	3.50 (0.84 -6.16)	58%	3.82 (2.95-4.69)	68%	Not reached	73%	NR	75%
2	11	3.80 (nil values generated)	64%	Not reached	70%	6.41 (0.25-12.58)	71%	6.41 (0.25 – 12.58)	73%
3	14	0.75 (nil values generated)	50%	Not reached	62%	3.54 (0.95 – 11.87)	70%	3.54 (1.50 – 5.58)	77%

Complete pathological response rates

The most robust assessment of locoregional response could be considered to be histological confirmation. Therefore, an exploratory analysis of subgroup of 26 patients where biopsy taken 3-4 months following completion of radiotherapy. There was no statistically significant difference when comparing biopsy results between the three subtypes ($p = 0.855$; Appendix A5.22)

5.4.3.2. 5 subtypes identified using 91-gene panel

Patterns of relapse

As the colorectal subtypes did not show any clear association with outcome in MIBC, the CRCAssigner-38 specific genes were removed from the panel and NMF was repeated with the remaining 91 genes, as pre-specified in my method. 5 subtypes were seen (Appendix A5.23). As before, patients with multiple samples tested were included only once and the most prevalent subtype allocated was used. Table 5.10 below shows the results of this analysis

Table 5.10 Distribution of cases across the 5 subtypes and relapse patterns. NB: 2007 labelled as subtype 2 which was majority (A1-3 = 2/2/1). 2009 labelled 4 (A1 = 4, A3 3).

NMF91	N	Inv LRR +/- M1	Non-inv LRR +/- M1	M1 relapse only	No relapse at last follow up or death from other cause
1	5 (12%)	3 (60.0%)	0 (0.0%)	1 (20%)	1 (20%)
2	10 (23%)	2 (20%)	1 (10%)	1 (10%)	6 (60%)
3	8 (19%)	3 (38%)	1 (13%)	1 (13%)	3 (38%)
4	7 (16%)	1 (14%)	2 (29%)	1 (14%)	3 (43%)
5	13 (30%)	0 (0%)	4 (31%)	1 (8%)	8 (62%)

It is interesting that subtype 1 appeared to have the highest proportion of invasive locoregional relapse (3/5 patients) although once again, the small numbers within each cohort make it difficult to draw any firm conclusions at this point. However, it is more striking that subtype 5, the largest subtype which included 30.2% of the cohort, had no cases of invasive LRR i.e. 0 out of 13.

Figure 5.7 and table 5.11 overleaf show the Kaplan-Meier analysis for the 5 NMF91 subtypes.

Visual inspection of the Kaplan-Meier curves show quite a contrast between subtype 1 and subtypes 4 and 5. Kaplan-Meier analysis with regards to invasive locoregional relapse-free survival was not performed as the number of events and subgroup numbers were too small to make this meaningful.

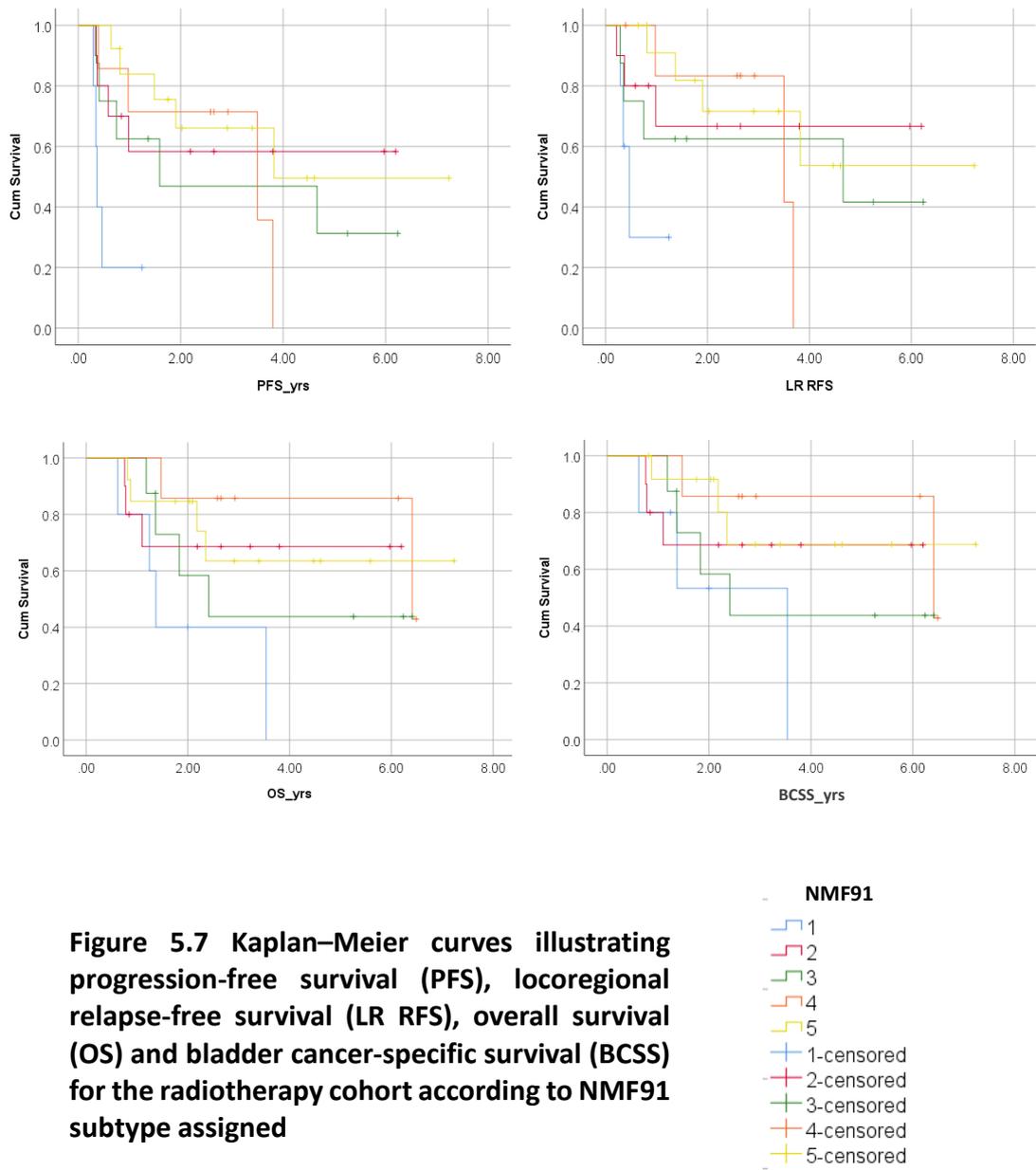


Figure 5.7 Kaplan–Meier curves illustrating progression-free survival (PFS), locoregional relapse-free survival (LR RFS), overall survival (OS) and bladder cancer-specific survival (BCSS) for the radiotherapy cohort according to NMF91 subtype assigned

Table 5.11 Table showing median and 2-year rates of progression-free survival (PFS), locoregional relapse-free survival (LR RFS), overall survival (OS) and bladder cancer-specific survival (BCSS) according to NMF91 subtype

NMF91	N	Median PFS (yrs) (95%CI)	2-yr PFS	Median LR RFS (95% CI)	2y LR RFS	Median OS (95% CI)	2-yr OS	Median BCSS	2-yr BCSS
1	5	0.37 (0.33- 0.41)	NR; 20% at 1 year	0.47 (0.28-0.66)	NR; 30% at 1 year	1.373 (1.096 – 1.649)	40%	3.54 (no values)	53%
2	10	5.63 (no values)	58%	NR	67%	NR	68%	NR	68%
3	8	1.49 (0.00 – 6.17)	47%	4.66 (0 – 12.21)	63%	3.422 (0.94 – 3.88)	58%	2.41 (0.94 – 3.88)	58%
4	7	3.50 (0.00- 7.19)	72%	3.50 (0 – 7.11)	83%	6.41 (0.00 – 13.41)	85%	6.41 (0.00 – 13.41)	85%
5	13	3.82 (no values)	66%	NR	72%	NR	84%	NR	92%

Pathological complete response rates

As before, an exploratory analysis of the post-radiotherapy assessment biopsies was performed and is shown in table 5.12 below.

Table 5.12 *Biopsy results at first check post-radiotherapy (3-4 months post completion of radiotherapy) for each subtype.*

NMF91	N	pT0 (No malignancy or atypia only)	pT1 (NMIBC)	≥pT2 (Residual invasive disease)
1	4	2/4	1/4	1/4
2	5	3/5	1/5	1/5
3	6	4/6	1/6	1/6
4	4	4/4	0/4	0/4
5	7	7/7	0/7	0/7
Overall	26	20	3	3

Across subtypes 4 and 5, there is a pathological complete response rate, defined as pT0, of 100% (11/11) compared to 60% (9/15) across groups 1-3 ($p = 0.0237$).

For the 13/39 patients undergoing a post-radiotherapy cystoscopy but with no biopsy result, all were documented to have no evidence of residual disease on cystoscopic appearances. If including these cases to the existing data above and labelling then as 'pT0; no malignancy/atypia only', comparison of complete response rates between subtypes 1-3 and 4-5 remained statistically significant ($p = 0.0267$). Furthermore, of the 4 patients not undergoing cystoscopy, none had a documented locoregional recurrence at the time of data analysis.

Invasive locoregional recurrence

Given the possible differential responses with regards to invasive recurrence between subtypes 1-3 and 4-5, further analysis was performed dividing the cohort into these two groups (figure 5.8). Combining the subtypes in this manner facilitated statistical comparison of survival outcomes, accepting the limitations of performing an unspecified analysis. I think that this is however reasonable in the context of a pilot study. There was no statistically significant difference between these groups for overall survival ($p = 0.130$), bladder cancer-specific survival ($p=0.108$) or locoregional relapse free survival ($p=0.431$). There was however a statistically significant difference in invasive locoregional relapse free survival ($p = 0.028$). 2 year invasive LR RFS for groups 1-3 was 71% compared to 100% for groups 4-5.

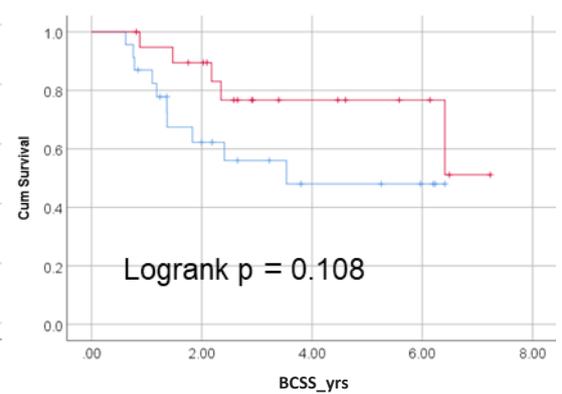
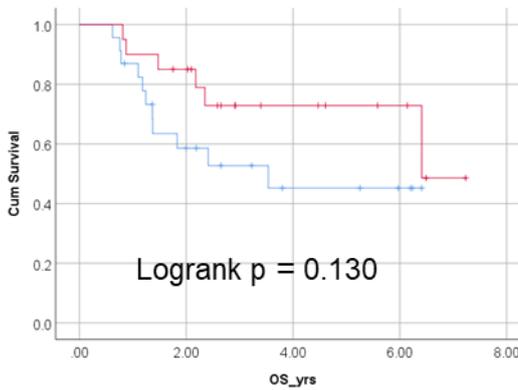
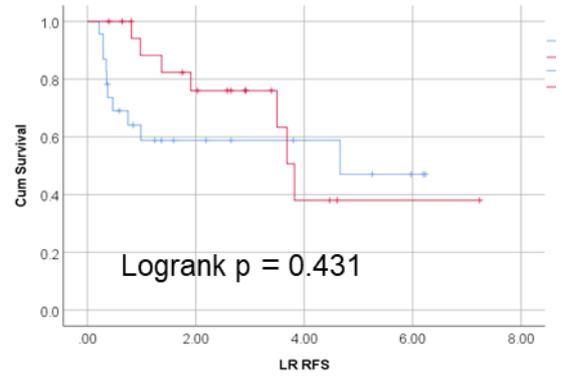
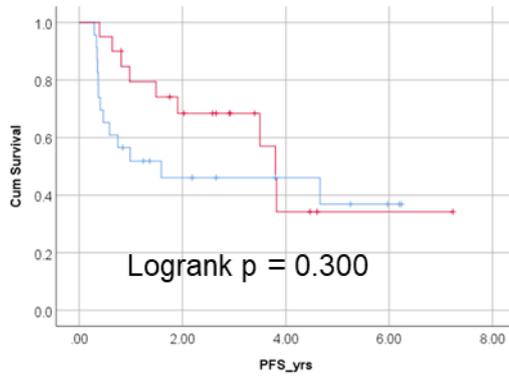
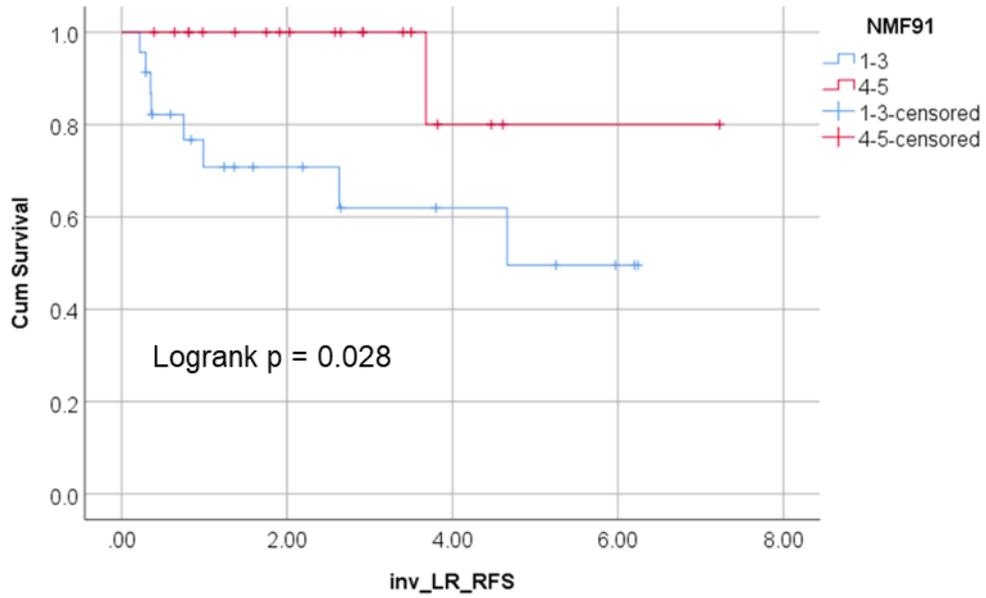


Figure 5.8 Kaplan–Meier curves illustrating invasive locoregional relapse-free survival (inv LR RFS), progression-free survival (PFS), locoregional relapse-free survival (LR RFS), overall survival (OS) and bladder cancer-specific survival (BCSS) for the radiotherapy cohort according to NMF91 subtype grouping (subtypes 1-3 vs 4-5)

5.4.3.3. 3 vs 5 subtypes

The classifier derived from the 91-gene panel with 5 subtypes appears to have more clinical relevance than that with 3 subtypes based on 134 genes, in that there is a statistically significant difference in pathological complete response rate when comparing outcomes of NMF subtypes 1-3 vs 4-5, which is not seen in NMF134. It may be that further work will demonstrate the possibility to simply divide patients into 2 subtypes i.e. radiosensitive and radioresistant, but at this early stage of work, I believe it is worthwhile to further explore the characteristics and features of the 5 subtypes presented. My analysis will therefore focus on this 5-subtype classifier.

Of note, after the initial analysis, I became aware that one of the probes on the Nanostring panel had not targeted protein kinase C as intended but due to ambiguity in gene name abbreviations, had instead targeted PRRT2. Removal of this gene and re-analysis with a 90-gene panel did not cause any change in the subtype groupings. 71/90 genes were found to contribute to the subtyping allocations. Therefore, moving forward, this subtyping classification will be referred to as NMF71.

5.4.4. The NMF71 classifier

5.4.4.1. Clinical characteristics

Table 5.13 below shows the clinicopathological features of each subtype identified. No significant difference was noted between the 5 groups although there was a trend towards subtype 1 having a lower tumour content. Of particular note, there was no statistically significant difference between the subtypes with regards to radiotherapy dose received.

Table 5.13 Clinico-pathological features of each NMF71 subtype.

**micropapillary differentiation; ^neuroendocrine differentiation in 3 patients and one with lymphoepithelioma-like histology*

	NMF71 Subtype					p-value
	1	2	3	4	5	
n	5	10	8	7	13	
Age (mean)	80.33	67.70	72.42	68.97	70.02	0.280
Female	1 (20%)	2 (20%)	2 (25%)	1 (14.3%)	3 (23.1%)	1
Stage T2	2 (40%)	7 (70%)	5 (62.5%)	1 (14.3%)	6 (46.2%)	0.225
Stage 3b	3 (60%)	3 (30%)	3 (37.5%)	5 (71.4%)	7 (53.8%)	0.610
Stage T4	0 (0%)	0 (0%)	0 (0%)	1 (14.3%)	0 (0%)	0.279

Table 5.13 (continued)

	NMF71 Subtype					p-value
	1	2	3	4	5	
Stage N0	2 (40%)	9 (90%)	6 (75%)	4 (57.1%)	11 (84.6%)	0.198
Stage N1	2 (40%)	0 (0%)	1 (12.5%)	3 (42.9%)	2 (15.4%)	
Stage N2	0 (0%)	0 (0%)	1 (12.5%)	0 (0%)	0 (0%)	
Stage N3	1 (20%)	1 (10%)	0 (0%)	0 (0%)	0 (0%)	
Stage M1	0 (0%)	1 (10%)	0 (0%)	0 (0%)	2 (15.4%)	0.842
Dose > 64GY	3 (60%)	7 (70%)	2 (25%)	3 (42.9%)	7 (53.8%)	0.426
Mean tumour content (%) (stdev)	55.0 (16.43)	71.8 (19.40)	77.4 (18.54)	75.71 (19.88)	75.08 (12.72)	0.192
Variant Histology	0 (0%)	2* (20%)	4^ (50%)	0 (0%)	0 (0%)	

5.4.4.3. Molecular characteristics of NMF71 subtypes

The heatmap below in Figure 5.9 illustrates the gene expression profiles for each of the subtypes. The full metagene list is shown in Appendix A5.24.

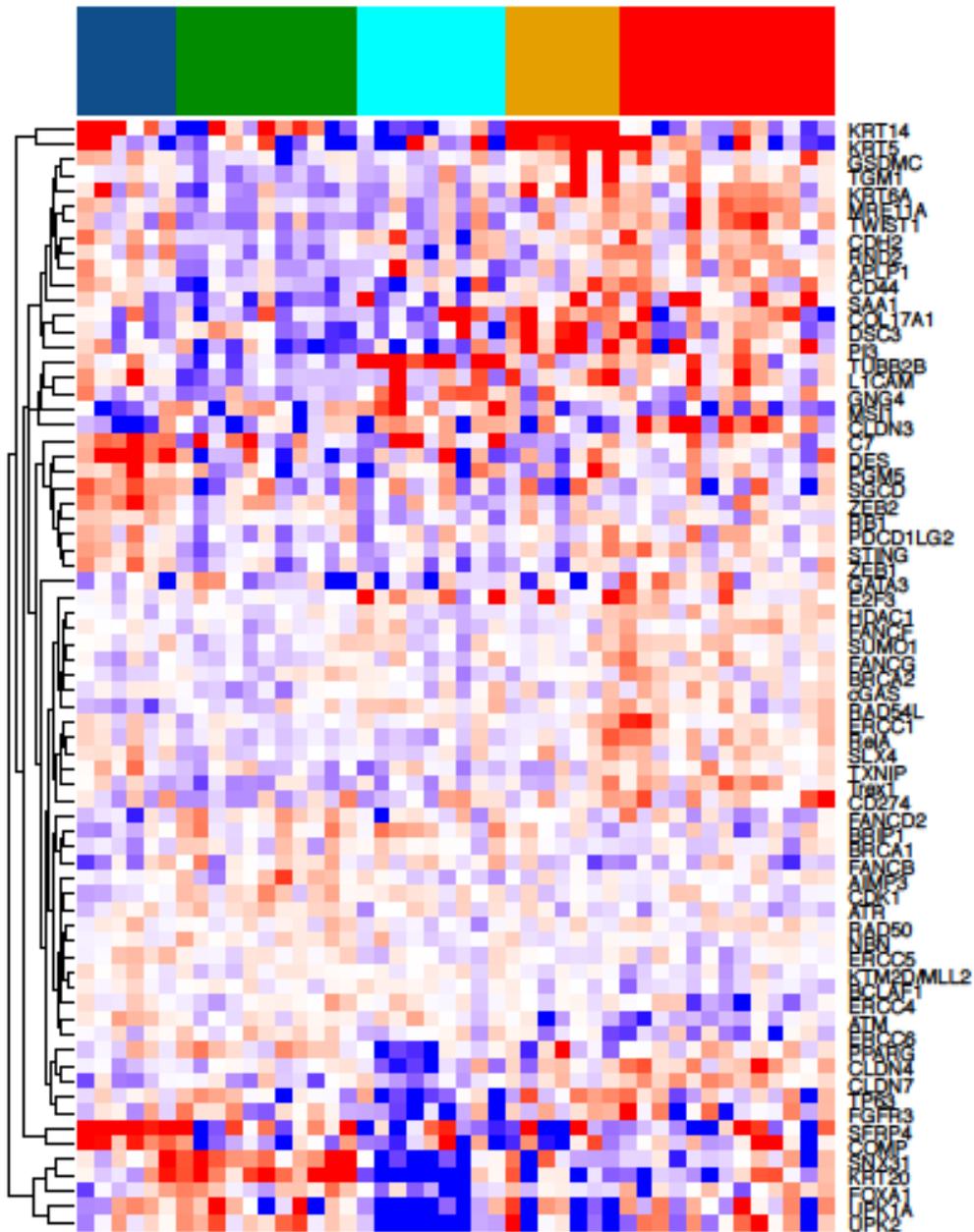
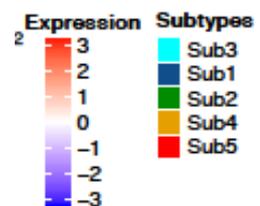


Figure 5.9 Heatmap illustrating the 5 NMF71 subtypes



Subtype 1 was enriched for genes within the epithelial-mesenchymal transition (EMT) pathway such as *SGCD*, *CDH2*, *SFRP4*, *ZEB2* and *COMP*, and also enriched in extracellular matrix genes such as *DES*. *CLDN3* and *CLDN7* were underexpressed which would be in keeping with a claudin-low subtype. Interestingly, *RAD54L*, *BRIP1* and *CDK1* were also underexpressed; *RAD54L* and *BRIP1* are involved in homologous recombination (repair of double stranded DNA breaks). There was a trend towards a lower tumour content compared to other subtypes which is something seen in TCGA luminal infiltrated cases⁶.

Subtype 2 was enriched for luminal markers such as *KRT20*, *PPARG*, *UPK2*. Of note, this subtype demonstrated higher levels of expression of *AIMP3*, *FANCB* and *NBN* compared to the other subtypes.

Subtype 3 displayed high expression of genes associated with the TCGA neuronal subtype such as *TUBB2* and *MSI1*. *RAD54L* and *FANCB* expression also featured. Luminal markers were underexpressed, in keeping with this being a basal subtype.

Subtype 4 demonstrated high levels of keratins expressed by basal cells (*KRT14* and *KRT5*, *KRT6A*). *ATM* was underexpressed although not to the same degree as that seen in subtype 5. Subtype 4 also demonstrated the highest levels of *L1CAM*, which was categorised as an immune marker in the TCGA report.

Subtype 5 was characterised by moderate expression of EMT genes (*CLDN3/4/7*, *TWIST1*). Of all the subtypes, this group had the highest expression levels of *Trex1* and *MRE11*. Of note, there was underexpression of *ATM*, *ERCC6*, *ERCC4*, *BCLAF1* and *ATR*. Subtype 5 also had the highest expression of immune markers *SAA1* and *CD274*.

5.4.4.4. Comparing NMF71 with CRCAssigner-38 and TCGA allocations

Table 5.14 below shows the distribution of CRCAssigner-38 and NMF71 subtypes in the radiotherapy cohort. With small numbers in each subtype, it is difficult to reliably report patterns of overlap. However, the colorectal enterocyte and goblet-like subtypes overlap most with NMF71 subtype 5, inflammatory with subtype 3, and stem-like with subtypes 2 and 1.

Table 5.14 Comparing CRCAssigner-38 and NMF71 subtype allocations in the radiotherapy cohort. Greatest overlap highlighted in orange

NMF71 CRC	1	2	3	4	5	Total
Enterocyte	0	2	1	2	4	9
Goblet-like	0	1	1	1	3	6
Inflammatory	1	1	4	1	1	8
Stem-like	4	5	2	2	2	15
TA	0	1	0	1	3	5
Total	5	10	8	7	13	43

Table 5.15 overleaf shows the overlap between NMF71 and TCGA (recreated centroids).

Table 5.15 Comparing TCGA (recreated centroids) and NMF71 subtype allocations in the radiotherapy cohort. Greatest overlap highlighted in orange

TCGA \ NMF71	1	2	3	4	5	Total
Basal-Sq	2	1	3	3	6	15
Lum_inf	3	1	0	0	2	6
Lum_pap	0	2	1	3	4	10
Luminal	0	6	0	0	0	6
Neuronal	0	0	4	1	1	6
Total	5	10	8	7	13	43

Luminal tumours were found in subtype 2 only and most of the neuronal tumours within subtype 3. Basal-squamous tumours tended to be in subtype 5.

Figure 5.10 shows the radiotherapy cohort arranged by response with allocated subtype according to NMF71, CRCAssigner-38 and TCGA (recreated centroids). Visually, there does not appear to be any overlap in the 3 subtypes allocated to each sample with respect to relapse status.

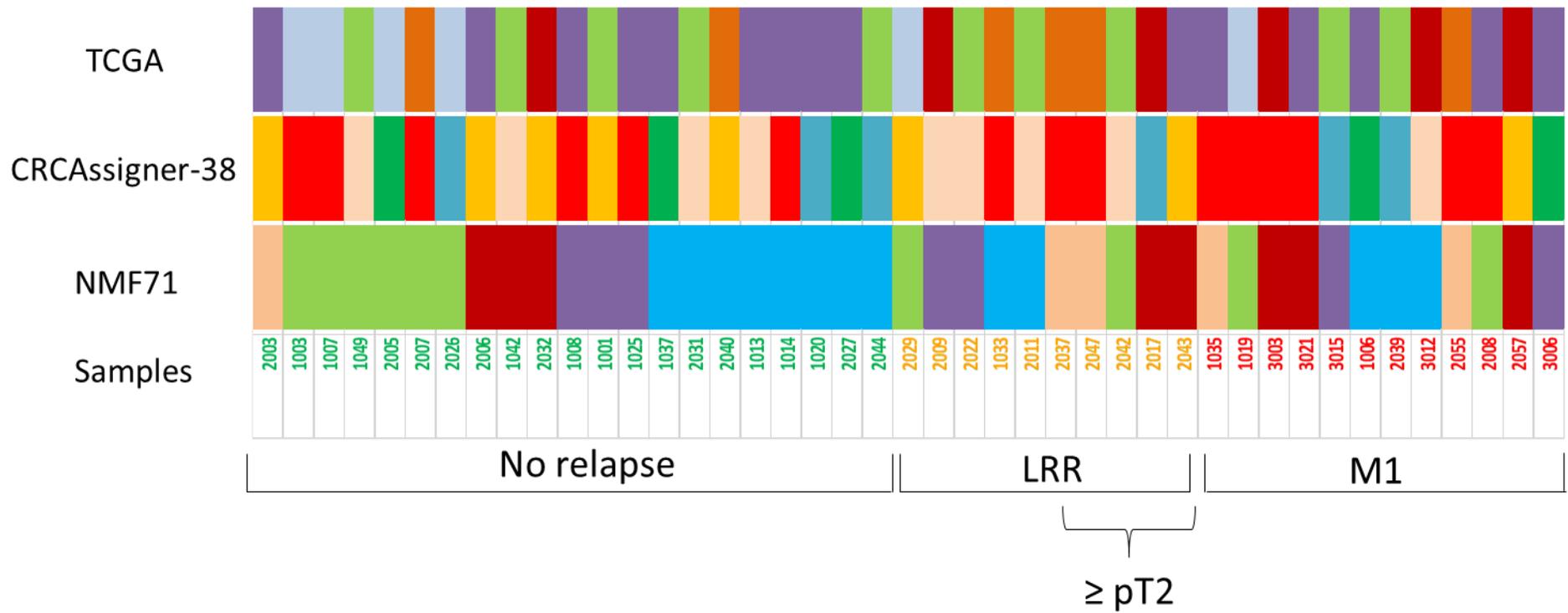


Figure 5.10 Diagram illustrating subtype allocations according to NMF71, CRCAssigner-38 and TCGA (recreated centroids)

CRCAssigner-38		TCGA		NMF71	
	Enterocyte		Basal squamous		1
	Goblet-like		Neuronal		2
	Inflammatory		Luminal infiltrated		3
	Stem-like		Luminal		4
	Transit-amplifying		Luminal papillary		5

5.4.4.5. Applying NMF71 classification to publicly available data

GSC data

On applying NMF71 to the GSC dataset, major response rates to NAC ranged from 29.6% (8/27) in subtype 1 to 50% (2/4) in subtype 5 (see Appendix A5.25). There was no significant association of NMF71 subtype with NAC response ($p=0.849$).

Figure 5.11 illustrates the subtypes assigned to each sample in the GSC dataset using the 4 different classifiers (GSC, CRCAssigner-38, TCGA-recreated centroids, and NMF71). Of note, no association was seen between TCGA subtype and NAC response ($p=0.128$).

It can be seen that there is broad agreement across the classifiers in cases being luminal or basal. Tables comparing the allocation of the different subtypes for each sample can be found in Appendix A5.26.

TCGA dataset

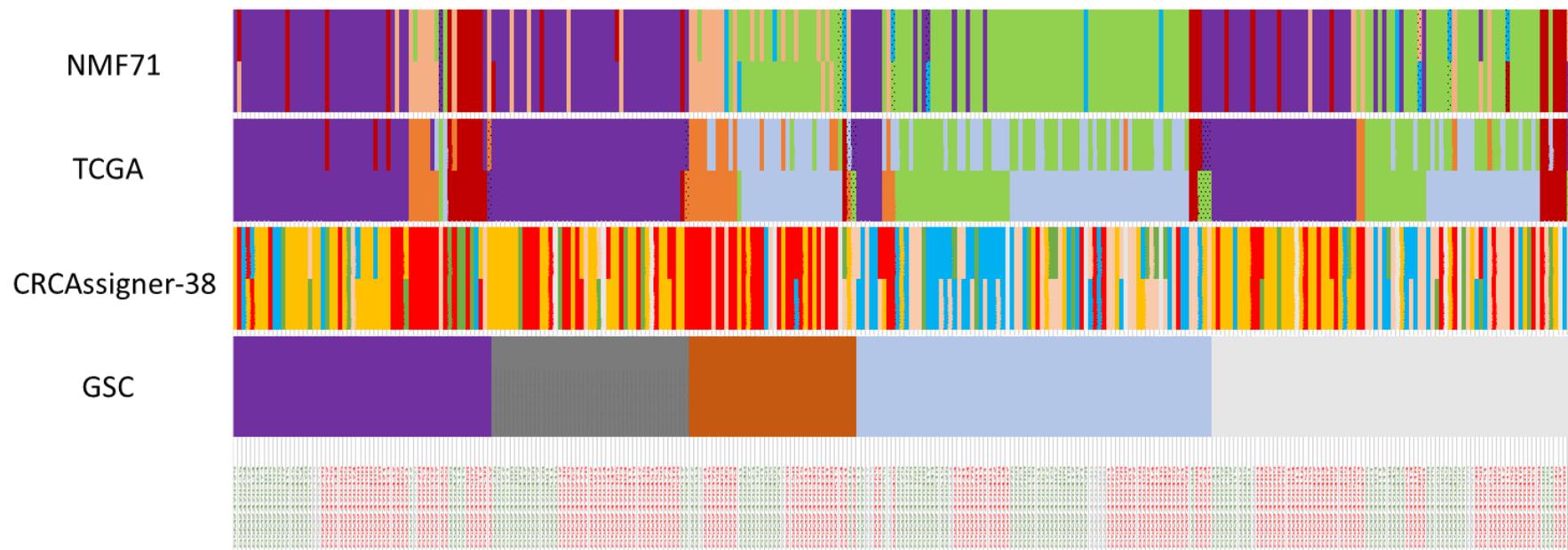
Figure 5.12 illustrates the subtypes assigned to each sample within the TCGA dataset using NMF71. The table in Appendix A5.27 summarises the subtype allocations.

When comparing the TCGA and NMF71 subtypes, it can be seen that 54/84 (64.3%) of the TCGA basal squamous samples were allocated NMF71 subtype 4 and 18/84 (21.4%) to group 3. 11/84 (13.1%) were deemed to have a mixed NMF71 subtype and further exploration revealed samples to be either a mix of subtypes 3 and 4, or of subtypes 3 and 1.

Of the samples labelled as TCGA luminal, 9/15 (60%) samples fell into subtype 2, and 3/15 (20%) were assigned to subtype 1. The remaining 3/15 (20%) of cases were all deemed to have a mix of NMF71 subtypes 1 and 3. Interestingly, 47/81 (58.0%) TCGA luminal papillary were also labelled as NMF71 group 2, suggesting that NMF71 is less sensitive to the proposed subdivision of luminal cases into luminal or luminal papillary.

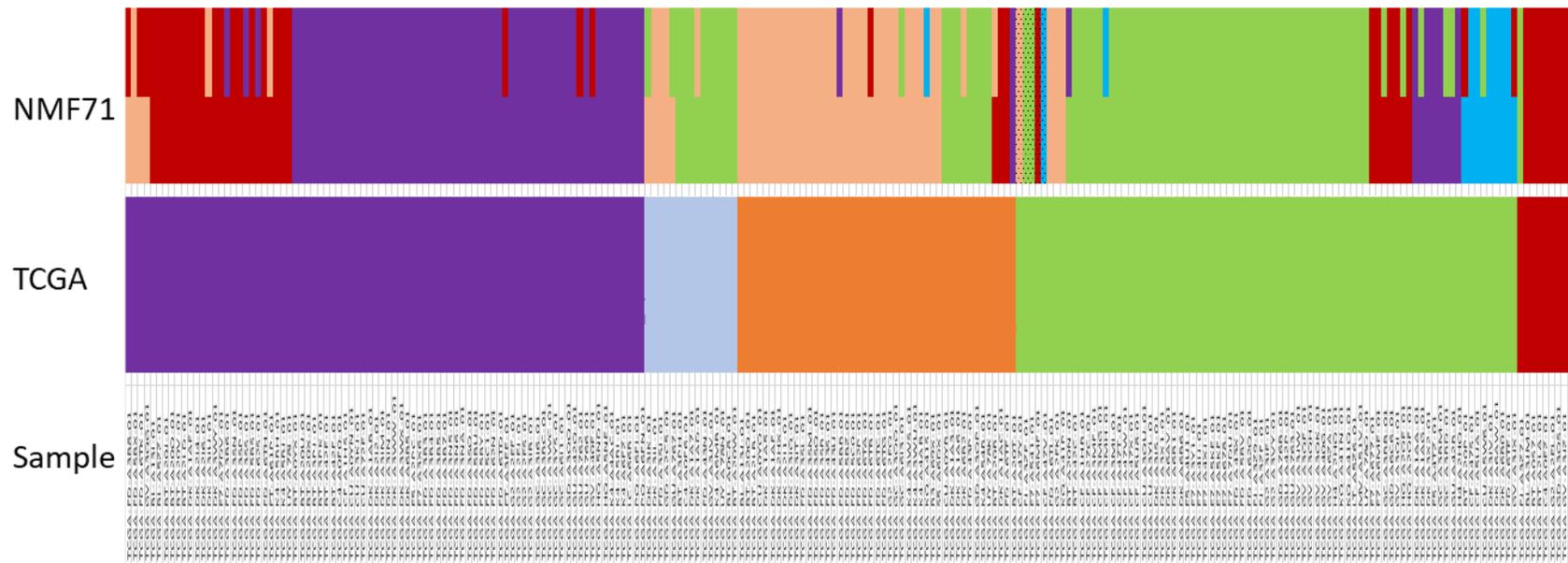
29/45 (64.4%) of the TCGA luminal infiltrated were labelled as NMF71 subtype 1. 8/9 (88.9%) of the neuronal samples fell into subtype 3.

Interestingly, only 6/234 TCGA cases were allocated to subtype 5, and these were all luminal papillary cases.



GSC		CRCAssigner-38		TCGA		NMF71	
	Basal		Enterocyte		Basal squamous		1
	Claudin-low		Goblet-like		Neuronal		2
	Infiltrated luminal		Inflammatory		Luminal infiltrated		3
	Luminal		Stem-like		Luminal		4
	Not assessable		Transit-amplifying		Luminal papillary		5
			Undetermined		Undetermined		Undetermined

Figure 5.11. Diagram illustrating distribution of subtypes allocated using the TCGA, CRC and NMF71 classifications on a dataset of 305 MIBC patients receiving NAC (GSC data set from Seiler et al)



TCGA		NMF71	
	Basal squamous		1
	Neuronal		2
	Luminal infiltrated		3
	Luminal		4
	Luminal papillary		5
			Undetermined

Figure 5.12 Diagram illustrating subtype allocation of 234 MIBC samples from TCGA dataset according to the TCGA allocated subtypes and NMF71

Survival outcomes according to NMF71 subtypes using TCGA dataset

Survival data for 233/234 of the TCGA dataset was available online. Figure 5.13 shows the Kaplan-Meier curves for overall survival for each NMF71 subtype

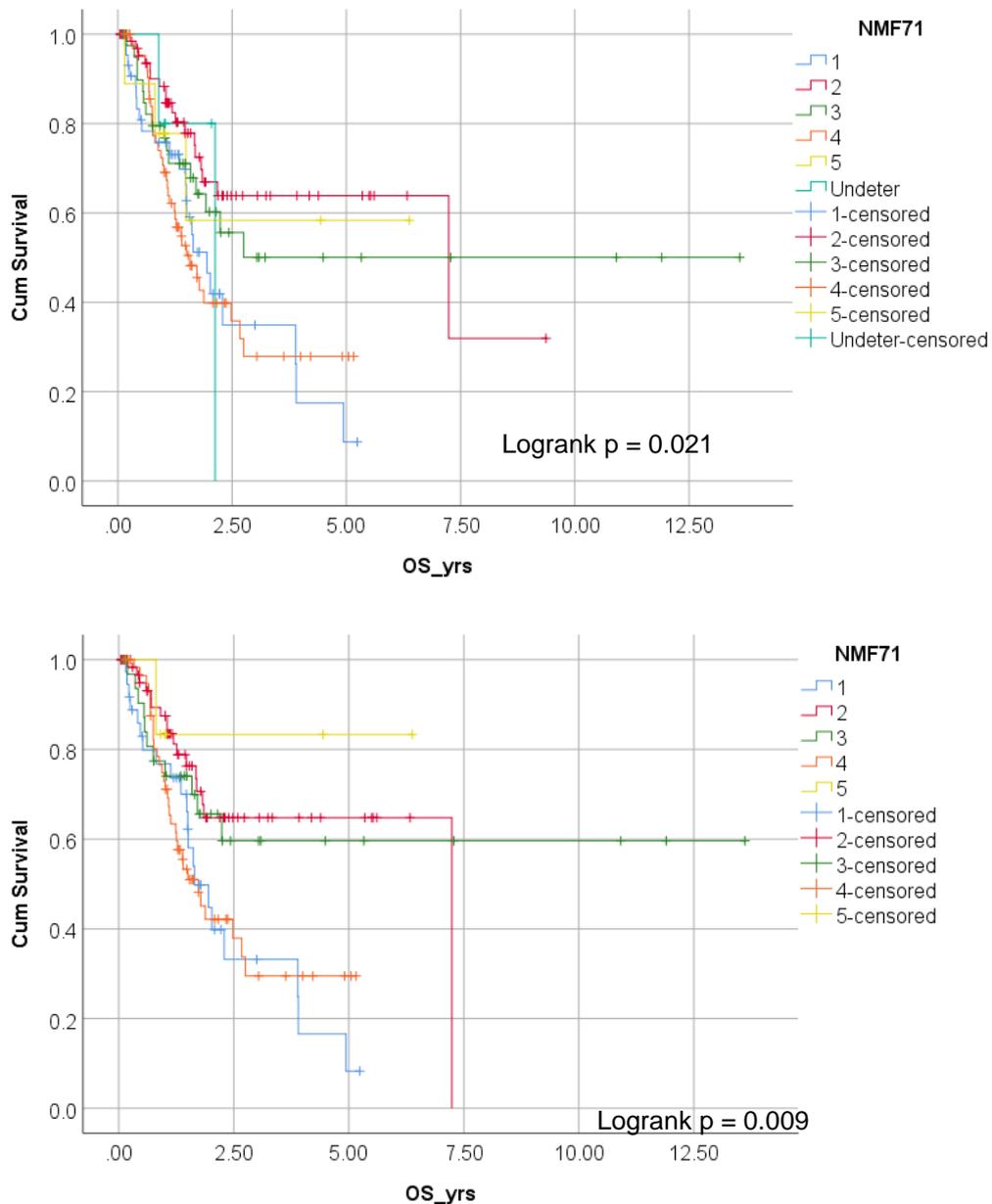


Figure 5.13 Kaplan-Meier curves showing overall survival (OS) in years of TCGA dataset according to a) primary NMF71 subtype i.e. dominant subtype of samples with mixed subtypes included (n = 233), and b) NMF71 for samples allocated a single subtype i.e. mixed and undetermined cases excluded (n = 199)

To allow fair comparison with the TCGA subtype survival outcomes, I performed Kaplan-Meier analysis using the same cohorts as above but based upon TCGA subtype (figure 5.14).

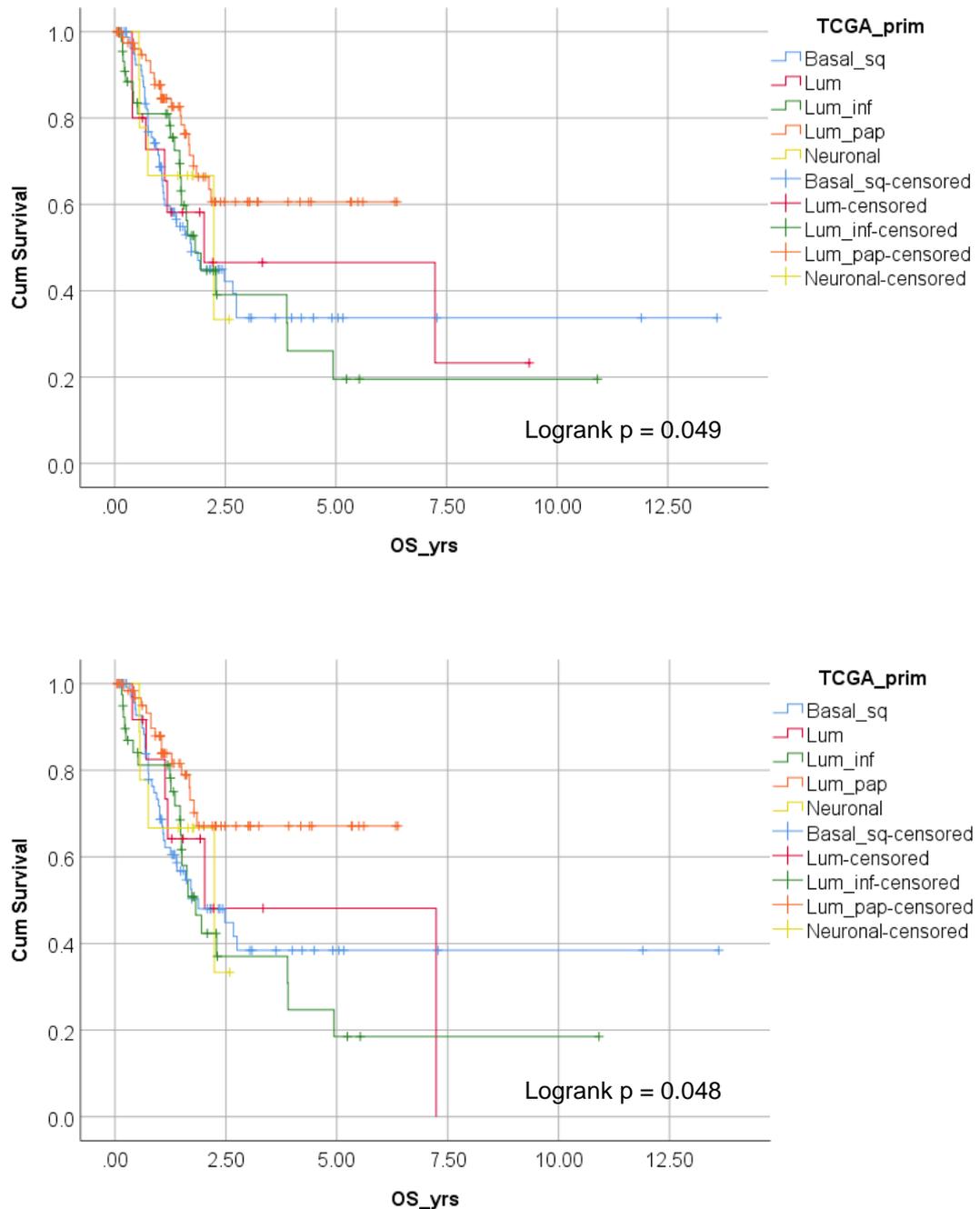


Figure 5.14 Kaplan-Meier curves showing overall survival (OS) in years of TCGA dataset according to TCGA allocated subtype for a) cohort of 233 subjects for whom subtyping and survival data available, and b) for subset of 199 subjects who were allocated a single subtype using NMF71 i.e. mixed and undetermined cases

Figure 5.15 below shows the survival curves for the TCGA dataset according to NMF71 and the RT dataset according to NMF71. It is interesting to note the difference in subtype 4 between the two cohorts.

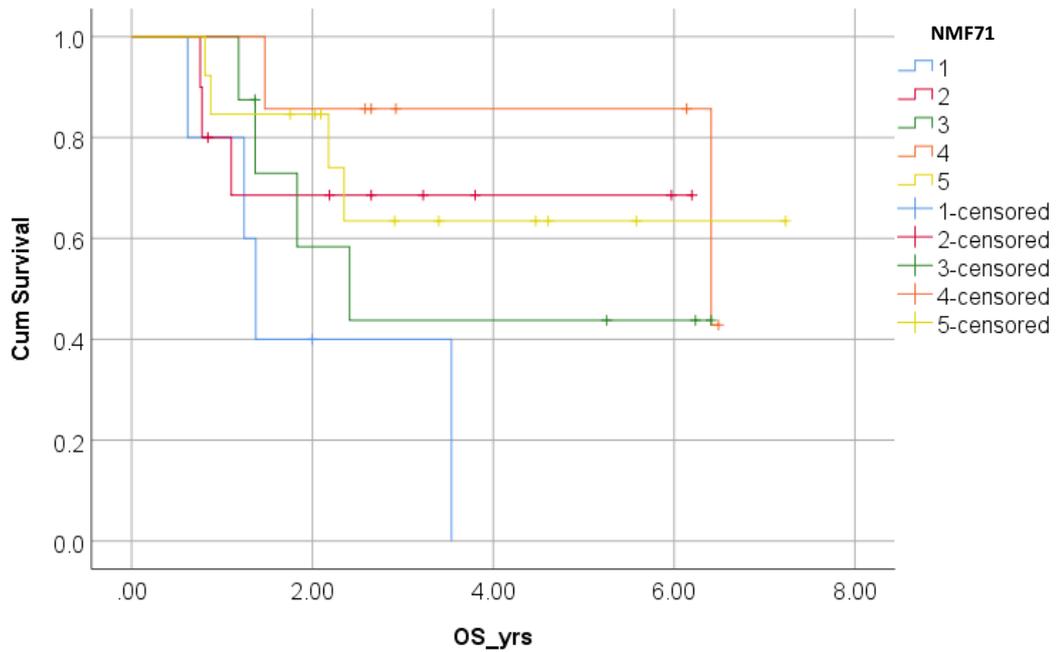
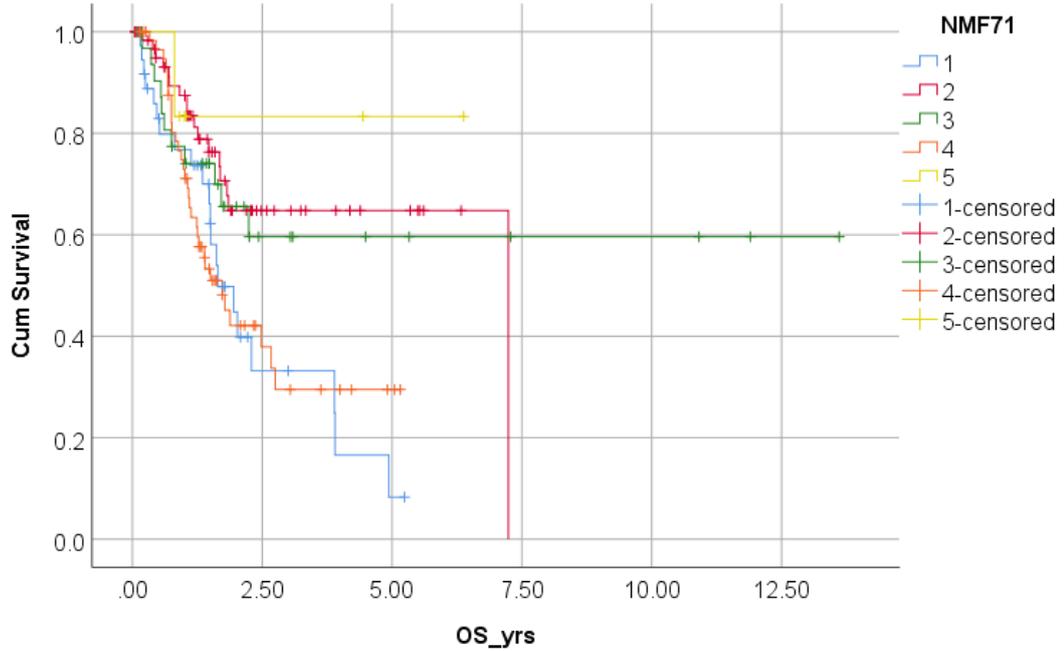


Figure 5.15 Comparing Kaplan-Meier curves for a) TCGA dataset and b) radiotherapy dataset, according to NMF71 subtype.

5.4.5. Inpatient heterogeneity within RT cohort

Three patients had RNA from more than one region from the same FFPE block tested. For 1 patient, the additional sample sent failed and so no result was available, thus leaving only 2 patients with more than one sample subtyped. Table 5.16 below summarises the subtypes allocated to each sample.

Table 5.16 Table summarising results from multiple samples sent from single individual patients

Patient	Sample	Histology	Tumour content	TCGA subtype	CRC	NMF71
2007	A1	TCC	90%	Luminal	Inflamm	2
	A2	TCC	80%	Lum-inf	Stem-like	2
	A3	TCC	60%	Lum-inf	Stem-like	1
2009	A1	TCC	90%	Basal-sq	Enterocyte + TA	4
	A3	TCC + neuroendocrine differentiation	70%	Neuronal	Enterocyte	3

5.4.6. Differentially expressed genes

Focussing on a subset of 36 genes associated with DNA damage repair or candidate radiosensitivity genes, Mann-Whitney tests comparing gene expression between patients with or without any locoregional recurrence, and additionally with or without invasive locoregional recurrence were performed.

Tables 5.17 and 5.18 below show the raw and adjusted p-values obtained for those with the most 'significant' raw p values. The full table of results can be found in Appendix A5.28. None of the genes reached statistical significance once adjustments were made for multiple testing.

A positive log₂ fold change value indicates higher levels of expression in patients with locoregional relapse vs those without, and a negative indicates lower expression in those with locoregional relapse i.e. a log₂ fold change of 1 indicates the gene expression level is twice as high in patients who had a locoregional relapse compared to those with no locoregional relapse, and conversely, a log₂ fold change of -1 indicates that the gene expression level in those with locoregional relapse is half of that that seen in patients with no locoregional relapse.

Table 5.17 Table summarising the top 5 most differentially expressed genes with respect to locoregional recurrence status; p-values adjusted for multiple testing using Benjamini-Hochberg correction with FDR 0.05

Gene	Locoregional Recurrence		
	Log ₂ Fold Change (Relapse – No relapse)	Raw p value	Adjusted p value
ERCC2	-0.46	0.002	0.072
PKC (PRRT2)	1.37	0.006	0.108
HDAC1	-0.37	0.043	0.516
MRE11	0.67	0.078	0.702
SLX4	0.43	0.104	0.7488

Table 5.18 Table summarising the top 5 most differentially expressed genes with respect to invasive locoregional recurrence status; *p*-values adjusted for multiple testing using Benjamini-Hochberg correction with FDR 0.05

Gene	Invasive LRR		
	Log2 Fold change (Relapse – No relapse)	Raw p value	Adjusted p value
<i>ATM</i>	1.05	0.002	0.072
<i>ERCC5</i>	0.63	0.002	0.072
<i>ERCC2</i>	-0.45	0.005	0.09
<i>BRCA2</i>	-0.58	0.008	0.096
<i>PKC</i> (<i>PRRT2</i>)	1.36	0.034	0.306
<i>HDAC1</i>	-0.38	0.105	0.756

Of note, the result for *PKC* expression failed for sample 1049. I therefore removed 1049 from the dataset and repeated the analysis. The raw p-values for *PKC* with regards to LRR and invasive LRR were 0.009 and 0.042 respectively and did not alter the adjusted p-values.

There has been a lot of interest in the potential association of *MRE11* levels at an immunohistochemical level with cause-specific outcome following radical radiotherapy in bladder cancer, and its potential role as a predictive biomarker. I therefore explored this in my dataset. I divided the cohort into high and low *MRE11* expression in accordance with previous thresholds used⁹ and performed a Kaplan Meier analysis. No difference in bladder cancer specific survival seen (Appendix A5.29).

5.4.7. Comparing gene expression and genomic aberration data

5.4.7.1. Overview

A total of 21/43 patients had both RNA expression data and genomic data (from Foundation Medicine F1 panel testing) available.

The *TERT* promoter was the most commonly seen mutation in this subset of patients with 16/21 (76.2%) affected. There was a trend towards subtype 3 having fewer *TERT* promoter mutations (33% compared to 75-100% of other subtypes) but this was not significant ($p=0.065$). The small numbers in each group limited any meaningful analysis with regards to aberrations seen in each subtype. The full list of aberrations reported can be found in Appendix A5.30.

5.4.7.2. Tumour mutational burden

Table 5.19 below shows the mean tumour mutational burden (TMB) for each subtype. There was no statistically significant difference in TMB between the subtypes ($p = 0.201$).

Subtype	N	Mean TMB	Range
1	3 (result available for n = 2)	6.5	4 - 9
2	6	10.2	3 - 24
3	3	14.2	8 - 17
4	5	9.2	5 - 13
5	4	17.25	11 - 21

Table 5.19 Mean tumour mutational burden (TMB) according to each NMF71 subtype; measured in mutations per megabase

5.4.7.3. ATM

In view of my previous results showing underexpression of *ATM* in NMF71 subtype 4 and 5, further exploration of *ATM* status in this group was performed. *ATM* was the 3rd most frequently mutated gene in this cohort with

9/21 (42.9%) demonstrating an *ATM* aberration. Of the 9 patients with an *ATM* aberration, one (11%) had invasive LRR. Five of the 12 with wildtype *ATM* (42%) had invasive LRR. This was not statistically significant ($p = 0.1778$ -fishers exact).

No difference in *ATM* status was seen between the 5 NMF71 subtypes ($p = 0.5444$; $n = 1, 4, 0, 2, 2$ in each subtype).

5.5. Discussion

5.5.1. Molecular subtype as a biomarker of radiosensitivity

5.5.1.1. Radiotherapy response and molecular subtyping

In this work, I have demonstrated that by combining TCGA subtype genes with those associated with DNA damage repair and radiosensitivity genes, 5 molecular subtypes can be identified. My results suggest that these subtypes may be associated with radiotherapy response in MIBC. Only one other study has reported on the relevance of molecular subtype in the context of radiotherapy in MIBC. In this paper, patients who had received bladder-preserving trimodality treatment were subtyped using the GSC subtypes; the authors reported a trend for luminal subtype to have improved OS compared to other subtypes¹⁰.

These results were not reproduced in my data where subtype 2, which displayed luminal features, appeared to have an intermediate survival outcome (2-year overall survival of 68%), whilst NMF71 types 4 (basal features) and 5 (neither strong luminal nor basal features) had better 2-year overall survival probabilities of 85% and 84% respectively. It would be interesting to apply NMF71 to the dataset from this group and explore whether my initial pilot findings are reproducible in a larger independent dataset.

It is interesting that only 57.1% of their cases with microarray data generated passed quality checks for inclusion in the final analysis. The reasons for this are not clear and are not addressed within the published abstract. It is therefore very satisfying that of my 51 patients from whom RNA was extracted, 44 had sufficient RNA for Nanostring analysis and the gene expression data from all but one of these patients passed quality checks (43/44; 97.7%). This success rate with the Nanostring platform is highly encouraging, especially in the context of using FFPE material which is recognised to cause artefact and nucleic acid degradation, particularly with RNA¹¹. My experience with the Nanostring platform in this work would support its applicability to the clinical setting given the high success rate in generating expression data for analysis, the relatively short turnaround from sample analysis to data, and relatively low

cost when compared to other profiling platforms. Other alternative platforms for RNA expression include Taqman Low Density Array (TLDA) but this requires cDNA synthesis and PCR, thus introducing potential PCR amplification bias, and represents a more complex workflow. Quantigene has the advantage of not requiring RNA extraction or target amplification, thus avoiding further inevitable RNA degradation during the extraction process. However, only up to 80 genes of interest can be interrogated at one time.

Of note, there were a subset of samples macrodissected for the intra-tumour heterogeneity substudy which were not suitable for the Nanostring platform due to low concentrations achieved despite good overall yields. The manufacturer's protocol states that lower RNA inputs can be used although at the time of my work, this had not been routinely performed within Dr Sadanandam's team. The team have however since achieved good results using just 25ng of RNA. With a lower input requirement, there would potentially be a total of 13 RNA samples available to test from 10 patients, 7 of whom have already had at least one sample tested in this pilot. Alternatively, other methods of concentrating RNA could be considered but such techniques are not currently routinely performed in FFPE samples due to the risk of further degradation.

5.5.1.2. Overlap with TCGA classification- is NMF71 novel?

While accepting the limitations of a small cohort size in my radiotherapy cohort, NMF71 subtype appeared to have an association with invasive locoregional relapse following radical radiotherapy+/-chemotherapy. No such association was seen with the TCGA subtypes allocated to this dataset.

Given the broad overlap between the TCGA and NMF71 subtypes, it is perhaps surprising that the TCGA subtypes did not show an association with radiotherapy outcome. There are several points to consider here.

Firstly, it was necessary to re-create the TCGA classification (centroid) using publicly available data, as the group did not make the 'official' classifier readily available. The finding therefore that the TCGA subtypes were not associated with radiotherapy outcome relies upon the re-created classifier being accurate.

One could suggest therefore that if using the 'official' TCGA subtypes, perhaps an association with radiotherapy outcome might have been seen. However, there are several reasons to have confidence in our recreated centroids. Firstly, the pipeline used to recreate the centroids has been previously used by the bioinformatician (Gift Nyamundanda) in other similar settings and so this approach is 'tried and tested'. Secondly, the application of the NMF71 classifier to the TCGA dataset which included the 'official' TCGA subtype allocation for each sample showed similar patterns of overlap with the results on the GSC dataset, thus giving confidence in our recreated centroids. The comparison table is available in Appendix 5.31. It would, however, be worthwhile contacting the TCGA group again to request the centroids, and repeating the analysis to explore any potential significant differences as in such a small cohort, the re-labelling of even a few cases could have a significant impact on the results and it may be that the TCGA classification does in fact have some association with radiotherapy response.

5.5.1.3. NMF71 subtype 4 and 5 appear to be associated with improved outcomes following radiotherapy+/-chemotherapy

Tumours with basal features are recognised to be associated with poorer outcomes^{1,5,12,13}. In my own analysis of the TCGA data which included those primarily treated with radical cystectomy, the 2-year overall survival of those with basal squamous tumours was <50%. Repeat analysis of the same dataset using NMF71 subtypes showed that subtype 4 had a similarly poor 2-year survival rate, and this reflects the broad overlap between the TCGA basal-squamous and NMF71 subtype 4 groups; 64% of TCGA basal-squamous tumours were allocated to NMF71 subtype 4. Subtype 1, which demonstrated features of a claudin-low subtype (previously reported to be of a basal origin¹²), also performed similarly poorly and this likely reflects the shared basal features in both subtypes 1 and 4.

However, in my radiotherapy-treated cohort, subtype 4 was associated with a 2-year overall survival of 85%, and using the TCGA recreated centroids, basal squamous tumours had a 2-year overall survival of 74% in this radiotherapy cohort. This result suggests that patients with subtype 4 tumours or basal-

squamous tumours derive greater benefit from radiotherapy +/- chemotherapy over that from surgery alone. This is worth further investigation in a larger cohort. To the best of my knowledge, the potential association of basal-like subtypes with improved outcomes following bladder preservation strategies over surgery has not been previously reported.

As the majority of the radiotherapy cohort received NAC and concurrent chemotherapy, it could be suggested that the improvement in overall survival observed in NMF71 subtype 4 and TCGA basal squamous patients in the radiotherapy cohort is primarily due to the use of chemotherapy, as opposed to radiotherapy over surgery. Certainly, Seiler et al reported in their NAC cohort that tumours assigned to GSC basal subtype appeared to derive the most benefit from NAC with 3-year OS rate of 77.8% compared to 49.2% in a non-NAC cohort. Reviewing the Kaplan-Meier curves, this is similar to the NMF71 subtype 4 3-year overall survival rates of 85% in the radiotherapy+/- chemotherapy group, and 32% in the TCGA surgical cohort where the majority did not receive NAC.

Given that MIBC is recognised to have high metastatic relapse rates and this is what limits a patient's survival, I suspect that that the benefit seen here in the basal-squamous/subtype 4 groups is most likely due primarily to the addition of chemotherapy, as opposed to being driven by the use of radiotherapy over surgery. It would be interesting to further assess this and compare outcomes according to subtype in surgical cohorts receiving NAC against bladder preservation cohorts receiving NAC +/- concurrent chemotherapy.

5.5.1.4. Overlap between basal and luminal tumours

An observation when comparing NMF71 and TCGA subtypes in the TCGA dataset is that while the majority of TCGA luminal papillary tumours were assigned to NMF71 subtype 2 which has luminal features, just under a quarter were assigned to NMF groups with basal features. This is surprising given that luminal and basal tumours have generally been considered to be entirely distinct and separate. It would be interesting to explore whether the subset of TCGA luminal tumours assigned to non-luminal NMF71 subtypes are

associated with poorer outcomes. Unfortunately due to small numbers when subdividing the TCGA dataset, it is not possible to perform any useful analysis on this dataset and exploration would require a much larger cohort.

This apparent blurring between luminal and basal features was also observed in subtype 1 which shares features with the TCGA luminal-infiltrated subtype but also displays claudin-low characteristics, which are thought to be derived from basal origins.

5.5.1.5. Claudin-low subtype in MIBC

The claudin-low subtype was first described in breast cancer but has also been reported in bladder cancer by the UNC group in 2014. They put forward claudin-low as a 3rd subtype alongside basal and luminal, although they considered it as a subset of basal tumour given the absence of luminal markers. The features they reported were that of enrichment of stromal, EMT, immune response and stem cell genes, along with underexpression of *CLDN3*, *CLDN4* and *CLDN7*; similar to that seen in breast cancer¹⁴. They further reported that claudin-low subtypes were found in the old TCGA clusters III and IV which were thought to be of basal origin. On reviewing the heatmap in the updated TCGA report⁶, *CLDN3* and *CLDN7* appear to be most underexpressed within the basal-squamous subtype.

These claudin-low features are in keeping with those seen in NMF71 Subtype 1. However, subtype 1 broadly overlaps with the TCGA luminal infiltrated tumours and both share the features of having a lower tumour content compared to fellow subtypes. It therefore appears that NMF71 subtype 1 encompasses both luminal and basal (albeit basal by virtue of absence of luminal markers) tumours. This goes against the generally recognised division of tumours into either luminal or basal. This may of course be an erroneous finding, perhaps a result of such a small cohort, but it would be interesting to explore this further. Another possible explanation for the co-existence of basal and luminal features is that of tumour heterogeneity. If, however, reflective of true underlying subtypes, this preliminary data supports UNC's work that claudin-low tumours should be considered as a 3rd subtype alongside basal

and luminal, and may further suggest that claudin-low tumours may not necessarily demonstrate strong basal features.

5.5.1.6. NMF71 Subtype 5 in Public Datasets

One marked observation is the number of cases assigned to subtype 5 in each dataset. In the radiotherapy cohort of 43 patients, subtype 5 formed the largest subgroup accounting for 30.2%. However, when applied to the GSC and TCGA datasets, only 4/305 (1.3%) and 6/234 (2.6%) were assigned to subtype 5 respectively.

To explore this further, I looked closer at the gene lists for each dataset as perhaps omission of key genes defining subtype 5 could account for the seemingly disproportionately lower numbers seen. There were 4 genes on the NMF71 panel (*cGAS*, *STING*, *KTM2D/MLL2* and *AIMP3*) that were not available in the GSC dataset, and another 4 genes (*cGAS*, *AIMP3*, *SLX4* and *STING*) in the TCGA dataset that were not tested/reported. Although none of these genes were strongly over or underexpressed in subtype 5, it would not be unreasonable to suggest that their absence may affect the overall allocations.

A further thought was whether there would be a difference in the expression data obtained from the different platforms. Seiler et al used expression data generated from FFPE arrays, the TCGA used RNA-seq on fresh frozen tissue and I used Nanostring on FFPE. Given the results from each set are normalised with respect to each other, I would not expect this to have been likely to cause any significant issues in applying one classifier to another.

Another potential factor that might influence distribution of cases to each subtype is that of population differences. The TCGA samples used in this analysis primarily came from Caucasian patients based in the USA. Seiler et al also based their work on expression data from patients in the USA and in Sweden. My expression data has been generated from patients based in the UK. Variations in subtype distributions between ethnic groups is not unheard of and is currently being explored¹⁵. I think it is unlikely however that this will

account for the marked differences noted in the proportion of NMF71 subtype 5 between datasets.

5.5.2. Molecular features of radiosensitivity

5.5.2.1. Candidate genes of interest

Of the 20 patients assigned to NMF71 subtype 4 or 5, only one had an invasive locoregional recurrence (5%), compared to 8/23 (34.8%) of patients allocated to subtypes 1-3 ($p=0.0243$). Of note, patients in subtypes 4 and 5 had similar median follow-up periods to those in subtypes 1-3 (3.40 vs 3.80 years; $p = 0.591$).

Furthermore, subtypes 4 and 5 together have a statistically significant higher pathological complete response rate after radiotherapy compared to the NMF71 subtypes 1-3 (100% vs 60%; $p = 0.0237$). If validated, this is potentially means NMF71 could function as a predictive biomarker with patients found to have NMF71 subtype 4 or 5 tumours being counselled towards bladder preservation strategies.

The molecular features of each subtype have already been summarised but there are some features and genes of particular interest.

ATM

The ataxia-telangiectasia mutated (*ATM*) gene was underexpressed in subtype 4 and particularly subtype 5.

ATM plays a key role in initiation of DNA damage repair pathways by interacting with the MRN complex which is composed of *MRE11*, *NBN* and *RAD50*¹⁶. Decreased levels of expression might therefore be hypothesised to result in decreased activation of DNA repair pathways with subsequent radiosensitivity. Of note, *ATR* was also underexpressed in subtype 5 and plays a similar role to that of *ATM* in sensing DNA damage and initiating repair pathways.

To the best of my knowledge, my data is the first to suggest that *ATM* mRNA level may be a predictor of radiotherapy response.

Of note, conversely, NMF71 subtype 1 had the highest levels of *ATM* expression and the highest incidence of invasive LRR (3/5; 60%) of the 5 groups. Although the numbers are small and no definitive conclusions can be drawn from this, the higher level of invasive local recurrence seen in NMF71 subtype 1 would fit with the hypothesis that *ATM* overexpression and underexpression is associated with radioresistance and radiosensitivity respectively.

These results described are also supported by the data from my differential gene expression analysis where *ATM* was noted to be expressed at twice the level in those with invasive locoregional relapse compared to those without (raw p value = 0.002), although this was no longer statistically significant after adjustment for multiple testing (p = 0.072). Correlation with genomic results did not show any significant association between *ATM* aberration status and outcome but numbers in the subset were small. On the basis of my pilot data, I think *ATM* warrants further investigation as a potential biomarker of response to radiotherapy, and as a potential target to modulate response to radiotherapy. Since analysing my data, published work has demonstrated that the *ATM* inhibitor, AZD1390, radiosensitises glioma and lung cancer cell lines²¹.

MRE11

MRE11 was most highly expressed in NMF71 subtype 5, and this gene has been the focus of several groups' work. It was first shown to be a potential predictive biomarker of radiotherapy response in MIBC by Choudhury et al⁹. They reported that higher *MRE11* levels at an immunohistochemical level were associated with better cause-specific survival following radiotherapy but not cystectomy. These results were validated in an independent cohort²⁰ although more recent work from 2 groups^{19,22} found no such association.

In my cohort of patients, higher expression levels of *MRE11* were seen in subtype 5, members of which had lower incidence of invasive locoregional recurrence and better overall survival. Dividing the cohort into those with high and low *MRE11* levels based upon the 25th percentile did not appear to show any difference in bladder cancer-specific survival, but this is not surprising given the lack of correlation reported between mRNA and protein levels²³.

AIMP3

Subtype 2 demonstrated higher levels of expression of *AIMP3*, a tumour suppressor and regulator of DNA damage, compared to the other subtypes. As previous work has demonstrated that *AIMP3* depletion results in an accumulation of double strand DNA damage²⁵ and radioresistance²⁶, one might therefore hypothesise that with higher expression levels of *AIMP3*, subtype 2 might be associated with radiosensitivity although it is not possible to confirm this within my small dataset.

NBN

NBN codes for the NBS1 protein which interacts with *MRE11* and *RAD50* to form the MRN complex which is key in detecting double strand DNA damage²⁷. *NBN* was expressed the most in subtype 2 although at relatively modest levels. There is, to the best of my knowledge, no data on the influence of *NBN* mRNA expression levels on radiotherapy response. However, interestingly, *NBN* gain has been demonstrated to be a predictor of poorer 5-year biochemical progression free survival in prostate cancer patients treated with image-guided radiotherapy but not in radical prostatectomy patients, leading the authors to suggest a possible role for *NBN* gain as a biomarker of radioresistance²⁸. Their in-vitro work also suggested that *NBN* gain was associated with intrinsic tumour radioresistance. There did not appear to be significantly increased relapse rates in subtype 2. However, it would be interesting to explore this in a larger cohort as one might hypothesise that the features seen in this subtype might be associated with poorer responses post radiotherapy due to increased DNA damage detection and subsequent repair.

Nucleotide excision repair genes

ERCC 1/2/4-6 were included in the Nanostring panel and are involved in nucleotide excision repair, which is the primary pathway by which adducts such as those from cisplatin or mitomycin C are repaired. It was somewhat surprising that *ERCC2* did not contribute to the final NMF71 panel despite being expressed at statistically significant lower levels in patients with locoregional relapse ($p = 0.002$) before multiple testing correction. There was also a trend towards lower expression levels of *ERCC2* in patients with invasive locoregional relapse ($p = 0.06$). Its omission from the NMF71 panel might in part be due to the fact that the fold change observed was small at -0.46 and -0.45 respectively, and that *ERCC2* primarily plays a role in the removal of platinum adducts rather than radiation-induced DNA damage repair.

Given the proposed role of *ERCC2* mutations with regards to chemosensitivity, it would be interesting to look at whether *ERCC2* status influences the effects of concomitant mitomycin-C which is used as a radiosensitiser in bladder radiotherapy. One hypothesis might be that those with reduced *ERCC2* function gain more radiosensitising effect from concomitant chemotherapy, while those with 'normal' or increased *ERCC2* effects may be better served by radiotherapy alone, or alternative radiosensitisers such as carbogen and nicotinamide.

In contrast to *ERCC2*, *ERCC4* and *ERCC6* however did form part of the NMF71 panel and were both underexpressed in NMF71 subtype 5. As far as I am aware, there is no relevant current data reporting on *ERCC4* or *ERCC6* expression in the context of bladder cancer or radiotherapy. Given the role of the ERCC gene family in nucleotide excision repair, it is not unreasonable to suggest that expression levels of *ERCC4* and *ERCC6* may influence the effect of chemotherapy given neoadjuvantly and/or concurrently. The majority of patients in the RT cohort received neoadjuvant chemotherapy and concurrent chemotherapy as they were treated following results from key trials^{31,32} demonstrating survival benefit for NAC and concurrent chemotherapy. The potential interaction between the *ERCC* genes expression and chemotherapy

is not something that can be investigated in the existing cohort as the subset of those not receiving chemotherapy is too small to make any worthwhile comparisons, and of course, the overall cohort size is small. Going forward, it would likely be difficult to prospectively accrue numbers for patients treated with radiotherapy alone as the standard of care is to use chemotherapy neoadjuvantly and concurrently if proceeding with a bladder-sparing strategy³¹⁻³⁶. Patients now treated with radiotherapy alone are likely to have significant comorbidities precluding the use of chemotherapy or surgery, and these would likely be confounding factors in data analysis. However, the capacity to retrospectively recruit patients to the CoMB protocol would allow expansion of a radiotherapy alone cohort, if focussing on those treated before neoadjuvant and concurrent chemotherapy became the standard of care in the UK. Analysis of these patients could provide further understanding of the mechanisms underlying response to chemotherapy in the neoadjuvant and concurrent settings.

Homologous recombination genes

As previously mentioned, subtype 1 had the highest incidence of invasive locoregional recurrence. This subtype had the lowest levels of *RAD54L* and *BRIP1*, which are components of the homologous recombination pathway, responsible for the repair of double-stranded DNA breaks (DSB).

One might therefore reasonably anticipate that underexpression of these genes would result in reduced DSB repair and subsequent radiosensitivity, but this does not fit with the previously proposed hypothesis that overexpression of *ATM* and *RB1* in subtype 1 has contributed to radioresistance. One possible explanation might be related to the type of DNA repair. DSB predominantly repaired by NHEJ in G1, HR in G2-M³⁷. Greater levels of *RB1* might therefore cause cells to arrest at G1/S checkpoint which is a less radiosensitive phase, and where the predominant repair mechanism is that of NHEJ. i.e. *RAD54L* and *BRIP1* would not contribute significant DNA repair and therefore the effect from their underexpression not significant.

Immune markers

It was interesting to note immune marker expression was overall highest in subtype 5 and an immune mechanism may underlie some of the results seen with respect to outcomes following radiotherapy. There is evidence that the presence of tumour-infiltrating lymphocytes is associated with improved disease-free survival and overall survival in MIBC patients managed surgically³⁸, and more recent work has reported that tumours with high immune infiltration have improved disease-specific survival after trimodality bladder sparing treatment⁴¹. These findings support the features seen in NMF71 subtype 5. I propose that the likely mechanism behind this is that of radiation inducing an immune response with the release of immunogenic tumour material following cell injury/death³⁹. This principle underlies the abscopal response whereby following radiotherapy to a lesion, response may be seen in a separate area of disease outside of the radiation treatment area. It is understood that tumour material released by the radiotherapy target triggers an immune response, allowing distant sites expressing the same tumour antigens to be targeted by the immune system, and underpins the rationale for combining radiotherapy with immunotherapy⁴⁰. One might expect that the ready presence of an immune infiltrate at the radiotherapy target may therefore result in a heightened immunogenic response, and potentially improved outcomes.

Tumour microenvironment

Recent work using the GSC classifier has reported that luminal-infiltrated tumours have a trend towards higher stromal signatures, and that tumours with stromal signatures had poorer outcomes following surgery (+/- NAC), but no such association was seen in the bladder preservation group⁴¹.

In my data, cases designated GSC luminal-infiltrated contained cases allocated to NMF71 subtypes 1 and 2; subtype 1 had a lower tumour content likely representing an increased stromal component, as supported by high levels of expression of extracellular matrix genes such as *DES*. Within my small cohort, subtype 1 had the highest rate of invasive locoregional relapse, and poorest survival outcomes of the 5 NMF71 subtypes. Although it is difficult

to comment further given that formal statistical comparison was limited by small subcohort numbers, the trend is interesting to note, and does not fit with the findings described above that no association of poor outcome was seen in luminal-infiltrated tumours treated with TMT.

5.5.3. Molecular subtype stratification for treatment

The concept of using molecular subtypes to guide treatment is not new. There are several tests available in breast cancer to guide decisions about whether or not adjuvant chemotherapy is indicated in patients with lymph node negative, ER positive, Her2 negative tumours⁴². Further exploration of molecular subtypes in MIBC may also lead to the development of clinically applicable assays to guide the management of bladder cancer, and identification of potential therapeutic targets.

5.5.3.1. NMF71 subtype 1

One key application would be the identification of chemoresistant patients. Eighty percent of the TCGA luminal infiltrated group was noted to have features similar to MDA's proposed chemoresistant p53-like subtype¹. One might reasonably hypothesise then that given the overlap between the TCGA luminal infiltrated and my NMF71 subtype 1 groups, a lower response rate to NAC might be seen. However, while both the luminal infiltrated and NMF71 subtype 1 groups each had the lowest proportion of NAC responders (18.8% and 29.6% respectively) in the GSC dataset compared to their fellow subtypes, this was not statistically significant. Given the trend seen however and the potential clinical significance, I think this is worth further exploring.

Of further interest, the TCGA luminal infiltrated group contained 23/24 of the previously named cluster 2 which was reported to benefit most from anti-PDL1 agents⁴³ with higher response rates seen in the metastatic setting. It would therefore be interesting to explore whether NMF71 subtype 1 could be a robust predictor of response to immunotherapy in the metastatic setting and neoadjuvant setting. The recently reported PURE-01 study⁴⁴ was a phase II trial which explored the use of pembrolizumab in the neoadjuvant setting. They

reported a complete pathological response rate of 42% of patients. Additionally, a significant association was noted between tumour mutational burden (TMB) and pT0 response with a TMB cut off at 15 mut/Mb. Similar findings have also been reported using neoadjuvant atezolizumab⁴⁵. Only 2 of the 3 patients in subtype 1 with additional genomic profiling had a TMB result and therefore it is not possible to comment any further on this in my dataset. Of note, the highest mean TMB was seen in subtype 5 (mean = 17.25), thus raising the hypothesis that subtype 5 might also demonstrate improved response to immunotherapy.

5.5.3.2. NMF71 subtype 2

Subtype 2 was enriched for luminal markers and encompasses the majority of those labelled as TCGA luminal and luminal papillary. This suggests that NMF71 does not distinguish between the two as the TCGA classifier does. The key question here is whether this lack of distinction would be of clinical relevance in terms of treatment strategy or overall prognosis. In the TCGA data, 44% of the luminal papillary cases were enriched for *FGFR3* aberrations making this a potential target whereas this was not seen in luminal cases. Saying that, luminal tumours accounted for only 6% of the TCGA cohort and given that less than half the TCGA luminal papillary group demonstrated potential targetable *FGFR3* aberration, I think it unlikely that the lack of distinction between the two luminal TCGA subtypes is of any clinical significance.

5.5.3.3. NMF71 subtype 3

Subtype 3 had high expression of genes associated with the TCGA neuronal subtype. Luminal markers were underexpressed in keeping with this being a basal subtype. On reviewing histology, this group included all 3 cases with neuroendocrine differentiation and 5 others with none reported at histological review. It would be worthwhile getting these cases re-reviewed by Dr Hazell but the lack of neuroendocrine histological features seen in the majority of samples within this subtype mirrors the findings of the TCGA group.

As previously discussed, this cohort of patients might benefit from 'neuroendocrine' type systemic regimens, regardless of histopathological findings, and this would be of interest to explore further.

Of note, the TCGA report did not highlight any novel targetable aberrations associated with the neuronal subtype. 17/20 cases did not have any histopathological features of neuroendocrine differentiation despite demonstrating high expression levels of neuroendocrine markers. In NMF71 subtype 3, all 3 cases with recorded neuroendocrine differentiation fell into this group with an additional 5 members without any histopathological neuroendocrine features seen. Identification of this subtype may therefore be useful in highlighting patients that might benefit from 'neuroendocrine-type' chemotherapy regimens i.e. cisplatin/etoposide instead of gemcitabine and cisplatin as is usually used in MIBC. This of course needs to be explored in a much larger cohort and in the context of a prospective trial. The rarity however of such cases would make this challenging.

5.5.3.4. NMF71 subtype 4

As previously discussed, this group appear to derive the most benefit from chemotherapy +/- radiotherapy and current data available⁵ would support the improvement being largely due to response to platinum-based chemotherapy. My data also show a higher rate of pathological complete response following radiotherapy and further work is warranted to further explore and validate my findings. Interestingly, the gene expression profile of subtype 4 did not demonstrate much contribution from the DNA damage repair genes added to the panel, other than *ATM* which was underexpressed. Subtype 4 also demonstrated overexpression of *L1CAM*, overexpression of which in various cancers has been reported to be associated with poor prognosis, tumour progression and lymph node involvement⁴⁶.

Work has also suggested that anti-EGFR agents may be effective in patients with basal tumours⁴⁷. 8/21 patients with genomic data available had an *EGFR* aberration and these cases were distributed across subtypes 2-5.

5.5.3.5. NMF71 subtype 5

My results suggest that this subtype may be associated with reduced invasive locoregional recurrence following radiotherapy. Much more work is required to validate this finding and explore whether this subtype is a predictive biomarker of response to radiotherapy. If proven to be so, this could be a valuable tool in counselling patients between surgery and radiotherapy, and could be applicable to other cancer sites. If confirmed to be radiosensitive, then perhaps there might even be potential to consider dose 'de-escalation' studies to find the optimal dose to achieve disease control and minimise normal tissue toxicity.

This subtype had the highest mean tumour mutational burden, thus suggesting that this group may be more responsive to immunotherapy. It would be very interesting to apply emerging data looking at combining immunotherapy and radiotherapy within the framework of molecular subtypes in MIBC.

5.5.3.6. A commercial panel to guide neoadjuvant chemotherapy use

Of note, a commercial panel, DECIPHER bladder (GenomeDx), has been released and marketed as a genomic subtyping tool for use in patients with MIBC considering NAC prior to surgery, with the claim of being predictive of survival benefit. This is based on the GSC classifier which divided patients into luminal, luminal infiltrated, basal and basal claudin-low subtypes, and reported that patients with basal tumours appeared to derive survival benefit with NAC although somewhat paradoxically, the basal subtype itself was not associated with a 'major response' to NAC⁵. GenomeDx state that their subtyping classifier has been developed and validated on 775 patients which is a small number compared to other tumour sites (e.g. breast), and to the best of my knowledge, all work has been performed on retrospective cohorts and has yet to be validated in a prospective clinical trial. Personally, this is not a panel that I would use for my patients as the evidence behind it is not robust enough to guide decisions regarding whether a patient should or should not receive NAC. Furthermore, 6 consensus subtypes have since very recently been published⁴⁸, with data suggesting that the consensus basal squamous and

luminal nonspecified subtypes may benefit from NAC. I will discuss these subtypes further in my final chapter.

5.5.4. Heterogeneity of molecular subtype

In this cohort, 2 patients had more than one sample tested from the same tumour. It is impossible to say whether the differences in subtype allocated to the individual samples from a single patient represent true heterogeneity, technical error or pure chance. However, work published during the course of my project has suggested that subtype heterogeneity is present in MIBC^{49,50}.

I had hoped to have a larger subset of patients with multiple samples tested from the same tumour, but unfortunately, my initial approach of performing macrodissection from individual tumour regions on each slide resulted in insufficient concentrations of RNA for Nanostring analysis (ideally minimum 10ng/ul). My hypothesis had been that areas of different histology may have different subtypes, but given the reported observation that TCGA luminal infiltrated cancers had a lower tumour content compared to other luminal subtypes, I also hypothesised that areas with the same histology but different tumour content would also be distinct in terms of molecular subtype. I was also interested to see whether regions appearing histologically identical in terms of morphology and tumour content might be assigned different subtypes at a molecular level.

One limitation of this part of my work, aside from the small numbers of viable RNA samples available, was the assumption that tumour areas marked on different tumour chips were spatially separate. As previously mentioned in Chapter 4, at TURBT the tumour is fragmented into chips using diathermy and the tissue pieces are put collectively in a pot. These are then set into paraffin blocks without any orientation and so it is impossible to know how one tumour area from one chip is geographically related to any of the other tumour areas marked. Prospective sampling and labelling of distinct areas of the tumour at TURBT for research purposes would be very interesting and useful and would provide a wealth of information for future work. For this exploratory work

however, I think the methods and assumptions made here were entirely reasonable and practical.

I would like to expand this part of my work by increasing my cohort size of those with multiple samples as above. As per my initial plan, I would aim to subtype different regions separately, including those with the same morphology and tumour content. The potential to use a lower RNA input now for Nanostring analysis would make this more readily achievable.

An alternative method to address heterogeneity of molecular subtypes might be to consider a completely different approach. As discussed in chapter 3, liquid biopsies such as ctDNA have the potential to overcome the barriers associated with tissue biopsy with regards to temporal and spatial sampling bias. Could analysis of circulating RNA be of use here? It would be interesting to conduct a small feasibility pilot to further explore this.

5.5.5. Radiosensitivity Index

Unfortunately due to a design error in the Nanostring panel, it was not possible to calculate the RSI as expression data for protein kinase C was not available. Instead of including PRKCB (also labelled PKC) on the panel, PRRT2 (also sometimes labelled PKC) was included. It would be worthwhile therefore adding PRKCB to any future panels so that the RSI can be assessed in MIBC.

Since performing this work, the RSI has been applied to bladder cancer as part of a large retrospective analysis involving over 8000 tumour samples. Bladder cancer samples accounted for 193 of this cohort. In this work, the authors combined the RSI algorithm with the linear quadratic equation and assessed outcomes across tumour types. They reported that the genomic-adjusted radiation dose (GARD) value varied across tumour sites and within tumour sites. They demonstrated that it predicted clinical outcomes in breast, pancreatic, lung cancers and glioblastomas, and suggested that this might be a tool to personalise radiation dose for each patient.

This could potentially be assessed within current MIBC radiotherapy trials e.g. RAIDER where patients are randomised to standard radiotherapy dose or dose-escalated arms. The potential to combine this also with imaging studies that have been ongoing within our department is an exciting prospect.

5.5.6. Differentially expressed genes

Following multiple testing correction, there were no statistically significant differentially expressed genes noted. However, given the exploratory nature of this pilot, I think it is worthwhile exploring the individual results.

It is interesting that although erroneously included in the panel, *PRRT2* was found (before multiple testing correction), to be significantly differentially expressed in patients with and without locoregional relapse. This of course could be a false positive result, and removal of this gene from the subtyping analysis did not change the allocations, suggesting that *PRRT2* does not play an instrumental role. Defects in *PRRT2* have been implicated in neurological conditions but not as a driver mutation in cancer. I suspect therefore that this result is not of relevance in the context of MIBC.

5.5.7. Limitations and future directions

My pilot study on molecular subtypes in MIBC has generated some interesting data and potentially clinically relevant results. The addition of DDR and radiosensitivity genes to key TCGA bladder subtype genes resulted in the identification of 5 subtypes- the NMF71 classifier. Although broad overlap was noted between the majority of TCGA and NMF71 subtypes, NMF71 subtype was associated with differential rates of invasive locoregional recurrence whereas no such association was seen with the TCGA subtypes. Whether this is a true discrepancy or the product of mislabelling samples using a recreated TCGA classifier is unclear but needs to be addressed. Future work should therefore include obtaining the TCGA classifier to re-analyse this radiotherapy dataset and explore any association with radiotherapy response. Saying that,

exploration of the NMF71 metagenes suggested a key role for *ATM* which was not one of the TCGA subtype genes so it would not be unreasonable to suggest that NMF71 offers a novel and potentially clinically useful classifier.

A key limitation of this work was the small numbers involved, which is reflected in the wide confidence intervals seen and limited the potential statistical analysis. However, this was a pilot study designed to identify potential biomarker candidates for further investigation, and perhaps more importantly, to explore the feasibility of extracting RNA from FFPE for interrogation using a custom designed Nanostring panel. With regards to these aims, I feel I have succeeded. The workflow and infrastructure I have established in this project will greatly support future work.

It goes without saying that both the NMF71 classifier and 'official' TCGA classifier should be applied to a much larger cohort of patients treated with radiotherapy +/- chemotherapy, and a cohort of patients managed surgically. We are currently liaising with another UK-based group to obtain expression data for a large cohort treated with radical radiotherapy. However, small cohort size aside, it is encouraging that in my data, the potential association between subtype and radiotherapy response appears to be underpinned by the plausible mechanism that *ATM* over- or under-expression might affect radiosensitivity. My differential gene expression results lend further support to this given that data suggested *ATM* levels may be differentially expressed between patients with or without invasive LRR. The current interest in *ATM* inhibitors only serves to further support my findings, and I strongly believe this classifier is worth further investigation.

One avenue to explore the existing data further would be to re-interrogate the data using different approaches e.g. re-analysis of the cohort to identify key genes distinguishing samples in subtypes 1-3 vs 4-5. It would also be interesting to take a more functional approach of dividing the cohort into those with or without invasive locoregional recurrence and characterising the 'subtype' of these two groups. This would however require expansion of the cohort to include a greater proportion of patients with invasive locoregional recurrence. While the low incidence of invasive locoregional recurrence made

it difficult to perform any more meaningful or powerful statistical analysis, it is on the other hand somewhat pleasing that despite trying to enrich the cohort for invasive locoregional recurrences, relatively few cases were found. Encouragingly, the locoregional recurrence rates in this cohort were the same as those reported in the landmark BC2001 trial.

If it transpires that the TCGA subtypes do indeed show differential responses to radiotherapy, then it would be interesting to assess which of the two subtyping classifiers is the most clinically relevant. I would hypothesise that NMF71 would be more specific with regards to radiotherapy response given the addition of DDR/radiosensitivity genes and my results suggesting that *ATM* expression may underpin the differences in radiotherapy response seen. If NMF71 were to be validated in a larger cohort, it would be interesting to then assess whether it might be of relevance in other cancer subtypes. I am not aware of any evidence to suggest different mechanisms of radiosensitivity between different cancer types, and would therefore anticipate that NMF71 could be useful to all cancer types with respect to radiotherapy response.

An alternative approach to assessing molecular subtype is that of using immunohistochemistry. Tanaka et al⁵¹ explored this using the Lund immunohistochemical-based classification to assign subtype in a cohort of 92 patients receiving induction chemoradiation for bladder cancer. They reported a greater proportion of pathological complete responses in patients with genomically-unstable (GU) and squamous-like subtypes (SCCL) compared to those with urobasal tumours (pCR rates of 52%, 45% and 15% respectively). This is an interesting paper although it is difficult to relate directly to my work given that levels at a protein and mRNA level do not necessarily correlate. One might suggest however that the SCCL subtype would be similar to NMF71 group 4 (basal features), and there did appear to be a trend in my data for group 4 patients to have reduced invasive local recurrence. Immunohistochemical stains however are subject to issues regarding reproducibility, as demonstrated in a recent paper assessing a potential MRE11 assay for clinical use²². In contrast, the Nanostring platform provides more robust and reproducible results.

It would be interesting to also explore the role of tumour infiltrating lymphocytes further. During the course of this work, I did explore with Dr Hazell (Consultant Histopathologist), the possibility of making an assessment of the degree of lymphocytic infiltration in my samples. However, this was not possible within the scope of this project but would certainly be an area of interest for the future. My expectation was that those patients with greater presence of tumour infiltrating lymphocytes would respond better to radiotherapy given the immunogenic response induced by radiation-induced cell death.

With regards to optimising the workflow, one potential improvement might be to use laser capture microdissection rather than the relatively crude macrodissection where it is more likely part of a tumour region may be missed or surrounding tissue included, or even to consider single cell gene expression analysis.

As previously discussed, I think it would be very informative to further explore heterogeneity in subtyping and whether circulating RNA might be a worthwhile avenue pursuing, and the addition of *PRKCB* to the panel to allow RSI and GARD analysis would be a worthy project. In addition, there has been a huge surge of interest in immune-oncology genes and it would be worth investigating this if a new panel were to be designed. The involvement of the immune system and the tumour microenvironment in radiotherapy is of huge interest and I will discuss this a little further in my final discussion chapter.

5.6. Conclusion

There is a real clinical need for biomarkers to guide treatment in MIBC. In this chapter, I have explored molecular subtype as a potential biomarker. Using a custom-designed Nanostring panel, I have demonstrated that expression data can be reliably generated from RNA extracted from archived FFPE blocks, and that the addition of DDR and radiosensitivity genes to key TCGA bladder genes results in a new 5-subtype classifier, which may help identify those patients more likely to achieve a complete pathological response to radiotherapy. Examination of the gene profiles and differential gene expression analysis point towards *ATM* playing a role in radiotherapy response but further work is needed to validate these findings and further explore underlying radiosensitivity mechanisms. The potential clinical implications, if validated, are significant and would help guide treatment strategy.

However, it is worth acknowledging that systemic relapse remains a huge problem in MIBC- ultimately, almost 30% of my cohort developed metastatic disease and these cases were distributed across the subtypes. This is in keeping with the often quoted concept that muscle-invasive bladder cancer is a systemic disease⁵², and while radiotherapy can provide local control, further progress on systemic treatment options is needed if outcomes are to be optimised.

5.7. References

1. Choi, W., *et al.* Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer Cell* **25**, 152-165 (2014).
2. Hoadley, K.A., *et al.* Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158**, 929-944 (2014).
3. Poudel, P., *et al.* Revealing unidentified heterogeneity in different epithelial cancers using heterocellular subtype classification. *BioRxiv* (2017).
4. Eschrich, S.A., *et al.* A gene expression model of intrinsic tumor radiosensitivity: prediction of response and prognosis after chemoradiation. *Int J Radiat Oncol Biol Phys* **75**, 489-496 (2009).
5. Seiler, R., *et al.* Impact of Molecular Subtypes in Muscle-invasive Bladder Cancer on Predicting Response and Survival after Neoadjuvant Chemotherapy. *Eur Urol* (2017).
6. Robertson, A.G., *et al.* Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell* **171**, 540-556 e525 (2017).
7. Sadanandam, A., *et al.* A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med* **19**, 619-625 (2013).
8. Ragulan, C., *et al.* Analytical Validation of Multiplex Biomarker Assay to Stratify Colorectal Cancer into Molecular Subtypes. *Sci Rep* **9**, 7665 (2019).
9. Choudhury, A., *et al.* MRE11 expression is predictive of cause-specific survival following radical radiotherapy for muscle-invasive bladder cancer. *Cancer Res* **70**, 7017-7026 (2010).
10. Efsthathiou, J.A., *et al.* Subtyping Muscle-Invasive Bladder Cancer to Assess Clinical Response to Trimodality Therapy. *Int J Radiat Oncol Biol Phys* **99**, s118 (2017).
11. Do, H. & Dobrovic, A. Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. *Clin Chem* **61**, 64-71 (2015).
12. Damrauer, J.S., *et al.* Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *Proc Natl Acad Sci U S A* **111**, 3110-3115 (2014).
13. Sjodahl, G., *et al.* A molecular taxonomy for urothelial carcinoma. *Clin Cancer Res* **18**, 3377-3386 (2012).
14. Perou, C.M. Molecular stratification of triple-negative breast cancers. *Oncologist* **15 Suppl 5**, 39-48 (2010).
15. Ding, Y.C., *et al.* Molecular subtypes of triple-negative breast cancer in women of different race and ethnicity. *Oncotarget* **10**, 198-208 (2019).
16. Rupnik, A., Grenon, M. & Lowndes, N. The MRN complex. *Curr Biol* **18**, R455-457 (2008).
17. Plimack, E.R., *et al.* Defects in DNA Repair Genes Predict Response to Neoadjuvant Cisplatin-based Chemotherapy in Muscle-invasive Bladder Cancer. *Eur Urol* (2015).

18. Yap, K.L., *et al.* Whole-exome sequencing of muscle-invasive bladder cancer identifies recurrent mutations of UNC5C and prognostic importance of DNA repair gene mutations on survival. *Clin Cancer Res* **20**, 6605-6617 (2014).
19. Desai, N.B., *et al.* Genomic characterization of response to chemoradiation in urothelial bladder cancer. *Cancer* **122**, 3715-3723 (2016).
20. Laurberg, J.R., *et al.* Expression of TIP60 (tat-interactive protein) and MRE11 (meiotic recombination 11 homolog) predict treatment-specific outcome of localised invasive bladder cancer. *BJU Int* **110**, E1228-1236 (2012).
21. Durant, S.T., *et al.* The brain-penetrant clinical ATM inhibitor AZD1390 radiosensitizes and improves survival of preclinical brain tumor models. *Sci Adv* **4**, eaat1719 (2018).
22. Walker, A.K., *et al.* MRE11 as a Predictive Biomarker of Outcome After Radiation Therapy in Bladder Cancer. *Int J Radiat Oncol Biol Phys* **104**, 809-818 (2019).
23. Martin, R.M., *et al.* Post-transcriptional regulation of MRE11 expression in muscle-invasive bladder tumours. *Oncotarget* **5**, 993-1003 (2014).
24. Park, B.J., *et al.* AIMP3 haploinsufficiency disrupts oncogene-induced p53 activation and genomic stability. *Cancer Res* **66**, 6913-6918 (2006).
25. Kim, S.M., *et al.* AIMP3 depletion causes genome instability and loss of stemness in mouse embryonic stem cells. *Cell Death Dis* **9**, 972 (2018).
26. Gurung, P.M., *et al.* Loss of expression of the tumour suppressor gene AIMP3 predicts survival following radiotherapy in muscle-invasive bladder cancer. *Int J Cancer* **136**, 709-720 (2015).
27. Wouters, B. & Begg, A.C. Irradiation-induced damage and the DNA damage response. in *Basic Clinical Radiobiology* (eds. Joiner, M. & van der Kogel, A.) 11-26 (Hodder Arnold, 2009).
28. Berlin, A., *et al.* NBN gain is predictive for adverse outcome following image-guided radiotherapy for localized prostate cancer. *Oncotarget* **5**, 11081-11090 (2014).
29. Van Allen, E.M., *et al.* Somatic ERCC2 mutations correlate with cisplatin sensitivity in muscle-invasive urothelial carcinoma. *Cancer Discov* **4**, 1140-1153 (2014).
30. Mouw, K. Investigating the effect of ERCC2 mutations on DNA repair capacity and chemo-radiotherapy response in muscle-invasive bladder cancer. Vol. 2019.
31. Advanced Bladder Cancer Meta-analysis, C. Neoadjuvant chemotherapy in invasive bladder cancer: update of a systematic review and meta-analysis of individual patient data advanced bladder cancer (ABC) meta-analysis collaboration. *Eur Urol* **48**, 202-205; discussion 205-206 (2005).
32. James, N.D., *et al.* Radiotherapy with or without chemotherapy in muscle-invasive bladder cancer. *N Engl J Med* **366**, 1477-1488 (2012).
33. Yin, M., *et al.* Neoadjuvant Chemotherapy for Muscle-Invasive Bladder Cancer: A Systematic Review and Two-Step Meta-Analysis. *Oncologist* **21**, 708-715 (2016).

34. Alfred Witjes, J., *et al.* Updated 2016 EAU Guidelines on Muscle-invasive and Metastatic Bladder Cancer. *Eur Urol* **71**, 462-475 (2017).
35. NCCN. Bladder Cancer. in *NCCN Guidelines* (2018).
36. Chang, S.S., *et al.* Treatment of Non-Metastatic Muscle-Invasive Bladder Cancer: AUA/ASCO/ASTRO/SUO Guideline. *J Urol* **198**, 552-559 (2017).
37. Branzei, D. & Foiani, M. Regulation of DNA repair throughout the cell cycle. *Nat Rev Mol Cell Biol* **9**, 297-308 (2008).
38. Sharma, P., *et al.* CD8 tumor-infiltrating lymphocytes are predictive of survival in muscle-invasive urothelial carcinoma. *Proc Natl Acad Sci U S A* **104**, 3967-3972 (2007).
39. Demaria, S. & Formenti, S.C. Role of T lymphocytes in tumor response to radiotherapy. *Front Oncol* **2**, 95 (2012).
40. Formenti, S.C. & Demaria, S. Combining radiotherapy and cancer immunotherapy: a paradigm shift. *J Natl Cancer Inst* **105**, 256-265 (2013).
41. Efstathiou, J.A., *et al.* Impact of Immune and Stromal Infiltration on Outcomes Following Bladder-Sparing Trimodality Therapy for Muscle-Invasive Bladder Cancer. *Eur Urol* **76**, 59-68 (2019).
42. Excellence, N.I.f.H.a.C. Adjuvant therapy for early and locally-advanced breast cancer. (2019).
43. Rosenberg, J.E., *et al.* Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet* **387**, 1909-1920 (2016).
44. Necchi, A., *et al.* Pembrolizumab as Neoadjuvant Therapy Before Radical Cystectomy in Patients With Muscle-Invasive Urothelial Bladder Carcinoma (PURE-01): An Open-Label, Single-Arm, Phase II Study. *J Clin Oncol*, JCO1801148 (2018).
45. Powles, T., *et al.* Atezolizumab versus chemotherapy in patients with platinum-treated locally advanced or metastatic urothelial carcinoma (IMvigor211): a multicentre, open-label, phase 3 randomised controlled trial. *Lancet* **391**, 748-757 (2018).
46. Altevogt, P., Doberstein, K. & Fogel, M. L1CAM in human cancer. *Int J Cancer* **138**, 1565-1576 (2016).
47. Rebouissou, S., *et al.* EGFR as a potential therapeutic target for a subset of muscle-invasive bladder cancers presenting a basal-like phenotype. *Sci Transl Med* **6**, 244ra291 (2014).
48. Kamoun, A., *et al.* A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. *Eur Urol* **77**, 420-433 (2020).
49. Thomsen, M.B.H., *et al.* Comprehensive multiregional analysis of molecular heterogeneity in bladder cancer. *Sci Rep* **7**, 11702 (2017).
50. Warrick, J.I., *et al.* Intratumoral Heterogeneity of Bladder Cancer by Molecular Subtypes and Histologic Variants. *Eur Urol* (2018).
51. Tanaka, H., *et al.* Impact of Immunohistochemistry-Based Subtypes in Muscle-Invasive Bladder Cancer on Response to Chemoradiation Therapy. *Int J Radiat Oncol Biol Phys* (2018).
52. Prout, G., Griffin, P. & Shipley, W. Bladder Carcinoma as a Systemic Disease. *Cancer* **43**, 2532-2539 (1979).

5.8. Appendix

A5.1 CRCAssigner-38 R-Script

```
rm(list = ls())

source("/Volumes/NO NAME/RNA/CRCAssigner Nov18 Correct
38centroids/correl_subtypes_Nano.R")
direc <- "/Volumes/NO NAME/RNA/CRCAssigner Nov18 Correct
38centroids"
file <- "20180914_BL_46samples_norm_10HK_log2_ready2.txt"
pam_centroids <- "/Volumes/NO NAME/RNA/CRCAssigner Nov18
Correct 38centroids/2018-08-09-38-genes-PAM-centroid_1%25.txt"
und_cutoff = 0.15
mix_cutoff = 0.06
correl_subtypes_Nano(direc, file, pam_centroids, und_cutoff,
mix_cutoff)
```

A5.2 Histological assessment of slides

CoMB	GTB	No of regions marked	Histology	Tumour content (%)	Notes
1001	15373	1	TCC	70	
1003	15390	2	TCC	60, 60	
1006	15385	1	Micropapillary TCC and solid TCC	80	
1007	15398	1	TCC with squamous differentiation	70	
1008	15430	1	TCC. Desmoplasia	50	
1010	15428	2	TCC	70, 100	
1013	15460	1	TCC	50	not good quality
1014	16017	2	micropapillary/plasmacytoid TCC	60, 50	
1016	15442	4	TCC	70	
1020	15432	1	TCC	80	
1024	15436	2	TCC	40, 70	
1025	15421	1	TCC/squamous	50	
1033	16036	1	TCC	70	
1035	16215	1	TCC	70	
1037	16279	1	TCC	80	
1039	16222	2	TCC	60, 95	
1041	16245	1	TCC	70	
1042	16236	1	TCC	60	
1049	16283	1	TCC	90	
2003	15375	1	TCC	70	
2005	15351	1	TCC	90	
2006	16039	1	TCC	90	
2007	16008	3	TCC	90, 80, 60	
2008	16027	3	TCC. Desmoplasia	40, 30, 40	
2009	16389	2	TCC; neuroendocrine differentiation	90, 70	
2011	16030	3	glandular + solid TCC	80, 80, 90	
2015	16398	3	squamous cell ca, TCC, TCC	80, 50, 80	
2017	16941	1	small cell	100	mangled
2022	16415	1	TCC	80	
2026	16394	1	TCC	80	
2027	16296	1	TCC	95	
2028	16946	1	surface ca only, no muscle invasion	NA	
2029	16383	1	TCC	90	
2031	16489	2	TCC	70	
2032	16943	1	almost small cell	90	
2035	16411	2	squamous cell ca, cartilage (carcinosarcoma)- no TCC	70, 100	
2036	15352	1	TCC	30	
2037	15384	1	TCC	30	
2039	16023	1	TCC	80	
2040	16954	1	TCC	70	necrosis ++
2042	16025	2	TCC	60, 60	
2043	NA	3	1. Sarcomatoid 2. lymphoplasmocytic TCC 3. lots of TILs	100, 30, 60	
2044	16385	3	TCC. Desmoplasia	80, 80, 50	
2046	16944	2	1. small cell. 2 TCC	100, 70	
2047	16246	3	TCC	40, 30, 60	
2055	16455	1	TCC desmoplastic	60	
2056	16475	1	sarcomatoid TCC	50	
3003	15402	3	TCC	80, 80, 70	
3006	15415	1	TCC	90	
3012	16414	2	TCC	90, 70	
3015	16212	3	TCC	100, 90, 90	
3021	NA	2	TCC	70, 80	
2057	16940	1	small cell	100	

A5.3 Deparaffinisation Protocol

- Put unstained slides on trays and warm in oven at 60C for 5 minutes
- Remove slides and transfer to coplin jars
- In the fume cupboard, add xylene to each coplin jar and leave for 5 minutes
- Discard xylene in to waste bottle.
- Perform a wash with 100% ethanol. Discard ethanol into waste bottle
- Repeat ethanol wash
- 100% ethanol for 5 minutes
- 2 washes in 96% ethanol
- 2 washes in water
- Keep in water while awaiting macrodissection

A5.4 Dual Extraction Protocol

- Add 100ul diluted digestion buffer and 4ul protease to each sample
- Ensure sample is immersed
- Incubate at 50C for 16 hours
- Centrifuge briefly to collect any condensation
- Incubate at 80C for 14 min 30s, then centrifuge briefly to collect any condensation
- Place a filter cartridge in a collection tube for each sample
- Add 120ul Isolation Additive to each sample and mix by pipetting up and down
- Add the sample to the filter cartridge and close the lid
- Centrifuge at 10 000xg for 1 min
- Place the filter cartridge in a new collection tube and store at 4C for DNA isolation later
- Place a new filter cartridge in a collection tube for each sample
- Add 1.25 volume (275ul) of 100% ethanol to the flow through from the previous step
- Mix by pipetting up and down. Transfer the sample to the filter cartridge
- Close the lid and centrifuge at 10 000xg for 1 min

- Discard the flow through and re-insert the filter cartridge into the same collection tube
- Add 700ul Wash buffer 1 to the filter cartridge
- Centrifuge at 10 000 xg for 1 min
- Discard the flow through and re-insert the filter cartridge into the same collection tube
- Add 500ul Wash Buffer 2/3 to the filter cartridge and centrifuge at 10 000 xg for 1 min
- Discard the flow through and re-insert the filter cartridge into the same collection tube
- Centrifuge at 10 000 xg to remove residual fluid
- Prepare a DNase Mix using the following quantities per sample- 4ul DNase, 6ul of 10X DNase Buffer, 50ul of nuclease free water
- Add 60ul of the DNase mix to each sample.
- Incubate at room temperature for 30 mins
- Add 650ul Wash Buffer 1 to the filter cartridge
- Incubate for 30-60s then centrifuge at 10 000 xg for 1 min
- Discard the flow through and re-insert the filter cartridge into the same collection tube
- Add 500ul of Wash Buffer 2/3 and centrifuge at 10 000 xg for 1 min
- Discard the flow through and re-insert the filter cartridge into the same collection tube
- Perform a second wash with Wash Buffer 2/3
- Centrifuge at 10 000 xg for 1 min to remove residual fluid
- Transfer filter cartridge to an elution tube
- Add 20ul pre-heated (to 50C) nuclease-free water to each cartridge
- Incubate at room temperature for 5 minutes
- Centrifuge at 12 000 xg for 2 minutes
- Transfer eluate to filter again (double elution)
- Incubate at room temperature for 5 minutes
- Centrifuge at 12 000 xg for 2 minutes to elute
- Keep on ice while extracting DNA
- Remove filter cartridges from fridge for DNA extraction

- Add 700ul Wash Buffer 1 to filter cartridge
- Centrifuge at 10 000 xg for 1 minute
- Discard the flow through and re-insert the filter cartridge into the same collection tube
- Add 500ul Wash Buffer 2/3 to the cartridge and centrifuge at 10 000 xg for 1 minute
- Discard the flow through and re-insert the filter cartridge into the same collection tube
- Centrifuge at 10 000 xg for 1 minute to remove residual fluid
- Prepare the RNase mix using the following quantities per sample- 10ul RNase, 50ul nuclease free water
- Add 60ul of RNase mix to each filter cartridge and incubate at room temperature for 30 minutes
- Add 700ul Wash Buffer 1 to the filter cartridge
- Centrifuge at 10 000 xg for 1 minute
- Discard the flow through and re-insert the filter cartridge into the same collection tube
- Add 500ul Wash Buffer 2/3 to the filter cartridge and centrifuge at 10 000 xg for 1 minute
- Discard the flow through and re-insert the filter cartridge into the same collection tube
- Perform a 2nd wash using Wash Buffer 2/3
- Centrifuge at 10 000 xg for 1 minute to remove residual fluid and transfer the filter cartridge to an elution tube
- Add 20ul pre-heated (at 50C) nuclease-free water to the filter cartridge
- Incubate at room temperature for 5 minutes
- Centrifuge at 12 000 xg for 2 minutes
- Perform a double elution as before
- Keep on ice
- Quantify all samples at Nanodrop. Store RNA and DNA samples at -80C

A5.5 RNA yields

Total of 70 dual extractions on 51 patients

Sample ID	Conc. (ng/ul)	A260	A280	260/280	260/230
1007A1	307.4	7.685	3.806	2.02	1.96
1020A1	23.3	0.583	0.299	1.95	0.32
1037A1	12.8	0.319	0.16	1.99	0.41
2022A1	19.1	0.477	0.24	1.99	0.4
2039A1	43.8	1.094	0.517	2.12	1.26
1035A1	12.9	0.322	0.156	2.07	0.92
1042A1	74.5	1.863	0.908	2.05	0.91
1024A1	2.3	0.058	0.013	4.39	0.41
1024A2	0.5	0.012	-0.002	-5.06	0
1001A1	49.2	1.231	0.588	2.09	1.63
1041A1	3.1	0.077	0.039	1.97	0.02
1008A1	163.3	4.083	1.983	2.06	1.55
1049A1	476.7	11.917	5.74	2.08	1.6
2003A1	29.4	0.735	0.362	2.03	1.49
2005A1	431.8	10.796	5.291	2.04	1.93
2005A2	4.5	0.114	0.055	2.08	0.16

Sample ID	Conc. (ng/ul)	A260	A280	260/280	260/230
2005A3	8.5 (rpt 11.2)	0.212	0.11	1.92	0.36
2006A1	969.5	24.238	11.687	2.07	2.05
2006A2	2 (rpt 12.1)	0.051	0.021	2.43	0.2
2006A3	7.8	0.195	0.101	1.94	0.85
2007A1	571.6	14.289	7.018	2.04	1.9
2007A2 rep	126.2	3.155	1.527	2.07	-1.77
2007A3	26.5	0.661	0.319	2.07	0.21
2042A1	24.7	0.618	0.296	2.09	0.23
2008A1	193.4	4.834	2.411	2	0.37
2008A2	4.9	0.122	0.051	2.37	0.16
2015A1	4.6	0.114	0.051	2.25	0.39
2015A2	2.4	0.06	0.023	2.55	0.12
2036A1	10.4	0.26	0.135	1.92	0.68
2037A1 rep	81.7	2.042	0.991	2.06	1.48
2043A1	5.2	0.131	0.058	2.25	0.98
2043A2	8.2	0.206	0.089	2.31	0.4
2043A3	7.8	0.195	0.091	2.15	0.15
2044A1	46.7	1.168	0.577	2.03	1.56

Sample ID	Conc. (ng/ul)	A260	A280	260/280	260/230
2044A2	5.5	0.137	0.066	2.06	0.44
3006A1	10.1	0.252	0.121	2.09	0.32
1019A1	64.9	1.624	0.778	2.09	1.74
1019A2	0.9	0.022	0.001	32.56	0.07
2009A1	16.8	0.421	0.209	2.02	0.64
2009A2	2.5	0.062	0.031	1.99	0.05
2009A3	11.7	0.293	0.14	2.1	0.69
2047A1	3.6	0.091	0.041	2.18	0.01
2047A2	11.1	0.277	0.139	1.99	0.53
2047A3	4.6	0.116	0.048	2.39	0.02
3012A1	14.6	0.365	0.183	1.99	0.34
3012A2	0.9	0.023	-0.002	-9.27	0.03
3015A1	9.4	0.234	0.113	2.07	0.47
3015A2	109.5	2.737	1.329	2.06	0.51
3021A1	7.6	0.191	0.101	1.89	0.03
3021A2	8.1	0.204	0.096	2.12	0.24
3012AR	479.6	11.989	5.887	2.04	1.46
1019AR	927.6	23.191	11.093	2.09	2.06

Sample ID	Conc. (ng/ul)	A260	A280	260/280	260/230
2009AR	936.6	23.416	11.484	2.04	2.14
3015AR	1475.4	36.885	17.658	2.09	1.63
2047AR	461.5	11.538	5.668	2.04	2.04
3021AR	276.4	6.909	3.389	2.04	1.79
2055A	170.8	4.27	2.151	1.99	1.83
2032A	337.1	8.427	4.139	2.04	2.11
1013A	12.7	0.317	0.165	1.92	0.26
2017A	149.1	3.727	1.927	1.93	1.47
1003A	46.1	1.151	1.183	0.97	0.65
1025A	115.8	2.895	1.449	2	1.56
3021A	36.1	0.904	0.475	1.9	1.09
1014A	23.3	0.582	0.285	2.04	1.56
2011A	25.9	0.647	0.317	2.04	0.93
2056A	5.1	0.126	0.069	1.84	0.49
2046A2	6.3	0.159	0.078	2.04	0.79
1010A	7.8	0.194	0.106	1.83	0.35
1039A	7.1	0.176	0.094	1.89	0.37
2046A1	1.6	0.04	0.014	2.84	0.04

Sample ID	Conc. (ng/ul)	A260	A280	260/280	260/230
1033A	13.6	0.339	0.166	2.04	0.27
2026A	216.7	5.418	2.617	2.07	2.01
2029A	59.3	1.482	0.714	2.08	0.22
3003A	254.5	6.363	3.118	2.04	2.09
2043A	98.3	2.458	1.226	2	1.36
2031A	20.6	0.514	0.254	2.03	0.87
2057A	-2.9	-0.072	0.007	-9.79	-0.23
2057A rep	83.1	2.076	1.018	2.04	1.69
1006A	43.2	1.08	0.531	2.03	1.25
2040A	13.8	0.344	0.166	2.07	1.78
2027A	53.5	1.338	0.655	2.04	1.79

A5.6 DNA yields

Sample	Conc (ng/ul)	A260	A280	260/280	260/230
1007A1	108.3	2.165	1.199	1.81	0.55
1020A1	161.7	3.235	1.767	1.83	1.19
1037A1	171.8	3.437	1.827	1.88	1.1
2022A1	62.4	1.248	0.662	1.89	0.99
2039A1	15.3	0.307	0.148	2.07	0.29
1035A1	7.6	0.152	0.064	2.37	0.17
1042A1	134.2	2.684	1.421	1.89	2.12
1024A1	11.6	0.233	0.112	2.07	0.14
1024A2	4.1	0.081	0.132	0.62	-0.3
1001A1	57.8	1.156	0.717	1.61	0.47
1041A1	12.7	0.253	0.102	2.49	0.02
1008A1	40.8	0.817	0.512	1.6	0.55
1049A1	95.4	1.909	1.01	1.89	0.17
2003A1	8.6	0.173	0.097	1.78	0.14
2005A1	51.1	1.023	0.62	1.65	0.25
2005A2	7	0.14	0.162	0.86	-0.14
2005A3	38.4	0.768	0.39	1.97	0.08
2006A1	571.6	11.433	6.252	1.83	0.64
2006A2	52.9	1.059	0.563	1.88	0.17

Sample	Conc (ng/ul)	A260	A280	260/280	260/230
2006A3	58.6	1.172	0.612	1.91	0.18
2007A1	175.8	3.516	1.937	1.81	0.2
2007A2	61.1	1.222	0.695	1.76	0.13
2007A3	15.9	0.318	0.192	1.66	0.07
2042A1	22	0.439	0.263	1.67	0.07
2008A1	17.5	0.351	0.2	1.75	0.65
2008A2	7.4	0.147	0.072	2.06	0.28
2015A1	42.7	0.855	0.438	1.95	0.3
2015A2	17.2	0.344	0.192	1.79	0.57
2036A1	12.4	0.248	0.208	1.19	0.25
2037A1	21.9	0.439	0.244	1.8	0.87
2043A1	442.1	8.841	4.741	1.86	1.81
2043A2	315.8	6.316	3.707	1.7	1.83
2032A3	122.8	2.457	1.311	1.87	0.26
2043A3	121.9	2.439	1.288	1.89	0.24
2044A1	231.4	4.628	2.569	1.8	1.54
2044A2	-5.5	-0.11	-0.124	0.89	0.97
2044A2 rep	16.5	0.331	0.177	1.86	0.9
3006A1	30.5	0.611	0.292	2.09	0.07

Sample	Conc (ng/ul)	A260	A280	260/280	260/230
1019A1	-10.3	-0.206	-0.106	1.93	0.75
1019A1 rep1	-2.5	-0.049	-0.004	13.28	7.62
1019A2	18.8	0.377	0.191	1.97	0.07
1019A1 rep2	32.3	0.645	0.34	1.9	1.01
2009A1	71.4	1.429	1.57	0.91	0.84
2009A2	24.2	0.483	0.238	2.03	0.05
2009A3	42.9	0.858	0.451	1.9	0.62
2047A1	4.8	0.097	0.053	1.83	0.07
2047A2	9.5	0.191	0.114	1.68	0.36
2047A3	13.9	0.277	0.167	1.66	0.5
3012A1	28.5	0.569	0.303	1.88	0.16
3012A2	2.3	0.045	0.023	1.97	0.09
3015A1	78.6	1.572	0.729	2.15	0.29
3015A2	163.6	3.272	1.847	1.77	0.88
3021A1	10.4	0.208	0.114	1.82	0.13
3021A2	19.6	0.392	0.219	1.79	0.25
3012AR	132.6	2.653	1.663	1.59	0.64
1019AR	204.4	4.088	2.459	1.66	0.9

Sample	Conc (ng/ul)	A260	A280	260/280	260/230
2009AR	275.4	5.508	3.267	1.69	0.27
3015AR	150.5	3.01	1.823	1.65	0.66
2047AR	36.8	0.737	0.468	1.58	0.37
3021AR	125.8	2.516	1.582	1.59	0.61
2055A	10.4	0.208	0.291	0.71	6.53
2032A	125.6	2.512	1.371	1.83	0.94
1013A	50.7	1.013	0.565	1.79	0.9
2017A	1.9	0.038	0.161	0.23	0.26
1003A	34.8	0.697	0.4	1.74	0.67
1025A	41.5	0.829	0.528	1.57	0.56
3021A	30.7	0.613	0.34	1.81	0.93
1014A	56.7	1.134	0.614	1.85	1.09
2011A	28.2	0.564	0.311	1.81	0.77
2056A	8.4	0.169	0.091	1.85	0.35
2046A2	13.5	0.27	0.153	1.76	0.27
1010A	43.4	0.868	0.566	1.53	0.55
1039A	32.2	0.643	0.364	1.77	0.82
2046A1	13.8	0.277	0.168	1.65	0.29

A5.7 Nanostring Panel Targets

Gene	Accession
ACSL6	NM_001009185.1
AMMECR1L	NM_031445.2
AQP8	NM_001169.2
AREG	NM_001657.2
AXIN2	NM_004655.3
BHLHE41	NM_030762.2
BIRC3	NM_182962.1
CA1	NM_001738.2
CA4	NM_000717.3
CEL	NM_001807.3
CFTR	NM_000492.3
CLCA4	NM_012128.3
CLDN8	NM_199328.2
COL10A1	NM_000493.3
CXCL13	NM_006419.2
CXCL9	NM_002416.1
CYP1B1	NM_000104.3
DHX16	NM_001164239.1
DNAJC14	NM_032364.5

Gene	Accession
EREG	NM_001432.2
FCF1	NM_015962.4
FLNA	NM_001456.3
GZMA	NM_006144.2
IDO1	NM_002164.3
IFIT3	NM_001031683.2
KRT20	NM_019010.1
KRT23	NM_015515.3
LINC00261	NR_001558.3
LY6G6D	NM_021246.2
MET	NM_001127500.1
MGP	NM_000900.2
MS4A12	NM_001164470.1
MSRB3	NM_198080.2
MUC2	NM_002457.2
PCSK1	NM_001177876.1
PLEKHB1	NM_021200.2
PPIA	NM_021130.3
PRPF38A	NM_032864.3

Gene	Accession
QPRT	NM_014298.3
RARRES3	NM_004585.3
REG4	NM_032044.3
RPL13A	NM_012423.2
SFRP2	NM_003013.2
SFRP4	NM_003014.2
SLC4A4	NM_001098484.2
SNAI2	NM_003068.3
SPINK4	NM_014471.1
STAT1	NM_139266.1
TAGLN	NM_003186.3
TCN1	NM_001062.3
TFF1	NM_003225.2
TFF3	NM_003226.3
TMUB2	NM_024107.2
TOX	NM_014729.2
TWIST1	NM_000474.3
ZEB1	NM_001128128.1
ZEB2	NM_014795.3
ZG16	NM_152338.2

Gene	Accession
ZNF143	NM_003442.5
ZNF384	NM_001039920.2
AR	NM_000044.2
AIMP3	NM_004280.2
APLP1	NM_005166.3
ARID1A	NM_006015.4
ATM	NM_138292.3
ATR	NM_001184.2
BRCA1	NM_007305.2
BRCA2	NM_000059.3
BRIP1	NM_032043.1
C7	NM_000587.2
cABL	NM_005157.3
CD274	NM_014143.3
CD44	NM_001001392.1
CDH2	NM_001792.3
CDK1	NM_001786.4
cGAS	NM_138441.2
cJun	NM_002228.3
CLDN3	NM_001306.3

Gene	Accession
CLDN4	NM_001305.3
CLDN7	NM_001307.3
COL17A1	NM_000494.3
COMP	NM_000095.2
CXCL11	NM_005409.3
DES	NM_001927.3
DSC3	NM_001941.3
E2F3	NM_001243076.2
ERCC1	NM_001983.3
ERCC2	NM_000400.2
ERCC4	NM_005236.2
ERCC5	NM_000123.2
ERCC6	NM_001277058.1
FANCB	NM_001018113.2
FANCD2	NM_033084.3
FANCF	NM_022725.2
FANCG	NM_004629.1
FGFR3	NM_022965.2
FOXA1	NM_004496.2
GATA3	NM_001002295.1

Gene	Accession
GNG4	NM_004485.2
GSDMC	NM_031415.2
HDAC1	NM_004964.2
HIF1alpha	NM_001530.2
IRF1	NM_002198.1
KAT5	NM_182710.1
KDM6A/UTX	NM_021140.2
KRT14	NM_000526.4
KRT5	NM_000424.2
KRT6A	NM_005554.3
KTM2D/MLL2	NM_003482.3
L1CAM	NM_024003.2
MRE11A	XM_011542837.2
MSI1	NM_002442.3
NBN	NM_002485.4
NFEL2L2	NM_006164.3
PALB2	NM_024675.3
PARP1	NM_001618.3
PDCD1LG2	NM_025239.3
PEG10	NM_001040152.1

Gene	Accession
PGM5	NM_021965.3
PI3	NM_002638.3
PKC	XM_011545716.1
PLEKHG4B	NM_052909.3
PPARG	NM_005037.5
RAD50	NM_005732.2
RAD54L	NM_003579.2
RB1	NM_000321.1
RelA	NM_021975.2
RND2	NM_005440.4
SAA1	NM_199161.1
BCLAF1	NM_001077440.1
SGCD	NM_000337.5
SLX4	NM_032444.2
SNAI1	NM_005985.2
SNX31	NM_152628.3
SOX2	NM_003106.2
STING	NM_198282.1
SUMO1	NM_003352.4
TGM1	NM_000359.2

Gene	Accession
TP63	NM_003722.4
Trex1	NM_016381.3
TUBB2B	NM_178012.3
TXNIP	NM_006472.3
UPK1A	NM_007000.3
UPK2	NM_006760.3

A5.8 Nanostring Panel Construction

Adapted from Nanostring's nCounter Elements XR User Manual

- Separate master stocks were created for Probe A and Probe B
- Pipette 5ul of each probe A (or probe B) into a microfuge tube
- Add TE to ensure final volume of 1ml
- Stored in aliquots at -80
- Separate working pools for 48 reactions were created from the master stocks
- 7ul from Probe A or probe B master stock combined with 51ul TE-Tween.

A5.9 Nanostring nCounter protocol- hybridisation

Adapted from Nanostring nCounter Elements XT User Manual

For 12 reactions

- Thaw a Tagset aliquot at room temperature and invert to mix
- Create a master mix by combining 70ul of hybridisation buffer and 7ul of the probe A working pool to the Tagset. Invert to mix and spin down
- Add 7ul probe B working pool to the master mix
- Add 8ul of master mix to each of the 12 microfuge tubes
- Add the RNA sample to each tube
- Add RNA-ase free water to each tube to bring the total volume to 15ul
- Cap tubes and invert to mix. Briefly spin to collect sample
- Incubate at 67C for 16 hours. Ramp down to 4C when complete
- Transfer samples to nCounter Prep Station

A5.10 Nanostring runs

RUN 1

Nanostring panel	Posn	- 80 posn	Sample	Final RNA INPUT	Water
1	1	A1	1007A1	3.77	3.23
2	2	A2	1020A1	4.29	2.71
3	3	A3	1037A1	7.81	0.00
4	4	A4	2022A1	5.24	1.76
5	5	A5	2039A1	2.28	4.72
6	6	A6	1035A1	7.75	0.00
7	7	A7	1042A1	1.34	5.66
8	8	B1	1001A1	2.03	4.97
9	9	B3	1008 A1	4.05	2.95
10	10	B4	1049A1	3.66	3.34
11	11	B5	2003A1	3.40	3.60
12	12	C6	2042 A1	4.05	2.95

RUN 2

Nanostring panel	Posn	- 80 posn	Sample	Final RNA INPUT	Water
13	1	C7	2008A1	4.98	2.02
14	2	G7	1003A	2.17	4.83
15	3	D3	2037A1	1.22	5.78
16	4	D7	2044A1	2.14	4.86
17	5	I6	1006 A	2.31	4.69
18	6	E1	1019A1	1.54	5.46
19	7	E7	2047A2	9.01	0.00
20	8	E9	3012A1	6.85	0.15
21	9	F3	3015A2	0.91	6.09
22	10	G3	2055A	4.76	2.24
23	11	G4	2032A	4.27	2.73
24	12	G5	1013A	7.87	0.00

RUN 3

Nanostring panel	Posn	- 80 posn	Sample	Final RNA INPUT	Water
25	1	G6	2017A	4.93	2.07
26	2	D2	2036A1	9.62	0.00
27	3	G8	1025A	4.39	2.61
28	4	G9	3021A	2.77	4.23
29	5	H1	1014A	4.29	2.71
30	6	H2	2011A	3.86	3.14
31	7	H8	1033 A	7.35	0.00
32	8	H9	2026 A	3.69	3.31
33	9	I1	2029 A	1.69	5.31
34	10	I2	3003 A	4.13	2.87
35	11	I3	2043 A	1.02	5.98
36	12	I4	2031 A	4.85	2.15

RUN 4

Nanostring panel	Posn	- 80 posn	Sample	Final RNA INPUT	Water
37	1	I5	2057 A	1.20	5.80
38	2	D9	3006A1	9.90	0.00
39	3	I7	2040 A	7.25	0.00
40	4	I8	2027 A	1.87	5.13
41	5	C3	2007A1	3.58	3.42
42	6	E3	2009 A1	5.95	1.05
43	7	B6	2005 A1	3.51	3.49
44	8	B9	2006 A1	4.46	2.54
45	9	C4	2007A2	4.02	2.98
46	10	C5	2007A3	3.77	3.23
47	11	E5	2009A3	8.55	0.00
48	12	B8	2005A3	8.93	0.00

A5.11 Sample size calculation

As provided by Gift Nyamundanda

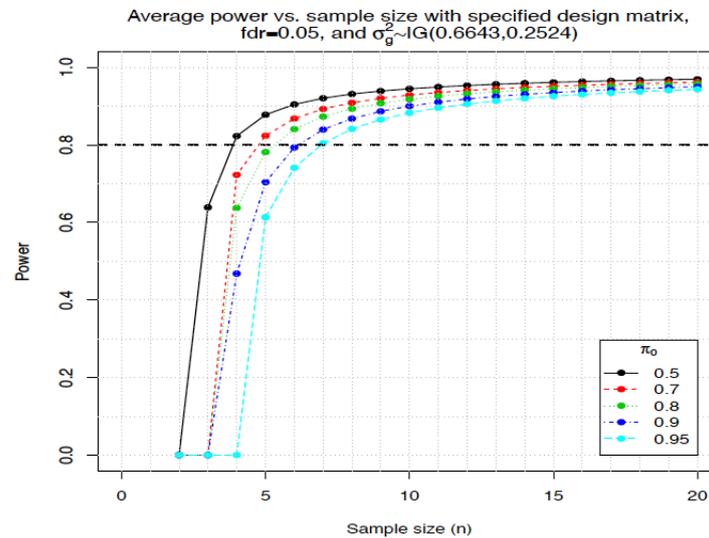


Figure 2: Power calculation for different proportions of non significant genes π_0 and gene expression variances following an Inverse Gamma(0.66,0.25) with FDR controlled set at 5%. Setting statistical power at 0.8, the minimum sample size required for each subtype is 7 since the proportions of non significant genes π_0 was estimated to be around 0.96. NOTE: n is the number of samples per group. Hence, the minimum number of samples required to achieve 80% statistical power to find five subtypes is 7×5 .

- X is linear contrasts of interest,
- β is the true parameter and can be estimated from analysis of pilot data using `limma`,
- a and b are shape and scale parameters of the inverse gamma distribution followed by variances of genes and can be obtained from `limma`,
- $fdr = \alpha$ is the level of false discovery rate to control usually set at 5%,
- $power$ is the desired power to be attained,
- π_0 proportion of non-differentially expressed genes and can be estimated using `limma`.

References

- Nyamundanda, G. *et al.* Probabilistic principal component analysis for metabolomic data. *BMC Bioinformatics* 11:571 (2010).
- Orr, M. *et al.* Sample size estimation while controlling false discovery rate for microarray experiments using `ssize.fdr` package. *The R journal* 1 (2009).

A5.12 Full table of RT cohort characteristics

	N	%
Age	Median = 71.9 years	Range: 46.1 – 90.9
Male	34	79.1
Female	9	20.9
Disease stage		
T2-4 N0 M0	29	67.4%
Any T N1-3 M0	11	25.6%
Any T Any N M1	3 (para-aortic nodes only)	7.0%
Concurrent NMIBC		
pTa	1	2.3%
pTis	13	30.2%
pT1	1	2.3%
Histology		
TCC	37	86.0%
Small cell	3	7.0%
TCC Variants		
Micropapillary	1	2.3%
Micropapillary/plasmacytoid	1	2.3%
Lymphoepithelial	1	2.3%

	N	%
Neoadjuvant chemo		
Yes	32	74.4%
Gem/Cis	26	81.3%
GCarbo	5	15.6%
Carbo/etop	1	3.1%
No	11	25.6%
No of NAC cycles		
1	2	6.3%
3	12	37.5%
4	9	28.1%
6	9	28.1%
Cystoscopy post NAC	21/32	65.6%
Biopsy taken?	15/32	46.9%
Concurrent chemo	42	97.7%
5FU/MMC	39	92.9%
Cape/MMC	1	2.4%
Gemcitabine	2	4.8%
RT Dose*		
64Gy in 32#	21	48.8%
68Gy in 34#	2	4.7%

	N	%
70Gy in 32#	19	44.2%
72Gy in 32#	1	2.3%
Cystoscopy 3-4 mth post RT	39	90.7%
Biopsy taken?	26/39	66.7%

A5.13 Undetermined CRCAssigner-38 cases in RT cohort

Subtype		n	Outcomes
Primary	Secondary		
TA	Goblet_like	2	NMIBC then M1 x1 NO relapse x 1
TA	Enterocyte	1	No relapse
Inflammatory	TA	1	CIS

A5.14 Mixed CRCAssigner-38 cases in RT cohort

Subtype		N	Outcomes
Primary	Secondary		
Stem-like	Enterocyte	1	No relapse
Stem-like	TA	1	NMIBC relapse
Stem-like	Inflammatory	2	M1 relapse in both (prior CIS relapse in one patient)
Goblet-like	Enterocyte	1	No relapse
TA	Enterocyte	1	Local and distant recurrence

A5.15 Table showing allocations and clinical outcomes according to primary CRC-Assigner38 subtype allocated

Subtype	N	Any Relapse	LRR	Invasive LRR	M1 relapses
Enterocyte	9/43 (21%)	5/9 (56%)	5/9 (56%)	1/9 (11%)	1 (11%)
Goblet-like	6/43 (14%)	3/6 (50%)	1/6 (17%)	1/6 (17%)	2/6 (33%)
Inflammatory	8/43 (19%)	3/8 (38%)	3/8 (38%)	2/8 (25%)	1/8 (13%)
Stem-like	15/43 (35%)	9/15 (60%)	6/15 (40%)	4/15 (27%)	6/15 (40%)
Transit-amplifying	5/43 (12%)	2/5 (40%)	2/5 (40%)	1/5 (20%)	2/5 (40%)
Total	43	22	17	9	12
fishers		0.8851	0.7276	0.9547	0.4539

A5.16 CRCAssigner-38 subtype and clinical features

T-stage

	T2	T3b	T4	Total
Enterocyte	3	6	0	9
Goblet	3	3	0	6
Inflamm	4	4	0	8
Stem	8	6	1	15
TA	3	2	0	5
Total	21	21	1	43

N-stage

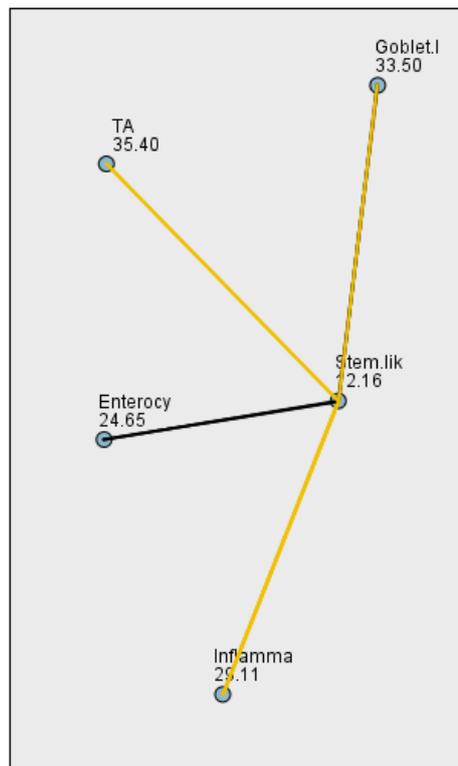
	N0	N1	N2	N3	Total
Entero	6	2	1	0	9
Goblet	5	1	0	0	6
Inflamm	6	2	0	0	8
Stem	11	2	0	2	15
TA	4	1	0	0	5
Total	32	8	1	2	43

M-stage

	M0	M1	Total
Entero	9	0	9
Goblet	5	1	6
Inflamm	8	0	8
Stem	14	1	15
TA	4	1	5
Total	40	3	43

A5.17 Post-hoc analysis of Tumour Content in CRC-Assigner38 groups

Pairwise Comparisons of CRC_primary



Each node shows the sample average rank of CRC_primary.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
Stem.liik-Enterocy	12.494	5.358	2.332	.020	.197
Stem.liik-Inflamma	16.955	5.539	3.061	.002	.022
Stem.liik-Goblet.I	21.344	6.363	3.354	.001	.008
Stem.liik-TA	-23.244	6.810	-3.413	.001	.006
Enterocy-Inflamma	-4.461	6.107	-.730	.465	1.000
Enterocy-Goblet.I	-8.850	6.864	-1.289	.197	1.000
Enterocy-TA	-10.750	7.281	-1.477	.140	1.000
Inflamma-Goblet.I	4.389	7.006	.626	.531	1.000
Inflamma-TA	-6.289	7.414	-.848	.396	1.000
Goblet.I-TA	-1.900	8.049	-.236	.813	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is .05. Significance values have been adjusted by the Bonferroni correction for multiple tests.

A5.18 Mixed and undetermined cases- TCGA subtypes and RT cohort

Subtype (mixed)		N	Outcomes
Primary	Secondary		
Basal-squamous	Neuronal	1	No relapse
Luminal	Luminal_infiltrated	3	No relapse x 2 M1 relapse x 1
Luminal_infiltrated	Luminal	1	Local (node) + M1
Luminal	Luminal_papillary	1	No relapse

Subtype (undetermined)		n	Outcomes
Primary	Secondary		
Basal Squamous	Luminal papillary	1	No relapse
Luminal papillary	Luminal infiltrated	1	NMIBC

A5.19 Clinicopathological features of RT cohort according to primary TCGA subtype

T-stage

	T2	T3	T4
Basal squamous	8	6	1
Luminal	4	2	0
Luminal infiltrated	3	3	0
Luminal papillary	4	6	0
Neuronal	2	4	0

Fishers exact: p = 0.9212

N-stage

	N0	N1	N2	N3
Basal squamous	10	4	0	1
Luminal	5	0	0	1
Luminal infiltrated	5	1	0	0
Luminal papillary	8	1	1	0
Neuronal	4	2	0	0

Fishers exact: p = 0.7594

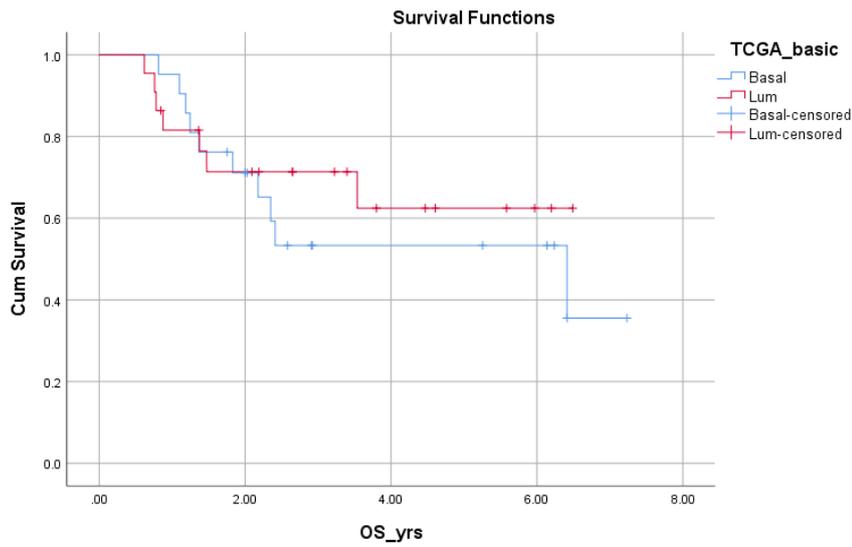
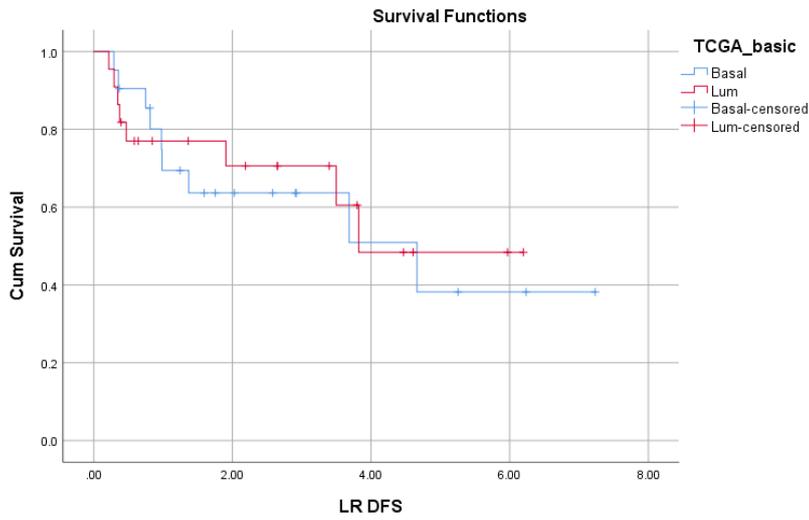
M-stage

	M0	M1
Basal squamous	13	2
Luminal	5	1
Luminal infiltrated	6	0
Luminal papillary	10	0
Neuronal	6	0

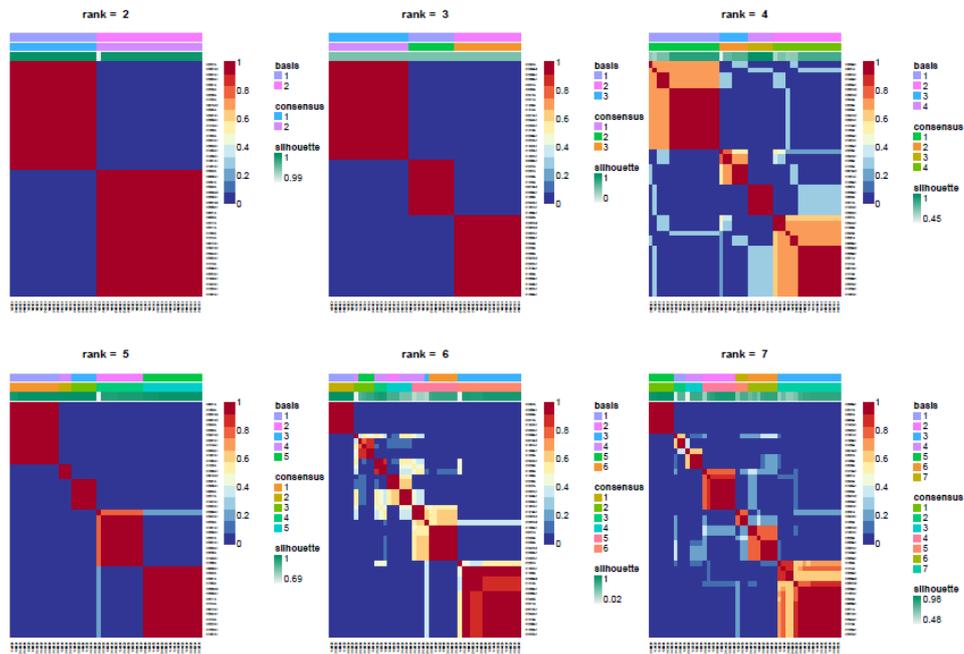
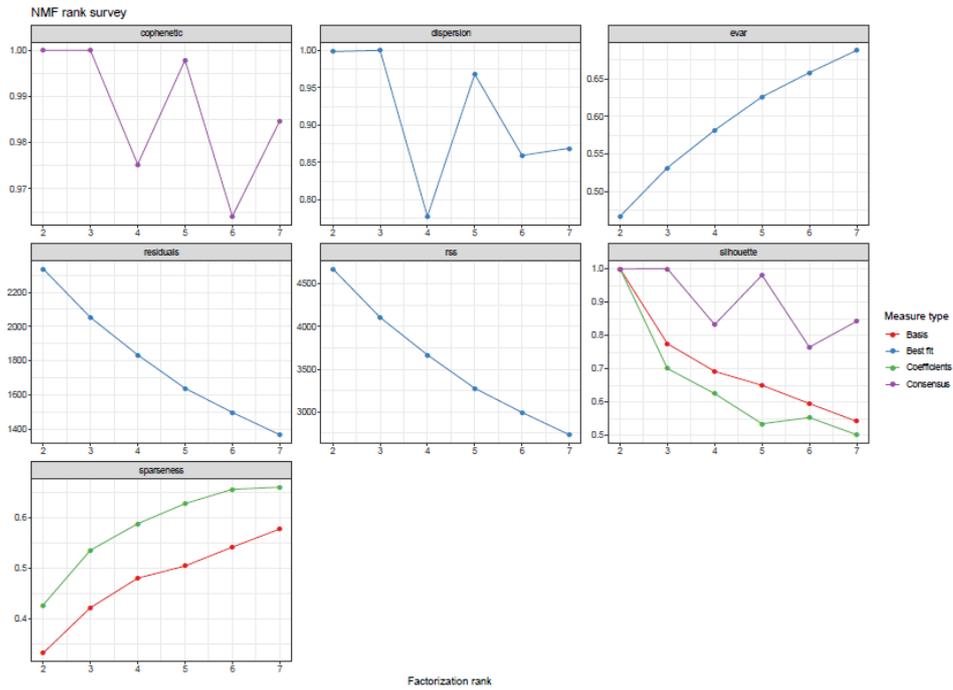
Fishers exact: $p = 0.6414$

A5.20 TCGA subtypes: basal vs luminal

	n	Any relapse	LRR	Invasive LRR	M1
Luminal	22/43	11	8	4	4
Basal	21/43	11	9	5	8



A5.21 NMF 134 pipeline outputs

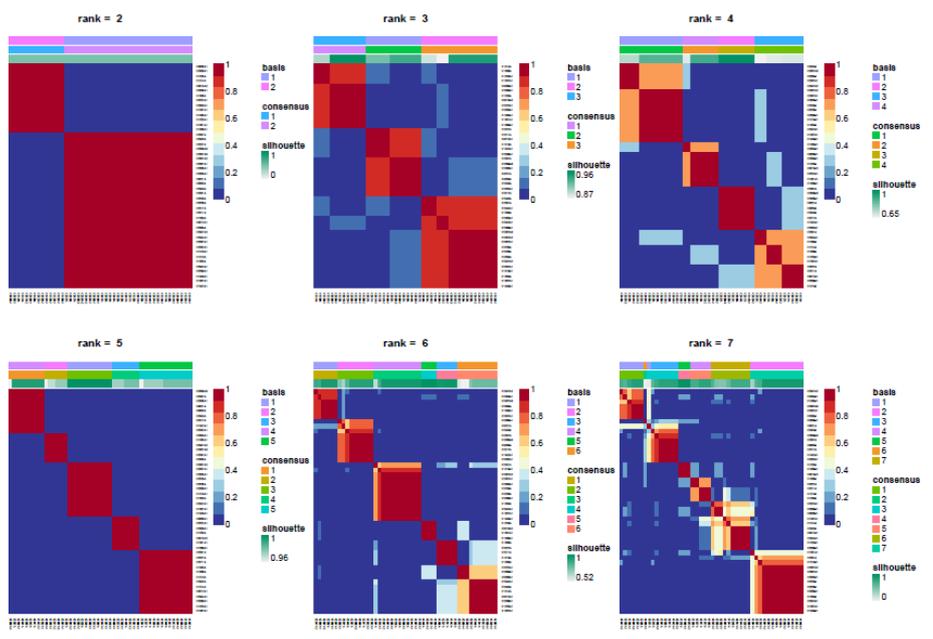
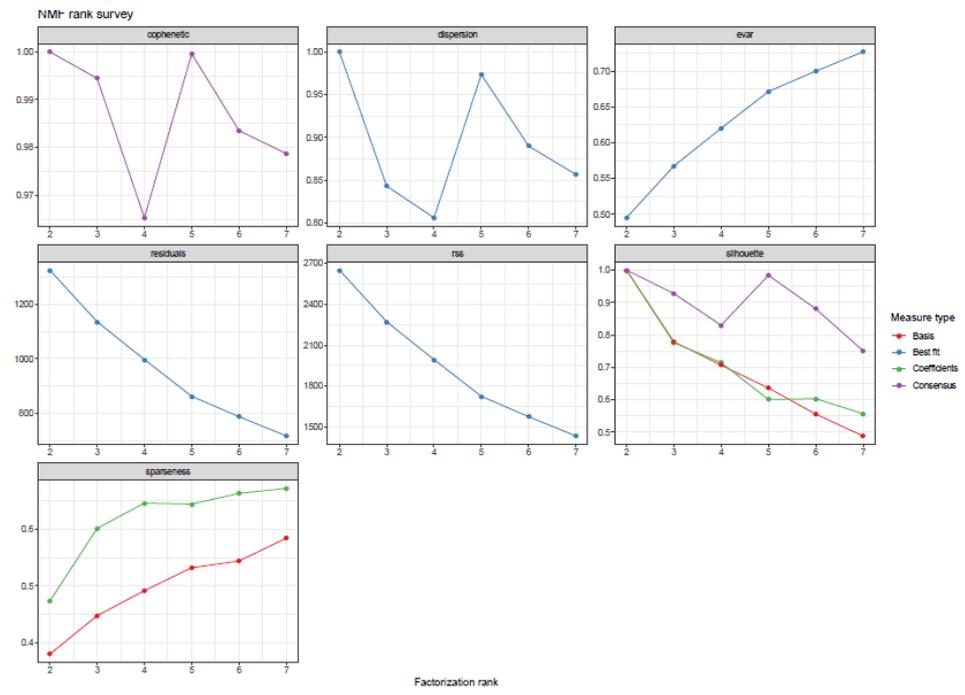


coph	silh	subtypes
0.999996	0.999563	2
1	1	3
0.975315	0.832483	4
0.997882	0.980667	5
0.964042	0.764597	6
0.98465	0.842323	7

A5.22 Post-radiotherapy biopsy results according to NMF134 subtype

NMF134	N	pT0 (No malignancy or atypia only)	pT1 (NMIBC)	≥pT2 (Residual invasive disease)
1	11	9/11	1/11	1/11
2	6	5/6	0/6	1/6
3	9	6/9	2/9	1/9
Overall	26	20	3	3

A5.23 NMF 91 pipeline outputs



coph	silh	subtypes
	1	2
0.994457	0.927623	3
0.965268	0.82992	4
0.999605	0.98387	5
0.983508	0.880406	6
0.978697	0.75121	7

A5.24 NMF71- Metagene list; PAM centroids

Genes	1-score	2-score	3-score	4-score	5-score
KRT14	0.3888	-0.0473	-0.4445	1.176	-0.3298
TUBB2B	0	-0.6179	1.1611	-0.0308	0
KRT5	0	-0.3959	-0.2414	1.0793	0
DES	0.9739	-0.2251	-0.0373	0	0
UPK2	0.0953	0.4493	-0.9182	-0.1384	0.179
SFRP4	0.8902	0.0169	-0.1718	-0.0933	-0.1525
KRT20	0.0352	0.8868	-0.8569	-0.1118	-0.0096
PI3	0	-0.4586	-0.3209	0.8157	0.1768
SNX31	0.0825	0.7214	-0.7389	-0.0046	-0.0309
ZEB2	0.6316	0	-0.1319	-0.0414	-0.0111
COMP	0.6137	-0.1354	-0.079	-0.1032	0.034
FOXA1	0	0.2425	-0.6036	0	0.0421
KRT6A	0.0112	-0.4294	-0.2069	0.5815	0.1128
UPK1A	0.1434	0.175	-0.5813	-0.0327	0.0987
RAD54L	-0.5504	0	0.1592	0	0.0831
CLDN4	-0.0526	0	-0.5467	0	0.4749
MSI1	-0.1647	0.0099	0.5309	0	-0.2336
SGCD	0.5194	-0.0302	0	0	-0.005
FGFR3	0	0.0253	-0.5105	0.1193	0.0572
CLDN3	-0.3992	-0.189	0	-0.079	0.5093
DSC3	-0.1435	-0.0718	-0.0183	0.5082	0
C7	0.4999	0	0.0673	-0.2731	-0.0525
PPARG	0	0.2301	-0.4933	0	0.051
ZEB1	0.4824	-0.0557	-0.2028	0	0
TWIST1	0	-0.3889	-0.1933	0.002	0.4736
TGM1	0	-0.3668	-0.0666	0.4471	0.1302
CLDN7	-0.313	-0.0855	-0.1561	0	0.4398
TP63	0	0	-0.4355	0.2474	0
LICAM	0.0516	-0.4311	0	0.2765	0.0308
GNG4	0	-0.3545	0.4219	0	0
COL17A1	-0.0531	-0.1889	0	0.404	0
PGM5	0.3966	-0.121	0	0	0
SAA1	0	-0.3873	-0.0446	0.1527	0.3329
Trex1	0	-0.3466	-0.0628	0	0.3852
MRE11A	0	-0.2508	-0.1323	0	0.3675
GSDMC	0.0252	-0.3167	-0.1325	0.3666	0.0752
ATM	0.3418	0.0218	0.0846	-0.1191	-0.2526
APLP1	0.08	-0.3358	0	0	0.1645
RND2	0.0519	-0.3312	-0.0694	0	0.3152
GATA3	0	-0.0494	-0.1144	0	0.3186
CD44	0.0705	-0.3104	-0.2129	0.1309	0.2314
RB1	0.2999	0	-0.3053	0.0255	0
RelA	0	-0.2001	-0.212	0.0166	0.292
STING	0.2911	-0.2353	-0.2094	0	0.1436
FANCD2	-0.1255	0.2897	-0.0025	0	0
CDH2	0.1548	-0.256	0	0	0.0087
ATR	0	0.2376	0	0	-0.1468
HDAC1	-0.0385	-0.0178	-0.0716	0	0.2336
CD274	0	-0.1894	0	0	0.2317
BRIP1	-0.2299	0.1049	0.2061	0	-0.0816
CDK1	-0.2265	0.2027	0.0127	0	-0.0151
ERCC6	0	0.218	0	0	-0.2148
ERCC4	0.0111	0	0	0	-0.2015
ERCC5	0.2004	0	-0.0322	0	-0.0137
FANCG	-0.0505	-0.1221	0	0	0.1917
E2F3	-0.1078	-0.1151	0.1877	0	0
cGAS	0	-0.1832	0	0	0.1268
FANCF	0	-0.0966	0	-0.0181	0.1772
PDCD1LG2	0.1295	-0.1742	0	0	0.0204
SLX4	0.0936	-0.1685	-0.0622	0.0308	0.0503
BCLAF1	0	0	0.0308	0	-0.1623
RAD50	0	0	-0.1615	0	0.0523
TXNIP	0.1351	-0.0537	-0.1524	0	0.1469
BRCA2	-0.0729	0	-0.0521	0	0.1495
ERCC1	0	-0.1371	0	0	0.1477
AIMP3	0	0.134	-0.1081	0	0
FANCB	-0.0583	0.1295	0.106	-0.091	-0.0315
BRCA1	-0.1295	0.035	0.0068	0	0
NBN	0	0.1222	0	0	0
SUMO1	0	-0.1022	0	0	0.101
KTM2D/MLL2	0.0362	0	0	0	-0.1015

A5.25 NAC response according to NMF71 subtype on GSC dataset

Response	1	2	3	4	5	Mixed	Undetermined
N	18	53	13	71	2	8	6
Y	8	46	10	43	2	6	3
NA	1	7	1	4	0	3	0

	Basal-sq	Lum-infil	Lum-pap	Luminal	Neuronal	Mixed	Undetermined
N	70	10	9	20	9	48	5
Y	44	3	17	18	6	27	3
NA	5	3	1	3	0	4	0

P=0.128

A5.26 Tables comparing subtype allocations in GSC cohort

NMF71 GSC	1	2	3	4	5	Mixed	Undeter	Total
Basal	8	1	10	36	0	3	1	59
Claudin low	4	0	2	37	0	2	0	45
Lum_inf	9	14	0	1	1	11	2	38
Luminal	1	60	3	13	2	0	2	81
NA	5	31	9	31	1	1	4	82
Total	27	106	24	118	4	17	9	305

NMF71 CRC	1	2	3	4	5	mixed	undeter	total
Enterocyte	2	24	3	8	1	1	4	43
Goblet-like	0	23	3	12	0	0	1	39
Inflammatory	5	11	6	52	0	4	1	79
Stem-like	20	20	3	19	0	9	1	72
TA	0	6	6	7	0	0	0	19
Mixed	0	18	2	16	0	2	1	39
Undetermined	0	4	1	4	3	1	1	14
Total	27	106	24	118	4	17	9	305

TCGA GSC	Bas-sq	Lum_inf	Lum_pap	Lum	Neuronal	Mixed	Undeter	Total
Basal	37	5	1	1	7	7	1	59
Claudin-low	43	0	0	0	0	1	1	45
Lum_inf	0	8	0	16	1	11	2	38
Luminal	6	1	14	15	2	40	3	81
NA	33	2	12	9	5	20	1	82
Total	119	16	27	41	15	79	8	305

A5.27 Table comparing NMF71 subtype allocation in TCGA dataset

TCGA	NMF71							Total
	1	2	3	4	5	Mixed	Undetermined	
Basal-sq	1.2% 1/84	0% 0/84	21.4% 18/84	64.3% 54/84	0% 0/84	13.1% 11/84	0% 0/84	84
Lum	20% 3/15	60% 9/15	0% 0/15	0% 0/15	0% 0/15	33.3% 3/15	0% 0/15	15
Lum-inf	64.4% 29/45	15.6% 7/45	4.4% 2/45	2.2% 1/45	0% 0/45	13.3% 6/45	0% 0/45	45
Lum-pap	3.7% 3/81	58.0% 47/81	6.2% 5/81	6.2% 5/81	7.4% 6/81	12.3% 10/81	6.2% 5/81	81
Neuronal	0% 0/9	11.1% 1/9	88.9% 8/9	0% 0/9	0% 0/9	0% 0/9	0% 0/9	9
Total	36	64	33	60	6	30	5	234

A5.28 Differentially expressed genes

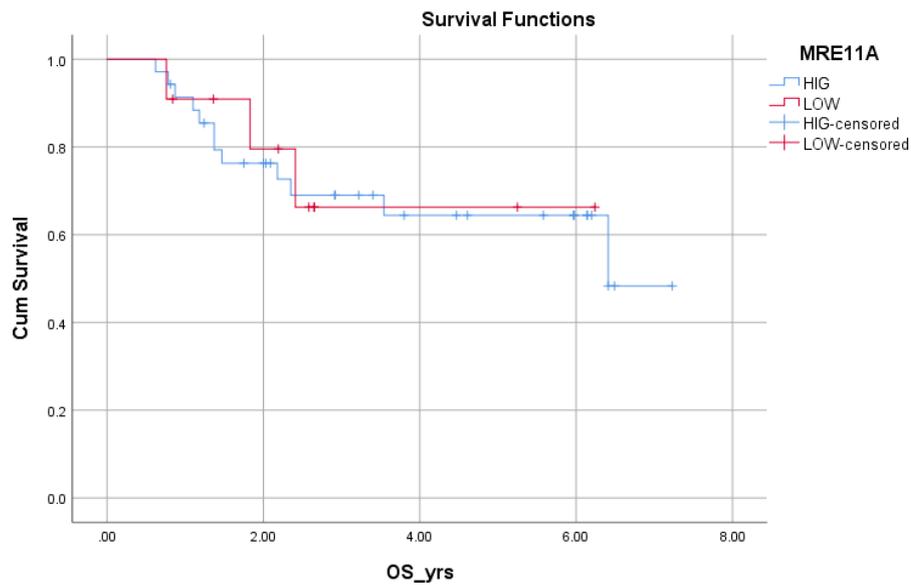
Critical value = $(i/m)Q$ where i = rank, m = number of tests, Q = false discovery rate (FDR). Significance is reached if the raw p-value is smaller than the critical value. For locoregional recurrence:

<i>Gene</i>	Raw p-value	i (rank)	BH with FDR 0.2	BH with FDR 0.05	q value ie. adjusted $p = \text{raw } p \times (n/\text{rank})$
<i>ERCC2</i>	0.002	1	0.006	0.001	0.072
<i>PKC</i>	0.006	2	0.011	0.003	0.108
<i>HDAC1</i>	0.043	3	0.017	0.004	0.516
<i>MRE11A</i>	0.078	4	0.022	0.006	0.702
<i>SLX4</i>	0.104	5	0.028	0.007	0.749
<i>ERCC5</i>	0.106	6	0.033	0.008	0.636
<i>PALB2</i>	0.136	7	0.039	0.010	0.699
<i>Ar</i>	0.248	8	0.044	0.011	1.116
<i>AIMP3</i>	0.258	9	0.050	0.013	1.032
<i>ATM</i>	0.269	10	0.056	0.014	0.968
<i>BRCA2</i>	0.297	11	0.061	0.015	0.972
<i>ERCC4</i>	0.385	12	0.067	0.017	1.155
<i>RB1</i>	0.391	13	0.072	0.018	1.083
<i>SUMO1</i>	0.398	14	0.078	0.019	1.023
<i>STAT1</i>	0.419	15	0.083	0.021	1.006
<i>FANCG</i>	0.456	16	0.089	0.022	1.026
<i>HIF1alpha</i>	0.456	16	0.089	0.022	1.026
<i>RAD50</i>	0.479	17	0.094	0.024	1.014
<i>FANCF</i>	0.585	18	0.100	0.025	1.170
<i>cJun</i>	0.602	19	0.106	0.026	1.141
<i>BRIP1</i>	0.637	20	0.111	0.028	1.147
<i>FANCD2</i>	0.655	21	0.117	0.029	1.123
<i>RelA</i>	0.655	21	0.117	0.029	1.123
<i>CDK1</i>	0.7	22	0.122	0.031	1.145
<i>NFE2L2</i>	0.766	23	0.128	0.032	1.199
<i>ERCC6</i>	0.794	24	0.133	0.033	1.191
<i>FANCB</i>	0.823	25	0.139	0.035	1.185
<i>cABL</i>	0.833	26	0.144	0.036	1.153
<i>IRF1</i>	0.842	27	0.150	0.038	1.123
<i>KAT5</i>	0.842	27	0.150	0.038	1.123
<i>BRCA1</i>	0.852	28	0.156	0.039	1.095
<i>ATR</i>	0.921	29	0.161	0.040	1.143
<i>NBN</i>	0.931	30	0.167	0.042	1.117
<i>ERCC1</i>	0.97	31	0.172	0.043	1.126
<i>RAD54L</i>	0.97	31	0.172	0.043	1.126
<i>TXNIP</i>	1	32	0.178	0.044	1.125

For Invasive LRR:

Gene	Raw p-value	i (rank)	Critical value with FDR 0.2	Critical value with FDR 0.05	q value ie. adjusted $p = \text{raw } p \times (n/\text{rank})$
<i>ATM</i>	0.002	1	0.006	0.001	0.072
<i>ERCC5</i>	0.002	1	0.006	0.001	0.072
<i>ERCC2</i>	0.005	2	0.011	0.003	0.090
<i>BRCA2</i>	0.008	3	0.017	0.004	0.096
<i>PKC</i>	0.034	4	0.022	0.006	0.306
<i>HDAC1</i>	0.105	5	0.028	0.007	0.756
<i>RAD50</i>	0.198	6	0.033	0.008	1.188
<i>AR</i>	0.281	7	0.039	0.010	1.445
<i>RelA</i>	0.323	8	0.044	0.011	1.454
<i>TXNIP</i>	0.353	9	0.050	0.013	1.412
<i>AIMP3</i>	0.401	10	0.056	0.014	1.444
<i>PALB2</i>	0.435	11	0.061	0.015	1.424
<i>STAT1</i>	0.435	11	0.061	0.015	1.424
<i>SUMO1</i>	0.49	12	0.067	0.017	1.470
<i>CDK1</i>	0.527	13	0.072	0.018	1.459
<i>RAD54L</i>	0.607	14	0.078	0.019	1.561
<i>ERCC1</i>	0.607	14	0.078	0.019	1.561
<i>ERCC4</i>	0.66	15	0.083	0.021	1.584
<i>SLX4</i>	0.692	16	0.089	0.022	1.557
<i>ATR</i>	0.7	17	0.094	0.024	1.482
<i>FANCB</i>	0.714	18	0.100	0.025	1.428
<i>IRF1</i>	0.736	19	0.106	0.026	1.395
<i>BRCA1</i>	0.758	20	0.111	0.028	1.364
<i>NBN</i>	0.758	20	0.111	0.028	1.364
<i>cJun</i>	0.758	20	0.111	0.028	1.364
<i>FANCD2</i>	0.78	21	0.117	0.029	1.337
<i>FANCG</i>	0.78	21	0.117	0.029	1.337
<i>ERCC6</i>	0.803	25	0.139	0.035	1.156
<i>MRE11A</i>	0.826	26	0.144	0.036	1.144
<i>RB1</i>	0.849	27	0.150	0.038	1.132
<i>cABL</i>	0.849	27	0.150	0.038	1.132
<i>BRIP1</i>	0.872	28	0.156	0.039	1.121
<i>FANCF</i>	0.872	28	0.156	0.039	1.121
<i>HIF1alpha</i>	0.872	28	0.156	0.039	1.121
<i>NFE2L2</i>	0.942	29	0.161	0.040	1.169
<i>KAT5</i>	0.988	30	0.167	0.042	1.186

A5.29 MRE11- Kaplan Meier Analysis of Bladder Cancer Specific Survival



Means and Medians for Survival Time

MRE1 1A	Estimate	Std. Error	Mean ^a		Estimate	Std. Error	Median	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
HIG	5.115	.485	4.164	6.065	6.410	.	.	.
LOW	4.733	.720	3.321	6.145
Overall	5.132	.430	4.288	5.975	6.410	.	.	.

a. Estimation is limited to the largest survival time if it is censored.

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	.010	1	.921

Test of equality of survival distributions for the different levels of MRE11A.

A5.30 Cross-referencing NMF71 subtype with DNA aberrations

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>TERT</i> promoter	100.00	100.00	33.33	100.00	75.00	16/21 (76.2%)
<i>TP53</i>	66.67	33.33	66.67	80.00	75.00	14/21 (66.7%)
<i>ATM</i>	0.00	33.33	66.67	40.00	50.00	9/21 (42.9%)
<i>ERBB2</i>	0.00	33.33	50.00	40.00	50.00	8/21 (38.1%)
<i>RB1</i>	100.00	0.00	16.67	60.00	25.00	8/21 (38.1%)
<i>ERBB3</i>	33.33	0.00	33.33	40.00	50.00	7/21 (33.3%)
<i>FGFR3</i>	33.33	0.00	66.67	20.00	25.00	7/21 (33.3%)
<i>MLL2</i>	33.33	0.00	50.00	20.00	50.00	7/21 (33.3%)
<i>PIK3CA</i>	0.00	33.33	33.33	40.00	25.00	6/21 (28.6%)
<i>SDHC</i>	0.00	0.00	66.67	0.00	50.00	6/21 (28.6%)

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>ARID1A</i>	33.33	0.00	33.33	20.00	25.00	5/21 (23.8%)
<i>ASXL1</i>	33.33	0.00	50.00	20.00	0.00	5/21 (23.8%)
<i>GNAS</i>	0.00	0.00	33.33	20.00	50.00	5/21 (23.8%)
<i>LRP1B</i>	33.33	33.33	16.67	0.00	50.00	5/21 (23.8%)
<i>NOTCH1</i>	66.67	0.00	0.00	0.00	75.00	5/21 (23.8%)
<i>ZNF703</i>	0.00	0.00	33.33	40.00	25.00	5/21 (23.8%)
<i>CREBBP</i>	66.67	0.00	0.00	0.00	50.00	4/21 (19.0%)
<i>DDR2</i>	0.00	0.00	33.33	0.00	50.00	4/21 (19.0%)
<i>GPR124</i>	66.67	0.00	16.67	20.00	0.00	4/21 (19.0%)
<i>KDM6A</i>	33.33	0.00	33.33	20.00	0.00	4/21 (19.0%)
<i>KMT2C</i> (<i>MLL3</i>)	33.33	33.33	16.67	0.00	25.00	4/21 (19.0%)

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>MSH6</i>	0.00	0.00	33.33	40.00	0.00	4/21 (19.0%)
<i>NSD1</i>	0.00	0.00	33.33	40.00	0.00	4/21 (19.0%)
<i>RAF1</i>	0.00	33.33	16.67	40.00	0.00	4/21 (19.0%)
<i>SDHA</i>	33.33	0.00	16.67	40.00	0.00	4/21 (19.0%)
<i>SETD2</i>	0.00	33.33	33.33	20.00	0.00	4/21 (19.0%)
<i>STAG2</i>	33.33	33.33	0.00	20.00	25.00	4/21 (19.0%)
<i>ARID1B</i>	0.00	0.00	33.33	20.00	0.00	3/21 (14.3%)
<i>ATRX</i>	0.00	0.00	16.67	0.00	50.00	3/21 (14.3%)
<i>AXL</i>	0.00	0.00	33.33	20.00	0.00	3/21 (14.3%)
<i>BCL2L1</i>	0.00	0.00	16.67	20.00	25.00	3/21 (14.3%)
<i>BRCA1</i>	0.00	0.00	16.67	40.00	0.00	3/21 (14.3%)

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>BRCA2</i>	0.00	0.00	0.00	20.00	50.00	3/21 (14.3%)
<i>CARD11</i>	0.00	0.00	16.67	20.00	25.00	3/21 (14.3%)
<i>CDKN2A</i>	0.00	66.67	0.00	0.00	25.00	3/21 (14.3%)
<i>CDKN2B</i>	0.00	66.67	0.00	0.00	25.00	3/21 (14.3%)
<i>EP300</i>	33.33	0.00	16.67	20.00	0.00	3/21 (14.3%)
<i>EPHA5</i>	0.00	0.00	33.33	0.00	25.00	3/21 (14.3%)
<i>FANCD2</i>	0.00	0.00	16.67	20.00	25.00	3/21 (14.3%)
<i>FANCL</i>	0.00	0.00	33.33	0.00	25.00	3/21 (14.3%)
<i>FGF19</i>	0.00	33.33	16.67	0.00	25.00	3/21 (14.3%)
<i>FOXP1</i>	0.00	33.33	0.00	20.00	25.00	3/21 (14.3%)
<i>FRS2</i>	0.00	0.00	50.00	0.00	0.00	3/21 (14.3%)

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>GATA3</i>	0.00	0.00	33.33	0.00	25.00	3/21 (14.3%)
<i>IL7R</i>	66.67	0.00	16.67	0.00	0.00	3/21 (14.3%)
<i>KDM5A</i>	33.33	0.00	16.67	20.00	0.00	3/21 (14.3%)
<i>LYN</i>	0.00	0.00	33.33	20.00	0.00	3/21 (14.3%)
<i>LZTR1</i>	0.00	33.33	0.00	40.00	0.00	3/21 (14.3%)
<i>MCL1</i>	33.33	0.00	16.67	0.00	25.00	3/21 (14.3%)
<i>MDM2</i>	0.00	0.00	50.00	0.00	0.00	3/21 (14.3%)
<i>MLL3</i>	0.00	33.33	0.00	20.00	25.00	3/21 (14.3%)
<i>MUTYH</i>	0.00	0.00	50.00	0.00	0.00	3/21 (14.3%)
<i>MYST3</i>	33.33	0.00	16.67	0.00	25.00	3/21 (14.3%)
<i>NF1</i>	33.33	0.00	16.67	20.00	0.00	3/21 (14.3%)

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>NOTCH2</i>	0.00	0.00	16.67	40.00	0.00	3/21 (14.3%)
<i>NOTCH3</i>	33.33	0.00	16.67	20.00	0.00	3/21 (14.3%)
<i>NTRK1</i>	0.00	0.00	16.67	40.00	0.00	3/21 (14.3%)
<i>PBRM1</i>	0.00	0.00	16.67	40.00	0.00	3/21 (14.3%)
<i>POLE</i>	0.00	33.33	16.67	0.00	25.00	3/21 (14.3%)
<i>PRKDC</i>	0.00	0.00	16.67	40.00	0.00	3/21 (14.3%)
<i>SPTA1</i>	0.00	0.00	16.67	20.00	25.00	3/21 (14.3%)
<i>TGFBR2</i>	0.00	33.33	33.33	0.00	0.00	3/21 (14.3%)
<i>TOP2A</i>	0.00	0.00	0.00	20.00	50.00	3/21 (14.3%)
<i>TSC2</i>	0.00	33.33	0.00	0.00	50.00	3/21 (14.3%)
<i>VHL</i>	0.00	0.00	33.33	0.00	25.00	3/21 (14.3%)

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>XPO1</i>	0.00	33.33	16.67	20.00	0.00	3/21 (14.3%)
<i>APC</i>	0.00	0.00	0.00	20.00	25.00	2/21 (9.52%)
<i>AR</i>	33.33	0.00	0.00	20.00	0.00	2/21 (9.52%)
<i>BAP1</i>	0.00	33.33	0.00	20.00	0.00	2/21 (9.52%)
<i>BRAF</i>	0.00	0.00	33.33	0.00	0.00	2/21 (9.52%)
<i>BRIP1</i>	0.00	33.33	16.67	0.00	0.00	2/21 (9.52%)
<i>CBL</i>	0.00	0.00	0.00	40.00	0.00	2/21 (9.52%)
<i>CCND1</i>	0.00	33.33	0.00	0.00	25.00	2/21 (9.52%)
<i>CDH1</i>	33.33	0.00	16.67	0.00	0.00	2/21 (9.52%)
<i>CDK12</i>	33.33	0.00	0.00	0.00	25.00	2/21 (9.52%)
<i>CEBPA</i>	0.00	0.00	33.33	0.00	0.00	2/21 (9.52%)

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>CHEK2</i>	33.33	0.00	16.67	0.00	0.00	2/21 (9.52%)
<i>CIC</i>	0.00	0.00	16.67	0.00	25.00	2/21 (9.52%)
<i>CRKL</i>	0.00	0.00	0.00	40.00	0.00	2/21 (9.52%)
<i>CYLD</i>	0.00	33.33	0.00	0.00	25.00	2/21 (9.52%)
<i>DOT1L</i>	0.00	0.00	0.00	40.00	0.00	2/21 (9.52%)
<i>FAM123B</i>	0.00	33.33	0.00	20.00	0.00	2/21 (9.52%)
<i>FANCA</i>	0.00	33.33	0.00	20.00	0.00	2/21 (9.52%)
<i>FAT1</i>	0.00	0.00	16.67	0.00	25.00	2/21 (9.52%)
<i>FBXW7</i>	0.00	0.00	0.00	20.00	25.00	2/21 (9.52%)
<i>FGF3</i>	0.00	33.33	0.00	0.00	25.00	2/21 (9.52%)
<i>FGF4</i>	0.00	33.33	0.00	0.00	25.00	2/21 (9.52%)

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>FGFR1</i>	0.00	0.00	0.00	20.00	25.00	2/21 (9.52%)
<i>FLT4</i>	33.33	0.00	16.67	0.00	0.00	2/21 (9.52%)
<i>GATA6</i>	33.33	0.00	0.00	0.00	25.00	2/21 (9.52%)
<i>GNA13</i>	33.33	0.00	16.67	0.00	0.00	2/21 (9.52%)
<i>GRIN2A</i>	33.33	0.00	0.00	0.00	25.00	2/21 (9.52%)
<i>HSP90AA1</i>	0.00	0.00	16.67	0.00	25.00	2/21 (9.52%)
<i>IRS2</i>	0.00	0.00	16.67	20.00	0.00	2/21 (9.52%)
<i>JAK2</i>	66.67	0.00	0.00	0.00	0.00	2/21 (9.52%)
<i>KDR</i>	0.00	0.00	16.67	0.00	25.00	2/21 (9.52%)
<i>KRAS</i>	33.33	0.00	0.00	0.00	25.00	2/21 (9.52%)
<i>MAGI2</i>	33.33	0.00	16.67	0.00	0.00	2/21 (9.52%)

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>MLL</i>	33.33	0.00	0.00	0.00	25.00	2/21 (9.52%)
<i>MTOR</i>	33.33	0.00	0.00	0.00	25.00	2/21 (9.52%)
<i>NFE2L2</i>	0.00	0.00	16.67	20.00	0.00	2/21 (9.52%)
<i>NUP93</i>	0.00	0.00	16.67	20.00	0.00	2/21 (9.52%)
<i>PDGFRA</i>	66.67	0.00	0.00	0.00	0.00	2/21 (9.52%)
<i>PIK3R2</i>	33.33	0.00	0.00	0.00	25.00	2/21 (9.52%)
<i>PLCG2</i>	0.00	0.00	16.67	0.00	25.00	2/21 (9.52%)
<i>PRDM1</i>	0.00	0.00	16.67	0.00	25.00	2/21 (9.52%)
<i>RANBP2</i>	0.00	33.33	0.00	0.00	25.00	2/21 (9.52%)
<i>RARA</i>	0.00	0.00	0.00	20.00	25.00	2/21 (9.52%)
<i>RBM10</i>	0.00	33.33	16.67	0.00	0.00	2/21 (9.52%)

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>RET</i>	33.33	0.00	16.67	0.00	0.00	2/21 (9.52%)
<i>RICTOR</i>	33.33	0.00	16.67	0.00	0.00	2/21 (9.52%)
<i>RUNX1</i>	0.00	33.33	0.00	0.00	25.00	2/21 (9.52%)
<i>RUNX1T1</i>	0.00	0.00	33.33	0.00	0.00	2/21 (9.52%)
<i>SLIT2</i>	0.00	0.00	16.67	20.00	0.00	2/21 (9.52%)
<i>SMARCA4</i>	0.00	0.00	0.00	20.00	25.00	2/21 (9.52%)
<i>SPEN</i>	0.00	0.00	16.67	20.00	0.00	2/21 (9.52%)
<i>TAF1</i>	0.00	33.33	0.00	20.00	0.00	2/21 (9.52%)
<i>TET2</i>	0.00	0.00	16.67	0.00	25.00	2/21 (9.52%)
<i>TOP1</i>	0.00	0.00	0.00	20.00	25.00	2/21 (9.52%)
<i>ABL1</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>ACVR1B</i>	33.33	0.00	0.00	0.00	0.00	1/21 (4.76%)
<i>AKT1</i>	0.00	33.33	0.00	0.00	0.00	1/21 (4.76%)
<i>AKT2</i>	33.33	0.00	0.00	0.00	0.00	1/21 (4.76%)
<i>AKT3</i>	0.00	0.00	0.00	0.00	25.00	1/21 (4.76%)
<i>ATR</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>AURKA</i>	0.00	0.00	0.00	20.00	0.00	1/21 (4.76%)
<i>AXIN1</i>	33.33	0.00	0.00	0.00	0.00	1/21 (4.76%)
<i>BRD4</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>C11orf30</i> (<i>EMSY</i>)	0.00	0.00	0.00	0.00	25.00	1/21 (4.76%)
<i>CD79B</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>CDKN1A</i>	0.00	0.00	0.00	20.00	0.00	1/21 (4.76%)

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>CRLF2</i>	0.00	0.00	0.00	20.00	0.00	1/21 (4.76%)
<i>CTNNB1</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>EPHA3</i>	0.00	33.33	0.00	0.00	0.00	1/21 (4.76%)
<i>EPHA7</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>EPHB1</i>	0.00	0.00	0.00	0.00	25.00	1/21 (4.76%)
<i>ERG</i>	0.00	0.00	0.00	20.00	0.00	1/21 (4.76%)
<i>ETV</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>ETV1</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>FANCE</i>	0.00	0.00	0.00	20.00	0.00	1/21 (4.76%)
<i>FANCF</i>	0.00	0.00	0.00	0.00	25.00	1/21 (4.76%)
<i>FGF10</i>	33.33	0.00	0.00	0.00	0.00	1/21 (4.76%)

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>FGFR4</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>FLT1</i>	0.00	0.00	0.00	20.00	0.00	1/21 (4.76%)
<i>FLT3</i>	0.00	0.00	0.00	20.00	0.00	1/21 (4.76%)
<i>GATA1</i>	0.00	33.33	0.00	0.00	0.00	1/21 (4.76%)
<i>GATA2</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>GRM3</i>	0.00	0.00	0.00	20.00	0.00	1/21 (4.76%)
<i>GSK3B</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>HNF1A</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>IDH1</i>	0.00	33.33	0.00	0.00	0.00	1/21 (4.76%)
<i>IGF1R</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>IGF2</i>	0.00	0.00	0.00	20.00	0.00	1/21 (4.76%)

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>INHBA</i>	0.00	33.33	0.00	0.00	0.00	1/21 (4.76%)
<i>JAK3</i>	33.33	0.00	0.00	0.00	0.00	1/21 (4.76%)
<i>MAP2K2</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>MED12</i>	0.00	33.33	0.00	0.00	0.00	1/21 (4.76%)
<i>MITF</i>	0.00	0.00	0.00	0.00	25.00	1/21 (4.76%)
<i>MLH1</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>MPL</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>MSH2</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>MYCL1</i>	0.00	33.33	0.00	0.00	0.00	1/21 (4.76%)
<i>MYCN</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>MYD88</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>NF2</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>NTRK3</i>	0.00	33.33	0.00	0.00	0.00	1/21 (4.76%)
<i>PALB2</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>PARK2</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>PDGFRB</i>	0.00	33.33	0.00	0.00	0.00	1/21 (4.76%)
<i>PIK3R1</i>	0.00	0.00	0.00	0.00	25.00	1/21 (4.76%)
<i>PMS2</i>	0.00	33.33	0.00	0.00	0.00	1/21 (4.76%)
<i>PRKAR1A</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>PTCH1</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>PTEN</i>	0.00	0.00	0.00	0.00	25.00	1/21 (4.76%)
<i>RAD50</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>RAD51</i>	0.00	33.33	0.00	0.00	0.00	1/21 (4.76%)
<i>ROS1</i>	0.00	0.00	0.00	20.00	0.00	1/21 (4.76%)
<i>RPTOR</i>	0.00	0.00	0.00	0.00	25.00	1/21 (4.76%)
<i>SDHB</i>	0.00	0.00	0.00	20.00	0.00	1/21 (4.76%)
<i>SF3B1</i>	0.00	33.33	0.00	0.00	0.00	1/21 (4.76%)
<i>SMO</i>	0.00	33.33	0.00	0.00	0.00	1/21 (4.76%)
<i>SPOP</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>SRC</i>	0.00	0.00	0.00	20.00	0.00	1/21 (4.76%)
<i>STAT3</i>	0.00	0.00	0.00	20.00	0.00	1/21 (4.76%)
<i>STAT4</i>	0.00	0.00	0.00	0.00	25.00	1/21 (4.76%)
<i>SUFU</i>	0.00	33.33	0.00	0.00	0.00	1/21 (4.76%)

Gene	NMF90 Subtype (% with aberration)					Total no patients with aberration
	1	2	3	4	5	
	n = 3	n = 3	n = 6	n = 5	n = 4	
<i>TNFRSF14</i>	0.00	0.00	16.67	0.00	0.00	1/21 (4.76%)
<i>TSC1</i>	0.00	33.33	0.00	0.00	0.00	1/21 (4.76%)
<i>ZNF217</i>	0.00	0.00	0.00	20.00	0.00	1/21 (4.76%)

A5.31 Table to assess performance of recreated TCGA centroids

Figure in each box shows how many of the TCGA subtype fell into the NMF71 subtype e.g. for the first row, 86.6% of those labelled as basal-squamous using the recreated TCGA centroids in the GSC dataset fell into NMF71 subtype 4 and 65.1% of those labelled the basal squamous in the TCGA dataset (ie. using 'official' classifier) were allocated to NMF71 subtype 4.

	1		2		3		4		5	
	GSC dataset	TCGA dataset	GSC dataset	TCGA dataset	GSC dataset	TCGA dataset	GSC dataset	TCGA dataset	GSC dataset	TCGA dataset
Basal-sq	5.9% 7/119	1.2% 1/83	0% 0/119	0% 0/83	5.0% 6/119	21.7% 18/83	86.6% 103/119	65.1% 54/83	0% 0/119	0% 0
Lum	0% 0/41	20% 3/15	80.5% 33/41	60% 9/15	0% 0/41	0% 0/15	0% 0/41	0% 0/15	4.9% 2/41	0% 0/15
Lum-inf	75% 12/16	64.4% 29/45	6.3% 1/16	15.6% 7/45	0% 0/16	4.4% 2/45	0% 0/16	2.2% 1/45	6.3% 1/16	0% 0/45

	1		2		3		4		5	
	GSC dataset	TCGA dataset	GSC dataset	TCGA dataset	GSC dataset	TCGA dataset	GSC dataset	TCGA dataset	GSC dataset	TCGA dataset
Lum-pap	0% 0/27	3.7% 3/81	51.9% 14/27	58.0% 47/81	0% 0/27	6.2% 5/81	33.3% 9/27	6.2% 5/81	3.7% 1/27	7.4% 6/81
Neuronal	0% 0/15	0% 0/9	0% 0/15	11.1% 1/9	93.3% 14/15	88.9% 8/9	0% 0/15	0% 0/9	0% 0/15	0% 0/9

Chapter 6. Discussion

6.1. Overview of findings

Within this MD(Res), I have achieved many of the aims I initially set out in 2015. It has been an interesting and dynamic field to work in, especially given the surge of relevant publications and new data published over the course of my work. I have completed pilot studies at a genomic and transcriptomic level to explore the role of circulating tumour DNA, multiregion whole-exome sequencing and molecular subtyping in MIBC patients primarily treated with radical radiotherapy +/- chemotherapy. Key findings from my work include-

- Baseline pre-treatment plasma DNA concentration is an independent prognostic factor for bladder cancer-specific survival, with an increase in plasma DNA concentration associated with an increasing risk of bladder cancer death
- Plasma tumour fraction, estimated from copy number data generated by low-pass whole genome sequencing of plasma DNA, is significantly positively correlated with radiological disease burden and may be a surrogate of radiological assessment; although detectable only in patients with a high tumour burden, analysis of sequential samples from those receiving first-line chemotherapy suggest that tumour fraction dynamics reflect disease status
- Multiregion high depth whole-exome sequencing of FFPE-derived DNA is feasible in MIBC and I have established a workflow for this process; further ways to optimise the workflow have been suggested and further analysis of the raw data generated is underway
- Even in a limited cohort, high depth WES has demonstrated potential intra-tumour heterogeneity
- Targeting sequencing results from FFPE tumour tissue suggest that *TP53* and *KDM6A* aberrations may be associated with poorer clinical

outcomes, and that in the subset of patients treated with radical radiotherapy, patients with a DNA damage repair gene alteration may have improved locoregional relapse-free survival

- Analysis of expression data from my custom-designed panel has demonstrated that patients can be divided into 5 molecular subtypes, with subtypes 4 and 5 demonstrating statistically significant higher pathological complete response rates following chemoradiation compared to those in subtypes 1-3. Improved invasive locoregional relapse-free survival was also observed in subtypes 4-5. Underexpression of *ATM* in subtypes 4 and 5, in addition to higher expression levels of immune markers in subtype 5 may have contributed to the results seen

The limitations of my work have already been discussed in detail within each chapter but one common issue was that the small cohort numbers limited statistical analysis. This was not unexpected as the focus of these pilot studies was to assess feasibility of performing various translational approaches, to explore hypotheses and provide preliminary data to build upon. A huge part of my MD(Res) was also to write and establish the CoMB protocol in order to support this work.

To this end, I believe my project has been successful and my results have provided some interesting preliminary data to build upon. Moving forward, there are many areas for further work, which have been discussed in each of the data chapters. One key approach will also be to integrate the data generated for each patient.

6.2. Cross-platform analysis

6.2.1. Data available within CoMB

Samples from a total of 88 patients were used to generate data for my MD(Res). Of those, 51 patients have sequencing data at a genomic and/or

transcriptomic level available (Table 6.1). 21 patients have both expression data and genomic data from the FoundationOne panel available. Three of these patients additionally have multiregion WES and plasma LPWGS data.

I think this represents a good opportunity to integrate data across platforms and could be basis of future work, perhaps with the addition of immunohistochemical data. One area where I think such an approach would be interesting is that of exploring the role of *ATM* further, as this was a feature common to both my DNA and RNA work.

Table 6.1 (overleaf) Summary of data held for each patient

COMB ID	Plasma LPWGS	Multiregion WES	FoundationOne	Molecular Subtyping
1003	✓	✓	✓	✓
1010	✓	✓	✓	
1014	✓	✓	✓	✓
1016	✓	✓	✓	
1017	✓	✓	✓	
1021	✓	✓	✓	
1025	✓	✓	✓	✓
1027	✓	✓	✓	
1039	✓	✓	✓	
1052	✓		✓	
1001			✓	✓
1008			✓	✓
1019			✓	✓
1020			✓	✓
1035			✓	✓
1042			✓	✓
1049			✓	✓
2005			✓	✓
2008			✓	✓
2036			✓	
2037			✓	✓
2039			✓	✓
2042			✓	✓
2043			✓	✓
2047			✓	✓
3006			✓	✓
3012			✓	✓
3015			✓	✓
3021			✓	✓
1006				✓
1007				✓
1013				✓
1033				✓
1037				✓
2003				✓
2006				✓
2007				✓
2009				✓
2011				✓
2017				✓
2022				✓
2026				✓
2027				✓
2029				✓
2031				✓
2032				✓
2040				✓
2044				✓
2055				✓
2057				✓
3003				✓

6.2.2. ATM

ATM is recognised to be a potential driver gene in MIBC¹. It plays a significant role in the DNA damage pathway, and is phosphorylated by the MRN complex (*MRE11*, *NBS*, *RAD50*) following detection of DNA damage.

Patients with alterations in at least one of *ATM*, *RB1* and *FANCC* have been reported to demonstrate response to neoadjuvant chemotherapy² in MIBC. Interestingly this work included a subset of samples that had been used in Choi's landmark paper³ describing 4 molecular subtypes in MIBC, one of which appeared to be associated with chemoresistance. When integrating the genomic and subtype results, no correlation between *ATM/RB1/FANCC* signature and subtype was seen. Alterations that were predicted to be deleterious did not predict for lack of expression. This may be due to different effects of the alteration from predicted or may represent post-translational modifications. Whatever the reason, the discrepancy seen highlights that we still have much to understand about these pathways. With regards to the lack of correlation between the *ATM/RB1/FANCC* signature and subtypes, I do not think it is surprising that no association was seen given that the genomic signature is associated with chemosensitivity while the subtype classifier appeared to identify those who were chemoresistant, but not those who were most chemosensitive.

In my work, *ATM* was one of 5 DDR genes I selected (from published data at the time⁴) when assessing response to radiotherapy. There was a suggestion in my small cohort of 25 patients that those with at least one alteration in *ATM*, *BRCA1*, *BRCA2*, *FANCD* and *PALB2* might have an improved locoregional relapse-free survival following radiotherapy. Fifteen of the twenty-five patients had a DDR gene alteration and of those, *ATM* alterations were present in 10/15. Only 2 of the *ATM* mutations were reported to be of clinical significance by the FoundationOne panel with the remaining considered of unknown significance. Of the 10 patients with *ATM* aberrations, 1/10 (10%) had an invasive locoregional recurrence compared to 5/15 (33%) of those without an *ATM* aberration.

Nine of the ten patients with an *ATM* aberration also have expression data available. One might have hypothesised that those with *ATM* aberrations might be allocated to subtypes 4 and 5 where *ATM* expression was lower, however this did not appear to be the case. Further work exploring the predicted impact of the *ATM* aberrations reported would be of interest here.

On the basis of my data generated, I think further exploration of *ATM* is warranted. An interesting first step would be to apply NMF71 to the MIBC TCGA RNA-sequencing data and then assess whether subtypes can be characterised by *ATM* aberrations or indeed any other genomic alterations. Using such a large dataset would allow more meaningful analyses that could then be applied to an expanded cohort of patients treated with radiotherapy. In the small cohort of 21 patients with both expression data and FoundationOne results available, there did not appear to be any obvious aberrations clustering with any particular subtype but the cohort was too small to investigate this fully.

6.2.3. DNA damage repair genes

Recent work has recognised that MIBC has a higher proportion of alterations in DNA damage repair genes than other solid cancers¹. Key driver genes of *ATM* and *ERCC2* have been recognised and it has been shown that patients with DDR genes have a higher number of somatic mutations⁴ and subsequently greater tumour mutational burden (TMB). TMB has been of interest as a possible biomarker of immunotherapy response^{5,6}, and has also been suggested to be associated with response to neoadjuvant chemotherapy in MIBC⁷. It would be interesting to explore whether TMB has any relevance as a biomarker with regards to response to radiotherapy. One could hypothesise that tumours with a greater TMB represent greater immunogenic potential which, following radiation-induced cell death and injury, results in the release of a greater number of tumour-associated antigens, which stimulate and direct the immune system against cancer cells, potentially resulting in improved tumour control. In my data, subtype 5 which was associated with higher rates of pathological complete response following chemoradiation and

improved invasive locoregional relapse free survival. It also had the highest mean TMB of the 5 subtypes. To the best of my knowledge, I am not aware of any published work exploring directly the role of TMB as a biomarker of radiotherapy response.

6.3. Further areas of interest

6.3.1. Immunotherapy

Since commencing this project, immunotherapy has emerged as a new standard of care in MIBC. Unfortunately I was not able to incorporate this into my work but the CoMB resource has samples from multiple patients treated with immunotherapy, including tissue and sequential plasma samples from patients treated with concurrent hypofractionated radiotherapy and pembrolizumab⁸. Although this phase I trial was stopped due to toxicity, the plasma samples taken are a valuable resource and this would be an exciting area to explore.

6.3.2. Tumour microenvironment and tumour-infiltrating lymphocytes

The role of the tumour microenvironment, and in particular, the role of tumour infiltrating lymphocytes has also been a focus of recent interest, and it would be interesting to explore this further. In my RNA work, subtype 4 and 5 appeared to demonstrate higher expression of immune markers and this is in keeping with recent work suggesting tumours with an immune signature have better outcomes following trimodality treatment⁹. With regards to the tumour microenvironment, subtype 1 showed high expression of extracellular matrix genes and appeared to have poorer outcomes in my chemoradiotherapy cohort. Interestingly, there is data in rectal cancer to suggest that tumours with high stromal signatures display radioresistance¹⁰. This would be interesting to further explore.

6.3.3.Consensus molecular subtypes

In April 2020, Kamoun et al¹¹ published consensus subtypes in MIBC. By applying a network based approach previously used in colorectal cancer, they analysed 1750 RNA expression profiles and established the presence of 6 subtypes- luminal papillary, luminal nonspecified, luminal unstable, stroma-rich, basal squamous and neuroendocrine-like. They reported that in patients not receiving NAC, subtype was associated with prognosis, with median overall survivals ranging from 4 years in patients with luminal papillary tumours to 1 year in those with neuroendocrine like tumours. Of note, outcome did not appear to be associated with consensus class in patients who had received NAC.

I believe that these consensus classes may help to push forward research in this area as up to now, different groups have established different classifiers but the small numbers and retrospective design have limited progress and there has been a lack of independent validation, with no work so far in prospective trials. However, to establish the consensus subtypes, the authors merged 6 existing classifiers which have all been developed in cystectomy cohorts and from the data available, do not appear to include genes associated with DNA damage or radiosensitivity. I am therefore uncertain as to how applicable these consensus subtypes might be to a chemoradiation cohort as it was the addition of DDR and radiosensitivity genes to NMF71 that appeared to identify a potential radiosensitive subgroup. I therefore believe there is merit in continuing to explore NMF71. Of note, the authors have made their classifier available as a R package and it would be of great interest to apply this to my radiotherapy cohort to explore whether there might be any association with radiosensitivity or radioresistance.

6.4. The CoMB Protocol: Future opportunities

Since opening in August 2015, CoMB continues to collect samples and recruit subjects. In addition to the work reported here, CoMB has provided the necessary framework to support a feasibility study on urinary-derived lymphocytes in MIBC patients receiving radiotherapy, which was run by one of my colleagues (Dr Anna Wilkins) in collaboration with University College London following their success in this field with respect to immunotherapy¹².

I have identified a number of opportunities to continue the development of CoMB:

1. Urine collection

Although pre-defined in the protocol and approved by the ethics review committee, we did not have the capacity to collect and store urine samples from all patients. During the course of this MD(Res), work was been published looking at circulating DNA in the urine and as previously discussed, circulating DNA appears more readily detectable in urine of patients with localised disease^{13,14}. Going forward, it would be good for the CoMB Trial Management Group to re-discuss the logistics of urine collection. Certainly one area for future work would be to apply the iCHorCNA pipeline used to estimate tumour fraction from circulating plasma samples to urine, in order to explore whether this might be a useful approach in patients with localised disease where plasma tumour fraction was undetectable using this method.

2. Maintaining accrual and maximising time point capture

A further important area for discussion is that of recruitment and sample collection. A key feature of CoMB is the collection of sequential plasma samples. It was therefore frustrating to find, during the course of my work, that recruitment had slowed and samples were not being captured as intended from subjects recruited. The fluctuations in recruitment and sample collection illustrate the challenges in collecting this material and the critical importance

of close oversight to ensure optimal results. This was particularly evident at times of changes in staff.

To help support the collection of samples, the proposed amendment allows bloods to be taken for research purposes only (as opposed to alongside clinically indicated bloods) in the hope that this will help to improve the sample collection rate. CoMB would also however benefit from a named research fellow to support the research team and take forward future projects.

3. Recruiting patients at the Chelsea branch

At the time of setting up CoMB, I had hoped to recruit patients from the RMH Chelsea branch in addition to those at Sutton. The majority of surgical patients are seen at Chelsea and I felt this was an important cohort to capture, not only as it would be interesting to assess circulating tumour DNA (ctDNA) dynamics in the peri-operative period, but also to form a surgical cohort to complement the radiotherapy cohort. Any future biomarker candidates would then be able to be assessed in both cohorts to give some information regarding whether the biomarker might have a prognostic or a predictive role. Furthermore, there would be the potential to store fresh frozen samples at the time of cystectomy which would bypass the inherent issues, previously discussed, of using FFPE-derived nucleic acids. To date, a small cohort of patients have been recruited and from a subset of these, pre-operative and post-operative plasma samples were stored from the day of surgery, 24-48 hours post-surgery and 1-2 weeks post-surgery. Previous work looking at peri-operative ctDNA samples in colon cancer have shown the dramatic reduction in ctDNA levels in the post-operative period and the potential for ctDNA as a biomarker in the post-surgical follow-up period¹⁵. This would be an elegant study to reproduce in MIBC. Although some similar work has since been published in MIBC¹⁶, samples from a surgical cohort would provide the opportunity to apply our own tumour fraction estimation techniques in this setting.

Recruitment at RMH Fulham Road commenced in early 2016 but was abandoned after 2-3 months due to lack of resources in Chelsea to support it

and unfortunately it has not been possible to successfully reinstate recruitment and sample collection at RMH Fulham Road, primarily due to lack of resources.

4. Discovery of a methylation based signature for plasma MIBC DNA detection

iChorCNA has the advantage of not requiring *a priori* knowledge regarding aberrations present. However, this approach is limited by a lower limit of detection of approximately 8-9% (dependent on the number and magnitude of copy number aberrations present). I continue to collaborate with Prof Attard's team to facilitate a pilot study to discover an MIBC methylation signature in plasma DNA for high sensitivity detection and estimation of MIBC tumour DNA. This approach was successfully implemented in prostate cancer and uses bisulfite next-generation sequencing on plasma samples of variable tumour fraction, as I have collected and characterised, to discover segments across the genome that exhibit greatest methylation variance (unpublished data at the time of writing). This has led to a patent application and data from my samples could be included in a patent extension allowable prior to November 2020. The aim is to achieve superior sensitivity whilst maintaining the broad applicability and scalability of copy number profiling by LP-WGS. I would anticipate that this approach could be particularly applicable to bladder cancer, given the dominance of chromatin regulatory gene aberrations in MIBC.

6.5. Conclusion

When I embarked on this MD(Res), I sought to establish a protocol to allow the collection of biological samples and clinical data from patients with MIBC to support pilot studies exploring plasma circulating DNA and molecular subtypes in bladder cancer with a view to identifying potential biomarker candidates. Pleasingly, I have been able to apply a range of approaches to samples collected and my results have highlighted areas worthy of further investigation and generated new hypotheses to test. Perhaps most importantly, I have created a valuable resource for future work, and have also established workflows and collaborations with teams. This infra-structure is essential in such a fast-moving field, and translational work is key if we are to improve patient care. I hope that it will be possible to allocate more resources to support CoMB, so that it will remain a useful resource and also to ensure that all samples collected from patients who have so kindly agreed to participate can be used to their full potential. I look forward to seeing the future output from studies using CoMB samples, and in particular, whether my initial pilot results are reproducible in larger cohorts and are of relevance to clinical practice.

6.6. References

1. Abbosh, P.H. & Plimack, E.R. Molecular and Clinical Insights into the Role and Significance of Mutated DNA Repair Genes in Bladder Cancer. *Bladder Cancer* **4**, 9-18 (2018).
2. Plimack, E.R., *et al.* Defects in DNA Repair Genes Predict Response to Neoadjuvant Cisplatin-based Chemotherapy in Muscle-invasive Bladder Cancer. *Eur Urol* (2015).
3. Choi, W., *et al.* Intrinsic basal and luminal subtypes of muscle-invasive bladder cancer. *Nat Rev Urol* **11**, 400-410 (2014).
4. Yap, K.L., *et al.* Whole-exome sequencing of muscle-invasive bladder cancer identifies recurrent mutations of UNC5C and prognostic importance of DNA repair gene mutations on survival. *Clin Cancer Res* **20**, 6605-6617 (2014).
5. Bellmunt, J., *et al.* Pembrolizumab as Second-Line Therapy for Advanced Urothelial Carcinoma. *N Engl J Med* **376**, 1015-1026 (2017).
6. Rosenberg, J.E., *et al.* Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet* **387**, 1909-1920 (2016).
7. Van Allen, E.M., *et al.* Somatic ERCC2 mutations correlate with cisplatin sensitivity in muscle-invasive urothelial carcinoma. *Cancer Discov* **4**, 1140-1153 (2014).
8. Tree, A.C., *et al.* Dose-limiting Urinary Toxicity With Pembrolizumab Combined With Weekly Hypofractionated Radiation Therapy in Bladder Cancer. *Int J Radiat Oncol Biol Phys* **101**, 1168-1171 (2018).
9. Efsthathiou, J.A., *et al.* Impact of Immune and Stromal Infiltration on Outcomes Following Bladder-Sparing Trimodality Therapy for Muscle-Invasive Bladder Cancer. *Eur Urol* **76**, 59-68 (2019).
10. Isella, C., *et al.* Stromal contribution to the colorectal transcriptome. *Nat Genet* **47**(4), 312-319 (2015)
11. Kamoun, A., *et al.* A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. *Eur Urol* **77**(4), 420-433 (2020)
12. Wong, Y.N.S., *et al.* Urine-derived lymphocytes as a non-invasive measure of the bladder tumor immune microenvironment. *J Exp Med* **215**, 2748-2759 (2018).
13. Birkenkamp-Demtroder, K., *et al.* Genomic Alterations in Liquid Biopsies from Patients with Bladder Cancer. *Eur Urol* **70**, 75-82 (2016).
14. Patel, K.M., *et al.* Association Of Plasma And Urinary Mutant DNA With Clinical Outcomes In Muscle Invasive Bladder Cancer. *Sci Rep* **7**, 5554 (2017).
15. Diehl, F., *et al.* Circulating mutant DNA to assess tumor dynamics. *Nat Med* **14**, 985-990 (2008).
16. Christensen, E., *et al.* Early Detection of Metastatic Relapse and Monitoring of Therapeutic Efficacy by Ultra-Deep Sequencing of Plasma Cell-Free DNA in Patients With Urothelial Bladder Carcinoma. *J Clin Oncol* **37**, 1547-1557 (2019).