Circulating Biomarkers as Prognostic and Predictive Markers in Rectal Cancer

Panayiotis Alexander Georgiades

Submitted in accordance with the requirements for the degree of Doctor of Philosophy (PhD)

Cardiff University

2019

Summary

Standard treatment for patients with locally advanced rectal cancer (LARC) is currently preoperative radiotherapy with or without concurrent chemotherapy followed by total mesorectal excision. Approximately 10-20% of patients who receive preoperative therapy currently achieve a complete pathological or clinical response to therapy (pCR and cCR respectively). These patients have been demonstrated to have improved long-term outcomes, such as disease-free survival. At present, there are no methods available to reliably predict which patients will achieve pCR or cCR before surgical intervention or clinical examination respectively. This thesis aims to explore the technical aspects relating to a range of circulating biomarkers that might be used to facilitate this understanding for future evaluation in larger data sets.

As part of this thesis, we developed an assay for the extraction and analysis of exosome-derived microRNA (exoRNA) in patients with LARC. Using this assay, we detected variable levels of Mir-31, Mir-99a* and Mir-125b in longitudinal plasma samples. No significant associations were observed between microRNA levels and patient clinical outcomes.

Circulating tumour DNA (ctDNA) detection in longitudinal plasma samples was lower than expected. The low rate of detection seen in patients with LARC may have been due to technical limitations. Alternatively, this may be indicative of limited ctDNA shedding in this cohort, bringing into question the potential utility of this biomarker in these patients as a future routine test.

We also investigated the ability of immune and derived systemic inflammatory ratios to predict patient response to therapy in an expanded cohort of 235 patients with LARC. Again, few significant findings were observed. Overall, our findings suggest that the use of these circulating biomarkers may have limited clinical efficacy in patients with LARC.

DECLARATIONS

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed (candidate)

Date

STATEMENT 2

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is it being submitted concurrently for any other degree or award (outside of any formal collaboration agreement between the University and a partner organisation).

Signed (candidate)

Date

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available on the University's Open Access Repository (or, where approved, to be available in the University's library and for inter-library loan), and for the title and summary to be made available to outside organisations, subject to the expiry of a University-approved bar on access if applicable

Signed (candidate)

Date

Declaration

This thesis is the result of my own independent work, except where otherwise states, and the views expressed are my own. Other sources are acknowledges by explicit references. The thesis has not been edited by a third party beyond what is permitted by Cardiff University's Use of Third Party Editors by Research Degreee Students Procedure

Signed

Date

Word Count: 61,557
(Excluding summary, acknowledgements, declarations, contents page, appendices, tables, diagrams and figures, references, bibliography, footnotes and endnotes).

Acknowledgments

I would like to thank the funders of this research project Velindre Charitable Funds. I would also like to thank the Wales Cancer Bank and our colleagues in the Leeds Institute of Cancer and Pathology, Dr Daniel Bottomley and Professor Nick West, for their assistance in acquiring patient samples for this project.

I am grateful for the help I received from my fellow researchers throughout my time at Cardiff University. Daniel Nelmes, Matt Summers, Elena Meuser, Zoe Hudson, Magda Meissner, Marc Naven and Gareth Marlow all helped me in and out of the laboratory, through team-work, discussion and regular coffee breaks.

I would also like to thank the technical teams at the All Wales Medical Genetics Laboratory for their help, training and guidance in the laboratory. I also thank my supervisors, Professors Richard Adams, Rachel Butler and Simon Reed for their guidance, support and advice.

Finally, I would like to thank my wife for all her support throughout this difficult time. She has been my rock and my motivation throughout this difficult time, and without her this thesis would not be.

Table of Contents

Contents

List of Figures

List of Tables

Commonly used Abbreviations

ACMG – American College of Medical Genetics
AMP – Association for Molecular Pathologists
APR – Abdominal Perineal Resection
ASCO – American Society of Clinical Oncology
AWMGS – All Wales Medical Genetics Service
BCA – Bicinchoninic Acid
BSA – Bovine Serum Albumin
CA19-9 – Carbohydrate Antigen 19-9
CADD – Combined Annotation Dependent Depletion
CAF – Cancer-Associated Fibroblasts
CAP – Council of American Pathologists
cCR – Complete Clinical Response
cDNA – Complementary DNA
CDX – CTC-Derived Xenograft
CEA – Carcinoembryonic Antigen
cfDNA – Cell free DNA
CHPv2 – Cancer Hotspot Panel version 2
CIMP – CpG Island Methylator Phenotype
CIN – Chromosomal Instability
CMS – Consensus Molecular Subtypes
COSMIC – Catalogue of Somatic Mutations in Cancer
CRC - Colorectal Cancer
CRCSC – Colorectal Cancer Subtyping Consortium
CRIS - Colorectal Cancer Intrinsic Subtypes
CT – Computerized Tomography
CTC – Circulating Tumour Cells
ctDNA – Circulating Tumour DNA
CTL – Cytotoxic T-Lymphocytes
DAMPs – Damage Associated Molecular Patterns
ddPCR – Droplet Digital Polymerase Chain Reaction
DFS – Disease Free Survival
dMMR – Mismatch Repair Deficient
dNLR – Derived Neutrophil-to-Lypmphocyte Ratio
DRE – Digital Rectal Examination
dsDNA – Double Stranded DNA
EGFR – Epidermal Growth Factor Receptor
ELAPE - Extralevator Abdomino Perineal Excision
EMT – Epithelial to Mesenchymal Transition
EMVI – Extramural Vascular Invasion
ERUS – Endorectal Ultrasound
ESCRT – Endosomal Sorting Complexes Required for Transport
ESMO – European Society of Medical Oncology
ESP – Exome Sequencing Project
exoDNA – Exosome-Derived DNA
exoRNA – Exosome-Derived RNA
FDA – Food and Drug Association
FFPE – Formalin Fixed Paraffin Embedded
IGV – Integrated Genomics Viewer
IHC – Immunohistochemistry

ILV – Interluminal Vesicles
ISEV – International Society of Extracellular Vesicles
LARC – Locally Advanced Rectal Cancer
LMR – Lymphocyte-to-Monocyte Ratio
LoH – Loss of Heterozygosity
MCR – Mutation Cluster Region
MMR – Mismatch Repair
MRI – Magnetic Resonance Imaging
mrTRG – MRI-based Tumour regression Grade
MSI – Microsatellite Instability
MSS – Microsatellite Stable
MVB – Multivesicular Bodies
NGS – Next Generation Sequencing
NICE – National Institute of health and Care Excellence
NK – Natural Killer
NLR – Neutrophil-to-Lymphocyte Ratio
NMR – Nuclear Magnetic Resonance
NSCLC – Non-Small Cell Lung Cancer
NTA – Nanosight Tracking Analysis
OS – Overall Survival
PARP – Poly-ADP Ribose Polymerase
pCR – Pathological Complete Response
PCR – Polymerase Chain Reaction
pCRT – Preoperative Chemoradiotherapy
PD-1 – Programmed Death 1
PDX – Patient Derived Xenograft
PFS – Progression Free Survival
PGM – Personal Genome Machine
PLR – Platelet-to-Lymphocyte Ratio
pMMR – Mismatch Repair Proficient
P/P – Particle to Protein Ratio
pTRG – Pathological Tumour Regression Grade
qPCR – Quantitative PCR
RISC – RNA-Induced Silencing Complex
SCLC – Small Cell Lung Cancer
SNP – Single Nucleotide Polymorphism
SNV – Single Nucleotide Variant
TAA – Tumour-Associated Antigen
TCGA – The Cancer Genome Atlas
TEI – Thermofisher Exosome Isolation
TGF-β – Tumour Growth Factor Beta
TIL – Tumour Infiltrating Lymphocytes
TEM – Transanal Endoscopic Microsurgery
TMB – Tumour Mutational Burden
TME – Total Mesorectal Excision
TRG – Tumour Regression Grade
UTR – Untranslated Region
VAF – Variant Allele Frequency
VCF – Variant Caller Format
WCB – Wales Cancer Bank
WCC – White Cell Count

1 Introduction

1.1 Incidence

Colorectal cancer (CRC) is the fourth most common cancer type in the UK, and third for both men and women. In 2016, approximately 42,000 new cases were diagnosed in the UK and the 10-year Overall Survival (OS) for patients with CRC is approximately 57% (CRUK 2016). Approximately 30% of CRCs are located in the rectum (Conde-Muino et al. 2015).

Rectal cancers are clinically defined as tumours which arise within 15cm of the anal margin, where more proximal tumours are colon cancers. Rectal tumours can be subcategorised based on location: as high (10-15cm from anal margin), middle (5-10cm) or low (<5cm) rectal cancer. There is growing evidence that distinct differences exist between colon and rectal cancers (Glynne-Jones et al. 2017). Patients with rectal cancer often have a poor prognosis due to high risk of the development of distant metastases and their anatomical position result in more complex surgical intervention and risks of local recurrence.

1.2 Differences Between Colon and Rectal Cancer

Although the two disorders are often combined under a single umbrella (colorectal or bowel cancer) numerous differences exist between colon and rectal cancers regarding development, commonly observed alterations, clinical management and prognosis.

Rectal tumours are define anatomically as growing in the rectum of a patient whilst colon cancers are located in the caecum, ascending, transverse, descending or sigmoid colon. Colon tumours which occur in the proximal to the splenic fixture (i.e. the ascending and transverse colons) are referred to as right-sided or proximal tumours, whilst tumours occurring distal to the splenic fixture (i.e. the descending and sigmoid colons) are referred to as left-sided or distal colon tumours (Li and Lai 2009).

The differences between colon and rectal tumours are not as superficial as location alone. There are also molecular differences between tumours which arise within the different regions of the colon as well as with the rectum. Approximately 24% of all proximal colon tumours were reported as having microsatellite instability (MSI), whilst only 2% of rectal tumours were MSI positive in comparison (Slattery et al. 2009) . Furthermore, studies have also reported that rectal tumours generally display greater involvement of *TP53*, the β-catenin pathway and COX2 overexpression, and lower frequencies of *KRAS* mutations (Dimberg et al. 1999; Kapiteijn

et al. 2001a; Li and Lai 2009) . These differences in the molecular pathogenesis of disease between the two anatomical locations are likely a result of the fact that proximal colon tissue develops embryogenically from the midgut, whilst the distal colon and rectum develop from the hindgut (Li and Lai 2009).

At this time, standard therapy for patients with LARC is neoadjuvant therapy followed by surgery (See section 1.3.2 for further details), whereas patients with colon cancer only receive adjuvant therapy. Surgery is performed in rectal cancer due to the relative ease of access anatomically, allowing a less invasive surgery. Although, treatment is currently similar, historically survival and relapse rates have been poorer in patients with rectal cancer (Paschke et al. 2018).

1.3 Current Clinical Management of Patients with Locally Advanced Rectal Cancer

1.3.1 Diagnosis and Staging

The treatment and clinical management for patients with rectal cancer has some distinct differences to that of patients with colon cancer.

1.3.1.1 Current Methods of Clinical Staging

Patients with rectal cancer are usually diagnosed with an endoscopy, Digital Rectal Examination (DRE), and a tumour biopsy for histopathological confirmation (Glynne-Jones et al. 2017). After diagnosis, Computed Tomography (CT) imaging is used to determine the extent of disease. The pelvis, abdomen and chest are all scanned to help identify distant disease and Magnetic resonance imaging (MRI) of the pelvis is used to provide locoregional assessment including T stage, lymph-node involvement and vascular invasion (Trakarnsanga et al. 2012).

Patient tumours are staged as part of treatment decision-making and risk stratification using CT imaging, MRI and histopathology (Glynne-Jones et al. 2017). The current standard for tumour staging is based on tumour size, lymph node involvement and the presence of metastases (T/N/M). Tumours can be staged clinically by clinical examination and imaging methods (cT/cN/cM), or tumour size and lymph node involvement can be determined pathologically at the time of surgery (pT/pN). T/N/M staging can be grouped into pathological TNM stages (stages I-IV) according to Brierely and colleagues (2016) for determining patient treatment. Use of this patient system is currently recommended by the European Society for Medical Oncology (ESMO).

Within our study, we focus primarily on patients with Locally Advanced Rectal Cancer (LARC). LARC is clinically defined as pathological stage II/III or T3b or greater with the presence of extramural vascular invasion (EMVI+) (Glynne-Jones et al. 2017).

1.3.1.2 Future Methods of Clinical Staging

In recent years, there have been attempts to improve how patients with rectal cancer are staged in order to incorporate targeted therapies and molecular biomarkers available for risk stratification. The Colorectal Cancer Subtyping Consortium (CRCSC) investigated gene-expression data from 4151 tumours (Guinney et al. 2015) and have since suggested a novel taxonomy for CRC subtyping consisting of four Consensus Molecular Subgroups (CMS1-CMS4). In the 2,651 patients with known tumour site (i.e. left colon, right colon or rectum) only 15% had rectal cancer and these were integrated combined with patients with colon cancer without any adjustments. This was performed because this study had previously seen no clear differences between patients with colon and rectal cancer based on downloaded TCGA RNA-Seq data (Guinney et al. 2015). However, this data was not shown within this report.

In a similar vein, Bertotti and colleagues (2016) investigated cancer cell intrinsic transcriptional traits by performing gene expression profiling analysis on 515 patient derived xenograft (PDX) models from 244 patients with CRC; removing stromal content. Unsupervised clustering on the resulting profiles identified 5 CRC intrinsic subtypes (CRIS) defined as CRIS-A–E.

Both systems incorporate biomarkers such as CpG Island Methylator Phenotype (CIMP) and Microsatellite instability (MSI) status among other molecular markers. Both taxonomies may assist by further dividing chromosomal instability (CIN) patients into smaller subgroups (CMS2-4; CRIS-C-E); as they represent 85% of the patients and are very heterogeneous (Bertotti et al. 2016; Isella et al. 2017), each having specific driving mechanisms. However, studies have previously demonstrated that MSI-high and dMMR tumours are much rarer in patients with rectal cancer than colon cancer (Ostwal et al. 2019). This may limit the usefulness of these specific subgroups in patients with rectal cancer which should be considered when moving forward.

Applying uniform taxonomies can assist in defining a consistent framework for subtyping in future studies and clinical analyses. It has the potential to help overcome inconsistencies and contradictory results between separate reports, facilitating the investigation of the effects of treatment and biomarkers in a more standardised fashion.

Consistency is valuable in both research and the clinic, playing a pivotal role in predicting survival outcomes and stratifying treatment options. Recently, MSI-High/MMR-deficient (dMMR) patients with CRC were reported to have an improved response to immune checkpoint inhibitors than their MMR-proficient counterparts (Fujiyoshi et al. 2017). As a result, the FDA have approved both Pembrolizumab and Nivolumab as third line therapies in MSI-High/MMR-deficient CRC patients (Research 2018a,b) highlighting the importance of consistent subtyping.

Although these subtyping systems are applied to CRC, further investigation specifically into patients with rectal cancer may facilitate the prediction of patient response to standard therapy.

1.3.2 Patient Treatment

For patients with LARC, treatment is centred around the completion of surgery with curative intent (Glynne-Jones et al. 2017). In order to facilitate curative surgery, pre-operative radiotherapy, with or without concurrent chemotherapy, is generally administered. Successful therapy aims to downsize and downstage patient tumours to enable curative surgery whilst also attempting to allow sphincter-saving procedures to improve patient quality of life (Sauer et al. 2004).

1.3.2.1 Surgery

The primary aims of surgery are to achieve local disease control (i.e. reduce local recurrence) and improve long-term survival (Glynne-Jones et al. 2017). Surgical methods can vary from local excision to radical resection, based on patient stage and response to neoadjuvant therapy. These different methods can have a variable benefit on survival and quality of life through sphincter preservation (Smith and Garcia-Aguilar 2015). Currently, available surgical methods include trans anal endoscopic microsurgery (TEM) and total mesorectal excision (TME).

TEMs are a local excision surgery (without lymph node resection) which are appropriate for very small and early tumours T1 or less (Glynne-Jones et al. 2017). These are normally performed as a single modality treatment, without the need of pre- or post-operative adjuvant therapy.

TME has become the standard surgical technique for treating patients with rectal cancer. Patients with LARC are generally excised using this method, rather than TEM, due to the increased risk of local recurrence and local lymph-node involvement in these patients. A TME

is performed to excise and remove mesorectal fat and lymph-nodes surrounding the primary tumour (Slater et al. 2016).

There are also more complex surgeries, such as the Abdominoperineal Resection (APR) or Extra-Levator Abdominoperineal excision (ELAPE) for low rectal tumours which result in a permanent stoma and perineal wound with complete removal of the anal sphincter complex (Shen et al. 2015).

1.3.2.2 Neoadjuvant Therapy

1.3.2.2.1 Radiotherapy

In the 1990s, clinical trials demonstrated that pre-operative radiotherapy followed by surgery provided improved local control and survival in patients with LARC in comparison to patients who only received surgery (Cedermark et al. 1997). Similarly, a recent systematic-review has demonstrated that overall mortality at 4-12 years was reduced in patients who received pre-operative radiotherapy compared to those who were treated with surgery alone (Abraha et al. 2018). Furthermore, local control was significantly improved with pre-operative radiotherapy. However, disease-specific mortality remained fairly consistent between the two treatment regimens (Abraha et al. 2018). This study also suggested that patient who received TME did not receive a significant benefit in overall survival from preoperative short course radiotherapy with immediate surgery (Abraha et al. 2018). Although patients do appear to benefit from neoadjuvant treatment, this improvement is likely to be bolstered by recent improvements in surgical techniques.

Radiotherapy can be administered as pre-operative long- or short-course therapy over five weeks or one week respectively (Glynne-Jones et al. 2017). Long-course radiotherapy exposes patients to low doses (1.8-2Gy per dose/fraction) over a 5-week period whilst short-course radiotherapy involves much higher levels (5Gy per dose/fraction) over a shorter period (1 week).

Studies have reported that no differences were seen in local control or survival between patients who received short vs long course preoperative radiotherapy (Erlandsson et al. 2017; Wang et al. 2018), whilst others have found that long-course radiotherapy was associated with improved OS and less toxicity in patients with LARC (Margalit et al. 2019). In contrast to this, Erlandsson and colleagues (2019) reported that short-course radiotherapy with delayed time to surgery was associated with an improved rate of pathological complete response (pCR) in comparison to long-course radiotherapy with delay. At present, there are no strict guidelines

for whether to use short- or long-course radiotherapy with or without chemotherapy for patients with LARC. Treatment regimens may be decided according to factors such as patient frailty or co-morbidities.

Clinical studies have also attempted to improve patient outcomes through the addition of chemotherapy or targeted therapies to concurrent radiotherapy.

1.3.2.2.2 Chemotherapy

Pre-operative radiotherapy may be administered with concurrent chemotherapy, which includes fluorourpyrimidines as standard. 5-FU is a fluorourpyrimidine, which is commonly used alongside radiotherapy pre-operatively in patients with LARC due to the radiosensitizing nature of 5-FU in tumour cells. 5-FU has been demonstrated to improve tumour response to radiotherapy in patients with LARC (Crane et al. 2003) but does not impact tumour growth in un-targeted locoregional lymph-nodes or distant metastases. Capecitabine (also known as Xeloda) is an orally ingested fluoropyrimidine which is metabolised into 5-FU within the tumour tissue and cells by thymidine phosphorylase (Pentheroudakis and Twelves 2002). An improvement was initially sought to combat the limitations of 5-FU administration, such as the drug's limited plasma half-life and the difficulties associated with regular intra-venous administration (Pentheroudakis and Twelves 2002). These issues were potentially solved using an oral pro-drug substitute.

Both 5-FU and Capecitabine have similar mechanisms for sensitizing tumour cells to radiotherapy through inhibiting DNA synthesis. Fluorouracils inhibit thymidine synthase and block the production of thymidylate, blocking downstream cellular DNA synthesis. Furthermore, alternative metabolites of 5-FU (FUTP) are incorporated into RNA in place of UTP. This interferes with RNA synthesis and downstream translation. The activity of thymidine phosphorylase, which is a key enzyme in the metabolism of capecitabine metabolites into 5-FU, was seen to be three times higher in tumour cells than adjacent tissue (Pentheroudakis and Twelves 2002). This explains the low levels of toxicity and adverse experience observed in patients undergoing capecitabine therapy over intravenous 5-FU treatment.

Irinotecan has also been used concurrently preoperatively with fluorouracils in patients with LARC. Irinotecan produces cytotoxic affects through inhibiting topoisomerase I. Topoisomerase I is a key enzyme in DNA replication and repair (Xu and Her 2015) this has been further explored in the UK ARISTOTLE trial- with results awaited (ISRCTN09351447).

1.3.2.2.2.1 The ARISTOTLE Clinical Trial

The ARISTOTLE clinical trial is currently investigating the potential benefit of adding Irinotecan to standard preoperative chemoradiotherapy in patients with LARC. The clinical trial has two separate arms (A and B), in both of which patients are treated with radiotherapy (45Gy in 25 fractions over five weeks) with concurrent capecitabine treatment (900mg/m^2 or 650mg/m^2 orally twice daily in arms A and B respectively). Patients in Arm B also received Irinotecan intravenously (60mg/m^2) once per week during weeks 1-4 of treatment. The primary outcome measured in this trial was 3-year disease-free survival after the completion of therapy. This is illustrated below in Figure 1.

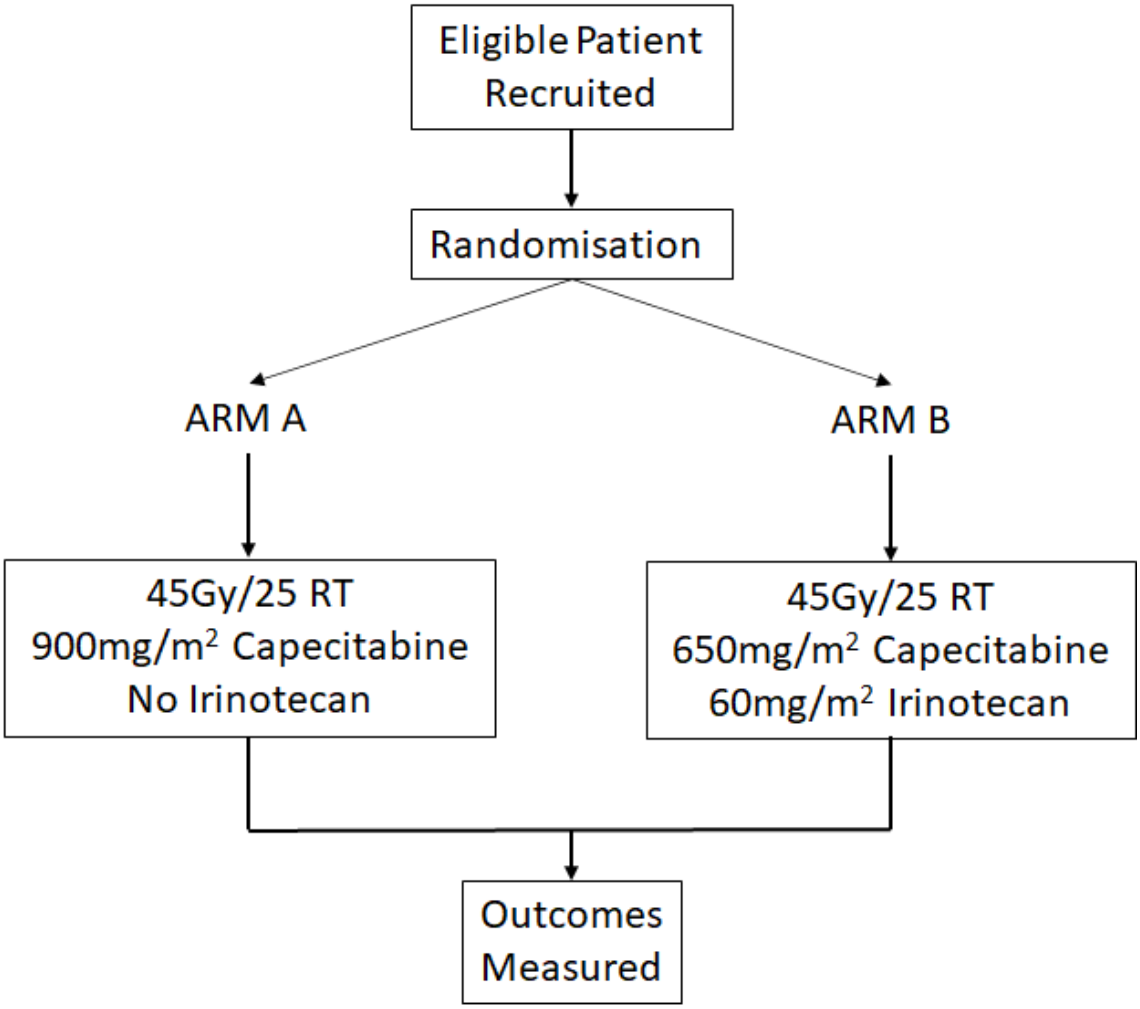

Figure 1: Illustration of the overall process and two treatment arms in the ARISTOTLE clinical trial.

There are numerous other cytotoxic agents and families used in the treatment of patients with rectal cancer; such as Oxaliplatin and Leucovorin, among others. Oxaliplatin is a platinum-based agent from the same family as cisplatin. Oxaliplatin binds to guanine and cytosine in

DNA, inhibiting DNA synthesis and transcription. Leucovin is derivative of folic acid and facilitates the metabolism of 5-FU downstream by stabilising the bond of 5-FUMP to thymidine synthetase. Therefore, it has been used to enhance the cytotoxic effects of Fluorouracils.

Studies have demonstrated that the addition of concurrent chemotherapy to preoperative radiotherapy improved pathological response and local control in patients with LARC (Ceelen et al. 2009; De Caluwe et al. 2013). However, these studies also demonstrated that adding chemotherapy did not improve disease free survival (DFS) or OS in patients with stage II/III LARC (Ceelen et al. 2009; De Caluwe et al. 2013).

In addition to the standard treatment described previously, other targeted therapies can be administered or are under current research for future utility. Among these are anti-Epidermal Growth Factor Receptor (EGFR) monoclonal antibodies (Cetuximab and Panitumumab) and anti-Programmed Cell Death-1 (PD-1) antibodies (Pembrolizumab, Durvalumab etc.).

1.3.2.3 Anti-EGFR Monoclonal Antibodies

Anti-EGFR antibodies are only used for treating patients with metastatic CRC. Anti-EGFR monoclonal antibodies, including cetuximab and panitumumab, bind to the extracellular domain of the EGFR protein. This stops the activation of EGFR by blocking its activating binding partners, such as EGF. However, variants in the *EGFR* gene or in genes downstream of *EGFR,* such as *KRAS, NRAS, BRAF or PIK3CA*, can circumvent the need for EGFR activation. Therefore, these monoclonal antibodies are only used to treat patients with wild-type *KRAS, NRAS* and *BRAF*.

1.3.2.4 Immune Checkpoint Inhibitors

Immune checkpoint inhibitors are still in the very early stages of trials in patients with rectal cancer in comparison to more standardised therapies. This family of cytotoxic agents enhances the patient's immune response to tumour cells. They perform this action by blocking the binding of immune checkpoints with corresponding receptors on T-cells. Immune checkpoints (such as programmed cell death-1; PD-1) are expressed on normal tissue to prevent auto-immune activity. However, these receptors are co-opted by tumour cells and inhibit or 'exhaust' the activity of T-cells by disrupting the recognition of non-self-cells and blocking the immune synapse which stimulate T-cell function (Ramsay 2013).

Immune checkpoint inhibitors have shown improved therapeutic efficacy in patients with dMMR tumours (Cohen et al. 2019), but these changes are very rare in patients with rectal cancers (Ostwal et al. 2019). Trials have also commenced in the combination of anti-PD-1

therapy with radiotherapy to enhance the immunogenic activity of such immunotherapies (Overman et al. 2017).

In addition to the treatments described there are other targeted therapies available, such as anti-angiogenic monoclonal antibodies (bevacizumab or cediranib) and Poly-ADP Ribose Polymerase (PARP) inhibitors (veliparib). PARP inhibitors have recently been explored in an attempt to enhance tumour response to radiotherapy in patients with rectal cancer (Shelton et al. 2013). There are also other approaches such as the 'watch and wait' or "active monitoring" approach described below.

1.3.2.5 The 'Watch and Wait' or Active Monitoring Approach

An alternative option to post neoadjuvant surgery which has recently been suggested is the 'watch and wait' approach. In this case, patients who achieved a complete clinical response (cCR) did not initially receive any surgery and are monitored closely instead. cCR is assessed both clinically and using medical imaging methods. cCR is defined by ESMO as the absence of any palpable tumour during DRE, no visible tumour lesions at rectoscopy and the absence of detectable tumour at primary site or in drained lymph nodes using imaging techniques such as MRI or endorectal ultrasound (ERUS).

cCR and pCR have been previously demonstrated to have poor concordance in patients with breast cancer (Smith et al. 2012) as pCR is defined by the lack of detectable tumour cells at the time of surgery.

The watch and wait approach has been suggested in order to avoid the adverse effects of surgery, such as the risk of perioperative complications (Marijnen 2015) and sexual dysfunction (Den Oudsten et al. 2012). Studies have been investigating the effect of surgery vs watch and wait on patient outcomes. Renehan and colleagues (2016) have previously reported that the watch and wait approach had similar DFS as surgery at 3 years and was, thus, oncologically safe. However, there has been wide variability in the local recurrence rate of patients treated with the watch and wait approach (5-60%) (Renehan et al. 2016). This wide variability may be due to numerous differences between these studies, such as follow-up time, and surveillance intensity.

Therefore, as investigations continue to analyse the oncological safety of this approach, this has not yet been widely implemented in the treatment of patients with LARC.

Although there are numerous options for the treatment of patients with LARC, the current standard remains neoadjuvant chemoradiotherapy followed by surgery. How this will develop with the emergence and development of immunotherapies remains to be seen.

1.3.3 Follow-Up and Surveillance

To date, there are no current biomarkers which can predict the efficacy of treatment or long-term clinical outcomes in patients with LARC undergoing preoperative chemoradiotherapy (pCRT). For the moment, close surveillance by clinical examinations, colonoscopies and regular CT or MRI imaging is applied to detect tumour recurrence in these patients. Biomarkers, circulating or otherwise, need to be developed in this group of patients to identify those at greater or lesser risk of local recurrence and distant metastases down the line.

1.3.4 Clinical Outcomes

1.3.4.1 Response and Survival

Currently, patient response to therapy can be gauged by MRI and/or at the time of surgery. Post-operative levels of carcinoembryonic antigen (CEA) may also be indicative of tumour burden or patient response to therapy.

After neoadjuvant treatment, approximately 10-20% of patients experience pCR where the tumour has been eliminated by therapy (Maas et al. 2010). The remaining patients can experience variable outcomes, from a good response with large amounts of regression to minimal or no response.

Patients with LARC who achieve a pCR may have a decreased likelihood of experiencing local recurrence but are still at risk of distant metastases (Sun et al. 2019). Therefore, patients achieving a complete pathological response may not necessarily receive any distinct benefits with regards to OS.

When patients with rectal cancer have localised disease and surgical resection is curative, survival is still relatively poor, with 10-year survival being ~50% (van Gijn et al. 2011). Disseminated disease is the most frequent cause of mortality in patients with rectal cancer, but local recurrence can also have severe consequences which may be fatal.

Current treatment in patients with LARC focus on the primary tumour lesion and locoregional lymph-nodes, not on distant lymph-nodes or metastases. Therefore, although local control may be achieved, the growth of distant micro-metastases can remain undetected and

unchecked until clinical symptoms arise or radiological imaging identifies distant metastases during patient post-operative follow-up.

1.4 Predictors of Long-Term Clinical Outcomes in Patients with Locally Advanced Rectal Cancer

There are a limited number of biomarkers which can facilitate treatment decision-making by prognosticating or predicting long- or short-term clinical outcomes in patients with LARC. At present, only tumour response at time of surgery and post-treatment circulating CEA levels are used clinically. Here, we review surgical, tissue-based and circulating biomarkers, expanding from what is currently available in the clinic to what is of growing interest in the literature in patients with CRC, focussing on rectal cancer where possible.

1.4.1 Surgical/Pathological Markers

1.4.1.1 Surgical Outcomes

After the completion of surgery, the response of the tumour to neoadjuvant therapy is quantified, usually applying a tumour regression grade (TRG) system. TRG systems quantify the response of neoadjuvant therapy from a pCR to no response or tumour progression. Various grading systems are available, including the Dworak's and Rodel's TRG systems (Trakarnsanga et al. 2012). Although the tumour growth/shrinkage occurs in these patients during neoadjuvant therapy, changes can be measured more accurately at the time of surgery.

The detection of pCR after neoadjuvant therapy can be suggestive of improved local control and disease-free survival (Molinari et al. 2015; Bottarelli et al. 2018). Studies have shown an improved DFS in patients who achieved pCR compared to those who did not (Rodel et al. 2005). TRG has been demonstrated to be a prognostic factor for the occurrence of distant metastases (Fokas et al. 2014). However, poor response to therapy may be indicative of tumour aggressiveness and advanced tumour stage rather than merely indicative of the occurrence of distant metastases.

Patients who achieved a complete response to neoadjuvant therapy could also be identified using radiological imaging and clinical examination if surgery is not to be undertaken owing to patient wishes or comorbidities/frailty. This is defined as a cCR but is a macroscopic evaluation only and is not as accurate without histopathological confirmation, requiring surgical resection (Glynne-Jones et al. 2017).

At this time, there are no methods which can accurately predict which patients will achieve pCR or a good response/TRG and those who will not before the initiation of therapy. Such a predictive biomarker would facilitate treatment-decision making for patients on an almost individual basis. The ability to predict which patients will or will not respond to neoadjuvant therapy could help improve patient treatment stratification and improve patient quality of life by limiting their exposure to toxic agents unnecessarily or preventing the need for surgery.

1.4.2 Tissue-Based Biomarkers

1.4.2.1 Tumour Infiltrating Lymphocytes

In recent years, our understanding of the role of immunity in tumour growth has grown significantly. This has progressed research towards the clinical implementation of immunotherapies in patients with various cancer-types. Immune-related biomarkers have begun to develop for prognostic and response prediction purposes in patients treated with immunotherapies. These can be either tissue-based, such as tumour-infiltrating lymphocyte (TIL) density levels, or circulating, such as white cell counts (WCCs), lymphocyte sub-classes, neutrophil (NLR), monocyte (LMR) or platelet (PLR) to lymphocyte ratios.

With regards to TILs, studies have primarily focussed on infiltrating T-lymphocytes in patients with various solid tumours. T-cells are CD3+ lymphocytes which can be divided into two subsets in TIL analysis; these being CD8+ cytotoxic T-lymphocytes (CTLs) and CD4+ regulatory T-cells (Ramsay 2013). Each of these two subsets have been shown to perform different roles with regards to patient tumour cells. CTLs perform an anti-tumour function by targeting tumour cells, via tumour specific antigens, for destruction. In contrast, CD4+ T-cells regulate the activity of CTLs and limit their activity to a more pro-tumour function, especially FoxP3+ regulatory T-cells. Consequently, recent studies have investigated different T-cell subsets in TIL analysis, as opposed to T-cells as a whole.

Previous reports have demonstrated that high densities of pre-treatment CD8+ T-cells (within the tumour micro-environment) were associated with improved response (Yasuda et al. 2011; Matsutani et al. 2018) and survival (Shinto et al. 2014; Teng et al. 2015) after neoadjuvant therapy in patients with rectal cancer. In agreement with this, a recent meta-analysis reported that high densities of CD3+, CD8+ and FoxP3+ were associated with improved OS in patients in CRC (Kong et al. 2019). In patients with LARC, high CD8+ TIL density, were predictive of improved tumour regression after chemoradiotherapy (Kong et al. 2019) and greater sensitivity to neoadjuvant therapy (Matsutani et al. 2018).

The influence of CD4+ density is more controversial in patients with rectal cancer. Whilst some studies have shown that high density CD4+ TILs are associated with improved response to neoadjuvant therapy (Yasuda et al. 2011), other studies have shown contradictory findings (Huang et al. 2019). Studies have also reported that general counts of CD8+ or FoxP3+ T-cells were not a significant prognostic indicator of outcomes, but changes in the CD8+/FoxP3+ ratio as a result of therapy were associated with OS and Progression Free Survival (PFS) (Mirjolet et al. 2018).

It has also been reported that the location of lymphocyte infiltration, either into the malignant epithelium or the surrounding stroma, is another important feature to consider in patients with LARC. Huang and colleagues (2019) demonstrated that intraepithelial CD45+ and CD8+ TILs were associated with an improved response to neoadjuvant therapy in patients with LARC, whereas stromal TILs had no significant association.

Tumours with a high density of pre-treatment lymphocyte infiltration are seen to be more immunogenic and attract anti-tumour CD8+ CTLs (Yasuda et al. 2011). Radiotherapy directly causes immunogenic cell death to cancer cells which causes a release of Tumour Associated Antigens (TAAs) and Damage-Associated-Molecular Patterns (DAMPs). Released TTAs, DAMPs and cytokines then stimulate anti-tumour immunity and CTL maturation (Wennerberg et al. 2017). However, TILs may be inactive or 'exhausted' due to the presence of expressed immune checkpoints, such as PD-1 or CTLA-4 by tumour cells and/or regulatory T-cells (Ramsay 2013). Therefore, the addition of immune checkpoint inhibitors, (such as pembrolizumab of nivolumab) may integrate well with pre-operative chemoradiotherapy.

The reporting of contradictory findings in the literature may be due to the presence and activity of CD4+ T-cell subsets. Although several studies used CD4+ as a marker for regulatory T-cells, the T-cells can be divided into subgroups with pro- or anti-tumour functional variability. FoxP3 is a specific marker for a subgroup of CD3+ T-cells, thus CD4+ and FoxP3+ TIL density are not necessarily the same. Furthermore, studies also vary in methodology, as not all studies were quantitative or semi-quantitative in the IHC analysis. Hence, direct comparisons may be misleading.

There has been interest in the infiltration of Natural Killer (NK) cells in patients with rectal cancer. Lim and colleagues (2014) reported that NK cells were significantly lower in density in rectal cancer tissue in comparison to normal tissue. High NK cell TIL density was associated with improved progression but not survival (Lim et al. 2014). In agreement with these findings,

Alderdice and colleagues (Alderdice et al. 2017) reported that a high expression NK cell signature in the gene expression profiling of patients with LARC was indicative of improved TRG.

With regards to T-cells, these findings enabled the development of the immunoscore and the initiation of clinical trials testing the efficacy of immunoscores in predicting high risk patients with colon cancer (NCT03422601). Immunoscores are acquired from tumour tissue using ImmunoHistoChemistry (IHC) to evaluate the density of CD3+ and CD8+ lymphocyte populations in the tumour margin and tumour core (Zeitoun et al. 2019). This is not currently in place for patients with rectal cancer but is certainly a biomarker of interest to monitor in clinical trials for future potential utility.

1.4.2.2 Protein Expression

Studies have previously investigated the levels of apoptotic and DNA damage repair proteins in rectal cancer tissue as possible predictors of patient response to neoadjuvant therapy. This is because chemoradiotherapy induces cell death through large scale double stranded DNA damage. The ability of a cell to repair DNA damage and resist apoptosis has previously been linked to response to ionising radiation. However, these studies have only achieved limited success in identifying proteins which can reliably predict patient response to neoadjuvant therapy.

Studies have previously investigated the p53 protein before or after radiotherapy to identify any associations between expression and response to therapy. However, these studies have general found no significant association to clinical outcomes in patients with LARC (Kobayashi et al. 2007; Terzi et al. 2008). In contrast, the proteins Ki-67 (Jakob et al. 2008) and Survivin (Yu et al. 2016), involved in proliferation and inhibiting apoptosis respectively, have both been found to significantly associate with response and recurrence rates respectively. Similar studies have seen no significant associations between these proteins and response or prognosis (Terzi et al. 2008).

Studies have also reported significant associations between clinical outcomes and proteins such as COX-2 (Kobayashi et al. 2007; Peng et al. 2016), APAF-1 (Peng et al. 2016) and β-catenin (Gomez-Millan et al. 2017) in patients with LARC. These are generally small studies ($n<100$) and will require further validation in larger cohorts. Furthermore, these studies have large differences between each of them, creating difficulties in identifying clinically significant proteins in patients with LARC. These differences can vary from treatment administered or the

stratification of patients as good or poor responders. Furthermore, these studies use IHC for the quantification of expression, which is very subjective and difficult to remove bias from.

These studies do not include genetic analysis; the presence of tumour specific variants in each of these genes may impact the importance of protein expression in patients with LARC. A missense variant in the DNA-binding domain of *TP53* may limit the functional capabilities of the p53 protein, but not alter protein expression.

Although these studies have provided some interesting findings, the quantification of protein expression by IHC is still limited. The coupling of this with genetic analysis and/or gene expression profiling may be beneficial in future studies.

1.4.2.3 Genomic Alterations

At present, the only genomic variants available to predict patient outcome to therapy are the presence of variants in *KRAS* and *NRAS* in the context of anti-EGFR monoclonal antibody therapy (Hsu et al. 2016). In the context of neoadjuvant therapy and surgery, no genomic alterations can predict patient response or long-term clinical outcomes.

As previously alluded to, studies have recently suggested that tumour specific genomic variants may influence tumour response to neoadjuvant therapy in patients with LARC. Many of these studies have focussed on the predictive or prognostic roles of *KRAS* and *TP53* among other frequently mutated genes in rectal cancer. However, these reports have lacked consistency.

In patients with rectal cancer, the presence of variants in *KRAS* have been previously reported to associate with poorer response to neoadjuvant therapy and a decreased likelihood of patients achieving pCR (Duldulao et al. 2013). In contrast, numerous studies have shown no difference in response to neoadjuvant therapy between patients with *KRAS* mutant or wild-type disease (Sclafani et al. 2019). As far as we are aware, no studies reported an association between *KRAS* status and survival in patients with rectal cancer.

Interestingly, a study which showed no significant differences in response to therapy between patients with wild-type and mutant *KRAS,* reported small differences in response between two different variants in *KRAS* (p.G13D and p.G12V) (Gaedcke et al. 2010). Similarly, other studies have reported that patients with a *KRAS* variant at p.G13 were less likely to experience tumour downstaging or pCR after therapy; even though both studies reported that *KRAS* status was not significantly associated with clinical outcome (Lee et al. 2015; Martellucci et al. 2015). The prognostic or predictive differences between different variants in *KRAS* may be worth

consideration in future studies. At present, the influence that mutant *KRAS* may impose on response to neoadjuvant therapy in not clear.

Variants in *TP53* have been reported to correlate with worse response after therapy (Sclafani et al. 2019) and decreased rates of pCR (Chen et al. 2012). Sakai and colleagues (Sakai et al. 2014) have reported that variants in *TP53* were enriched in tumour tissue after the completion of chemoradiotherapy. Based on this finding, Sakai and colleagues hypothesised that the administration of therapy created an evolutionary advantage of mutant *TP53* tumour cells to grow within. Due to the role of p53 in DNA repair and apoptosis, this hypothesis is feasible and in line with reports demonstrating the poor prognostic significance of *TP53* variants.

Interestingly, numerous studies have reported that the presence of variants in both *KRAS* and *TP53* are indicative of worse PFS (Sclafani et al. 2019) and decreased likelihood of achieving pCR (Chow et al. 2016). Larger scale studies are required to determine whether this predictive value is true and significant in a wider cohort of patients with rectal cancer.

1.4.2.4 Micro Satellite Instability and DNA Mismatch Repair Status

In patients with CRC, ~15% of patients have dMMR functions within their tumour. Consequently, this creates MSI, where intergenic microsatellites are found at highly variable and unregulated lengths throughout tumour tissue. Patients with CRC can have MSI- High (MSI-H), -Low (MSI-L) or microsatellite stable (MSS) disease. In the clinical context, MSI is analysed as a marker of dMMR. In patients with CRC, dMMR is a good prognostic indicator and associated with improved OS and DFS after neoadjuvant therapy (Hong et al. 2012).

In patients with rectal cancer, dMMR status is considerably rarer (~1-2%) and the impact that MMR status can have on response to DNA damaging agents is currently not understood (Ostwal et al. 2019). Studies have reported that patients with MSI rectal cancer had a reduced rate of pCR after pCRT (Hasan et al. 2018) and high levels of MSH6 protein was indicative of improved OS and local control (Huh et al. 2016). This would suggest that proficient MMR (pMMR) function improved patient outcomes after treatment, which is contrary to reports in patients with colon cancer (Hong et al. 2012). However, previous reports have demonstrated that patients with MSI rectal cancer had an improved DFS and OS after neoadjuvant therapy, which contradicts other studies (Colombino et al. 2002; Du et al. 2013). As previously stated, with regards to neoadjuvant therapy, the current effects of MMR status remain unclear.

In patients with colon cancer, dMMR has been shown to be predictive of patient response to immune checkpoint inhibitors. However, if this remains relevant in patients with rectal cancer,

it will be difficult to determine due to the rarity of patients with dMMR rectal cancer. It will be of interest to investigate the presence of any interactions between MMR/MSI status and checkpoint inhibitor efficacy in future clinical studies.

Like MSI/MMR status, Tumour Mutational Burden (TMB) is the quantitative measurement of how frequently variations occur in the genome of cancer tissue. This is generally quantified using large gene panels or whole exome/genome sequencing as reports have demonstrated that >1Mega base pair (Mbp) of DNA was required for accurate TMB identification (Allgauer et al. 2018). TMB is generally higher in patients with dMMR and MSI-H tumours. TMB has also been suggested as a predictive biomarker for response to radiotherapy with immune checkpoint inhibitors in patients with non-small cell lung carcinoma (NSCLC) (Wilkins et al. 2019). However, Wilkins and colleagues (2019) found no association between levels of TMB and response to immune checkpoint inhibitors.

1.4.2.5 Tumour Heterogeneity Markers

Tumour heterogeneity markers have been investigated for their predictive and prognostic values in patients treated with neoadjuvant therapies. Tumour heterogeneity scores and indexes have been used to quantify levels and changes in intra-tumour heterogeneity in tumour tissue. Tumour heterogeneity can be calculated according to the number of variants and median variant allelic frequency (VAF) across a tumour sample. This also requires large sequencing panel (>400 genes) or whole exome/genome sequencing for accurate quantification

The clonal and subclonal composition of tumour tissue can also provide further information regarding patient response to therapy. In patients with CRC, studies have demonstrated that whether resistance-associated variants are contained in small tumour subclones or across all tumour cells can influence time to relapse after targeted therapies in patients with CRC (Normanno et al. 2015). These values can be quantified based on VAF and tumour cellularity in patient tissue to generate a tumour heterogeneity score. If variants or signatures are found to be associated to radiotherapy resistance, heterogeneity scores could provide valuable information as part of treatment-decision making.

Overall, tumour heterogeneity may provide valuable prognostic or predictive information with regards to patient prognosis, response to therapy or long-term clinical outcomes. This field is still in its infancy and must undergo more rigorous testing in larger patient cohorts. However, with the general trend moving towards tumour profiling with panel-based or whole genome

sequencing, large scale metrics such as TMB and heterogeneity scores/indexes may have the potential for frequent clinical application.

Consequently, a larger amount of DNA may be required, and small biopsies may no longer be sufficient for such analysis. Furthermore, such analyses may also be limited by their application on a tumour biopsy representing a small region of the tumour, rather than the whole tumour.

1.4.3 Circulating Biomarkers

As an invasive CRC evolves, cancer cells and stromal products are shed into the circulation of an individual. Novel technologies have enabled the identification and tracking of these readily accessible tumour "envoys" as potential disease biomarkers.

Minimally invasive testing with liquid samples (such as whole blood, plasma, urine, cerebral-spinal fluid, ascites or sputum samples) for disease monitoring has become particularly attractive for researchers and clinicians in recent years.

Circulating biomarkers are of significant interest in the clinical setting due to the potential for regular serial sampling taken in 'real-time' as a patient undergoes various therapies. Tumour biopsies are an invasive procedure and are thus impracticable to perform for sequential tumour analysis in patients. Circulating biomarkers provide a minimally invasive alternative which can allow frequent sequential sampling and analysis. However, circulating biomarkers may be limited by high background levels being secreted or released by non-tumour tissue which can hamper the identification of tumour-specific signatures. Previous studies have also demonstrated that tumour-derived circulating biomarkers can vary depending on tumour size and stage, and therefore have limited potential in early stage pre-metastatic cancer patients (Bettegowda et al. 2014).

1.4.3.1 Clinical Circulating Biomarkers

There are currently very few established circulating biomarkers commonly used in in the clinical setting for patients with rectal cancer. Among these are CEA and Carbohydrate Antigen 19-9 (CA19-9).

Most notably, CEA levels can be quantified before the initiation of pCRT as part of treatment decision making/risk stratification. Alternatively, CEA can be measured post-operatively during follow-up for patients monitoring and surveillance to assess tumour response and detect recurrence.

CEA is an oncofetal antigen which is secreted by epithelial tumour cells in the digestive tract. It is usually analysed in patient serum using immunosorbence assays, such as Enzyme-Linked Immunosorbance Assay (ELISA) (Sorensen et al. 2016). Because it is secreted by epithelial tissue under normal physiological conditions, CEA is determined to be high or low by using cut-offs as recommended by guidelines from the ESMO (Glynne-Jones et al. 2017). In research, this cut-off value is often variable from 2.5– 10ng/mL (Nicholson et al. 2015; Sorensen et al. 2016).

In patients with LARC, pre- and post-treatment CEA concentrations and changes there within were associated with pathological response to neoadjuvant therapy and disease recurrence (Peng et al. 2018; Saito et al. 2018). This is hypothesised to be the case because consistently high CEA concentrations after surgery can be indicative of distant undetected micro-metastases (Saito et al. 2018).

With regards to clinical utility, the ambition is to reliably predict which patients will and will not respond well to neoadjuvant therapy before administering treatment. High CEA levels may therefore be used to do so. However, studies have also demonstrated that CEA has significant limitations as a biomarker in disease monitoring; including low sensitivity and specificity in CRC diagnosis and prognostication (Primrose et al. 2014; Sorensen et al. 2016). CEA is also secreted in benign conditions of the colon and is detected more frequently in heavy smokers, which can influence the potential efficacy of CEA in risk stratification or prognostication (Nicholson et al. 2015).

CA19-9 is another tumour marker which has previously been demonstrated to predict survival after therapy in patients with rectal cancer. Recent reports have demonstrated that large decreases in CA19-9 concentration were associated with increased rates of pCR (Song et al. 2018) whilst elevated pre-treatment CA19-9 was predictive of reduced OS (Miki et al. 2018). However, as previously stated with CEA, the presence of contradictory findings in the reliability of pre-treatment CA19-9 to predict patient response (Buijsen et al. 2014) makes the use of CA19-9 a controversial matter.

Studies have demonstrated that the combined utility of CEA and CA19-9 could identify patients with worse survival in the context of both rectal cancer (Zhang et al. 2015) and CRC (Stiksma et al. 2014). Furthermore, Buijsen and colleagues (2014) reported that CA19-9 and CEA concentrations correlated well together in patients with LARC. Therefore, the combination of these two biomarkers in a pre- or post-therapeutic setting may increase sensitivity and specificity of response and outcome prediction in patients with rectal cancer.

1.4.3.2 Systemic Inflammatory Ratios

1.4.3.2.1 Neutrophil Lymphocyte Ratio

With the emergence of immune checkpoint inhibitors among other immunotherapies, there has been increasing interest in immune-related biomarkers. The use of NLRs, identified from routine differential white cell counts in the peripheral circulation has re-emerged as an area of interest; both prognostically and predictively. Neutrophils and lymphocytes were selected for their respective roles in inflammation and anti-tumour functions. Inflammation and inflammatory markers have been demonstrated to promote tumour growth or protect tumour tissue from a host immune response, of which lymphocytes are central.

At present, outside of MSI status, there are no clinically validated biomarkers to predict patient response to immune checkpoint inhibitors (Friedman and Postow 2016) Theoretically, the development of an immunological biomarker could be used to predict patient response to immunotherapies. In agreement with this, various groups have reported that low NLR (<5) is indicative of improved survival in patients undergoing ipilimumab therapy for metastatic melanoma (Ferrucci et al. 2015).

Pine and colleagues (2015) reported that high NLR (≥5) predicted decreased OS and DFS for CRC patients undergoing surgery with curative intent and that NLR correlated to tumour stage and could predict disease recurrence. This study further reported that a low NLR was associated with increased lymphocytic reaction in the invasive margin and a better prognosis.

NLRs may be relevant to CMS1 and 3 subgroups, which are characterized as enriched for MSI-High, hyper-mutated genotypes with increased tumour infiltrate (Guinney et al. 2015). Similarly, CRIS-A tumours are either MSI or 'MSI-like' and CRIS-B tumours presented with an inflammatory phenotype with strong Tumour Growth Factor Beta (TGF-β) activity (Isella et al. 2017); though both presented inflammatory traits. Patient response to immunotherapies within these subtypes will require further exploration. As previously stated, high MSI and/or dMMR tumours are much rarer in patients with rectal cancer compared to colon cancer. Therefore, this brings into question the efficacy of analysing these subtypes in patients with rectal cancer and, perhaps, immune-derived systemic markers such as NLRs. This is a factor we aim to investigate further in patients with LARC as part of this thesis.

As these tumour types display increased immune or 'MSI-like' profile and immune infiltration, these patients are potentially more likely to respond to immunotherapies. The potential utility of NLR or derived NLR (dNLR) has been hinted by reports that dNLR levels were generally good

indicators of response to chemotherapy across mutations; (Wood et al. 2017) especially so for *BRAF* mutant tumours which are commonly MSI-High (Dienstmann et al. 2017b). Unfortunately, this group also reported that NLR was not associated with MMR status, but neutrophil counts and c-reactive protein levels may have been (Pine et al. 2015).

NLR does show promise in that, even with relatively low patient numbers, consistent results have been seen with regards to checkpoint inhibitor response. Furthermore, this assay is relatively cheap, simple to perform and part of routine practice; a large advantage with regards to embedding a biomarker into a clinical setting (Wood et al. 2017).

1.4.3.2.2 Other Markers

Similar inflammatory index markers have been investigated in patients with rectal cancer undergoing neoadjuvant therapy but apart from NLRs, systemic inflammatory markers have not performed consistently. Studies have previously reported that low pre-treatment LMRs were associated with worse OS (Deng et al. 2017) and DFS (Yamamoto et al. 2019a). Similarly, high PLRs have been demonstrated to predict worse response to therapy (Kim et al. 2018). Platelets have previously been associated with promoting tumour growth via increased angiogenesis and metastatic spread. Similarly, monocytes have been reported to induce tumour growth, migration and invasion from within the tumour microenvironment (Condeelis and Pollard 2006).

A different study demonstrated that there was no significant association between PLR and survival and that CEA was superior when predicting DFS and OS than PLR in patients with rectal cancer (Portale et al. 2018). These ratios are limited by using inconsistent cut-off values when determining high or low levels. Therefore, the assignment of high or low and resulting outcome analysis is highly variable between studies. This factor will require standardisation before similar inflammatory index markers can become effective in guiding clinical practice. At this time, these markers have provided valuable information regarding the role of systemic inflammation in tumourigensis and tumour response to neoadjuvant therapy.

These studies are currently limited by small patient numbers with regards to those comparing systemic inflammatory ratios with patient response. Further research also needs to be undertaken in order to determine whether such ratios are predictive of therapeutic response or general outcome, due to the correlation with tumour staging (Pine et al. 2015). Finally, how these ratios compare in predicting outcome to other circulating biomarkers also remains to be seen.

1.4.3.3 Circulating Tumour DNA

1.4.3.3.1 Cell Free DNA Secretion

Circulating tumour DNA (ctDNA) is a tumour derived sub-population of circulating free DNA (cfDNA) in patients with various solid tumours, including CRC. There are two hypothesised mechanisms of cfDNA release, these are passive release through cell death (Diehl et al. 2008) and active release via extracellular vesicles, such as exosomes (Figure 2) (Wang et al. 2017).

During passive cfDNA release, apoptotic or necrotic activity causes the fragmentation of DNA into ~180-200bp fragments (Thierry et al. 2010). Studies have suggested that these kinds of cell death invoke an immune response that release cfDNA through T-cell-mediated cell destruction (Thierry et al. 2010). Alternatively, other studies have suggested that oncogenic DNA is selectively packaged into exosomes and exocytosed to promote local or distant oncogenic activity (Wang et al. 2017). Alternatively, other studies have reported that exosomes are used to secrete harmful DNA in response to DNA damage in tumour cells (Takahashi et al. 2017). DNA has been demonstrated to migrate through nuclear pores into the cytoplasm in response to large amounts of DNA damage. The accumulation of nuclear DNA in the cytoplasm can then induce an immune response. Researchers have consequently hypothesised that tumour cells can circumvent this process by actively excreting cytoplasmic DNA via exosomes (Takahashi et al. 2017).

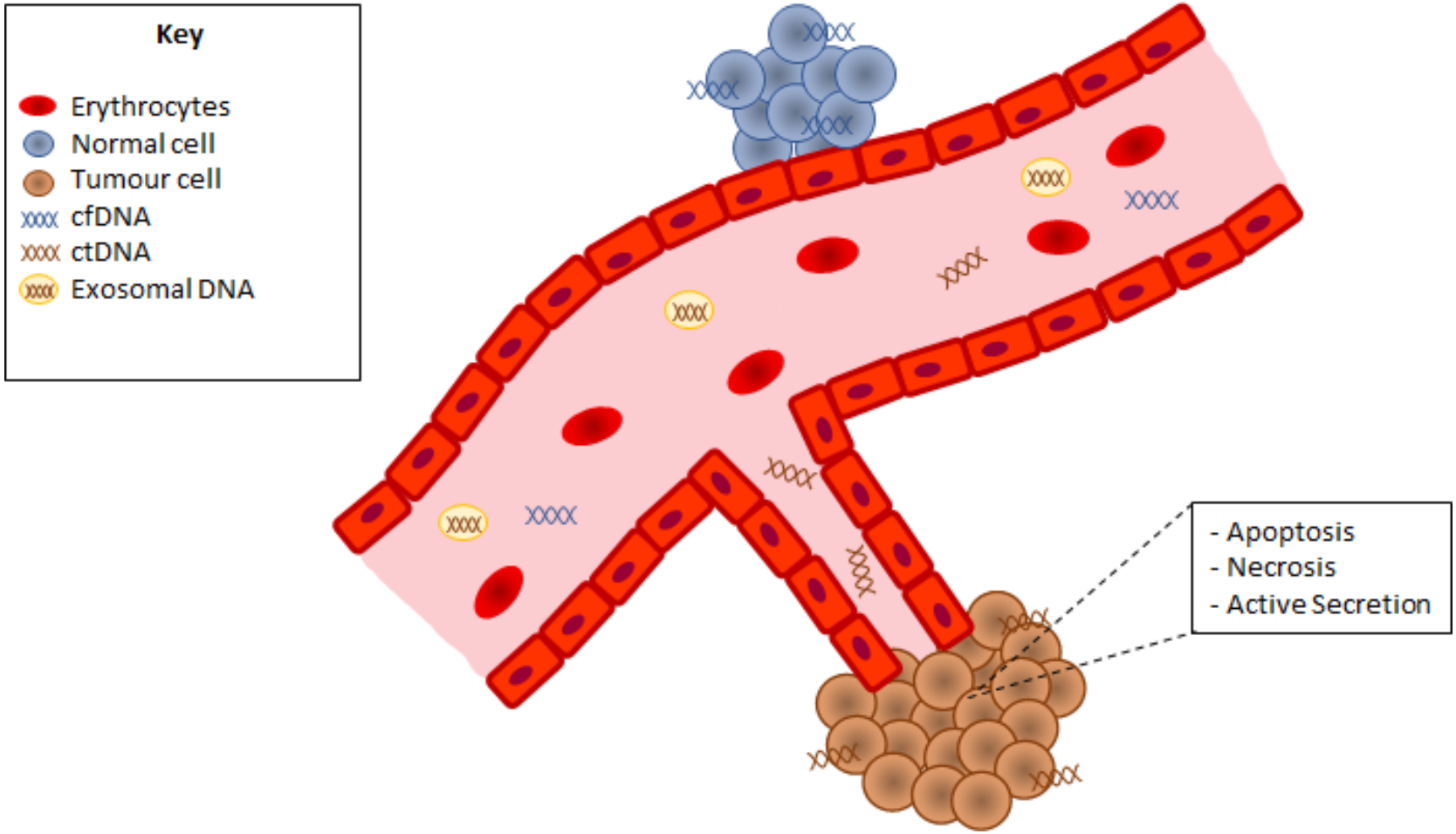

Figure 2: The various mechanisms of cfDNA release into the circulation from tumour and non-tumour tissue in patients with LARC.

1.4.3.3.2 Clinical Utility of Cell Free DNA as a Circulating Biomarker

cfDNA can be released into the circulation by either tumour or non-tumour tissue (Thierry et al. 2010). Distinguishing the fraction of tumour-derived from non-tumour DNA in plasma has been accomplished by detecting tumour specific genetic alterations. This can be achieved using a variety of methods, such as next-generation sequencing (NGS), digital PCR or pyrosequencing. CtDNA has been demonstrated to predict progression, relapse, monitor levels of disease and detect the emergence of treatment-resistant clones (Jia et al. 2017).

CtDNA levels have been reported to vary depending on clinical factors, such as tumour stage and the presence of metastases (Bettegowda et al. 2014) in patients with CRC and a number of other solid tumours. In patients with rectal cancer, ctDNA levels were also demonstrated to correlate with tumour size (Tie et al. 2015).

Quantitative changes in ctDNA levels during first line chemotherapy could act as early biomarkers of therapeutic response in patients with CRC before current methods allow, preventing the futile use of expensive and toxic therapies. Reports suggest that ctDNA was a more sensitive method for detecting disease progression than current methods, as ctDNA frequency changes occur significantly earlier than changes in CEA or radiological imaging (Carpinetti et al. 2015; Tie et al. 2015; Sun et al. 2018).

Post-operative ctDNA variant allelic frequencies were reported to help predict 3-year response free survival (RFS) after neoadjuvant therapy in patients with LARC (Tie et al. 2019). In patients with metastatic CRC, significant early decreases in ctDNA were indicative of improved PFS (Tie et al. 2015). In comparison, CEA was reported to have no positive predictive value to predict patient response to chemotherapy, and was particularly poor in comparison to ctDNA fold-changes (Tie et al. 2015). Similar reports have suggested that post-surgical ctDNA levels in plasma are a significant predictor for later relapse in patients with CRC undergoing adjuvant therapy (Reinert et al. 2015).

Russo and colleagues (2015) reported that they could detect *KRAS* mutations in patient ctDNA which were absent in the primary tumour biopsy of a patients with CRC. These variants were later detected within metastatic lesions. This is an agreement with findings from a separate study which reported that ctDNA had an improved concordance with metastatic tissue than primary tumours (Brannon et al. 2014). This finding indicates the potential use of a ctDNA to detect the presence or emergence of sub-clones/metastatic lesions before the initiation of

treatment and facilitate risk stratification and treatment decision making, both of which are being explored in prospective clinical trials.

Developing CRC subtyping systems will likely influence how ctDNA is utilised in the future. Though ctDNA currently looks likely to monitor and detect *KRAS/NRAS* mutations, the emergence of these novel systems may lead to alternative mutations being analysed as part of subtype allocation/validation. One possibility could be that monitoring of CRIS-D patients may include levels of *IGF2* copy number variations, or mutations within other genes vital to WNT signalling (Isella et al. 2017). Similarly, CMS1 may include tracking *ATM* or *PTEN* mutations throughout therapy (Dienstmann et al. 2017b). The numbers of genes of interest for tracking will likely expand as our knowledge of genetic alterations associated with each subtype increases.

There remain several limitations associated with ctDNA, such as the generally low cfDNA concentrations extracted from plasma which can hinder analysis. Furthermore, ctDNA is often found fragmented into short lengths of DNA, limiting the molecular alterations which can be reliably detected; though evolving techniques are swiftly overcoming this issue. Other features have also been suggested to cause an increase in cfDNA/ctDNA levels, including inflammation and trauma (Thierry et al. 2010). These are factors which must be considered when investigating patients being treated with cytotoxic agents and surgery.

Furthermore, a significant proportion of cfDNA can originate from non-tumour cells. This can cause a high background level of non-tumour cfDNA and hinder tumour DNA detection (Bettegowda et al. 2014) and the confident quantification of changes in allelic frequency (Diehl et al. 2008). This can also be exacerbated through improper pre-analytical handling (Ignatiadis et al. 2015).

Analysis of ctDNA is rapidly approaching routine clinical use due to its speed, simplicity and relatively low cost. In fact, screening of resistance associated mutations in ctDNA is already being applied in non-small cell lung cancer (FDA 2016) and kits for *RAS* screening in ctDNA have recently been CE-marked for utility in CRC (Sysmex-Inostics 2016). Though the information provided may be relatively limited, ctDNA is detectable in a relatively high proportion of patients and analysis can be screened for mutations of interest. These advantages coupled with our current knowledge of CRC genetics and clonal evolution make ctDNA a very appealing biomarker.

These examples are in the context of targeted therapies with known resistance-associated variants or signatures. Chemoradiotherapy in patients with LARC is yet to have an identified variant or signature to associate with treatment resistance or sensitivity. Therefore, for patients with locally advanced disease ctDNA can hypothetically be used in patient monitoring and surveillance to detect disease recurrence. However, ctDNA cannot yet be used to identify the progression of chemoradiation resistant disease. Further research is required to identify variants or signatures which are indicative of therapeutic resistance to fully exploit ctDNA analysis in this capacity.

1.4.3.4 Circulating MicroRNA

1.4.3.4.1 MicroRNA Biogenesis

MicroRNAs are small non-coding RNAs (~18-25 nucleotides long) which have been found to regulate targeted mRNA expression in both healthy and cancer tissues. MicroRNA sequences are either located in non-coding intergenic regions of DNA or within gene introns (Ameres and Zamore 2013). In the nucleus, primary microRNA (pri-microRNA) is synthesised by RNA polymerase II enzymes from genetic microRNA sequences in nuclear DNA. These are initially synthesised into long (~60nt) stem loop structures. Pri-microRNAs are then cropped into shorter intermediates by Drosha (an RNase III enzyme) in combination with Pasha or DCGR8 (Ameres and Zamore 2013).

Short pri-microRNAs are then transported into the cytoplasm via the nuclear pore by the nuclear transport receptor Exportin 5. In the cytoplasm, the RNase II protein Dicer liberates microRNAs into ~22nt microRNA-microRNA* duplexes (Ameres and Zamore 2013). This process is referred to as dicing. MicroRNAs then bind to the Argonaute (AGO) protein in the cytoplasm. Here the microRNA* strand of the duplex is cleaved and released. The remaining microRNA strand and AGO protein recruit other proteins to form the RNA-Induced Silencing Complex (RISC). The microRNA then proceeds to guide the RISC complex to complementary mRNA sequences in order to regulate expression (Figure 3) (Ameres and Zamore 2013).

In microRNAs, the major determinant of complementarity is a 5’ 6-8 nucleotide sequence on the microRNA which is known as a ‘seed sequence’. This complementary sequence is normally found in the 3’ untranslated region (UTR) of the mRNA. Upon binding of the RISC to complementary mRNA, the complex represses translation through mRNA degradation (Ameres and Zamore 2013).

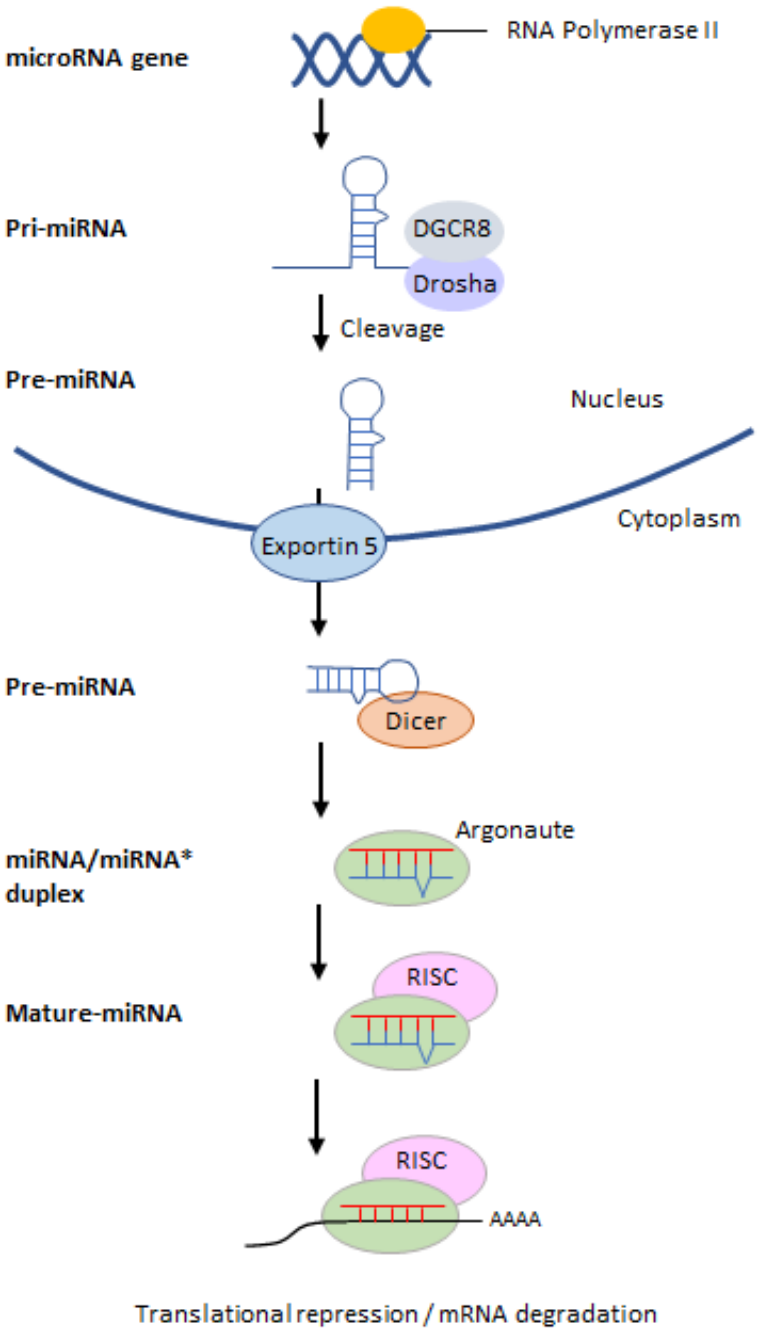

Figure 3: MicroRNA biogenesis from transcription from DNA by RNA polymerase II in the nucleus to mRNA degradation in the cytoplasm

1.4.3.4.2 MicroRNAs in Cancer

Since microRNAs were originally associated with cancer (Calin et al. 2002), they have been linked with numerous roles in tumourigenesis, including migration, invasion and metastasis (Ma et al. 2007) and growth, proliferation and inhibition of apoptosis (Cekaite et al. 2015). The function of microRNAs can either be tumour suppressing or driving; varying between different tissues and cancer types.

Numerous microRNAs have been investigated for their predictive or prognostic values in patients with CRC. Mir-31 was previously demonstrated to be down-regulated in the tumour tissue of patients with LARC in comparison to adjacent normal mucosa (Jo et al. 2017). In colon cancer cells, Mir-31 was demonstrated to down-regulate the MMR gene *MLH1* (Kim et al. 2014). Mir-31 has also been reported to play a role in cetuximab resistance in *KRAS* wild-type CRC patient tumours (Pugh et al. 2017); though studies still dispute that this may simply be an indicator of generally poor prognosis (Carames et al. 2016).

1.4.3.4.3 Circulating MicroRNAs

MicroRNAs are relatively stable in circulation due to their protection from RNase activity through their association with carrier proteins or inclusion within exosomes (Lindner et al. 2015). Considerable research has been performed in order to devise cell-free microRNA panels for clinical utility; with results varying considerably. Research has been directed at early diagnosis, predicting relapse, survival and treatment response and as a general prognostic marker (Toiyama et al. 2013; Chang et al. 2016). Various circulating microRNAs have been investigated as potential biomarkers, though none have yet been validated. Although several studies have been performed, few microRNAs have been consistently identified as diagnostic indicators of CRC when compared to healthy control subjects. These few include Mir-21 (Kanaan et al. 2012; Ogata-Kawata et al. 2014), Mir-29a (Huang et al. 2010) and Mir-19a (Zheng et al. 2014; Matsumura et al. 2015). However, these have been in generally small cohorts (up to 300 patients) and further investigation will be required to determine clinical efficacy as well as why variations exist in the literature.

Greystoke and colleagues (2015) recently utilised small cell lung cancer (SCLC) circulating tumour cell (CTC) Derived Xenograft (CDX) and PDX models to identify a panel of 10-microRNAs for diagnosis and post-treatment follow-up. The use of these models allowed the identification of tumour-specific human miRNAs and limit background interference from non-tumour DNA in the circulation. The 10-plex miRNA panel could distinguish stage 3 and 4 SCLC

patients from healthy volunteers with 82% and 98% sensitivity respectively and both having 100% specificity.

This study has established a potential pipeline for developing expression-based panels/signatures for clinical use, although the significance of removing the stromal influence is not fully understood (Greystoke et al. 2015).

Although microRNAs do not currently influence subtyping, the CRCSC has reported that CMS2 tumours up-regulate the MiR-17-92 cluster (including Mirs-17, -19a, -19b-1 and -92a), whilst CMS3 and CMS4 down-regulate let-7 and MiR-200 respectively (Guinney et al. 2015). Each of these can theoretically be quantified and monitored in plasma or plasma exosomes to predict response or stratify patient treatment in the future. The CRIS system has not yet reported incorporation of microRNAs.

As our knowledge of the roles of microRNAs in cancer expands, so will potential applications in a clinical setting; though this is still in relatively early stages of research. MicroRNAs may also play a role in determining patient treatment, either through targeting miRNAs directly or through inferring tumourigenic drivers. We further predict changes in microRNA expression may be incorporated into a uniform subtyping infrastructure in the future.

1.4.3.5 Circulating Exosomes

1.4.3.5.1 Exosome Biogenesis

In recent years, there has been increasing interest in the investigation of extracellular vesicles as circulating biomarkers in patients with solid tumours (Lotvall et al. 2014). Exosomes are a subset of extracellular vesicles which are ~30-120nm in diameter and have various roles; including cell-to-cell communication (Valadi et al. 2007).

Exosomes are formed from early endosomes in the cell cytoplasm. Early endosomes are formed through the invagination of the plasma membrane, which swallows extracellular ligands and intracellular components. Endocytic vesicles can then share cargo and combine forming intraluminal vesicles (ILVs) within early endosomes. Proteins are then sorted in a process facilitated by exosome-sorting complex required for transport (ESCRT) proteins or other similar pathways (Wan et al. 2018b).

These mature endosomes are then referred to as multivesicular bodies (MVBs). Some MVBs are degraded by the lysosome whilst others proceed to merge with the cell membrane in a process facilitated by the SNARE complex. As MVBs merge with the cell membrane, ILVs are

released by exocytosis into the extracellular milieu (Figure 4). Once exocytosed, these ILVs become exosomes (Wan et al. 2018b). Exosomes have been reported to play key roles in several different aspects of molecular biology, including immune cell modulation, apoptosis and angiogenesis (Gurunathan et al. 2019).

Exosomes facilitate intercellular communication by transporting their molecular cargo from their parent cell and depositing their cargo into a target cell by fusing with the lipid membrane. Exosomal cargo has been reported to include proteins, various RNA species and DNA. It has been hypothesized that the contents of circulating tumour exosomes may either give an indication as to the genetic profile of parent tumour cells or the interaction of tumour cells with surrounding stroma.

Both the internal (cargo) and external molecular (membrane-bound proteins) features of exosomes have been characterised and can be variable between different types of tissues of origin. Exosomes may contain various endosome-associated proteins, such as TSG101, which is involved in endosome biogenesis (Lotvall et al. 2014). However, as far as we are aware, there are some tetraspanins (e.g. CD9 and CD81) which are located and significantly overexpressed in the membranes of exosomes from a large variety of tissues in comparison to their cells of origin (Escola et al. 1998; Gurunathan et al. 2019). These proteins are often analysed to ensure that exosomes have been successfully isolated. There are also proteins which are known not to be associated with exosomes , such as *AGO* (Argonaute/RISC complex) or *GM130* (Golgi complex) (Lotvall et al. 2014). The presence of such proteins can be analysed to ensure the purity of exosomes isolated, rather than exosomes being isolated alongside other contaminating extracellular vesicles or microvesicles.

Exosomes have been demonstrated to be selectively packaged with their molecular cargo by their cells of origin to influence local or distant target cells (Ratajczak et al. 2006; Valadi et al. 2007). The molecular cargo of exosomes has been reported to include proteins and different species of RNA (including mRNA and microRNA), however, the presence of exosome-derived DNA is still a disputed subject. Previously, Valadi and colleagues (2007) demonstrated that both mRNA and microRNA which was packaged into exosomes was found to be transcribed in their target cells, highlighting a key function of exosomes and their molecular cargo. This transfer of molecular cargo aims to induce a specific cellular function in their target cell. Recently, one study reported that mesenchymal stromal cells could secrete exosomes which influenced the proliferation and differentiation of Th1 T-cells through TGF-beta signalling (Cunha et al. 2020).

With the increasing interest in exosomes in recent years, numerous new technologies have been developed to facilitate their isolation. Originally, the only accepted way to isolate pure exosomes involved differential ultra-centrifugation which was laborious and required access to an ultra-centrifuge. However, newly developed methods aim to isolate exosomes is several different ways, some by precipitating them out of solution using polymers, others by isolating based on the presence of specific receptors (such as CD9 or CD81) or chromatographically, based on physical properties such as size and density. However, each of these methods has their own respective limitations regarding abundance or purity, which must be overcome as circulating exosomes approach clinical utility.

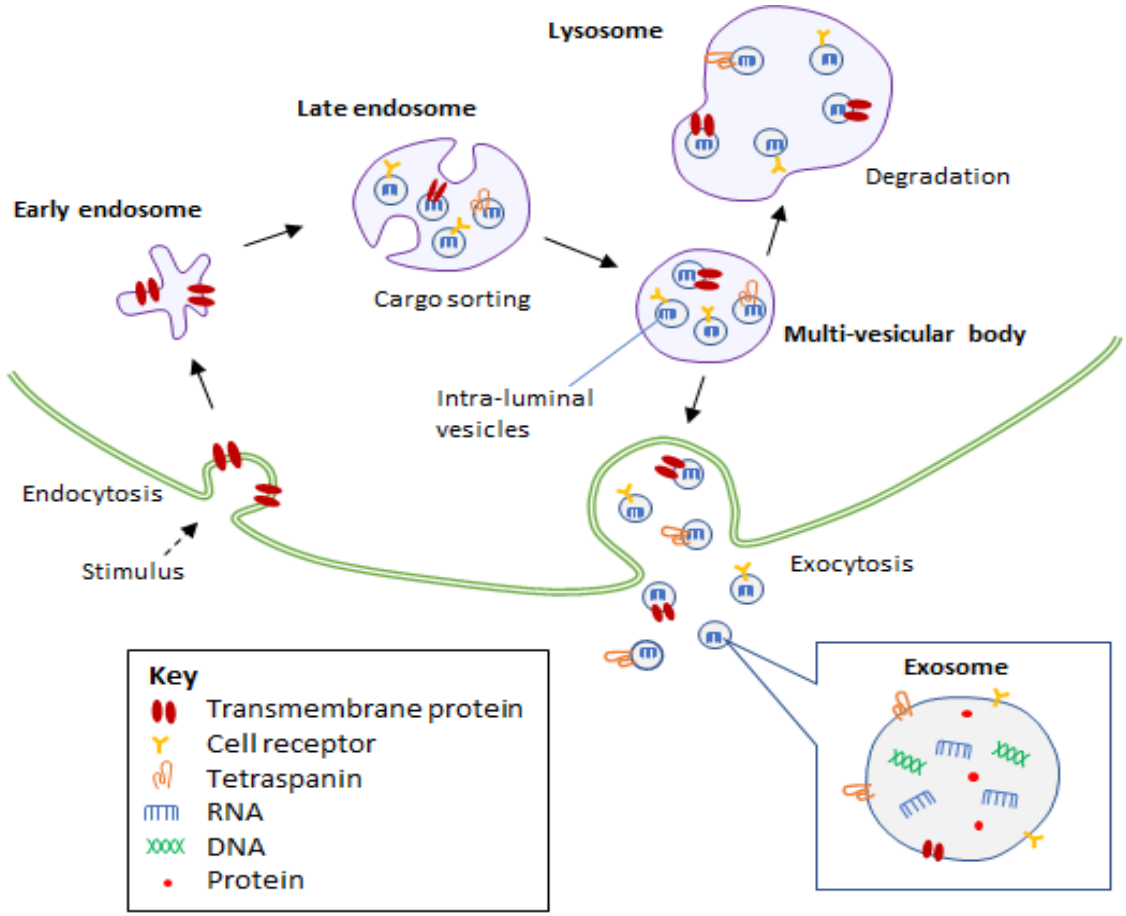

Figure 4: Exosome biogenesis and secretion from stimulated endocytosis to the release of exosome into the extracellular milieu by exocytosis

1.4.3.5.2 Exosomes in Cancer

The functions of exosomes have been demonstrated to be co-opted by tumour cells in order to promote pro-tumourigenic characteristics, such as proliferation, vasculogenesis and migration (Ruiz-Lopez et al. 2018).

In cancer cell lines, exosomes were demonstrated to transport oncogenic DNA which would promote endothelial cell proliferation (Huang and Feng 2017). Another study reported that exosomes secreted by colon cancer cell lines containing TGF-B, promoted the differentiation of fibroblasts into cancer-associated fibroblast (CAF)-like fibroblasts (Rai et al. 2019). Transformed CAFs have also been demonstrated to induce metastasis and chemotherapy resistance through the secretion of exosomes (Hu et al. 2019), both of which are hallmarks of cancer (Hanahan and Weinberg 2011).

Studies have also suggested that exosomes transfer oncogenic DNA with similar affects to those seen for mRNA and microRNA. Studies have demonstrated the presence of tumour-specific variants in tumour-derived exosomes from cancer cell lines (Thakur et al. 2014) and patients with pancreatic cancer (Kahlert et al. 2014). Exosomes were also demonstrated to transfer oncogenic DNA to target cells *in vitro (Lee et al. 2014).* Cellular uptake of exosomes carrying oncogenic DNA resulted in stimulated cellular proliferation *in vitro*.

In various cancer types, there is evidence to suggest that oncogenic crosstalk between tumour tissue and the surrounding tumour microenvironment is mediated by exosomes. Studies have also reported that exosomes can enter the circulation and form pre-metastatic niches in distant sites (Costa-Silva et al. 2015). Overall, exosomes appear to be co-opted in cancer and play an impactful role in tumour development.

1.4.3.5.3 Exosomes as a Circulating Biomarker

Exosomes have been detected in the circulation of healthy individuals (Vlassov et al. 2012) but have been reported to be at increased levels in cancer patients and during disease progression (Ko et al. 2015). Studies have investigated the molecular cargo of circulating exosomes as biomarkers in a variety of cancers, including CRC (Matsumura et al. 2015); with greater emphasis on exosomal miRNA.

One study reported that increased expression of exosome-derived Mir-19a has been suggested to be a prognostic biomarker for recurrence in patients with colorectal cancer (Matsumura et al. 2015), which is in agreement with other studies quantifying mir-19a in serum, serum exosomes and cancer tissue (Zhu et al. 2017).

Studies have recently investigated tumour-specific alterations in exosome -derived RNA (exoRNA) or DNA (exoDNA) for monitoring patients and stratifying therapy, in a similar vein to ctDNA. One such study investigated sequencing concordance, reporting that 14/19 (73.7%) of *KRAS* mutations and 6/8 (75%) of *BRAF* mutations detected in patient tissue were also detected in patient exoRNA (Hao et al. 2017).

The emergence of cancer-specific exosomal markers may facilitate the detection of subtler molecular signatures and limit non-tumour background (Melo et al. 2015). Several studies have reported that, using a panel of tumour-exosomal markers, they were able to detect an increased frequency of *KRAS* mutation in exoDNA when compared to generic exosome isolation (Melo et al. 2015; Castillo et al. 2017); suggesting that they were able to isolate a purer yield of tumour-derived exosomes using a specific panel of surface markers.

Recently, studies have reported the identification of proteins which are specific to exosomes secreted from pancreatic cancer and CRC cells. These tumour-specific exosomal markers may facilitate the isolation of an increasingly pure population of tumour exosomes to limit the presence of wild-type background of DNA, RNA or proteins and further facilitate analysis. With the future emergence of CRC-specific exosomes, it may also be possible to quantify tumour-exosome levels to hypothesise tumour burden or aggressiveness before in-depth analysis begins.

As suggested previously for CTCs, exosomes may be incorporated into a novel molecular subtyping system; as exosomes carry mRNA and microRNA, they may be a possible surrogate source for mutation detection or gene expression profiling where tumour tissue is unavailable. Concentration of tumour-specific exosomes may also be incorporated into staging, like that predicted for CTC numbers.

Limitations with circulating exosomes do exist and will continue to be encountered as this field progresses. For example, as seen with ctDNA and microRNA previously a small concentration of exosomes will originate purely from the tumour, and low numbers will likely cause issues with molecular analysis, certainly in hypothesis-generating studies. Furthermore, a low concentration of exosomes extracted may be from the circulation of locally advanced non-metastatic patients. Therefore, the presence of non-tumour exosomes and their respective molecular cargo may mask subtle signatures and low-level results from tumour-derived exosomes. We may yet overcome these issues as the technologies continue to develop, and simultaneously, even more limitations will almost certainly arise.

The study of cancer exosomes is a rapidly developing field, with an increasing number of publication focussing on exosomes and extracellular vesicles since the early 2000's (Lotvall et al. 2014). As a result, several different technologies have recently emerged which facilitate exosome isolation from plasma/serum. Exosomal cargo may then be analysed using several methods.

1.4.3.6 Circulating Tumour Cells

CTCs are cells detected in the circulation of cancer patients which are $CD45^{-ve}$ whilst expressing epithelial markers (Ignatiadis et al. 2015). CTCs are understood to be tumour cells which have disseminated into peripheral blood and are precursors and/or direct cause of distant tumour metastasis (Janni et al. 2016).

To date CellSearch is the only Food and Drug Association (FDA) approved method for capturing CTCs in a clinical setting (Sastre et al. 2012). This method uses the epithelial markers EpCAM and cytokeratins to isolate and enumerate CTCs for further investigation.

Huang and colleagues (2015) investigated the potential of CellSearch-based CTC detection in 11 studies and a total of 1847 patients with CRC. This study reported that high CTC counts at baseline and throughout treatment were associated with poorer disease control, disease relapse/progression PFS and OS. This is in agreement with studies with Bork et al. (2015) and Tan et al (2018) who demonstrated this in stage I-III and metastatic CRC patients respectively. Further studies have suggested that patients who converted from being CTC-high to CTC-low during therapy had better disease control than patients who changed from low to high, or remained persistently high (Huang et al. 2014). In patients with rectal cancer, studies have also shown that patients with high numbers of CTCs (>5) were more likely to develop distant metastases within one year of curative surgery (Tsai et al. 2016). Similarly, another study reported that patients with rectal cancer who responded well to neoadjuvant therapy had low post-treatment CTCs (Sun et al. 2016).

Despite initial promise, this biomarker continues to display limitations in the CRC arena. The specificity of these techniques has been widely debated; whilst some studies have reported no detectable CTCs in benign conditions or healthy individuals (Sun et al. 2013b) others have detected CTCs in benign colon disorders (Pantel et al. 2012). Furthermore, detection of one or more CTCs per 7.5mL has only been reported in 30-50% of CRC patients (Hardingham et al. 2015); limiting potential clinical efficacy. There is also the possibility of false negatives due to

loss of epithelial markers during epithelial-mesenchymal-transition (EMT); further limiting potential utility (Hardingham et al. 2015).

CTCs have provided greater insight into intra-tumour heterogeneity than biopsies alone. Gasch and colleagues (2013) reported that separate CTCs within a single patient had different mutational profiles, illustrating that separate regions of the invasive tumour can have individual driving mechanisms. Additional studies have also detected relevant mutations in CTCs when absent in tumour biopsies, influencing patient-response (Mostert et al. 2013).

CTCs have been investigated *ex vivo* in a number of solid tumours; being culminated into cell-lines (Cayrefourcq et al. 2015) and /or transplanted into immune-deficient mice (Lallo et al. 2017). Grillet and colleagues (2017) reported that, in CRC, CDX mirrored the response of the patient to platinum and etoposide therapy; illustrating the potential to truly personalise therapy. Though still in its infancy, this gives an indication into the potential of CTCs as an *ex vivo* microcosm of the tumour itself.

As CTCs were predictive of the occurrence of distant metastases in patients with rectal cancer (Tsai et al. 2016), CTCs may be able to identify patients with micro-metastases after therapy or before surgery. Therefore, CTCs may be used for post-therapeutic risk-stratification and treatment-decision making with regards to the administration of post-operative adjuvant therapy.

We hypothesise that CTC enumeration could be incorporated into staging or predicting the likelihood of metastatic occurrence, further facilitating in treatment decision making. CTCs also have the potential to be used as surrogate markers for subtyping where tissue samples are unavailable. Reports have stated that CTC-derived microRNA did not correlate with CTC count, illustrating that we still have a limited understanding of the roles of both CTCs and microRNA in the context of CRC (Tan et al. 2018). Both the CRIS and CMS systems have an EMT-associated group with enhanced TGF-β signalling (CRIS A/B and CMS4) possibly allowing for future incorporation. As CTCs will be free of adjacent stroma, it will allow for a purer cancer cell intrinsic signature, potentially corresponding well with CRIS subtyping.

CTCs and ctDNA have often been compared for their potential clinical impact. Concordance of CTCs and ctDNA with mutations detected in primary tumour or metastases has been reported at approximately 78% (Kidess-Sigal et al. 2016). In this study, each biomarker was detected in one patient where the other was absent; moreover, each biomarker detected a mutation previously unseen in the primary tumour or metastasis. This suggests both biomarkers may be

equally useful when investigating or considering intra-tumour heterogeneity; although both are thus limited and may be best used in tandem. One advantage of ctDNA over CTCs is that ctDNA has been reported as detectable in 100% of metastatic CRC, whereas only 35% had detectable CTC levels from 10mL whole blood (Germano et al. 2017). However, due to CTCs detection not requiring knowledge of existing tumour-specific changes, it can fulfil a role of screening and early detection in a superior fashion, having a different niche than that predicted for ctDNA.

Despite recent advances, reported false positives and low CTC numbers are a potential cause for concern. New technologies, such as CellCollector, are in development which involve inserting probes into directly into patient blood vessels for an extended period of time in order to collect larger numbers directly from blood, as opposed to collecting CTCs from a limited volume of whole blood (Theil et al. 2016). Other methods, such as the AdnaTest, are relatively simple and consistent techniques, having been designed for clinical utility and incorporate both isolation and mRNA extraction (Gorges et al. 2016). Future clinical studies can assist in creating the basis for clinical utility of CTCs in terms of subtyping, prognostication, response/metastasis prediction or treatment stratification. The potential use of CDXs to test numerous treatments is also an exciting future direction.

1.4.3.7 Circulating Metabolites

Metabolomics is defined as a post genomics research field and the study of metabolic profiles. Like miRNA and ctDNA, several groups have used hypothesis-free methods such as mass spectrometry (MS) and nuclear magnetic resonance (NMR) imaging to determine metabolic signatures in CRC patient subgroups. These signatures were developed for utility in early diagnosis (Surinova et al. 2015), monitoring disease progression (Zhu et al. 2015) or inflammatory status (Bertini et al. 2012). These signatures include proteins, glycoproteins, amino-acids and lipids among other metabolites.

Zhu and colleagues (2015) used Liquid Chromatography with Mass Spectrometry (LC-MS) to analyse sequential patient serum samples from 70 patients with CRC in an attempt to distinguish between progressing and stable disease. Using five core-metabolite levels, this group demonstrated that they could distinguish patients with progressing and stable disease with sensitivity and specificity of 83% and 94% respectively. Furthermore, this study also reported that this metabolomic profile was superior to CEA concentrations in predicting disease progression (Zhu et al. 2015).

Metabolomics has great clinical potential for largely different reasons to the biomarkers discussed previously. As previously mentioned, CMS3 and CRIS-A have been defined as consisting of metabolic dysregulation (Guinney et al. 2015; Isella et al. 2017).

The CRCSC has reported a number of metabolic pathways which are deregulated within CMS3; some of which agree with a systematic review published by Zhang and colleagues (2017). Zhang and colleagues reported deregulated levels of amino acid and carbohydrate metabolism in various tissues; organised into pathways by use of the Kyoto Encyclopaedia of Genes and Genomes (Goto et al. 1997). The CRCSC reported overall increased expression of these two metabolic pathways in CMS3 (Guinney et al. 2015), highlighting the driving mechanism behind this subgroup of tumours. This is also in agreement with the CRIS A subtype, which was reported to have a sustained glycolytic metabolism (Isella et al. 2017), and which, like CMS3, also had a high rate of mutations in *KRAS*. These factors combined might suggest therapeutic targeting of the metabolic deregulation within this subset of CRC tumours. These patients might also benefit from staging or monitoring with circulating metabolites rather than other biomarkers, as one would be directly utilising the pathway driving tumour development.

Though these studies may not be directly comparable due to the different questions each report poses, the existence of overlap may help patient sub-classification in the future, and possibly lead to treatment stratification. Thus, metabolomic analysis of patient tumour and/or plasma/urine may become integral for patient subtyping into a future uniform system, with the further possibility of treatment stratification.

In addition to the numerous molecular and circulating biomarkers discussed here, there are many others which are in earlier stages or growing in popularity in patients with rectal cancer. These include radiomics, the use of radiological imaging to investigate parameters such as vascular growth and tumour heterogeneity. In patient circulation, circulating nucleosomes and proteomics are also growing fields of interest but we could not discuss all biomarkers of interest in such depth here.

1.5 Technical Evaluation

Recent technological advances in the field of sequencing have helped to usher in a new era in the potential clinical utility of circulating biomarkers. Methods such as NGS and digital Polymerase Chain Reaction (PCR) have been an integral part of this evolution.

1.5.1 Next Generation Sequencing

Since 2005, the emergence of commercially available massively-parallel bench top sequencers has caused a significant increase in the widespread availability of targeted sequencing and analysis, in both a research and clinical setting (Voelkerding et al. 2009). NGS technologies, such as the Lifetechnology's Ion Torrent Proton and the Illumina Miseq, can allow consistent and reliable detection of tumour variants in tumour biopsies or ctDNA.

1.5.1.1 Scientific Basis of Next Generation Sequencing

NGS technologies involved a large shift from the limited and laborious low-throughput first generation sequencing methods which they succeeded. Using NGS, many randomly arrayed DNA molecules on a sequencing chip are sequenced and captured in parallel. In most cases, thousands of identical DNA strands are bound to a fixed position, sequenced with sequential nucleotide washes and scanned in order to detect the addition of new nucleotides. The main commercial sequencers apply reversible terminator or semiconductor sequencing for NGS analysis.

Reversible terminator sequencing technology is used in Illumina sequencers such as the Miseq. In this process, DNA is amplified, barcoded, washed and bound to a sequencing chip where another round of amplification occurs to form amplicon clusters. These clusters are washed with nucleotides which contain sequencing terminators, which blocks the addition of >1 nucleotides and allows the augmentation of a single base at a time. Each added based contains a fluorescent tag. Once added, excess nucleotides are washed away, and the added base is scanned and recorded. The terminators are then enzymatically cleaved and new nucleotides are washed over the sequencing chip and the cycle continues.

Semiconductor sequencing is applied by LifeTechnologies sequencing such as the Ion Torrent Personal Genome Machine (PGM) and the Ion Torrent Proton. This technology also included the amplification, barcoding and sequential washing of template DNA. Individual DNA molecules are then bound to silicon sphere through their respective barcodes. DNA is then amplified so that thousands of copies of each amplicon are then bound to each silicon sphere. Spheres are then loaded into a sequencing chip containing tens of millions of wells, each designed to contain one sphere. Sequencing chips are then loaded into the sequencer. Each chip is then sequentially washed with a single nucleotide at a time. When a nucleotide is added to sphere-bound amplicon, a single hydrogen ion is released. The sequencers measure the pH after each nucleotide is added, and where pH changes are detected (as a result of the released

hydrogen ion) and the sequencers records the addition of said nucleotide. The number of nucleotides added is calculated according to the change in pH detected by the sequencer.

NGS allows massively parallel high-throughput sequencing ranging from small targeted sequencing panels to whole exome or genome sequencing and is ideal for hypothesis free variant detection.

1.5.2 Digital PCR

1.5.2.1 Scientific Basis of Digital PCR

Digital PCR is the result of advancements in microfluidics and emulsion PCR chemistry to create a highly precise, reproducible and sensitive method for rare variant detection. Digital PCR was first described by Vogelstein and Kinzler in 1999 in a study which displayed the high sensitivity of digital PCR to known *RAS* variants in patients with CRC (1999). Digital PCR shares similarities with real-time quantitative PCR (qPCR) in how the assays use different coloured fluorescent probes to signal the detection of variants of interest and wild-type DNA. However, technical aspects of digital PCR provide distinct advantages with regards to sensitivity, precision and quantification (Cao et al. 2017).

There are two main types of digital PCR applied, these being chip-based and emulsion-based digital PCR. In both cases, digital PCR analysis is based upon three principles, compartmentalisation, single-molecule PCR and Poisson statistics (Cao et al. 2017). The only difference between chip-based and emulsion-digital PCR is the method by which reactions are compartmentalised. Chip-based digital PCR separates a single PCR reaction (containing primers, fluorescent probes DNA polymerase enzyme, sample DNA etc.) into a large number of wells on a microchip, whereas emulsion digital PCR separates reactions into oil-based droplets. Emulsion-digital PCR is also referred to as droplet digital PCR (ddPCR).

The number of compartmentalised reactions can vary between different technologies but will usually be >10,000. Each reaction is designed to contain a single molecule of DNA. Each compartmentalised reaction is then amplified under normal qPCR conditions and amplified reactions are scanned for fluorescence. As each reaction only theoretically contains up to one molecule of DNA, each separate compartment should only have a single detectable fluorescence, mutant or wild type. The number of mutant and wild-type compartment can then be quantified, allowing reliable VAF calculation and absolute quantification without the need of a standard curve.

The high sensitivity and precision of this technology makes ddPCR an ideal technique for ctDNA analysis. ddPCR is however limited as an individual assay, requiring prior knowledge for variant detection. For this reason, among others, ddPCR is often coupled with NGS analysis. NGS allows the sensitive analysis of tumour-specific variants across a wider range of genes, from smaller targeted sequencing panels to whole exome or genome sequencing. ddPCR can then be applied for fast and precise sequential variant detection and quantification.

1.6 Thesis Overview

This thesis will focus on the technical aspects of a range of potential prognostic and predictive circulating biomarkers in patients with LARC. These circulating biomarkers will specifically include ctDNA, cfDNA, circulating exosomes and systemic immunological biomarkers; such as NLR, PLR and LMR. This project will involve the design, optimisation and validation of assays for the reliable extraction and analysis of ctDNA and circulating exosomes in these patients. This study will also involve determining the ability of ctDNA, circulating exosomes and systemic immunological biomarkers to predict patient pathological response to pre-operative radiotherapy in patients with LARC. The specific aim of this thesis is expressly to determine whether these circulating biomarkers can be reliably extracted and analysed in patients with LARC.

1.7 Thesis Aims

- To choose and validate appropriate methods for the analysis of ctDNA in patients with LARC
- To analyse sequential ctDNA and cfDNA samples in the context of clinical characteristics and outcomes in patients with LARC
- To develop an assay for the extraction and analysis of exoRNA in patients with LARC
- To analyse sequential exoRNA samples in the context of clinical characteristics and outcomes in patients with LARC
- To analyse standard clinical blood counts and related inflammatory indices in the context of clinical outcomes in patients with LARC

2 Materials, Methods and Assay Validation

2.1 Introduction

In recent years, circulating biomarkers have moved to the forefront of research in patients with cancer. This is at least in part, a result of recent technological advances, including the emergence of technologies such as NGS and ddPCR. The development of NGS has facilitated large scale high-throughput sequencing, whilst ddPCR has allowed highly sensitive and specific analysis of low-concentration circulating biomarkers.

As a circulating biomarker, ctDNA is rapidly approaching common clinical utility for patients with cancer. This is DNA which has been shed by tumour tissue into the circulation of patients with cancer (Diehl et al. 2005). CtDNA is a subgroup of cfDNA which is shed into the circulation by either tumour or non-tumour tissue (Thierry et al. 2010). CfDNA can contain variably high or low levels of tumour-specific variants, depending on a number of factors, such as patient stage (Bettegowda et al. 2014). In patients with locally advanced disease, ctDNA levels and VAFs can be very low and difficult to reliably detect.

Circulating exosomes and their molecular cargo is another biomarker which has been growing in interest and popularity (Lotvall et al. 2014). The role of circulating exosomes in the clinical context is currently not well understood, and thus, is a significant distance away from potential clinical utility. Within our study, we investigate the clinical potential of circulating exoRNAs alongside cfDNA and ctDNA.

As part of this study, we intend to design an analytical workflow in order to investigate cfDNA and ctDNA using NGS and/or ddPCR across longitudinal plasma samples in patients with LARC. In order to achieve this, we needed to decide upon a targeted NGS panel and associated sequencing technology and then determine the limit of detection for this assay using reference standards. We then proceeded to validate the ability of our targeted NGS panel to identify tumour-specific variants in ctDNA. We also determined the limit of detection for ddPCR, which would be used in tandem with NGS throughout this thesis. Finally, we designed an analytical workflow to be used in this study for variant detection and circulating biomarker analysis in patients with LARC.

2.2 Materials and Methods

2.2.1 Patients and Ethics

Throughout this study, patient samples were acquired from two separate sources. Firstly, we consented patients with LARC for our own study through the Wales Cancer Bank (WCB). Secondly, additional samples (both tissue and plasma) from patients with LARC were acquired from the ARISTOTLE clinical trial.

2.2.1.1 Wales Cancer Bank

A research protocol, patient information and subject review was submitted to the WCB to recruit patients with LARC undergoing pre-operative radiotherapy. Patients were invited to give informed consent to participate in this project, to allow access to previously taken formalin fixed and paraffin embedded (FFPE) tumour biopsy tissue, future tumour resection tissue, access to clinical case notes and to provide sequential blood samples during the course of treatment (See Appendix Section 9.1).

Ethical approval was acquired from the WCB in January 2016 for consenting and investigating up to 40 patients with LARC for the investigation of circulating biomarkers (Project 16/001). All patients were seen and consented in Velindre Cancer Centre by the lower gastrointestinal team.

Twenty-one patients were recruited to this study from January 2016 to March 2018. All patients were locally advanced at diagnosis and had >1 sequential plasma timepoint available for circulating biomarker analysis.

All patients underwent radiotherapy with the intention of receiving curative surgery following recovery. However not all patients were able to undergo surgery after the completion of radiotherapy. Patients had either short course (25Gy/5) or long course (45Gy/25) radiotherapy over one week or five weeks respectively (Figure 5).

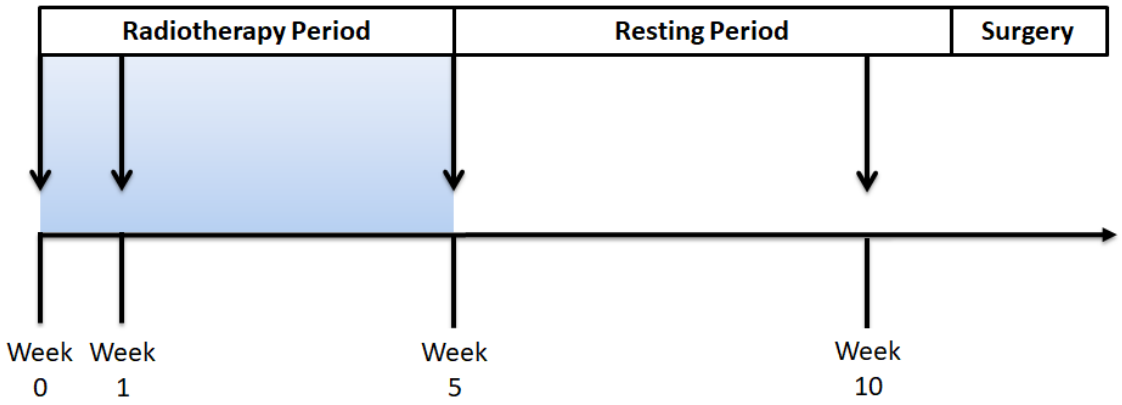

Figure 5: Patient treatment regimen of patients with LARC from the WCB undergoing preoperative radiotherapy plus surgery. Arrows represent hypothesised blood collection timepoints for circulating biomarker analysis

No patients had primary tumour tissue available for DNA extraction from the WCB from this cohort, however, ten patients had excess tumour biopsy DNA available after the completion of clinical genetic testing from the All Wales Medical Genetics Service (AWMGS). Therefore, excess DNA was available for molecular testing from these 10 patients.

Patient clinical data was collected by Professor Richard Adams at Velindre Cancer Centre. Patient identifiers were anonymised by the WCB and given a unique ID before sample extractions and molecular analyses were performed. The full list of clinical factors acquired from these patients is in the Appendices Section 9.2.

2.2.1.2 ARISTOTLE

Patient samples were also acquired from the ARISTOTLE clinical trial for molecular and ctDNA analysis. Each patient was consented for the trial and treated at their respective treatment site. Each patient was made aware that DNA and plasma samples may be used for translational research. For this study, patients were selected according to the availability of patient biopsy, resection and plasma samples.

A total of 19 patient samples were acquired from the ARISTOTLE trial for sequencing and circulating biomarker analysis.

Patients in the ARISTOTLE clinical trial received five weeks of chemoradiotherapy followed by surgery. Surgery was recommended to occur 8-10 weeks after the completion of chemoradiotherapy within this clinical trial (Figure 6). All patients received 45Gy of radiotherapy in 25 doses over five consecutive days (Monday to Friday) alongside capecitabine, which was taken twice daily concurrent with radiotherapy. Patients within a single arm of the trial were also treated with irinotectan once per week for the first four weeks

of radiotherapy. Details on the method of administration and mechanism by which Capetiabine and Irinotecan function were previously described in Chapter 1.

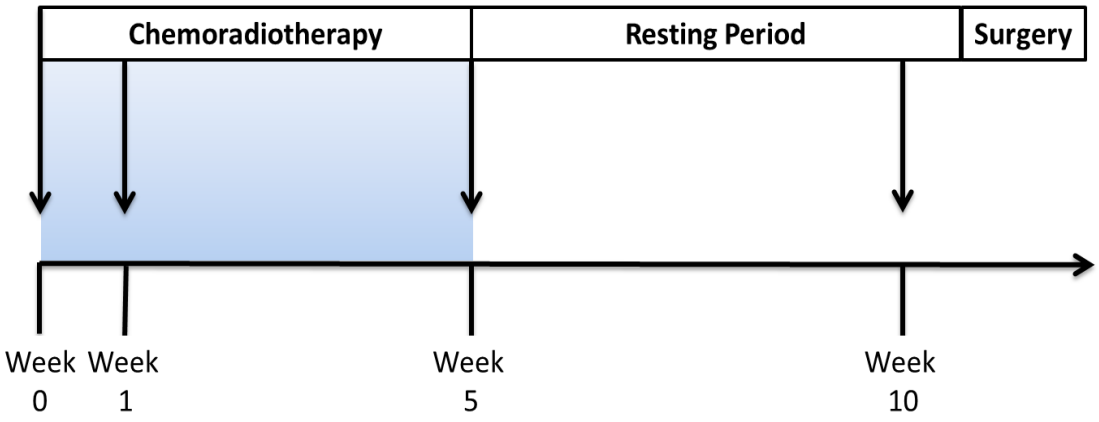

Figure 6: Patient treatment regimen of patients with LARC from the ARISTOTLE clinical trial undergoing preoperative radiotherapy plus surgery. Arrows represent specified blood collection timepoints from the clinical trial for circulating biomarker analysis

2.2.2 Pre-Analytical Sample Handling

2.2.2.1 Wales Cancer Bank

2.2.2.2 FFPE Tissue or Pre-Extracted DNA Collection

No patients had excess FFPE tissue available for non-clinical analysis, however, excess FFPE tumour-derived DNA was available after clinical testing for ten patients. This DNA was extracted by the AWMGS using the Promega Maxwell 16 FFPE Plus Low Elution Volume Purification Kit (AS1130).

2.2.2.2.1 Blood and Plasma Collection

Whole blood (20mL) from WCB patients was collected into EDTA tubes and processed within one hour of collection. Whole blood underwent a double spin protocol to minimalize cell lysis and resulting cellular DNA contamination into plasma.

Double spin protocol: Blood samples were initially centrifuged at 2000xg for 10 minutes at 4°C. The resulting plasma underwent a second centrifugation step at ~4000xg for 10 minutes at room temperature. Plasma was then separated into 1mL aliquots for storage at -80°C. This process was carried out by the WCB.

For all patients, up to 2 x 10mL of peripheral blood was collected for circulating biomarker analysis at each timepoint. Blood samples were collected before treatment (week 0), during radiotherapy and after the completion of therapy during follow up (Figure 5). Collection timepoints were variable due to clinical scheduling.

Plasma samples were transported frozen on wet ice to the Institute of Medical Genetics, Cardiff and Vale University Hospital.

2.2.2.3 The ARISTOTLE Clinical Trial

2.2.2.3.1 Tumour Tissue Collection

Patient FFPE tumour tissue was acquired from the ARISTOTLE clinical trial and processed into 4x5µm thick sections at the Leeds Institute of Cancer and Pathology, St James' University Hospital. A total of four sections on slides from each patient were sent to the AWMGS for NGS analysis.

2.2.2.3.2 Blood and Plasma Collection

Blood samples were collected from each patient at weeks 0 (pre-treatment), 1 (first week of radiotherapy), 5 (last week of radiotherapy) and 10 (during follow-up) as in the timeline displayed in Figure 5. Week 0 (pre-treatment) samples were collected up to two weeks before the initiation of treatment. At week 0, a second whole blood sample was collected for germ-line DNA extraction and analysis.

Samples were collected in either EDTA or Streck tubes (see Table 1 for details). Blood samples collected in EDTA tubes were processed into plasma at each respective site before being collectively shipped on dry ice to the AWMGS. Blood samples collected in Streck tubes were transported at room temperature to the AWMGS and processed into plasma upon receipt. All Streck tubes which took more than three days to arrive were discarded. Samples which were collected in Streck tubes and processed by the AWMGS used the same protocol as described in Section 2.2.2.2.1. The protocol used to process whole blood into plasma from EDTA tubes may have varied between laboratories.

Whether collected in EDTA tubes and shipped as plasma or in Streck tubes and transported at room temperature, all samples which took >3 days to arrive at the AWMGS were excluded from analysis in this study. This factor was a major requirement for patient selection in this study, to try and provide consistent high-quality plasma samples for circulating biomarker analysis.

No information was provided regarding the time interval between blood collection and processing into plasma for samples collected in EDTA at respective treatment sites. Large variations in this time interval can affect sample quality during ctDNA analysis. This is less of a concern for samples collected in Streck tubes, as these can be processed up to three days after collection without significantly impacting sample quality.

Table 1: List detailing which type of collection tube that sequential plasma samples were collected in for each patient from the ARISTOTLE clinical trial

Patient ID	Blood Collection Tube
ARI-166	EDTA
ARI-182	EDTA
ARI-202	EDTA
ARI-239	EDTA
ARI-295	EDTA
ARI-297	EDTA
ARI-306	EDTA
ARI-316	EDTA
ARI-341	EDTA
ARI-346	EDTA
ARI-366	Streck
ARI-373	Streck
ARI-378	Streck
ARI-400	Streck
ARI-403	Streck
ARI-408	Streck
ARI-412	Streck
ARI-413	Streck
ARI-436	Streck

All plasma samples were separated into 1mL aliquots and stored long-term at -80°C.

2.2.3 Sample Extraction

2.2.3.1 FFPE DNA Extraction

DNA was extracted from FFPE tumour tissue using the Promega Maxwell 16 FFPE Plus Low Elution Volume Purification Kit (AS1130) according to the manufacturer's instructions.

The tumour tissue was assessed and macro-dissected to ensure tumour content was >20%. Tumour assessment was performed by Professor Richard Adams. FFPE tissue was scraped from slides using sterile scalpel blades, added to Proteinase K and Promega Incubation Buffer for protein degradation and de-paraffinization. Samples were incubated overnight (approximately 16-18 hours) at 70°C whilst being shaken at 1300rpm in an Eppendorf Mixer C (15158953).

After incubation, Lysis Buffer was added to each sample and mixed thoroughly. Samples were transferred to Promega Maxwell Cartridges and underwent the automated FFPE/Cells DNA protocol on the Maxwell 16 Instrument. The automated process binds DNA to silica clad paramagnetic particles which undergo sequential washes in ethanol before being eluted. All samples were eluted in nuclease free water and quantified using the Invitrogen High Sensitivity Qubit Fluorometer (see Section 2.2.5.1). Extracted DNA was stored at 4°C.

2.2.3.2 Germline DNA Extraction

Genomic DNA was extracted from blood samples by the AWMGS. 1mL of 'buffy coat' was collected from each patient blood sample and stored at -80°C. Genomic DNA was extracted using the Maxwell RSC Instrument (Promega). In this instrument, nucleic acids are captured using paramagnetic Magnesil® particles, which are washed in a series of ethanol washes before being eluted into nuclease free water according to manufacturer's instructions.

Resulting DNA was quantified using the Invitrogen Qubit Fluorometer (Section 2.2.5.1) and stored long-term at 4°C.

2.2.3.3 Cell free DNA Extraction

Plasma was thawed and warmed to room temperature before cfDNA extraction using the QiaAmp Circulating Nucleic Acid kit (Qiagen; Cat No. 55114) according to manufacturer's instructions. CfDNA was liberated from protein carriers within plasma by Proteinase K treatment. Treated plasma was then mixed with Binding Buffer, containing guanidine thycyanate and trometamil, in order to help bind DNA to a QiaAmp MiniElute Silica column. The column was washed consecutively with wash buffers and ethanol, dried and eluted into elution buffer (an RNase free buffer containing a very low concentration of sodium azide). CfDNA was stored long-term at 4°C.

Isolated cfDNA was quantified using the Thermofisher Scientific Qubit Fluorometer (Section 2.2.5.1).

2.2.4 Reference Standards

Reference standards were purchased from Horizon as pre-extracted DNA from cultured cell lines (see Table 2). These were used during experimental validations and limit of detection assays for NGS in Section 2.3.1.2.

Table 2: Product numbers and expected VAFs for variants contained within reference standard (Horizon 2019)

Product ID	Expected VAF for Variants in *EGFR* (%)	Expected VAF for Variants in *KRAS*, *NRAS* and *PIK3CA* (%)
HD777	5.0	6.3
HD778	1.0	1.3
HD779	0.1	0.1
HD776	0.0	0.0

2.2.5 DNA Quantification

2.2.5.1 Invitrogen Qubit 3.0

DNA samples were quantified using the Invitrogen High Sensitivity (HS) double stranded DNA (dsDNA) kit (Q32851) and the Invitrogen Qubit 3.0. The HS dsDNA kit utilises a fluorochrome which specifically intercalates and binds with dsDNA to quantify DNA concentration. Fluorescence levels are then quantified by the qubit 3.0 and normalised to known and supplied standards before quantifying sample DNA. The reported detection range for the HS kit is 0.2 – 100ng/µL.

All DNA samples were quantified immediately after extraction using the qubit according to the manufacturer's instructions. Qubit dsDNA HS buffer was added to dsDNA HS reagent in the ratio 199:1 respectively. The Qubit was always standardised for each batch of extracted samples. For standardisation, 10µL of each Standard was added to 190µL of Qubit buffer/reagent mix as per manufacturer's instructions. These standards were provided within the Qubit kit.

For each patient sample, 2µL of DNA was added to 198µL of Qubit buffer/reagent mix. Each sample/mixture was mixed thoroughly and incubated for two minutes at room temperature before quantification. The sample/mixture was placed into the Qubit 3.0 and the DNA was quantified according the level of fluorescence detected by the Qubit and converted into ng/µL.

2.2.6 Exosome Isolation

Three methods were directly compared for exosome isolation purity and yield in this study. For each method, 1mL of pre-treatment plasma was used for exosome isolation.

2.2.6.1 Total Exosome Isolation Kit

The Total Exosome Isolation (TEI) Plasma Kit (Invitrogen; 4484450) was used as per manufacturer's instructions. Plasma was warmed to room temperature and centrifuged at

2,000xg for 20 minutes followed by another centrifugation at 10,000xg for a further 20 minutes. 5µL of Proteinase K was added to the plasma sample and incubated at 37°C for 10 minutes. The resulting mix was added to 300µL of the Exosome Precipitation Reagent. The mixture was mixed thoroughly and incubated on ice for 30 minutes. The sample mix was centrifuged at 10,000xg for 5 minutes and the supernatant was discarded. The pellet, containing the exosomes, was resuspended in 100µL of PBS and stored at 4°C until processed further.

2.2.6.2 ExoEasy Kit

Exosomes were isolated using the ExoEasy Maxi kit (Qiagen; 76064) according to the manufacturer's instructions. Briefly, 1ml of plasma was filtered using a 0.22µm filter (Millipore; SLGP033RS) and centrifuged at 16,000xg for 10 minutes at 4°C. Resulting plasma was mixed with 1mL of binding buffer and warmed to room temperature. The sample mixture was added to a membrane affinity column and centrifuged at 500xg for 5 minutes. The column was then washed with 3.5mL of wash buffer (XWP) at 3,000xg for 5 minutes. Finally, the exosomes were eluted into 400µL of elution buffer (XE) and stored at 4°C until processed further.

2.2.6.3 ExoSpin Columns

Exosomes were isolated using ExoSpin columns (CellGS: EX04-20) according to the manufacturer's instructions. ExoSpin columns were warmed to room temperature and washed through with 2 x 10mL of PBS. 1mL of plasma was thawed and filtered using a 0.22µm filter. Filtered plasma was added to the ExoSpin column and allowed to drain through by gravity. 3mL of PBS was washed through the column to remove contaminating microparticles. 3.5mL of PBS was then added to the ExoSpin column to elute the column-bound exosomes. Resulting exosomes were collected and stored at 4°C until processed further.

Exosomes were processed within 24 hours of isolation following all three methods.

2.2.7 Exosome Purity Analysis

Resulting exosomes from each method were analysed for sample yield and purity using a particle to protein (P/P) ratio as suggested by Webber and Clayton (2013). In addition to the P/P ratio, we also compared vesicle size and protein concentrations to determine the optimal method with which to proceed. The exosome size and particle concentrations were quantified using Nanosight Tracking Analysis (NTA) from Malvern Panalytical. Protein concentrations were quantified using a micro-Bicinchoninic Acid (micro-BCA) assay (Thermofisher; 23235).

2.2.7.1 Nanosight Tracking Analysis (NTA)

All suspended exosomes were diluted 1/1000 in PBS for particle size and concentration analyses.

The NTA is a commonly used method to calculate vesicle size and concentration. Suspended particles pass through a chamber in which a laser beam shines through. The particles scatter light which is captured by a mounted camera. The particles move under Brownian motion and the NTA uses Stokes Einstein equation to calculate particle diameters based on the speed of movement. The camera recorded, measured and quantified particles which were passed through a flow system at a known speed for a minute in triplicate.

The particle concentration was calculated in particles (P) per mL (P/mL) and particle size was recorded in nm.

2.2.7.2 BCA Assay

A micro-BCA assay calculates sample protein concentrations using a working reagent (WR) to cause a sample colour change based on the protein concentration when incubated at 37°C. The protein concentration of a sample in question is then calculated by comparing levels to a standard curve Bovine Serum Albumin (BSA). BSA standards range from 0–2000ng/mL (See Appendices Section 9.6).

The WR was made up by combining BCA reagents A and B in the ratio 50:1. 150µL of WR was added to 150µL of each standard and 150µL of isolated exosomes. The samples and standards were incubated at 37°C for 30 minutes and read using a spectrophotometer at 562nm within 10 minutes of incubation.

Protein concentrations were recorded as µg/mL.

2.2.7.3 Particle to Protein Ratio

Particle to protein ratios were calculated using the particle concentrations (P/mL) from the NTA and the protein concentration (µg/mL) from the micro-BCA assay. The particle to protein ratio was calculated using the following formula; (P/mL)/(µg/mL). Purity levels of resulting exosomes were compared to recommended concentrations by Webber and Clayton (2013).

2.2.8 Exosome-Derived DNA Extraction

For exoDNA extraction, exosomes were isolated from 1mL of plasma using the ExoSpin Kit as described in Section 2.2.6.3.

2.2.8.1 Vesicle Treatment

Exosomes were treated using Proteinase K and DNase I to ensure any extracted nucleic acids were intracellular, rather than extracellular.

Isolated exosomes were incubated at 60°C for 60 minutes in 0.2X Proteinase K (Thermofisher: 25530049). Proteinase K was heat degraded by incubation at 95°C for 10 minutes before cooling down to room temperature. Following this, 0.01X Units of DNase I (Thermofisher; ENO521) was added to the sample and incubated at 37°C for 60 minutes. DNase I was denatured by heating exosomes at 75°C for 15 minutes before cooling back down to room temperature.

ExoDNA was extracted using the QIAamp Circulating Nucleic Acids Kit as previously described in Section 2.2.3.3 and quantified using the Invitrogen dsDNA HS Qubit (see Section 2.2.5.1).

2.2.9 Exosome-Derived DNA Analysis

2.2.9.1 Droplet Digital PCR for Variant Detection

Variant detection was performed using the Bio-Rad ddPCR as described previously in the Materials and Methods.

2.2.9.2 PCR for the Analysis of Variable Lengths of DNA

ExoDNA length was analysed using primers pairs to generate amplicons of variable lengths in the *BRCA1* gene by PCR. The primer pairs are described in Table 3. The *BRCA1* gene was used in this study as it is not relevant in the context of rectal cancer and previously optimised primers were available.

Table 3: Primer sequences and amplicon length for the analysis of variable lengths DNA extracted from exosomes

Amplicon Length (bp)	Direction	Primer Sequence
224	Forward	AGCCTTCATCCGGAGAGTGTAG
224	Reverse	CCAGTCTTGCTCACAGGAGAGA
500	Forward	TGTCTGTTGCATTGCTTGTGTT
500	Reverse	CCGCACATTTCTCATGTTGTAGC
827	Forward	CTACTTTGGATTTCCACCAACACTG
827	Reverse	GGTAAATTCACCCATGTGAGACAAG

The reaction was set up using the appropriate primers as described in Table 4 and the thermocycler according to the conditions depicted in Table 5.

Table 4: Reagents and reagent volumes for the amplification of exoDNA and cfDNA

Reagent	Volume (µL)
10X Anglian Buffer	2.5
dNTPs (5mM)	1
Primers	1
Water	15.4
Taq	0.1
DNA (5ng)	5
Total	25

Table 5: Thermocycler condition for PCR amplification of ExoDNA and cfDNA

Temperature (°C)	Time (Minutes)	Cycles
95	7	1
95	1	35
X	1	
72	1	
72	7	1
4		1

X – Primers which produced 224bp amplicons worked optimally at 62°C, whilst the primers producing amplicons of 500bp and 827bp worked optimally at 63°C.

2.2.9.2.1 Bioanalyzer 2100

Amplified exoDNA was analysed using the Agilent High Sensitivity (HS) DNA Assay on the Agilent Bioanalyzer 2100 according to the manufacturer's instructions. Briefly, Gel Dye mix was created by combining the HS DNA dye concentrate with HS DNA gel matrix. This was mixed and centrifuged at 2240xg for 10minutes through a spin filter.

The loading dye mix was then loaded into an Agilent HS DNA chip followed by the HS DNA marker and then ladder/samples in respective wells. The chip was then run on the Agilent Bioanalyzer 2100.

2.2.10 Systematic Reviews

A systematic review was performed to identify microRNAs of interest in patients with CRC and to predict response to radiotherapy.

The search engine 'Pubmed' was used to look for the terms 'Circulating microRNA rectal cancer' and 'Circulating microRNA colorectal cancer'.

We excluded reviews and meta-analyses from inclusion. We also excluded studies which focussed on the application of novel methodologies, were not in English, or looked at non-colorectal cancers (See results Section 5.2.3.1).

2.2.11 Exosome-Derived RNA Extraction

2.2.11.1 Developed ExoRNEasy Protocol

Extraction of exoRNA was performed using the ExoRNEasy Plasma Kit (Qiagen; 77064) from 3mL of plasma according to the manufacturer's instructions. This kit is an extension of the Qiagen ExoEasy kit described previously in Section 2.2.6.2 to include exosome lysis and RNA extraction. The protocol has been adapted based on suggestions from Enderle and colleagues (2015) and in-house changes as described in the Appendices Section 9.7.

2.2.11.1.1 Exosome Isolation

Plasma was pre-filtered (0.22µm) and centrifuged at 16,000xg for 10 minutes at 4°C. After centrifugation, the plasma was mixed with 3mL of binding buffer (XBP) and warmed to room temperature.

The mixture was warmed to room temperature and added to an ExoEasy Spin column and centrifuged at 500xg for 1 minute. The column was then washed by centrifuging 10mL of wash buffer XWP through the column at 3,000xg for 5 minutes. Column-bound exosomes were treated with RNase buffer which consisted of 6.25µg/mL of RNase A diluted into 400µL of wash buffer XWP. The RNase mix was briefly centrifuged through the column before the flow through was reapplied and incubated on the column for 30 minutes at room temperature. The RNase was washed through the column using 18mL of wash buffer XWP and spun at 3,000Xg for 5 minutes. Finally, 400µL of elution buffer (XE) was added to the column and incubated for 1 minute before being centrifuged at 500xg for 5 minutes. The eluent was re-applied to the column incubated for 1 minute at room temperature before being centrifuged for 5 minutes at 3,000xg.

2.2.11.1.2 Exosome-Derived RNA Extraction

Treated exosomes were lysed in 1mL of Qiazol. Qiazol is based on Trizol and is made up of phenol and guanidine isothiocynthate (Chomczynski 1993) and is used to lyse cells whilst maintaining RNA integrity.

The lysed sample was vortexed briefly and incubated at room temperature for 5 minutes and 180µL of chloroform was added to the lysed sample. The sample was briefly shaken, incubated at room temperature for 3 minutes and then centrifuged at 12,000xg for 15 minutes at 4°C.

The top aqueous phase (approximately 800uL) was aspirated and added to 2X volume of 100% ethanol and mixed briefly. The mixture was then spun through a Qiagen RNeasy MiniElute spin column to extract RNA. The column was serially washed with wash buffers (1 x RWT, 2 x RPE and 1 x 100% ethanol.)

The RNEasy column was then dried by full-speed centrifugation and eluted into 14µL of RNase free water.

Extracted RNA was stored on ice and reverse transcribed immediately after extraction.

2.2.12 Reverse Transcription

2.2.12.1 MicroRNA Reverse Transcription

2.2.12.1.1 TaqMan MicroRNA RT Kit

MicroRNA was reverse transcribed using the TaqMan MicroRNA RT Kit according to the manufacturer's instructions. Reagents were added to total RNA as shown in Table 6.

Table 6: Reagents volumes for microRNA reverse transcription using the TaqMan MicroRNA RT Kit

Reagent	Reagent Volume (µL)
RT Primer Pool	12
dNTPs (100mM)	0.6
Reverse Transcriptase (50U/µL)	6
10X RT Buffer	3
RNase Inhibitor (20U/µL)	0.38
NFW	2.02
Sample RNA	6
Total	30

The RT Primer Pool is a combination of all genes of interest making up a total volume of 12µL.

2.2.13 Next Generation Sequencing

2.2.13.1 Principles of Next Generation Sequencing with the Ion Torrent Proton

Whole genome, exome and targeted sequencing methods have become much more common and easy to apply, largely due to the emergence of next generation sequencing (NGS) technologies. In 2011, several bench-top sequencers became commercially available, including

the Ion Torrent PGM and the Illumina Miseq (Liu et al. 2012). The following years saw numerous studies comparing the accuracy of single nucleotide polymorphism (SNPS) calling between these two methods (Liu et al. 2012; Misyura et al. 2016).

Ion Torrent Sequencers, including the PGM and the Proton, use semiconductor technology to sequencing DNA. In contrast, Illumina sequencers, including the Miseq, use fluorescence-tagged DNA bases with reversible terminators (Rothberg et al. 2011; Liu et al. 2012). Semiconductor sequencers measuring the change in pH, if any, which occurs when a base is added to a string of oligonucleotides, releasing hydrogen ions (Liu et al. 2012). This method was developed with the intention of overcoming certain limitations of competitor technologies; such as the need for imaging technologies to detect fluorescence and specialised nucleotides (Bentley et al. 2008).

However, semiconductor sequencers have their own limitations. Though a change in pH can be read simply with the addition of a small number of identical bases (1-4), when larger strings of homopolymers are sequenced, there is an increased error-rate in homopolymer length quantification. Thus, the false calling of indels may be observed. Furthermore, in comparison to competitors, the Ion Torrent methods generally have a higher cost per Gb of data and require more hands-on time, increasing the chances of human error (Garrido-Cardenas et al. 2017). The Ion Torrent has a low machine and operational cost in comparison to other technologies with a significantly shorter sequencing time (Garrido-Cardenas et al. 2017).

In this study, the Ion Torrent Proton platform (LifeTechnologies, Thermofisher Scientific, UK) was applied to identify tumour-specific variants in patient tumour tissue and/or matching ctDNA. Detected variants could then be used for sequential ctDNA analysis. The process of decision-making for sequencing technology and targeted panel are described in Section 2.3.1.1 in this chapter. NGS analysis using the Proton comprises of three main phases; these are library preparation, template preparation, and sequencing and data analysis. Here we describe the principles of each phase of NGS analysis, with a more detailed protocol to follow.

2.2.13.1.1 Library Preparation

Specific genes and variant hotspots of interest are amplified based on primers and probes included in a targeted NGS panel. Samples are initially cleaned and barcoded before being bound to magnetic beads. Once bound to these magnetic beads, DNA amplicons are cleaned in sequential washes of ethanol. A second-round of DNA amplification also occurs before a

second-round of sequential ethanol washes and the final DNA libraries are eluted and combined into a single pool for template preparation.

2.2.13.1.2 Template Preparation

Pooled DNA libraries are then separated into oil droplets with PCR reagents and ion sphere beads (similar to the ddPCR process). Within these droplets, DNA undergoes emulsion PCR and is bound to a bead directly by their respective barcodes. Where only a single DNA template is present within a droplet, the sphere is deemed monoclonal and appropriate for sequencing. Where more than one different template exists within a droplet, the ion sphere is deemed polyclonal and no sequencing occurs for these spheres. This highlights the important balance of fully loading spheres with DNA without significantly increasing polyclonality. Spheres are then loaded onto a sequencing chip which contains 70,000,000 individual wells, one for each ion sphere.

2.2.13.1.3 DNA Sequencing

Loaded microchips are then inserted into the Ion Torrent Proton for sequencing. The Proton washes the chip sequentially with a single base at a time and reads any pH changes, indicating the addition of said washed bases. As a nucleotide is incorporated onto a chain, a hydrogen ion is released, decreasing the pH, which is detected by the sequencer. The raw sequencing files are generated into FASTQ files for each respective barcode for downstream bioinformatic analysis.

2.2.13.2 Next Generation Sequencing Protocol

2.2.13.2.1 Library Preparation

Patient DNA samples were amplified using the Ion Ampliseq Cancer Hotspot Panel Version 2 (CHPv2; Catalogue Number: 4475346). For the first amplification step, 20-30ng of FFPE DNA or the maximum volume of 12µL of cfDNA was amplified using the CHPv2 primer pool. This difference is specified because cfDNA samples were unlikely to contain such a large amount of DNA, whereas this was reliably achievable from FFPE tumour tissue.

For the first amplification step, two different numbers of cycles were used depending on DNA concentration. For FFPE Tumour DNA where 20ng was extracted, 21 amplification cycles were used, and for cfDNA samples 25 cycles were used due to low concentrations. The remaining library preparation steps were performed using reagents from the Ion Ampliseq Library Kit 2.0 (Catalogue Number: 4475345) as per manufacturer's instructions.

Amplified DNA was treated for primer digestion with the FuPa reagent before being barcoded and bound to magnetic beads for purification in sequential ethanol washes. DNA libraries then underwent a second round of DNA amplification before being purified in another round of sequential washes in ethanol. The resulting DNA libraries were eluted into Lo-Tris EDTA (Lo-TE), diluted to 100pM and combined into a single DNA pool as part of template preparation.

2.2.13.2.2 Template Preparation

Prepared libraries were quantified using the Qubit 3.0 Fluorometer. 5µL of eluted DNA library was added to 195µL of Qubit buffer/reagent mix for quantification. Individual samples were diluted down to 100pM (15ng/mL) in Lo-TE and pooled together as per the manufacturer's instructions. The DNA concentration of NTCs were quantified using the Qubit but excluded from sequencing as long as they were negative. Once DNA libraries from all samples had been pooled together, 10µL of the DNA libraries was added to 190µL of buffer/reagent mix for quantification with the Qubit.

The pooled libraries were diluted further down to 75pM and loaded into the Ion Torrent Chef with the Ion Pi Hi-Q Chef Kit for further template preparation as per manufacturer's instructions. Within the Ion Torrent Chef, samples were amplified, purified and bound to ion spheres which were loaded into an Ion 318 sequencing chip.

2.2.13.2.3 Sequencing

2.2.13.2.3.1 Ion Ampliseq Cancer Hotspot Panel Version 2 Sequencing Optimisation

A Run Plan was generated for each individual sequencing run. The run plan was set up on the Cardiff University Ion Torrent Server (IP Address: 131.251.159.10). The Ion Torrent Proton was washed with water and bleach before preparing for sequencing. Wash buffers and sequencing reagents were prepared as per manufacturer's instructions. The Ion Torrent Proton was then calibrated to ensure the pH levels were optimal for sequencing. Once the sequencer had been calibrated, loaded sequencing chips were taken from the Ion Torrent Chef were inserted into the sequencer within four hours of Chef protocol completion. Sequencing occurred over 150 minutes and resulting data was uploaded to the Ion Torrent Server for variant analysis.

2.2.14 Next Generation Sequencing Analysis

2.2.14.1 Ion Torrent Variant Caller

In this study, sequencing data was processed and converted into a Variant Call Format (VCF) file using the Ion Torrent Variant Caller. Ion Torrent Variant Caller is a plugin connected directly to the Ion Torrent Server and designed specifically to analyse Ion Torrent Proton NGS data.

The Ion Torrent Variant Caller will use bioinformatic tools to align sequencing data to the reference genome, calculate the quality of data under specific parameters before identifying variants. The tools used to achieve this would include Samtools and bcftools. Variants are then filtered by standard and specified features and thresholds using VCFfilter. Such features would include VAF and sequencing quality (EdgeBio 2019).

2.2.14.2 Variant Filtration and Determining Clinical Significance

2.2.14.2.1 Variant Filtration Principles

In both a research and clinical setting, generated VCF files are examined to ensure all called variants are genuine and clinically relevant in the context of the disease of interest; in this case rectal cancer. To determine whether our variants were genuine, as opposed to sequencing artefacts, we analysed numerous properties of the VCF file, such as sequencing depth and VAF. We also annotated variants before analysing their clinical significance.

Sequencing artefacts are found where a recorded DNA sequence does not represent the actual genotype of the sample due to the accumulation of errors throughout the NGS process. Such artefacts can occur through errors in PCR amplification or during sequencing. Studies have previously demonstrated that the DNA polymerase enzyme Taq has an approximate total error rate of 1 /10,000 bases (Potapov and Ong 2017). Therefore, as our NGS protocol includes a pre-amplification step with Taq, some sequencing errors may have occurred at this point. Studies have also demonstrated that all NGS methods, including the semi-conductor sequencing, have a sequencing error rate; this is the likelihood that a base can be read incorrectly (Quail et al. 2012). Although the sequencing quality begins relatively high for the first few bases, the quality of each base sequenced will naturally decline as the amplicon becomes longer, thus errors can occur after a certain length. For this reason, NGS methods are generally paired with shorter PCR amplicons. Sequencing artefacts can also be caused by primer dimer or the formation of similar biproducts of PCR and can inappropriately align to the reference genome, appearing to be genuine DNA reads. This aspect is less of an issue with targeted NGS panels, in comparison to whole exome or genome sequencing.

A greater concern when analysing FFPE tissue is that sequencing artefacts may also be caused by sample fixation. This is more common in FFPE tissue, where cross-links can occur between DNA strands as a result of the formalin fixation process. This can cause both DNA fragmentation and cytosine deamination, where cytosine bases appear as an uracil base during sequencing. Therefore, there is an increase frequency of C>U/T and G>A changes in these

sample types (Ivanov et al. 2017). This should be considered in all future NGS analyses on FFPE tissue samples.

To combat these potential issues, there are bioinformatic filters which can be applied during analysis to help avoid mistaking sequencing artefacts for genuine variants. Factors by which data is filtered include sequencing quality, the number of total and mutant reads, VAF and direction bias. The Ion Torrent Variant Caller will filter out any bases with a quality (Q) score below a certain value (Q20 in this pipeline), and trim sequences to remove the low-quality bases at the end of long amplicons before generating a VCF file. This pipeline will also compare the sequenced DNA to a BED file for specific targeted panel, which shows where DNA should be amplified and detected, and will exclude any sequencing not matching these expected regions.

We can also manually filter out variants from the VCF file based on factors such as sequencing coverage and VAF. Sequencing depth is the number of times that a base in a specific location in the genome is sequenced. Whenever a base is sequenced, this is referred to as a read and is denoted by an X (e.g. a genomic location read ten times has 10X coverage). There are guidelines as to the number of reads which are recommended for variant detection, for both the mutant base and the number of total reads at a single location when using NGS (Jennings et al. 2017). These are described further in Section 2.2.14.2.2.

VAF is the proportion at which a base change occurs in a specific location compared to the total number of reads at that location and is often displayed as a percentage. As low numbers of sequencing artefacts can filter through to the VCF file, a minimum VAF is often recommended to exclude the 'background noise' of sequencing artefacts. The minimum VAF applied for this analysis was based on the limit of detection performed in Section 2.3.1.2.

Once sequencing artefacts have been removed and only genuine variants remain, variants are then annotated. This process involves aligning a variant's genomic location to the reference genome and identifying whether this gene occurs within a gene, in the coding region of said gene, whether this causes a synonymous or non-synonymous change to the coding sequence and so on. Many programs will even highlight if a variant has been identified in the germline of general populations or alternatively, frequently occurring somatically in certain cancer types before. This annotation process can remove variants which are unlikely to influence gene function. Variants which pass all filtration processes were investigated for clinical significance as described in Section 2.2.14.2.3 of this chapter.

2.2.14.2.2 Variant Filtration Protocol

Variants were initially filtered as described in Figure 7 before being classified as clinically significant or benign. Firstly, variants were filtered by VAF according to the sample type. In cfDNA variants <1.0% and >90% were removed, due to the likelihood of these being sequencing artefacts and homozygous SNPs respectively. Variants at 40-60% were also excluded as these were also likely to be heterozygous germline variants. For FFPE tumour tissue, variants <10% and >90% were removed for the same reasons. In FFPE tissue, variants at 40-60% VAF were investigated further and validated as somatic using germline tissue where possible. If no germline tissue was available, these variants were included and would be analysed further for variant clinical significance.

Next, variants were filtered by coverage, with ctDNA samples and FFPE tissue requiring ≥1000 and ≥250 reads to pass respectively. A minimum of ten mutant reads (five in each direction) were deemed necessary for further investigation. This initial minimum value is based on guidelines from Jennings and colleagues (2017) for somatic variant detection using targeted NGS panels. This value would vary based on the total number of sequencing reads acquired and limit of detection as described in Section 2.3.1.2 in this chapter. Thus, the minimum of ten total reads was applied to facilitate the removal of sequencing artefacts in regions of poor coverage. Thereafter, variants with ≥90% direction bias were also removed, due to an increased likelihood of these containing artefacts.

Remaining variants were then annotated using ANNOVAR (Annotate Variation), and non-coding and non-splice site affecting variants were also removed, as these were less likely to be pathogenic and/or clinically actionable. Furthermore, once variants had passed each of these filtration steps, each variant was visually inspected in the BAM file using the Integrated Genome Viewer (IGV) software to confirm they were not sequencing artefacts. To this end, samples and variants were compared to the wild type and positive controls on the same sequencing chip to identify the presence of any commonly occurring sequencing artefacts in each sequencing run.

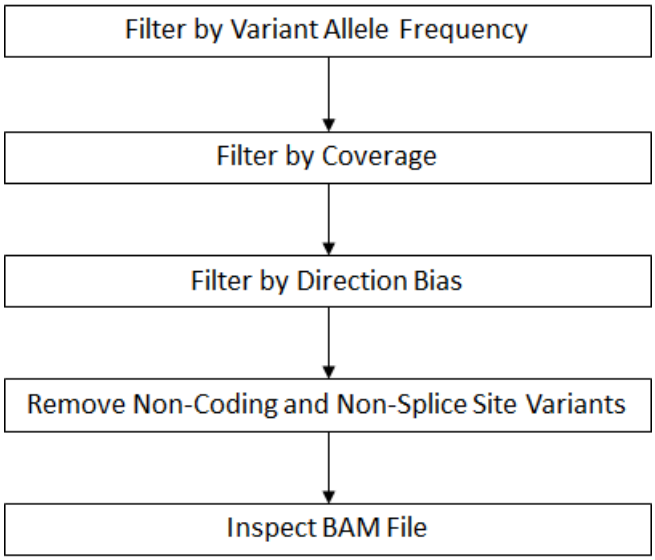

Figure 7: The Filtration Process for Variants Selected in the Next Generation Sequencing Analysis Pipe-Line

After this, variants were classified as clinically significant or benign based on recently published clinical guidelines as detailed below.

2.2.14.2.3 Determining Clinical Significance

With the rapid development of NGS assays and their implementation in the clinical setting in recent years, the decision as to which variants were to be reported also had to adapt and develop. For clinical laboratories or assays investigating disease-causing germline variants, guidelines have recently been published by the American College of Medical Genetics (ACMG) (Richards et al. 2015). These guidelines provided a tiered system by which germline variants can be classified as very strongly pathogenic (PVS1) or stand-alone benign (BA1) or a wide range in between based on specific evidence. Using these guidelines, variant classification requires two pieces of evidence. Firstly, there must be previously demonstrated evidence of the gene's relevance or involvement in the disorder of interest. Secondly, there must be evidence demonstrating that the variant in question detrimentally impacts gene function. This can either be in the form of predictive *in-silico* tools or *in vitro* or *in vivo* functional assays.

More recently, a collaboration between the Association for Molecular Pathology (AMP), the American Society of Clinical Oncology (ASCO) and the College of American Pathologists (CAP) have published similar tiered guidelines for the classification of somatic variants in cancer (Li et al 2017). These tiers vary from Tier I (strong clinical significance) to Tier IV (benign/unlikely to have clinical significance). Unlike germline variant classification, the somatic variant guidelines are based on the clinical 'actionability' or relevance of a variant or gene in the context of a disease, rather than the pathogenic effect of variants on gene function alone. Similar published

institutional practice guidelines from SVC (Sukhai et al 2016) and PHIAL (Van Allen et al 2016) have also based their tiered classification systems on clinical actionability rather than variant functionality alone. Although there were distinct similarities between the three systems mentioned, we evaluated detected variants based upon the guidelines published by Li and colleagues (2017).

In accordance with these guidelines, any of our detected variants which can be used to predict response or resistance to any FDA-approved therapies to rectal cancer were classified as Tier I. In contrast, Tier II variants (potential clinical significance) had to be indicative of diagnosis or prognosis (from 'convincing' published data but not necessarily published guidelines) or approved as a therapeutic biomarker for a different tumour type (Li et al. 2017a). For our variants of interest, this was performed using MyCancerGenome or screening active clinical trials using ClinicalTrials.gov.

For variants to be classified as either tier, they also had to be absent or at very low frequencies in population databases, such as Exome Sequencing Project (ESP), the 1000genomes project and Exome Aggregation Consortium (ExAC). We also ensured these variants were frequently detected in patients with rectal cancer using data from the Catalogue of Somatic Mutations in Cancer (COSMIC), cBioportal and The Cancer Genome Atlas (TCGA) Network. We further ensured that each variant was predicted to be damaging to protein function using the *in-silico* tools Polyphen2, Combined Annotation-Dependent Depletion (CADD) and MutationTaster. Finally, all variants were verified as being related to rectal cancer according to MyCancerGenome (Figure 8).

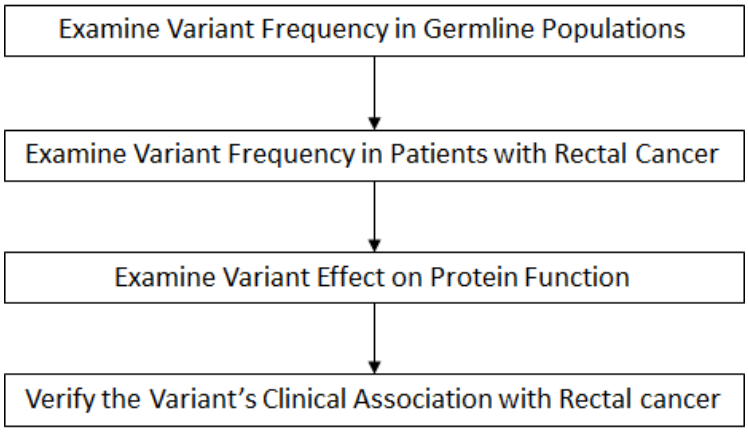

Figure 8: Flowchart describing the process of variant classification in accordance with guidelines from Li and colleagues (2017)

Any variants which did not fit these criteria will either have unknown significance (Tier III) or be benign (Tier IV) according to the guidelines from Li and colleagues (2017). Only variants which had strong clinical significance (Tier I) or potential clinical significance (Tier II) were used in future ctDNA or tumour heterogeneity analysis in subsequent chapters. Due to the lack of evidence supporting their role in tumourigenesis, variants classified as Tier III or IV were excluded as these are more likely to be 'passenger' rather than 'driver' variants. This process will have also removed germline variants which inappropriately passed the initial variant filtration process.

2.2.14.3 Molecular Analysis

2.2.14.3.1 Matching Paired Samples

For all patients with >1 sample being analysed using NGS, the presence of hetero- and homozygous germline SNPs were used to ensure paired samples are derived from the same patient by ensuring all germline SNPs were detected in both samples at equal VAFs in their respective VCF files.

2.2.14.3.2 Calculating Heterogeneity Scores

Tumour heterogeneity scores were calculated as previously described by Normanno and colleagues (2015) using the following formula.

$$Tumour\ Heterogeneity\ Score = 2\ X \frac{100\ X\ VAF(\%)}{Neoplastic\ Cell\ Content}$$

All VAFs were acquired NGS data unless specified otherwise. Neoplastic cell content was calculated using a hemotoxin and eosin stain slide by Professor Richard Adams.

2.2.15 Droplet Digital PCR

2.2.15.1 Principles of Droplet Digital PCR

Emulsion-based ddPCR was chosen for this study due to the reportedly greater sensitivity, reproducibility and precision offered over similar methods such as qPCR (Huggett and Whale 2013). This method has built upon the same principles as qPCR, a method which enables real-time quantification of amplified DNA by using oligonucleotide probes which are tagged with a fluorescent reporter and a quencher (Gut et al. 1999). This oligonucleotide probe can be ~20bp long and is complementary to a specific DNA sequence of interest.

Oligonucleotide probes contain fluorophore reporters which naturally emit light at a specific wavelength which varies between different fluorophores. The two molecular reporters used in

this study were 6-carboxyfluorescein (FAM) and hexachlorofluorescein (HEX) which emit light at a wavelength up to 525nm and ~556nm respectively. The different wavelengths at which light is emitted and consequently absorbed from each fluorophore can be used to distinguish between the two different reporters. Thus, FAM and HEX probes are normally associated with separate genes or genotypes in molecular analysis, in this case they were each used to represent mutant and wild-type DNA respectively. These molecular reporters are often coupled with a quencher on their respective oligonucleotides.

A black hole quencher is a molecular beacon which efficiently absorbs fluorescence. When close to FAM or HEX, the black hole quencher absorbs the emitted fluorescent light, making it undetectable to the user. The quenchers are thus generally located on the opposite end of a hairpin loop from their respective fluorophores on oligonucleotide probes. However, during PCR, when a probe containing both the reporter and quencher binds to complementary DNA, reporters and quenchers are separated from each other through the activity of Taq polymerase enzyme. This separation of the fluorophore from the quencher allows the reliable detection of fluorescence from the reader. Thus, when a mutant (FAM) or wild-type probe (HEX) binds to DNA, the probe is broken down by the enzyme and the reporter is separated from the quencher, allowing the detection of FAM or HEX, and thus, mutant or wild-type DNA (see Figure 9).

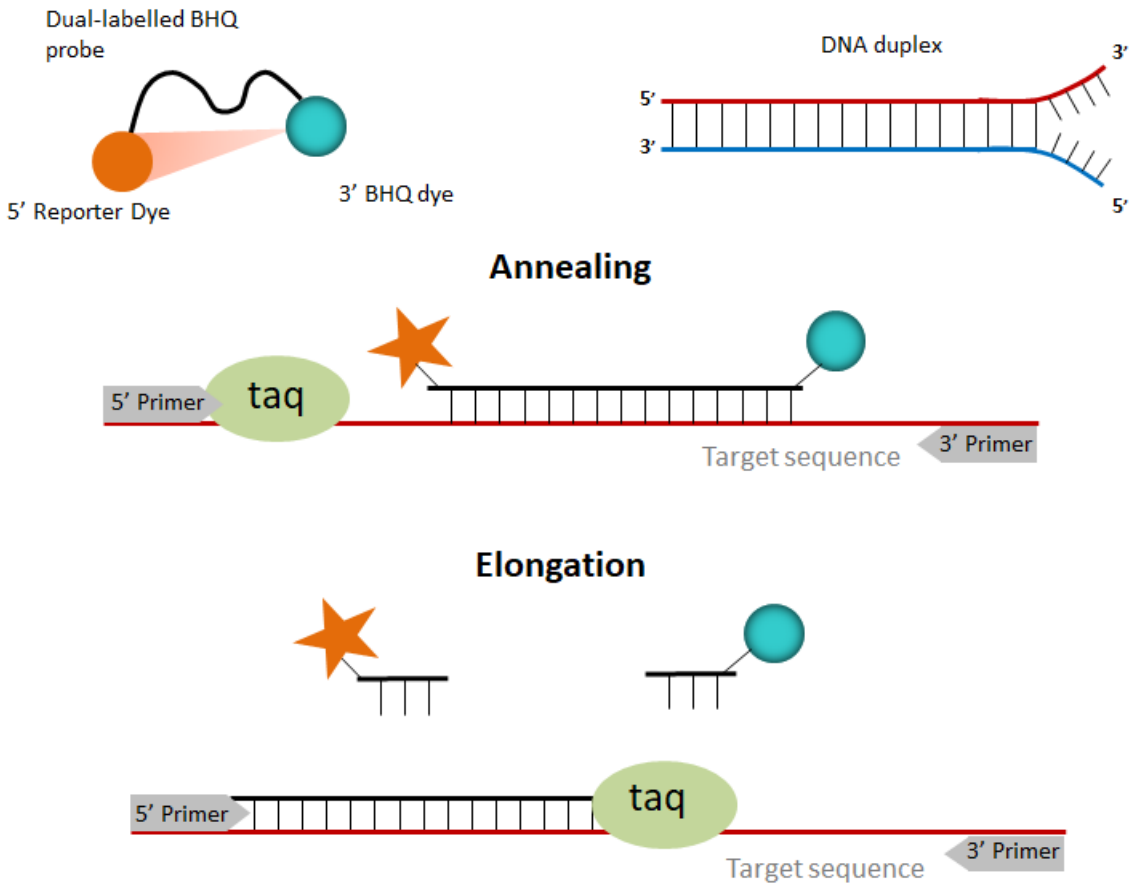

Figure 9: Illustration of the mechanism of DNA quantification using fluorescent reporter labelled probes and black hole quenchers during ddPCR analysis.

As stated previously, digital PCR can provide improved sensitivity, precision and reproducibility as a result of the reaction partitioning which defines it. The Bio-Rad ddPCR system partitions a single qPCR reaction into ~15,000-20,000 smaller compartments/droplets using emulsion chemistry. Alternative forms of digital PCR compartmentalise reactions by using solid partitions on chips (Hindson et al. 2011; Kreutz et al. 2011). Furthermore, emulsion-based ddPCR can provide a greater number of partitions at a generally lower cost. Overall, ddPCR can allow greater sensitivity due to the increased number of partitions and is less prone to being affected by PCR inhibitors (Huggett and Whale 2013).

ddPCR does have inherent limitations, largely involving the concentration of input DNA. A lower input DNA concentration can result in a smaller number of DNA-positive reactions and decreased confidence in the ability of the user to detect or quantify ctDNA. Overloading a ddPCR assay limits the reliability of quantified ctDNA. As each reaction is designed to only

contain one single DNA droplet, when the ddPCR is overloaded more than one DNA molecule may be contained in each compartment. Therefore, the quantification of ctDNA can become less reliable.

Throughout this project, ddPCR was used to quantify VAF in ctDNA and gene expression levels of exosome-derived microRNA exoRNA due to the high sensitivity of the assay and the low concentrations of each of these biomarkers. Primers and probes for specific tumour variants were designed in-house and ordered from Sigma-Aldrich.

In ctDNA samples, the Bio-Rad ddPCR system was applied to validate NGS findings in FFPE tumour tissue and track variants of interest in sequential plasma samples. Details regarding the primers and probes used for ddPCR analysis of exoRNA are described in Chapter 5.

2.2.15.2 Primer and Probe Design

Primers and probes were designed using Primer3plus (Untergasser et al. 2007) based on the human reference genome (hg19) from Ensembl.org (Hubbard et al. 2002). Genetic sequence surrounding the variant in question (>70bp in both directions) was acquired from Ensembl.org and pasted into Primer3plus. We then allocated specific conditions for primer design (such as optimal annealing temperature and length of amplicon). Pimer3plus then generated numerous suggestions for primer and/or probe designs using the information provided. We then confirmed the predicted primer pair conditions and specificity using the *in-silico* browser from UCSC.org (Kent et al. 2002). All probes were designed to have a melting temperature >3°C greater than associated primers. All oligonucleotides were designed with melting temperatures between 50-60°C. Each probe contained a 3' black hole quencher and a 5' fluorescent probe. All probes detecting variants had 5' FAM probes whilst wild-type probes contained a 5' HEX probe. For the full list of designed primers, see Table 7.

2.2.15.3 Primer and Probe Preparation

For DNA analysis, primers and probes were made up to 100µM in TE according to manufacturer's instructions. Eluted oligonucleotides were then diluted to a concentration of 10µM and 5µM respectively and combined into a 20X primer/probe mix (Table 7). For exoRNA analysis, primers and Probes were purchased from Thermofisher Scientific for four microRNAs of interest and a reference gene (RNU6B). See Table 8 for product details.

Table 7: Sequences of all primers and probes for sequential ctDNA analysis designed using Primer3plus

Gene	Variant	Forward Primer	Reverse Primer	Mutant (FAM Probe)	WT (HEX) Probe
NRAS	p.G13R†	GTGTGAAATGACTGAGTAC	TCTGGATTAGCTGGATTG	TCC[+C]AA[+C]AC[+G]ACC[+T]GCTCC	TCC[+C]AA[+C]AC[+C]ACCT[+G]CTCC
APC	p.E1379X	CTCAGACACCCAAAAGTC	GACAGAAGTACATCTGCTAA	CACTATGTTCAGTAGACCCCACTC	CACTATGTTCAGGAGACCCCACTC
KRAS	p.G12V†	AGGCCTGCTGAAAATGACTG	TTGGATCATATTCGTCCACAA	TGCCTACGCCAACAGCTCCAACT	TGCCTACGCCACCAGCTCCAACT
KRAS	p.G12C	AGGCCTGCTGAAAATGACTG	TTGGATCATATTCGTCCACAA	TGCCTACGCCACAAGCTCCAACT	TGCCTACGCCACCAGCTCCAACT
KRAS	p.Q61H†	CCTGTCTCTTGGATATTCTC	AGTCCTCATGTACTGGTC	CTGT[+A]CT[+C]CTC[+A]TGA[+C]CTGCTGT	CTGT[+A]CT[+C]CTC[+T]TGA[+C]CTGCTGT
TP53	p.Q165X	CTGTGGGTTGATTCCA	CTCACAACCTCCGTC	CTGTG[+A]CT[+A]CT[+T]GT[+A]GAT	CTGTG[+A]CT[+G]CT[+T]GT[+A]GAT
TP53	p.R175H	CCATCTACAAGCAGTCAC	GAGCAATCAGTGAGGAATC	TTGTGAGGCACTGCCCCCAC	TTGTGAGGCGCTGCCCCCAC
TP53	p.R213X	CTCCTCAGCATCTTATCC	TCAGGCGGCTCATAG	CACCACACTATGTAGAAAAGTGTTTCT	CACCACACTATGTCGAAAAGTGTTTCT
TP53	p.C242Y	GCCTGTGTTATCTCCTAG	CAGTGTGATGATGGTGAG	CCGCCCATGTAGGAACTGTT	CCGCCCATGCAGGAACTGTT
TP53	p.G245S†	CACCATCCACTACAACTAC	GAGTCTTCCAGTGTGATG	CGGT[+T]CAT[+G]C[+T]GCC[+C]ATGCAG[	CGGT[+T]CAT[+G]C[+C]GCC[+C]ATGCAG
TP53	p.G266V	GACCTGATTTCCTTACTG	GGAGATTCTCTTCCTCTG	AATCTACTGGTACGGAACAGCTT	AATCTACTGGGACGGAACAGCTT

Table 8: Product codes for each microRNA analysed, including Gene Name, Assay ID, Product Number and Probe Colour for MicroRNAs purchased from ThermoFisher Scientific

Manufacturer	Gene	Assay ID	Product Number	Probe Colour
Thermofisher	Mir-21	Hs04231424_s1	4331182	FAM
Thermofisher	Mir-31	002279	4427975	FAM
Thermofisher	Mir-99a*	002141	4427975	FAM
Thermofisher	Mir-125b	000449	4427975	FAM
Thermofisher	RNU6B	Custom Product	P/NCCU001S	VIC

The RNU6B VIC probe was a custom order which had to be generated specifically for this project. It combines the sequence for RNU6B (FAM probe; Assay ID 001093) with a VIC probe.

2.2.15.4 DdPCR Protocol

ddPCR reaction mixes were set up to a total volume of 25µL. Forward and reverse primers were diluted to 20mM each whilst each set of probes was diluted down to 5mM per individual ddPCR reaction. All reagents were warmed to room temperature before use. A total of 20ng of DNA from FFPE tissue was analysed using this assay, with the remaining volume made up of PCR-grade water. For cfDNA, the maximum volume (10µL) was added to the reaction mix due to low sample concentrations(1-2ng). All cfDNA samples were analysed in duplicate and pooled for analysis to try and increase the total DNA concentration whilst allowing enough remaining sample for repeated analysis.

For exoRNA analysis, 3µL of cDNA was added to each reaction for analysis due to limited sample volume as well as to limit the possibility of contamination-associated PCR inhibition. It would also allow for sample re-testing if necessary. ddPCR reactions were set-up as illustrated in Table 9 for ctDNA analysis.

Table 9: Reagent and Sample Volumes and Concentrations for Droplet Digital PCR Analysis of Circulating Tumour DNA or Exosome-Derived RNA

Reagent	Volume (µL)	Concentration
Supermix	12.5	2X
Variant Primers and Probe (FAM)	1.25	20X
Wild-type Primers and Probe (HEX/VIC)	1.25	20X
Sample DNA/cDNA	χ	*
PCR Grade Water	10- χ	n/a

χ = The volume of DNA required to add 20ng of tumour tissue DNA. For cfDNA and cDNA, 10µL and 3µL of sample were added respectively.

*DNA concentrations for FFPE tissue were approximately 20ng.

From this mixture, 24µL of sample mix was added to a Bio-Rad DG8 Cartridge (1864008) with 70µL of Bio-Rad Droplet Generation Oil for Probes (1863005), covered with a Bio-Rad DG8 Gasket (1863009) and placed into the QX200 Droplet Generator. This process dispersed the reaction mix into approximately 20,000 droplets, each theoretically containing ≤1 copy of DNA or complementary DNA (cDNA). The generated droplets (approximately 45µL) were carefully transferred to a new plate. The plate was sealed using the Bio-Rad DX1 Plate Sealer, and the reaction amplified using the program detailed in Table 10.

Table 10: Droplet Digital PCR thermocycler conditions of PCR amplification in generated droplets

Time	Temperature	Ramp Rate	Number of Cycles
10 minutes	95°C	~2°C/second	1
30 seconds	94°C	~2°C/second	40
60 seconds	*	~2°C/second	
10 minutes	98°C	~2°C/second	1
∞	4°C	~1°C/second	1

*Annealing temperature varied from 50-60°C depending on optimal primer/probe conditions

The reaction was amplified within individual droplets using the Bio-Rad T100 Thermocycler. Droplets were read using the Bio-Rad QX200 Droplet Reader and analysed using QuantaSoft.

In all ctDNA ddPCR runs, positive, wild-type and no template controls (NTC) were used to ensure the assay was accurately calling positive and wild-type samples, whilst also being free of contamination.

For exoRNA analysis, an RNA positive control and NTC were used to ensure the assay worked and was clean of contamination.

2.2.15.5 ddPCR Analysis

2.2.15.5.1 Droplet Digital PCR Analysis using QuantaSoft

QuantaSoft is software provided by Bio-Rad which can be used to quantify gene expression or VAF in template cDNA or DNA respectively. The Bio-Rad droplet reader quantifies the fluorescence levels of FAM or HEX/VIC within individual droplets. These fluorescence levels were used to determine which droplets contain mutant or wild-type DNA respectively or genes of interest and reference genes respectively.

In QuantaSoft, amplified targets within droplets were analysed using 1D and 2D amplitude graphs to determine droplet positivity with thresholds acquired using positive control samples and definetherain.org. Once the thresholds had been set, the ratio of FAM and HEX positive droplets were used to calculate VAF or Gene Expression for ctDNA and exoRNA respectively. VAF, gene expression and respective standard deviations were calculated using Poisson distribution.

The number of total droplets generated during ddPCR analysis was also collected to ensure the process of droplet generation occurred efficiently and that results were not influenced through droplet lysis at any stage. A minimum of 10,000 droplets were required for a ddPCR run to be deemed successful. We also required a minimum of four mutant droplets for a variant to be confidently detected in ctDNA. The reason for this minimum number of droplets required is described in Section 2.3.2.2 in this chapter.

2.2.16 Clinical Data Analysis

Clinical data was acquired from 235 patients with LARC who were treated with neoadjuvant therapy at Velindre Cancer Centre between 2010-2019. Data collection was performed by Luca Galvani using the Cancer Network Information System Cymru (CaNISC) under the supervision of Professor Richard Adams. This dataset also includes the initial 16 patients recruited by the WCB who also underwent ctDNA, cfDNA and exoRNA.

All clinical data analysis was completed under the supervision of Professor Richard Adams. Statistical analysis was performed using the basic R software. This largely involved two-tailed unpaired two-sample T-tests. Resulting p-values were not corrected for according to the number of factors investigated. All analysis performed was univariate.

2.2.17 Statistical Analysis

During this study, we applied unpaired T-Tests, Chi-Squared and Fisher's Exact Tests at various points during analysis. Statistical analysis was performed using the basic R package with the advisement of Dr Matthew Summers, Division of Cancer and Genetics, Cardiff University.

2.2.17.1 R Packages

CtDNA VAFs and cfDNA concentration data were plotted using ggplot2 in R.

2.3 Assay Validation and Limit of Detection Analysis

2.3.1 Next Generation Sequencing

2.3.1.1 Choosing a Target Next Generation Sequencing Panel

Objective: Which is the most appropriate and efficient targeted NGS panel to detect tumour specific variants in patients with LARC?

In this study, our objective was to use NGS in order to detect tumour-specific variants from the tumour tissue of patients with LARC both efficiently and reliably for downstream analysis. These variants could then be used to analyse circulating biomarkers, such as ctDNA, in these patients. Circulating biomarker analysis has previously been achieved using either NGS or ddPCR in patients with LARC (Tie et al. 2018). For variant detection in tumour tissue, whole exome or genome sequencing of tumour and germline tissue would have been ideal to ensure tumour-specific variants were detected. However, due to limitations in funding, this was not feasible and, thus, we chose to proceed with a targeted NGS panel.

A targeted panel was to be chosen for NGS analysis in order to efficiently detect variants within patient tumour tissue. In order to choose the most appropriate targeted panel, we considered multiple factors, including the genes and mutational hotspots covered by each panel, cost and sequencing platform.

2.3.1.1.1 Identifying Genes of Interest in Rectal Cancer Patients

In the first instance, a literature search was performed, using cBioPortal, in order to highlight the most frequently mutated genes in patients with rectal cancer.

2.3.1.1.1.1 cBioPortal

A review of larger publications via cBioportal (Gao et al. 2013) revealed that >13,000 genes were found to be mutated in 140 patients with rectal cancer. These were then filtered to remove variants which had been detected in <5% of patients. Of the remaining 473 genes, 61 were related to cancer, according to OncoKB (See Appendices Section 9.3 for full gene list).

Four targeted gene panels were initially selected for potential utility within this study. These panels included; the Qiagen Colorectal Cancer Panel, the Qiagen Clinically Relevant Panel, the LifeTechnologies Ion Ampliseq CHPv2 and the Illumina TruSeq Amplicon Cancer Panel. Full gene lists for these panels can be seen in Table 11.

Table 11: Genes included within each of the four targeted sequencing panels. The genes in each column descend in order of the most frequently mutated in patients with rectal cancer. The % values in each cell represent how much of the gene is sequenced by each respective targeted panel.

Qiagen CRC Panel	Qiagen Clinically Relevant Panel	Ion Ampliseq CHPv2	Truseq Cancer Panel
APC (100%)	*TP53 (100%)*	*APC (7.7%)*	*APC (18.0%)*
TP53 (100%)	*KRAS (100%)*	*TP53 (35.0%)*	*TP53 (66.5%)*
KRAS (100%)	*PIK3CA (100%)*	*KRAS (6.1%)*	*KRAS (10.3%)*
FBXW7 (100%)	*NRAS (100%)*	*FBXW7 (8.7%)*	*FBXW7 (16.2%)*
PIK3CA (100%)	*PTEN (100%)*	*PIK3CA (11.9%)*	*PIK3CA (13.6%)*
SMAD4 (100%)	*CTNNB1 (100%)*	*SMAD4 (12.8%)*	*SMAD4 (16.3%)*
ATM (100%)	*BRAF (100%)*	*NRAS (5.8%)*	*NRAS (8.5%)*
NRAS (100%)	*ERBB2 (100%)*	*ATM (11.7%)*	*ATM (22.1%)*
DMD (100%)	*RET (100%)*	*ERBB4 (7.0%)*	*ERBB4 (11.8%)*
TCF7L2 (100%)	*KIT (100%)*	*PTEN (9.9%)*	*PTEN (15.0%)*
PIK3R1 (100%)	*AR (100%)*	*CTNNB1 (3.2%)*	*CTNNB1 (4.7%)*
DCC (100%)	*ALK (100%)*	*KDR (14.1%)*	*KDR (27.6%)*
CTNNB1 (100%)	*DDR2 (100%)*	*BRAF (3.8%)*	*BRAF (5.5%)*
SMAD2 (100%)	*MET (100%)*	*ERBB2 (8.0%)*	*ERBB2 (12.1%)*
BRAF (100%)	*PDGFRA (100%)*	*RET (10.9%)*	*RET (16.3%)*
ERBB2 (100%)	*EGFR (100%)*	*GNAS (5.4%)*	*GNAS (4.3%)*
MAP7 (100%)	*MAP2K1 (100%)*	*KIT (16.5%)*	*KIT (35.3%)*
MSH6 (100%)	*IDH2 (100%)*	*ALK (3.7%)*	*ALK (5.8%)*
MAP2K4 (100%)	*FGFR3 (100%)*	*CDH1 (6.8%)*	*CDH1 (10.9%)*
MSH2 (100%)	*STK11 (100%)*	*JAK2 (11.3%)*	*JAK2 (16.6%)*
PTPN12 (100%)	*GNA11 (100%)*	*MET (9.3%)*	*MET (13.1%)*
WBSCR17 (100%)	*GNAQ (100%)*	*MLH1 (4.6%)*	*MLH1 (6.7%)*
CASP8 (100%)	*IDH1 (100%)*	*PDGFRA (7.2%)*	*PDGFRA (11.4%)*
EP300 (100%)	*AKT1 (100%)*	*PTPN11 (12.6%)*	*PTPN11 (19.6%)*
MYO1B (100%)		*EGFR (9.0%)*	*EGFR (14.6%)*
MLH1 (100%)		*RB1 (17.3%)*	*RB1 (29.6%)*
TCERG1 (100%)		*HNF1A (5.1%)*	*HNF1A (10.3%)*
FZD3 (100%)		*ABL1 (7.3%)*	*ABL1 (16.6%)*
SLC9A9 (100%)		*FGFR1 (5.7%)*	*FGFR1 (8.7%)*
ATP6V0D2 (100%)		*FGFR2 (10.3%)*	*FGFR2 (21.0%)*
CDC27 (100%)		*FGFR3 (12.0%)*	*FGFR3 (21.4%)*
GPC6 (100%)		*FLT3 (8.2%)*	*FLT3 (18.5%)*
MIER3 (100%)		*IDH2 (4.2%)*	*MPL (9.6%)*
MSH3 (100%)		*MPL (6.2%)*	*NOTCH1 (3.8%)*
ACVR1B (100%)		*NOTCH1 (2.5%)*	*SMO (23.3%)*
TGFBR2 (100%)		*SMO (13.1%)*	*STK11 (41.5%)*
AKT1 (100%)		*GNA11 (2.2%)*	*GNA11 (30.6%)*
BAX (100%)		*CSF1R (4.5%)*	*CSF1R (9.0%)*
		CDKN2A (19.9%)	*CDKN2A (14.6%)*
		EZH2 (3.2%)	*GNAQ (25.7%)*
		GNAQ (1.8%)	*IDH1 (7.5%)*
		IDH1 (4.2%)	*JAK3 (6.7%)*

Qiagen CRC Panel	Qiagen Clinically Relevant Panel	Ion Ampliseq CHPv2	Truseq Cancer Panel
		JAK3 (5.5%)	*NPM1 (10.6%)*
		NPM1 (6.5%)	*SMARCB1 (41.5%)*
		SMARCB1 (27.7%)	*VHL (14.9%)*
		VHL (8.5%)	*AKT1 (6.0%)*
		AKT1 (7.8%)	*HRAS (29.6%)*
		HRAS (19.3%)	*SRC (4.1%)*
		SRC (2.2%)	

We initially investigated which genes of interest were covered by each targeted panel (Table 11). Of all the genes mutated in ≥5% of patients with rectal cancer, the Qiagen Clinically Relevant Panel included only six. Furthermore, this panel did not sequence *APC*, which is the most frequently mutated gene in this cohort (78.6%). The Qiagen CRC Panel contained 14 of the most frequently mutated genes whilst both the Illumina Truseq Cancer Panel and the LifeTechnologies CHPv2 sequenced 12. All of these panels covered hotspots or entire coding regions with *KRAS, TP53,* and *PIK3CA,* which are three of the four most frequently mutated genes in patients with rectal cancer, after *APC.* At this time, we chose to exclude the Qiagen Clinically Relevant Panel due to the reduced number of genes of interest sequenced within this gene panel.

We next chose to investigate the efficiency of sequencing of each of the remaining targeted panels. This included costings and number of analysable samples per sequencing cell/chip at set theoretical coverages and DNA concentrations required for sequencing (see Table 12). The LifeTechnologies Ion Ampliseq CHPv2 and Illumina Truseq Pan-Cancer panels could only be sequenced using their respective technologies; whereas the Qiagen CRC panel could be applied using either sequencer. For the purposes of this comparison, the efficiency of this panel was treated as if sequenced on the Ion Torrent Proton, due to its increased sequencing capacity over the Illumina Miseq. This decision was also influenced by the very limited availability of Illumina sequencers to our research group, in comparison to the less frequently used but well-maintained Ion Torrent Proton.

Potential sample numbers and coverage were calculated for sequencing cfDNA and FFPE tumour DNA. Coverages for cfDNA and tumour tissue DNA were calculated for a minimum of 5,000 and 1,000 reads respectively. These values are consistent with previous analyses from our research group.

The number of samples which can be sequenced for each NGS panel using the Proton was calculated using the following formula (provided by LifeTechnologies);

$$\text{Number of Samples} = \frac{70{,}000{,}000}{Total\ Number\ of\ Amplicons\ x\ Average\ Sequencing\ Depth\ per\ Amplicon}$$

Equation 1: Formula provided by ThermoFisher Scientific to calculate number of samples which can be sequenced at specific read depths

The number of samples which can be sequenced for the Illumina Truseq Cancer Panel was calculated using the CoverageCalculater from Illumina based on the MiSeq V3 (25,000,000 read flow-cell).

Table 12: Required DNA concentration, sample number calculation per chip/flow-cell at specified read depths for four targeted sequencing panels

Sequencing Panel	Sample Type	Minimum DNA Input (ng)	Minimum (Average) Coverage (X)	Samples per Chip
Ion Ampliseq CHPv2	CfDNA	10	5,000 (25,000)	13
CRC Panel	CfDNA	20	5,000 (25,000)	1
Truseq Cancer Panel	CfDNA	250	5,000 (25,000)	3
Ion Ampliseq CHPv2	FFPE	10	1,000 (5,000)	67
CRC Panel	FFPE	20	1,000 (5,000)	7
Truseq Cancer Panel	FFPE	250	1,000 (5,000)	18

Applying the Qiagen CRC Panel to cfDNA was deemed impractical, as a single cfDNA sample and up to five FFPE samples (alongside two control samples) could be sequenced with optimal coverage (Table 12). This same deduction was made for the Illumina Truseq Cancer Panel, which could sequence three cfDNA (not including controls) per flow cell, although the number of FFPE samples was more acceptable. Furthermore, the Illumina Truseq required unrealistic DNA concentrations for reliable cfDNA or FFPE DNA analysis in patients with LARC, as we are unlikely to have large amounts of tissue or plasma available for analysis.

Therefore, we decided to proceed with the Ion Ampliseq CHPv2. The Ion Ampliseq CHPv2 consisted largely of variant hotspots such as *KRAS*/*NRAS* codons 12, 13 and 61, and exon 15 within *APC* and, alongside the Truseq Cancer Panel, included the highest number of genes of interest according to cBioPortal. Furthermore, as this panel largely consists of hotspots, rather than entire genes, it contains fewer primer pairs and amplicons, allowing more efficient sequencing and analysis.

Although we chose to sequence our patient samples using the CHPv2 on the Ion Torrent Proton, this panel and sequencing technology are limited by several factors. Although this panel provides superior efficiency in terms of coverage and sample numbers, it only targets gene hotspots and, therefore, variants outside of common hotspots may be missed. However, the identification of novel tumour-specific variants which influence patient response to radiotherapy was not the aim of this project. This panel was chosen as it is the most efficient panel and sequencing technology which allows reliable detection for tumour-specific variants in patients with LARC.

Like all NGS panels and sequencing technologies, there are limitations to the CHPv2 and Ion Torrent Proton which are important to be aware of during sample analysis. Semi-conductor sequencing has previously been demonstrated as having limitations when analysing both small and large indels (Quail et al. 2012), as we also demonstrate here within *EGFR* in Section 2.3.1.2 in this chapter. Therefore, we may find false findings when investigating indels within this panel. This technology has also been shown to have difficulties sequencing large strings of homopolymers (Quail et al. 2012). As this is targeted at specific regions, and not whole exome or genome, this limitation should not impact our ability to confidently identify and call variants using this technology.

Overall, we chose to apply the Ion Ampliseq CHPv2 based upon the list of genes sequenced within this targeted gene panel. Furthermore, this panel provides potential benefits regarding sequencing efficiency ahead of our other panels of interest. This panel should allow the frequent detection of tumour specific variants in patient FFPE tumour DNA and/or ctDNA.

2.3.1.2 NGS Limit of Detection

Objective: What is the limit of detection for our chosen NGS panel?

Once a targeted NGS panel had been chosen, we proceeded to validate the ability of this panel to detect known variants in reference standards.

Reference standards (see Section 2.2.4) were sequenced at variable depths, defined as high, medium and low coverage (expected minimum coverages of 5000X, 3000X and 1000X respectively). At each of these depths, variants at 6.3%, 3.2%, 1.3% and 0.1% VAFs were sequenced alongside wild-type (0%) controls. A total of 10ng of reference standard DNA was used for NGS analysis (Table 13).

The reference standards used here are widely used and commercially available cell-line derived DNA, which are well characterised using several methods. This also gives us knowledge

as to exactly which variants we should be detecting and at which VAFs. Any anomalies can therefore be queried, whereas in uncharacterised patient samples this is not the case as they will not be as well characterised.

We investigated three different sequencing depths in this study for purposes of efficiency. Although we expect to see a difference in sensitivity between our three coverages, we aim to find out how much sensitivity is lost from one read depth to another. This is because we expect to identify ctDNA variants at approximately 1% and wish to identify at which coverages we can do so robustly. Decreasing sequencing coverage without harming sensitivity could also allow the analysis of a larger number of samples on a sequencing chip and improve the efficiency of any NGS-analysis.

During this analysis, variants were filtered and passed based on our coverage and the number of mutant reads as described previously in Section 2.2.14.2.2. Briefly, for a variant to be called, ≥250 total reads and ten mutant reads (five in each direction) were required.

Table 13: Results of NGS validation using reference standards with known variants at variable sequencing coverages.

				Observed vs Expected VAF (%)				
Depth	**Gene**	**Variant**	**Type of Variant**	**6.30%**	**3.20%**	**1.30%**	**0.10%**	**WT (0%)**
5,000X	*EGFR*	p.L858R	SNV	4.0	1.4	0.5	ND	ND
5,000X	*EGFR*	p.T790M	SNV	4.4	2.2	1.2	ND	0.1
5,000X	*KRAS*	p.G12D	SNV	6.7	3.6	1.3	ND	ND
5,000X	*NRAS*	p.Q61K	SNV	6.3	2.5	1.1	0.2	ND
5,000X	*NRAS*	p.A59T	SNV	6.6	2.8	0.8	0.3	0.1
5,000X	*PIK3CA*	p.E545K	SNV	5.3	3.6	1.2	0.3	ND
5,000X	*EGFR*	p.ΔE746 - A750	Indel	10.7	2.7	1.6	ND	ND
5,000X	*EGFR*	p.V769 - D770insASV	Indel	ND	ND	ND	ND	ND
3,000X	*EGFR*	p.L858R	SNV	3.2	2.2	0.7	ND	ND
3,000X	*EGFR*	p.T790M	SNV	4.8	ND	ND	ND	ND
3,000X	*KRAS*	p.G12D	SNV	2.5	2.5	ND	ND	ND
3,000X	*NRAS*	p.Q61K	SNV	5.6	2.6	0.8	ND	ND
3,000X	*NRAS*	p.A59T	SNV	6.4	4.5	0.9	ND	0.2
3,000X	*PIK3CA*	p.E545K	SNV	6.5	2.9	1.7	ND	ND
3,000X	*EGFR*	p.ΔE746 - A750	Indel	11.6	5.4	2.4	ND	ND
3,000X	*EGFR*	p.V769 - D770insASV	Indel	ND	ND	ND	ND	ND
1,000X	*EGFR*	p.L858R	SNV	3.4	1.9	ND	ND	ND
1,000X	*EGFR*	p.T790M	SNV	5.1	ND	ND	ND	ND
1,000X	*KRAS*	p.G12D	SNV	5.1	2.5	ND	ND	ND

Depth	Gene	Variant	Type of Variant	Observed vs Expected VAF (%) 6.30%	3.20%	1.30%	0.10%	WT (0%)
1,000X	*NRAS*	p.Q61K	SNV	5.8	1.8	2.9	ND	ND
1,000X	*NRAS*	p.A59T	SNV	6.0	2.7	0.6	ND	ND
1,000X	*PIK3CA*	p.E545K	SNV	6.3	2.2	1.4	ND	ND
1,000X	*EGFR*	p.ΔE746 - A750	Indel	6.4	4.1	ND	ND	ND
1,000X	*EGFR*	p.V769 - D770insASV	Indel	ND	ND	ND	ND	ND

ND- The variant of interest was not detected in this sample.

Indel – Insertion or deletion

The average coverages for high, medium and low depths were 13,886X (10,443X – 16,475X), 6,935X (5,573X – 9,334X) and 1,250X (16.17X – 3,492X) respectively.

At a calculated minimum coverage of 5,000X (high coverage calculated using Equation 1), all six Single Nucleotide Variants (SNVs) were confidently detected at 6.3%, 3.2% and 1.3% VAF (Table 13). At an expected average depth of 3,000X (medium coverage), only 4/6 SNVs were detected at 1.3% VAF. At this coverage, 5/6 and 6/6 SNVs were confidently detected at 3.2% and 6.3% VAF respectively. Finally, at a predicted read-depth of 1,000X (low coverage), 3/6 SNVs were detected at 1.3% and 5/6 SNVs were detected at 3.2% VAF. In contrast, all six SNVs were detected at 6.3% VAF at this coverage.

As predicted, variant detection was less consistent and sensitive for small indels in comparison to SNVs. The small insertion in *EGFR* (p.V769 - D770insASV) was not detected at any VAF or read-depth. In contrast, the small deletion in *EGFR* was confidently detected down to 1.3% VAF at both high (5,000X) and medium coverage (3,000X), and 3.1% VAF at low coverage (1,000X; Table 13). In this validation only a single deletion was analysed, which can be a major issue, especially considering the flaws that this technology has with sequencing indels. However, such indels are relatively rare in the context of rectal cancer in comparison to SNVs. Looking forward towards future clinical implementation, this technology and targeted panels will require more validation regarding the robust detection of small indels.

Our limit of detection for variants using this NGS panel was dependent on both the type of variant analysed and the sequencing coverage. Using 10ng of reference standard DNA the limit of detection for SNVs was 1.3% VAF when applying 5,000X coverage. At 3,000X and 1,000X coverage, the limit of detection was only 6.3%. It should be noted that only one variant (*EGFR*

p.T790M) was not reliably detected using these two coverages at 3.2% VAF. This was unsurprising as this region is known to be difficult region to sequence (Fahoum et al. 2018).

Few studies have investigated the limit of detection of the Ion Torrent Proton, or other semiconductor-based sequencing, to a degree below 5%. Lin and colleagues (2014) reported that SNVs in *KRAS* (p.G13D), *BRAF* (p.V600E) and *EGFR* (p.T790M and p.L858R) were consistently detected at 1% VAF in cell-line-derived DNA using the PGM. This limit of detection is in agreement with previous studies reporting a limit of approximately 1% using custom panels (Aloisio et al. 2016) on the Ion Torrent PGM using oligonucleotides. Although results generally suggest a 1% limit of detection, some studies have suggested that the Ion Torrent Variant Caller may miss variants with a VAF <1% which can be observed using the integrated genomics viewer (IGV) software (Mehrotra et al. 2018). This is performed to look for hotspot variants in *KRAS,* for example, but is not a practical solution for all samples moving forward. In addition, Tsongalis and colleagues (2014) reported variable limits between single nucleotide changes and indels, further reinforcing our findings of unreliable indel detection using this targeted panel.

In agreement with our findings, and in contrast to those mentioned previously, studies have reported a limit of detection at ~5% using this technology. Zhang and colleagues (2014) demonstrated that variants with VAF <5% were associated with a poor variant accuracy (18%) in comparison to variants with >5% VAF (94%). These studies also applied up to 30ng of DNA for NGS analysis, rather than the 10ng applied here. Nevertheless, our study was able to achieve a similar sensitivity for SNV detection. However, we have only analysed high-quality DNA extracted from cell-lines with known variants, rather than investigating FFPE-derived tumour DNA with no foreknowledge of exactly which variants will be unearthed. Therefore, we may see that background noise levels increase, and our limit of detection decreases with the application of FFPE tumour DNA and/or cfDNA.

Although increase sequencing depth could improve the determined sensitivity of this assay, at low VAFs it became increasingly difficult to confidently distinguish tumour-specific variants from sequencing artefacts.

The level of background artefacts was initially examined at each point using WT reference standards. At high and medium coverage, artefacts were observed between 0.01-0.13% and 0.01-0.18% respectively, which overlapped with the 0.1% VAF SNVs in the reference standards.

Therefore, we determined that variants at 0.1% VAF could not be confidently called, due to the potential for false positive results.

These background levels are consistent with previous reports. Lin and colleagues (2014) reported that the background noise in FFPE tissue was seen up to 1.3% VAF. Similarly, McCall and colleagues (2014) also reported the presence of false positives using the CHPv2 on the PGM. Both studies saw an increased number of false positive variants within 10-20bp of either the 5' or 3' end of relatively short amplicons. Lin and colleagues (Lin et al. 2014) also demonstrated that these false positives generally had a very large direction bias, whilst McCall and colleagues (2014) showed that increased false positive rates were associated with poor DNA quality and/or quantity and resulting primer mismatching. Therefore, samples with limited DNA concentrations and /or poor sample quality had an increased likelihood of producing such false positives. These reports contributed to the variant filtration process described previously in Section 2.2.14.2.

Overall, our limit of detection for NGS using high sequencing coverage was 1% VAF, although this was variable depending on sequencing depth.

2.3.1.3 Targeted Sequencing Panel Validation on Patient cfDNA

Objective: Can we validate that our NGS panel can be used to sequence cfDNA samples extracted from low plasma volumes from patients with LARC?

In this section, we apply the previously calculated limit of detection from Section 2.3.1.2 and aim to analyse patient cfDNA and validate our targeted NGS panel for this purpose.

When considering ctDNA in patients with LARC, VAFs are likely to be around or below the 1% level if detectable at all (Sclafani et al. 2018). We were able to detect variants at this level using high sequencing coverage in 10ng of reference standard DNA, but we were unable to achieve this sensitivity using a lower read depth. As this does not fully represent the quality of patient DNA samples, further validation assays are required to be performed on cfDNA extracted from patient plasma to ensure previous validations remained consistent. We also investigated how the volume of plasma from which cfDNA is extracted may influence variant detection and the quality of resulting NGS analysis.

Plasma samples from patients with LARC consented by the WCB were used for this pilot run. CfDNA was extracted from 2mL of pre-treatment plasma from 14 patients and sequenced with a calculated minimum coverage of 5,000X as performed previously in Section 2.3.1.2 of this chapter. Samples were sequenced alongside a positive (Horizon HD777; 4,000X), wild-type

(Horizon HD776; 4,000X) and a no template control (1,000X). Sequencing was simultaneously performed on cfDNA extracted from 1mL of pre-treatment plasma from a separate group of 13 patients with LARC Table 14. These were sequenced with a minimum calculated coverage of 5,000X alongside three control samples (1,000X reads each).

Table 14: NGS analysis of ctDNA in 1mL or 2mL or pre-treatment plasma from patients with LARC consented by the WCB

Patient ID	Plasma Volume (mL)	Variant Detected	VAF (%)
WCB1203	2	*KRAS* p.G12D	7.7
WCB1251	2	WT	n/a
WCB1260	2	WT	n/a
WCB1261	2	*KRAS* p.Q61H	2.0
WCB1262	2	WT	n/a
WCB1263	2	WT	n/a
WCB1265	2	WT	n/a
WCB1266	2	*TP53* p.G245S	19.5
WCB1268	2	WT	n/a
WCB1269	2	WT	n/a
WCB1367	2	WT	n/a
WCB1369	2	WT	n/a
WCB1415	2	WT	n/a
WCB1416	2	WT	n/a
WCB1295	1	WT	n/a
WCB1368	1	WT	n/a
WCB1419	1	WT	n/a
WCB1441	1	WT	n/a
WCB1473	1	*NRAS* p.G13R	1.0
WCB1476	1	WT	n/a
WCB1477	1	WT	n/a
WCB1478	1	*NRAS* p.Q61L	1.3
WCB1479	1	WT	n/a
WCB1493	1	WT	n/a
WCB1602	1	WT	n/a
WCB1603	1	WT	n/a
WCB1604	1	WT	n/a

WT - Wildtype

We detected tumour-specific variants in ctDNA in 2/13 (15.4%) and 3/14 (21.4%) patients with LARC from cfDNA extracted from 1mL and 2mL of plasma respectively (Table 14). As these two groups are unpaired (i.e. they contain different patients) and our sample numbers are very small, we cannot determine whether these values are significantly different or a result of sample plasma volume. Although we may have observed a small difference in the frequency at

which variants were detected in ctDNA between the two volumes of plasma, we could not perform any meaningful statistical analysis at this time. Therefore, we chose to investigate technical sequencing metrics from the respective sequencing runs for each plasma volume to see if any significant differences were observed here. These metrics include the number of total reads, average on-target reads, average read depth and uniformity. These metrics are described in Table 15 and illustrated in Figure 10.

Table 15: Sequencing metrics analysing the difference in read quality and quantity between sequenced cfDNA extracted form 1mL or 2mL of plasma from patients with LARC.

Plasma Volume	Total Reads	Average on Target Reads (%)	Average Read Depth (X)	Uniformity (%)
1mL	75,617,145	49.90	4,055.00	61.80
2mL	62,397,999	96.91	13,213.64	61.79

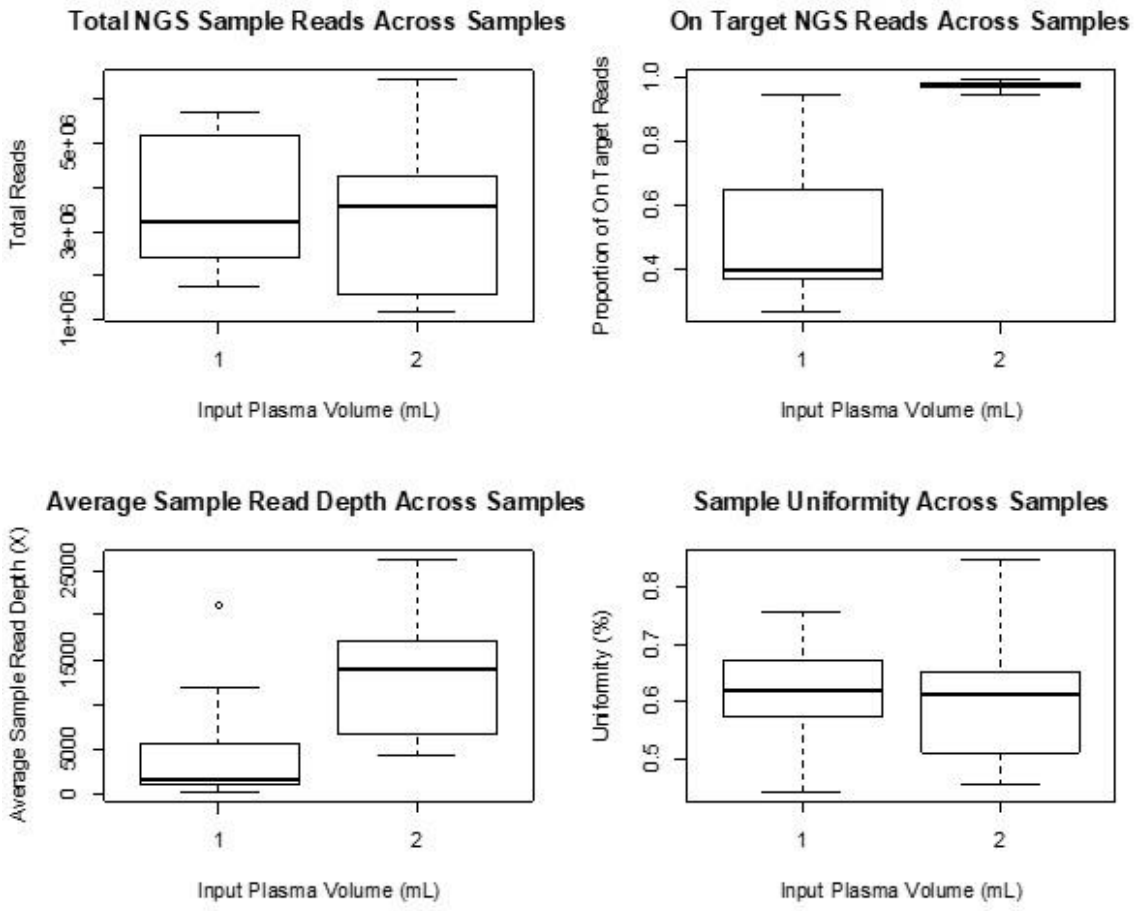

Figure 10: Comparison of numerous metrics of Next Generation Sequencing analysis across sequencing runs analysing cfDNA extracted from 1mL or 2mL plasma from patients with LARC

Read-depth refers to the number of times a single base was sequenced for each individual sample and on-target reads refers to the proportion of regions/amplicons sequenced which

match directly to the human genome reference or, in this case, a BED-file containing our regions of interest for the CHPv2.

Uniformity refers to the distribution of coverage over sequenced regions, with the intent being that all sequenced bases within a sample have relatively equal coverage (Quail et al. 2012). This, however, means that there can be inter-sample variability which this metric does not consider. Uniformity has been reported to vary between sequencing technologies, though the Ion Torrent PGM and Illumina Miseq were reported to have similar levels, with Illumina platform performing slightly better (Quail et al. 2012). Therefore, the figure we acquired from the Ion Torrent Server is a level of relative uniformity. Previous reports have shown that poor uniformity can significantly influence sequencing efficiency, more so than other metrics such as on-target read rates (Quail et al. 2012).

Here we saw a significant difference between the number of on target reads and average read depth between the two cohorts of samples ($p<0.01$ in both cases). For average read depth we aimed to achieve a minimum of 5,000x, which we achieved in the cfDNA samples extracted from 2mL of plasma (~13,000X) but not those extracted from 1mL (~4,000X). Furthermore, this run had a higher number of overall reads, suggesting that a significant number of reads were lost or were primer-dimer and not genuine patient DNA sequences. This is highlighted further by the significant decrease in the frequency of on-target reads from the same patients. Previous studies have suggested that 80% of all reads should be on-target during a successful sequencing run (de Leng et al. 2016). Although we achieved this in the 2mL plasma samples, the sequencing of the 1mL plasma samples fell far short of this target.

Studies have also recommended that uniformity should be consistently 90% across a sequencing run, which we were not able to achieve in either case here (Vanni et al. 2015). However, as our sequencing runs contained low concentrations of cfDNA, there may have been significant dropout in certain genomic regions by chance. Furthermore, as this metric is quantified between samples, the use of reference standards alongside cfDNA may have caused a stark contrast and consequential decrease in sequencing uniformity. Uniformity has also been reported to be relatively low using semiconductor sequencers (Vanni et al. 2015). Thus, our relatively low uniformity may not be purely a consequence of our samples or protocols.

Studies have also previously suggested that the number of total reads should be >100,000,000 which is much greater than our current levels (de Leng et al. 2016). In order to improve this factor, the protocol may have to be optimised further, particularly regarding the concentration

of DNA libraries loaded onto the sequencing chip. This should be performed with caution, as overloading the chip may cause a large increase in polyclonality, which can cause an overall loss of usable reads.

Both the frequency of on-target reads and the average sample read depth can be influenced by sample quality and quantity, whereas the previous two metrics mentioned are not. A low level of on-target reads could suggest that the majority of DNA sequenced for this sample consisted of primer-dimer or non-specific off-target primer-binding, both of which can be a result of low DNA quality and/or quantity. If coverage is evenly spread and a large amount of sequenced data is off-target, this then influences read depth, as usable read-depth is wasted on inappropriate DNA amplicons. This is highlighted further by the fact that total sample reads were not significantly different between the two volumes.

Apart from these two significant metrics, the technical sequencing of both runs was consistent, according to the total reads and uniformity which did not significantly differ (Figure 10). This further suggests that the volume of plasma used for cfDNA extraction may have contributed to the differences observed here. However, there are other factors which may have contributed to this outcome.

The two runs were analysed at separate times and separate sequencing runs, so there may be some variation in the quality of each of the library preparation and sequencing runs. Furthermore, we only analysed a small number of patients over two sequencing. Continuing this analysis in the future may help to improve our understanding as to whether plasma volume does have a significant impact on sequencing quality and outcomes. Finally, it is vital to remember that these are unpaired samples, and this variability in sequencing quality and variant detection may be due to biological differences between the two groups of patients. Although we have observed significant differences between the two plasma sample volumes, these findings are limited and warrant further investigation.

The difference in sequencing quality between the two plasma volumes (1mL vs 2mL) is likely to result from the difference in total cfDNA concentrations. Increased plasma volumes are likely to produce higher concentrations, especially as all samples are eluted in the same volume using the QiaAmp Circulating Nucleic Acids Kit. Higher DNA concentrations will then lead to an increased amount of primers binding to DNA, and, therefore an increase in the proportion of on-target reads and sample coverage.

Overall, we have demonstrated that this targeted NGS panel can be used to detect variants in pre-treatment ctDNA from patients with LARC; although the sensitivity of this panel requires further validation. Furthermore, our findings might suggest that the volume of plasma used for cfDNA extractions may have an impact regarding downstream sequencing quality and variant detection in ctDNA. Although these findings would imply that DNA extraction from 2mL of plasma may improve analytical sensitivity, this has improved our understanding that plasma volumes may be a limitation in our investigations later on, especially since most of our patients only had 1mL of plasma available for sequential analysis.

2.3.2 Limit of Detection Analysis for Droplet Digital PCR

Objective: What is the limit of detection for our chosen ddPCR assay?

2.3.2.1 Optimising Conditions

All primers and probes were initially optimised and validated on positive control material from FFPE patient tumour biopsies with pre-determined variants.

Primers/probe sets for ddPCR were optimised using a temperature gradient (50-60°C) to determine optimal annealing temperatures. Resulting ddPCR data was analysed using QuantaSoft as previously described. Primer/probe assays were validated using positive, wild-type and negative control samples to ensure assay sensitivity and specificity.

2.3.2.2 Limit of Detection

Once each assay had been optimised and validated, reference standards (as described previously in Table 2) were used to investigate the limit of detection using 1ng, 5ng and 10ng of total DNA (Table 16). Variants at 1.3%, 3.2% and 6.3% VAF were analysed at each DNA concentration in triplicate. Variant detection, VAFs and error bars were calculated by QuantaSoft using Poisson Distribution.

DNA concentrations were analysed down to 1ng as this is the approximate concentration that we expect to acquire during cfDNA extractions. Similarly, VAFs were analysed down to 1.3% as we expect to see low ctDNA levels in patients with LARC. We did not merge our triplicate data as this combination would not accurately reflect patient ctDNA samples during analysis.

We initially ensured that DNA concentrations were consistent within all reactions with the same input DNA (1-10ng). We saw no significant differences in DNA concentrations between paired ddPCR reactions using ANOVA (Analysis of Variance: data not shown), so we proceeded to investigate the limit of detection for this assay (Table 16).

Using QuantaSoft, at least five FAM (mutant) droplets are required for confident variant detection and accurate quantification in accordance with standard operating procedures used by the AWMGS (LP-GEN-ddPCR). Fluorescence thresholds in QuantaSoft were manually set based on droplet fluorescence from 10ng of positive control DNA.

Table 16: Limit of Detection analysis of variants at 1%, 3% and 6% using 1ng, 5ng and 10ng of DNA. Reactions were performed in triplicate but analysed separately. Five or more mutant droplets were required for confident detection of an individual variant.

Number of Mutant (FAM Positive) Droplets

DNA Concentration	**1.3% Triplicate**			3.2% Triplicate			6.3% Triplicate		
1ng	3	0	0	5	4	1	14	13	13
5ng	1	2	4	7	9	8	61	47	43
10ng	3	4	**6**	15	15	7	104	92	n/a

n/a – Variants at 6% in 10ng of DNA was only analysed in duplicate.

Values in bold are where error bars did not cross 0, thus this assay confidently detected the variant in question

As this analysis was performed using Horizon reference standards (Horizon 2019), the reactions where variants were not confidently detected represent the likelihood of encountering false negative results. In all accompanying wild type and NTC reactions, no mutant droplets were detected, providing this assay a specificity of 100%.

From these findings, our limit of detection using ddPCR depends on the concentration of input DNA and the number of reactions analysed (Table 16). Our limit of detection for a single reaction using 1ng of DNA is 6.3% VAF. Therefore, ctDNA is unlikely to be detected at 1-3% VAF from 1ng of DNA but variants at 3.2% may be identified using 5-10ng. These concentrations are unlikely to be consistently and reliably extracted from 1mL of plasma and, thus, we are likely to observe a high rate of false-negatives, and poor sensitivity under these conditions.

According to the Bio-Rad Best Practice Guidelines for rare mutation detection, Poisson distribution only requires the detection of three mutant droplets for the statistical estimation of detected VAF. These guidelines also describe a variable limit of detection depending on DNA concentration as observed here. The differences in limit of detection between a sample containing 300 and 3,000 total DNA molecules is, therefore, ten times lower (1% vs 0.1% respectively according to the rule of three detected droplets).

As far as we are aware, there are currently no clinical guidelines to standardise ctDNA detection across patients with cancer or cancer subtypes. Furthermore, the value used in the literature is highly variable; with some studies only requiring two droplets (Hrebien et al. 2016; Riva et al. 2017) and others needing five (Huang et al. 2016). We proceeded to analyse our data using a minimum of five mutant droplets required for confident ctDNA detection assuming all controls were free of contamination. This cut-off was chosen as it is currently applied in clinical practice by the AWMGS (ddPCR Standard Operating Procedure; LP-GEN-ddPCR). Though this may compromise the sensitivity of ctDNA detection, this will provide confidence in our findings.

The difficulty in identifying variants using up to 1ng of DNA at 1-3% highlights some of the issues we may encounter during ctDNA analysis. This may impact how frequently we detect ctDNA in our patient cohort on top of the biologically relevant factors (such as patients having non-metastatic disease).

This limit of detection for ddPCR is not as sensitive as demonstrated by previous reports, which have suggested a limit of detection between 0.0005% - 0.4% under various conditions (Pender et al. 2015; Reid et al. 2015; Huang et al. 2016; Yang et al. 2016). Reid and colleagues (2015) were able to demonstrate a very low sensitivity of 0.0005% for ddPCR by combining it with whole genome amplification in DNA extracted from circulating tumour cells (CTCs). This genomic amplification can provide some benefits to analysis by increasing the total yield of DNA and improving assay sensitivity. However, this can cause potential issues with absolute quantification of VAF by creating amplification bias (Weber et al. 2003).

The limit of detection reported by Yang and colleagues (2016) was 0.04% from ctDNA in patients with stage IV non-small cell lung carcinoma. The limit of 0.01% from Huang and colleagues was calculated using 50ng of patient cfDNA (2016) from patients with hepatocellular carcinoma. Finally, Pender and colleagues (2015) reported a limit of detection down to 0.045% using >500pg of reference standard DNA. These sensitivities are generally greater than our findings here, but for several reasons; including use of whole genome amplification, analysing ctDNA from patients with advanced diseases producing higher quantities of cfDNA and having better quality of DNA using reference standards. In comparison we predominantly investigated low quantities (1-10ng) of poor quality FFPE DNA. This may explain discrepancies in our limit of detections. We chose to apply such conditions as it would better reflect cfDNA samples extracted from patients with LARC.

In order to improve assay sensitivity and specificity, in future analyses, all samples will be performed in duplicate with a positive, wild-type and NTC for each assay; to ensure assay specificity remains high. The presence of ≥1 inappropriate DNA molecule in a control sample will void any results acquired from cfDNA during these analyses.

2.4 Analytical Workflow

Objective: What is the optimal process for the detection and tracking of circulating biomarkers in patients with LARC in future studies?

As part of this study, we have chosen a targeted NGS panel which we predicted would allow efficient and frequent variant detection in patients with LARC. We then proceeded to validate and demonstrate the limit of detection for our targeted NGS panel and ddPCR assay. We have also chosen to use the Invitrogen Qubit to quantify total cfDNA in sequential plasma samples from our patient cohort.

Due to limited financial resources, plasma volumes and sample availability, performing NGS on all plasma samples from our patients would not be feasible. Therefore, we designed an analytical workflow which will still allow us to analyse circulating biomarkers in these patients by analysing tumour-specific variants in ctDNA using ddPCR.

Initially, we will apply the Ion Ampliseq CHPv2 to detect tumour specific variants in tumour tissue from each patient with LARC. Where tumour tissue is unavailable, we will aim to acquire excess extracted tumour DNA from the AWMGS or patient RAS status from clinical records. Where available, we will also attempt to sequence pre-treatment plasma in patients with LARC to examine the ability of this panel to detect ctDNA.

Once tumour variants are detected, we will proceed to design and optimise tumour-specific primers and probes for ddPCR analysis. This will include a mutant (FAM) and wild-type (HEX) probe to allow the VAF quantification. Each probe will be optimised on a temperature gradient and validated using a positive, wild-type control and NTC. Some primer/probe pairs may also be purchased for Bio-Rad in certain circumstances.

ddPCR primer/probes will then be used to validate the original variant in tumour tissue where possible and then detect/quantify ctDNA in sequential plasma samples. Simultaneously, total cfDNA concentrations will also be quantified in each plasma sample using the Qubit. These findings will then be compared to patient clinical data to identify or discern any clinically relevant associations. A flow chart of this workflow can be seen in Figure 11.

For exoRNA, we did not have access to tumour tissue for hypothesis-free testing of microRNAs using either NGS or microarrays. Therefore, a small number of microRNAs were chosen based on a review of the literature at the time. These microRNAs were analysed using ddPCR only. The analytical workflow for exoRNA analysis is described further in Chapter 5.

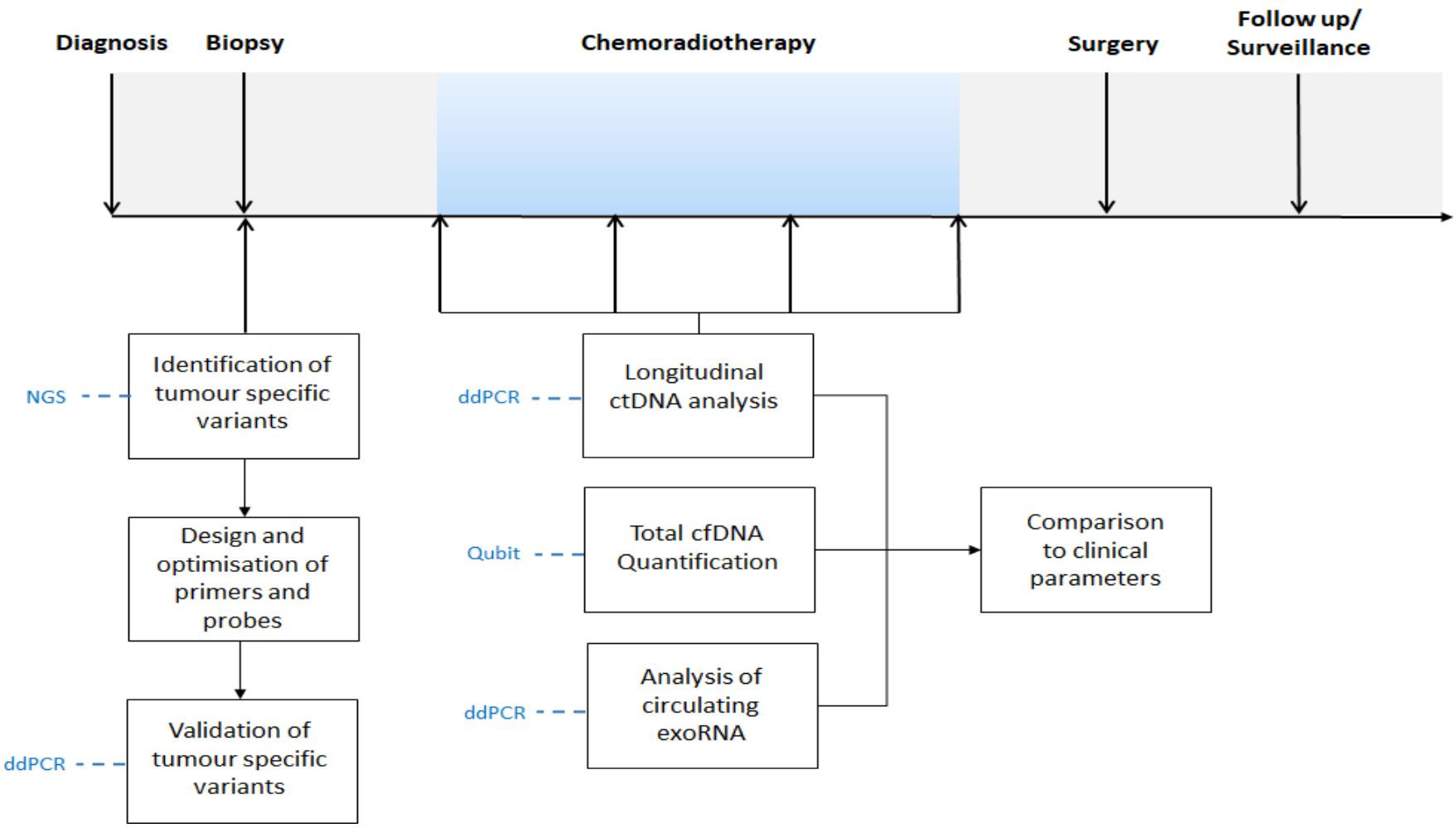

Figure 11: Flow chart illustrating our optimised workflow for circulating biomarker analysis in patients with LARC

3 Investigating the Utility and Feasibility of Circulating Tumour DNA in Locally Advanced Rectal Cancer

3.1 Introduction

CtDNA has emerged as a circulating biomarker of clinical interest in solid tumours, due to its ability to facilitate monitoring tumour burden (Reinert et al. 2016) and patient response to therapy (Carpinetti et al. 2015; Provencio et al. 2017). This has been investigated frequently in patients with metastatic CRC receiving targeted therapies such as cetuximab (Misale et al. 2012). A comparatively small number of studies have looked at ctDNA in relation to pCRT in patients with LARC.

Due to the increasing clinical interest, numerous technologies have been rapidly developed to detect and quantify low levels of ctDNA, including NGS and ddPCR. The ability to detect variants at low frequencies is important for patients with non-metastatic disease, where low VAFs are commonly reported (Sclafani et al. 2018). Using NGS and ddPCR will allow us to identify and monitor tumour-specific variants in longitudinal patient plasma (Tie et al. 2018).

Reports have also illustrated that longitudinal changes in total cfDNA can be predictive of clinical outcomes in patients with solid tumours (Zitt et al. 2008; Schou et al. 2018). The quantification of total cfDNA is not reliant on the presence of tumour-specific variants and can be analysed quickly.

We have a limited understanding of the importance of detecting sequential changes in ctDNA and cfDNA in patients with LARC undergoing non-targeted therapies. Therefore, we do not understand how to predict outcomes and influence patient therapy in the patients.

3.1.1 Study Objectives

This study intends to determine the feasibility and limitations of using NGS and ddPCR to detect and monitor low-frequency mutations in ctDNA. In this chapter, the aims are:

- Can ctDNA be detected in the circulation of patients with LARC before the initiation of radiotherapy using NGS?
- Can ctDNA be reliably detected in sequential plasma samples before, during or after radiotherapy in patients with LARC using ddPCR?
    - Is ctDNA released into the circulation in relation to radiotherapy in patients with LARC?
    - Do longitudinal ctDNA levels correlate with clinical outcomes in patients with LARC?

- Can we detect and measure quantitative changes in total cfDNA before, during and after radiotherapy in patients with LARC?
    - Is cfDNA released into the circulation in relation to radiotherapy in patients with LARC?
    - Do longitudinal cfDNA levels correlate with clinical outcomes in patients with LARC?

3.2 Results

3.2.1 Next Generation Sequencing Sensitivity in Detecting Circulating Tumour DNA in Pre-Treatment Plasma Samples

Objective*: Can ctDNA be detected in the circulation of patients with LARC before the initiation of radiotherapy using NGS?*

Tumour-specific variants were first identified in tumour biopsies using NGS in order to facilitate the detection of ctDNA in pre-treatment plasma from patients with LARC. Variants were analysed and filtered as previously described in 2.2.14.2.

Ten patients with LARC who were consented by the WCB had paired FFPE tumour biopsies and pre-treatment blood samples (Table 17). DNA extracted from FFPE tumour biopsies was provided by the WCB for sequencing. CfDNA was extracted from 1-2ml of patient plasma and patient DNA samples were sequenced using the Cancer Hotspot Panel V2 on the Ion Torrent Proton as previously described in the Materials and Methods.

Table 17: Somatic variant discovery using Next Generation Sequencing in paired patient tumour biopsies and cell free DNA. All patients with locally advanced rectal cancer were consented by the Wales Cancer Bank

Patient ID	Tumour Specific Variants	FFPE Biopsy VAF (%)	CtDNA VAF (%)	Plasma Volume (mL)
WCB1203	*KRAS;* p.G12D	4.5	2	2
	NRAS; p.Q61L	9.3	WT	
	TP53; p.G245D	1.8	WT	
WCB1251	None Detected	WT	WT	2
WCB1260	*BRAF*; p.V600E	11	WT	2
WCB1262*	*KRAS;* p.G12D	17	WT	2
	TP53; p.R248Q	16	WT	
WCB1266	*TP53;* p.G245S	20	3	2
WCB1269	None Detected	WT	WT	2
WCB1441*	*KRAS* p.G12R	8.3	WT	1
WCB1478*	*NRAS* p.Q61L	15	1.3	1
	APC p.E1379X	33	1.9	
WCB1479*	*KRAS* p.G12A	26	WT	1
	TP53; p.Y220C	52	WT	
WCB1493*	*KRAS* p.G13D	10	WT	1
	TP53; p.Y234H	44	WT	

*These patients had additional sequential plasma samples available and were used later in longitudinal ctDNA analysis.

VAF – Variant Allele Frequency.

WT – Wild-Type; variants in question could not be detected in this sample.

Overall, we were able to detect variants in the tumour biopsies from 8/10 patients with LARC using our targeted NGS panel. At least one corresponding mutation was successfully detected in pre-treatment plasma from 3/8 (37.5%) patients with known tumour-specific variants

Of the three variants detected in the DNA extracted from the tumour biopsy of patient WCB1203, only one was detected in pre-treatment plasma. In all other cases, either all or none of the tumour-specific variants were detected in cfDNA. This finding highlights the potential benefits of sequencing plasma rather than analysis of a single tumour-specific variant.

Relatively few studies have investigated the sensitivity of ctDNA analysis in patients with LARC. In 35 patients with *KRAS* mutant tumours, Sclafani and colleagues (2018) reported that 12 (34.3%) had detectable levels of ctDNA, based on the detection of said *KRAS* variants. In another cohort of 159 patients with LARC with known tumour variants, Tie and colleagues (2018) were able to detect pre-treatment ctDNA in 122 patients (76.7%).

The discrepancy between these two reports is likely to be impacted by analytical differences between the studies. Tie and colleagues (2018) applied a small gene panel for variant detection in tumour tissue before detecting one of these variants in cfDNA. In contrast, Sclafani and colleagues (2018) investigated *KRAS* variants in ctDNA using ddPCR. Furthermore, Tie and colleagues extracted ctDNA from 10mL of plasma, whereas Sclafani and colleagues used 2mL. Finally, each study applied a different method for ctDNA analysis, where Tie and colleagues applied Safe-SeqS, Sclafani used ddPCR, with each method having respective strengths and limitations. A combination of these factors may have contributed to the variation in analytical sensitivities observed here.

Our study found comparable sensitivity to Sclafani and colleagues (2018) (37.5% vs 34.3%). We also extracted ctDNA from a similar volume of plasma to this study in comparison to the 10mL used by Tie and colleagues. We applied a gene panel to detect ctDNA similar to that of Tie and colleagues (2018), although our sensitivity was closer to that of Sclafani (2018). Moreover both

our study and that by Sclafani and colleagues (2018) are limited by relatively small sample sizes in comparison to 159 patients with LARC (Tie et al. 2018).

The sensitivity of variant detection in plasma using NGS in our study is also similar to that reported by Chang and colleagues (2018), who demonstrated a sensitivity of 42.9% across a range of stages in four solid tumour types (including 50% in 4 patients with stage IIIb or IVa CRC) using NGS. Similarly, Beije and colleagues (2016) were able to detect 39% of mutations in cfDNA using a 21-gene NGS panel in patients with metastatic CRC. However, this group was able to improve the sensitivity of ctDNA detection using NGS and ddPCR to 80% and 93% respectively in the same cohort. Similarly, Yang and colleagues (2018) were able to improve the sensitivity of ctDNA detection in patients with stage IIIb/IV lung cancer from 63.5% to 83.2% by adding molecular barcodes to their NGS protocols. Rachiglio and colleagues (2016) were able to detect *KRAS/NRAS* mutant ctDNA in 12/19 (63.2%) of patients with metastatic CRC.

Overall, our results are within the expected range based on the literature, though towards the lower end of the spectrum. For some of these studies, this may due to differences in patient cohorts. The patients in our cohort were predominantly non-metastatic locally advanced, whereas some of the previous studies investigated ctDNA in patients with metastatic disease. This is based on reports from Bettegowda and colleagues (2014) who showed that ctDNA was detected significantly more frequently in patients with metastatic disease than locally advanced solid tumour patients across multiple cancers (including CRC). Differences between studies in patients with LARC may also be due to plasma volumes and/or analytical methods applied in each case.

3.2.1.1 Concordance between Circulating Tumour DNA in Pre-Treatment Plasma Samples and Tumour Biopsies

From the data illustrated in Table 17, we observed 37.5% (6/16) concordance between ctDNA and tissue genotyping in the previous cohort.

There are a limited number of studies which have examined concordance between pre-treatment ctDNA and tumour tissue in patients with LARC. Sclafani and colleagues (2018) reported a concordance of 68.9% between tumour tissue and ctDNA in patients with LARC. Whether this is significantly greater than our observed frequency is difficult to determine with only two studies. This difference may also be impacted by our small sample size.

As this is one of the only studies to investigate ctDNA concordance in patients with LARC, it is difficult to determine whether our results are within the expected range. As previous, we then compared our study to those investigating patients with metastatic disease.

Our concordance rate was lower than those reported between pre-treatment ctDNA and tumour tissue from patients with metastatic CRC, which have been reported between 80-90% (Rachiglio et al. 2016; Germano et al. 2017; Grasselli et al. 2017; Hsu et al. 2018). Another study by Siravegna and colleagues (2015) reported a concordance of 97% in a cohort of patients with CRC across a range of tumour stages. This is supported by reports from Bettegowda and colleagues (2014), who demonstrated that ctDNA was more frequently detected in patients with metastatic disease. Therefore, we expected to see a lower concordance between tumour tissue and pre-treatment plasma in patients with LARC.

We did not detect any tumour-specific variants in pre-treatment plasma which were absent in matching patient tumour biopsies at this time.

3.2.2 Investigating Longitudinal Circulating Tumour DNA using Droplet Digital PCR

Objective: *Can ctDNA be reliably detected in sequential plasma samples before, during or after therapy in patients with LARC using ddPCR?*

In Section 3.2.1 we demonstrated that ctDNA could be detected in the circulation of patients with LARC before the initiation of therapy. Five of these patients (WCB1262, WCB441, WCB1478, WCB1479 and WCB1493) had additional plasma timepoints available for longitudinal ctDNA analysis. A further 11 patients with LARC with two or more plasma timepoints were also consented for sequential ctDNA and cfDNA analysis. For these 11 patients, no FFPE tumour biopsies nor extracted tumour-derived DNA were available for research purposes. In this case each patient's *KRAS/NRAS/BRAF* status were acquired from the WCB. For the original five patients sequenced previously, variants were used from prior NGS analysis. Clinical characteristics of this cohort are described in Table 18.

Table 18: Clinical characteristics of patients with locally advanced rectal cancer consented by the Wales Cancer Bank.

Patient ID	T	N	M	Disease Stage	Radiotherapy (Gy in Doses)	Time Period of Radiotherapy	Tumour-Specific Variant for ctDNA Analysis
WCB1262	3	0	0	II	45 in 25	5 Weeks	*KRAS*; p.G12D
WCB1263	3	1	0	III	45 in 25	5 Weeks	*KRAS*; p.G13D
WCB1268	3	2	1	IV	45 in 25	5 Weeks	*KRAS;* p.G13R
WCB1295	3	2	0	III	45 in 25	5 Weeks	WT

Patient ID	T	N	M	Disease Stage	Radiotherapy (Gy in Doses)	Time Period of Radiotherapy	Tumour-Specific Variant for ctDNA Analysis
WCB1368	1	0	0	I	20 in 5	1 Week	WT
WCB1419	3	1	0	III	25 in 5	1 Week	WT
WCB1441	4b	0	0	II	45 in 25	5 Weeks	*KRAS;* p.G12R
WCB1473	0	0	1	IV	30 in 5	1 Week	*NRAS;* p.G13R
WCB1476	3	2	0	III	45 in 25	5 Weeks	*KRAS*; p.G13D
WCB1477	3	0	0	II	45 in 25	5 Weeks	*KRAS*; p.G12D
WCB1478	4	1	0	III	25 in 5	1 Week	*NRAS;* p.Q61L *APC;* p.E1379X†
WCB1479	3	2	1	IV	45 in 25	5 Weeks	*KRAS;* p.G12A
WCB1493	3	0	0	II	45 in 25	5 Weeks	*KRAS;* p.G13D
WCB1602	3	2	0	III	45 in 25	5 Weeks	WT
WCB1603	4	0	0	II	45 in 25	5 Weeks	*KRAS;* p.G12A
WCB1604	3	1	0	III	45 in 25	5 Weeks	*KRAS;* p.G12A

T - Clinical staging of tumour size.

N – Clinical staging of lymph node invasion.

M – Clinical staging of metastases.

Gy in Doses – The amount of radiotherapy received by a patient in Gy and the number of doses in which radiotherapy was distributed.

Disease Stage – Patient staging calculated using TNM stages according to the American Join Committee on Cancer (AJCC) as recommended by ESMO guidelines (Glynne-Jones et al. 2017).

WT – Wild-Type – No tumour specific variants were detected in tumour biopsies from these patients.

As tumour-specific variants were identified in 12/16 (75%) patients with LARC, longitudinal ctDNA analysis was only possible in these 12 patients. NGS analysis for ctDNA was not possible at this time and was, therefore, not available for the four patients with wild-type tumours. Sequential analysis was, therefore, performed using ddPCR. ddPCR assays were designed and validated for each tumour-specific variant in all 12 patients for sequential ctDNA analysis.

CtDNA was successfully detected in at least one timepoint from 2/12 (16.7%) patients with LARC whilst the remaining ten patients were negative at all times (Table 19). Both patients had ctDNA detected in pre-treatment plasma whilst one patient also had ctDNA detected post-treatment.

Table 19: Variant Allelic Frequencies of tumour-specific variants detected in cell free DNA in longitudinal plasma samples collected throughout therapy in patients with Locally Advanced Rectal Cancer.

Patient ID	Week of Sample Collection	Treatment Stage at Time of Collection	CtDNA VAF (%)
WCB1262	0	Pre-Treatment	0
	1	On-Treatment	0
WCB1263	1	On-Treatment	0
	3	On-Treatment	0
	9	Post-Treatment	0
	23	Post-Treatment	0
WCB1268	0	Pre-Treatment	0
	1	On-Treatment	0
WCB1441	-2	Pre-Treatment	0
	-1	Pre-Treatment	0
	1	On-Treatment	0
WCB1473	0	Pre-Treatment	1.0
	2	Post-Treatment	2.3
	4	Post-Treatment	0
	13	Post-Treatment	2.8
WCB1476	0	Pre-Treatment	0
	12	Post-Treatment	0
WCB1477	0	Pre-Treatment	0
	12	Post-Treatment	0
WCB1478	0	Pre-Treatment	3.5
	4	Post-Treatment	0
WCB1479	0	Pre-Treatment	0
	5	On-Treatment	0
	6	Post-Treatment	0
	8	Post-Treatment	0
	12	Post-Treatment	0
WCB1493	0	Pre-Treatment	0
	1	On-Treatment	0
	3	On-Treatment	0
	4	On-Treatment	0
	6	Post-Treatment	0
WCB1603	0	Pre-Treatment	0
	1	On-Treatment	0
	5	On-Treatment	0
	14	Post-Treatment	0
WCB1604	-8	Pre-Treatment	0
	-3	Pre-Treatment	0
	0	Pre-Treatment	0
	6	Post-Treatment	0

Detected ctDNA VAFs varied between 1.0 %– 3.5% in this study, which are above our threshold for positive variant detection (1% VAF and five or more mutant droplets; See Section 2.3.2.2).

Pre-treatment samples were available for 10 patients with known tumour-specific variants, within which, ctDNA was detected in two (20%). No patient had ctDNA detected whilst on-treatment, whilst 1/9 patients with available post-treatment sample (11.1%) had detectable levels of ctDNA. Due to the low number of ctDNA positive samples, very few patterns could be discerned. Statistical analysis was not possible between patients, stage of treatment or week of sample collection.

Notably, one patient with ctDNA detected in pre-treatment plasma had metastatic disease (WCB1473), whilst the other patient (WCB1478) displayed signs of tumour invasion into surrounding lymph nodes. However, these numbers are too small to make any meaningful deductions from.

It is currently unclear whether the decline of ctDNA levels in patient WCB1478 between weeks 0 – 4 is an immediate decline in ctDNA or a gradual decrease. Unfortunately, no patterns after this were observed as this patient had no later timepoints available for further analysis.

Patient WCB1473 had detectable ctDNA at both weeks 0 (pre-treatment) and 2 (post-treatment). CtDNA was then undetectable at week 4 (post-treatment) before being detected again at week 17 (post-treatment). This data may suggest that pCRT was not effective in causing significant tumour shrinkage or regression in this patient, as ctDNA can still be detected long after the completion of therapy. The undetectable levels of ctDNA at week 4 may suggest a small initial response to radiotherapy. This may, alternatively, be a result of natural biological variation or ctDNA levels being below our technical limit of detection. As this finding is only observed in a single patient, no significant conclusions can be drawn without expansion with a larger patient cohort.

We did not detect ctDNA at any time in patient WCB1479 who had metastases detected at diagnosis. This might be due to a number of other limitations of the study such as the sensitivity of ddPCR or low input plasma volume.

No ctDNA was detectable in any patient samples whilst on-treatment. Other reports have also shown that ctDNA is less frequently detected after pCRT and/or surgery (Tie et al. 2018). Provencio and colleagues (2017) showed similar results when they reported that ctDNA decreased significantly after the onset of therapy in 7/8 patients with NSCLC. We hypothesised that the decrease in ctDNA observed from pre- to on-treatment samples may be indicative of

initial patient response as a result of pCRT. However, this is a small number of patients with detectable levels of ctDNA, therefore no meaningful analyses can be made with regards to this observation.

Overall, we detected ctDNA in 2/12 (16.7%) patients with LARC using ddPCR. Although this may simply be the nature of this cohort, such low ctDNA detection is a potentially large and problematic limitation of this study. If this expected frequency is consistent within our patient cohort, a larger number of patients will be required for analysis to achieve ctDNA detection in enough patients with LARC for statistically meaningful analysis.

3.2.3 Analysing Sequential Changes in Total Cell Free DNA Throughout Therapy

3.2.3.1 Cell Free DNA Concentrations

Objective: Can we detect and measure quantitative changes in total cfDNA before, during and after treatment in patients with LARC?

For all 16 patients with LARC who were previously investigated for sequential ctDNA analysis, 1mL of plasma was used for total cfDNA extraction and analysis. Total cfDNA concentrations were quantified using the Invitrogen Qubit (ng/µL). Plasma samples were separated into pre-, on- and post-treatment according to sample collection and treatment dates (See Table 19).

Total cfDNA was successfully extracted and detected for all patient samples except one (patient WCB1479 week 5). No differences were observed between cfDNA samples which had detectable ctDNA levels and those which did not. Total cfDNA concentrations varied from 0.119ng/µL - 0.758 ng/µL (average: 0.29ng/µL), see Appendices Section 9.4 for full dataset.

We observed a general increase in total cfDNA concentration from pre- to on-treatment samples and a decrease from on- to post-treatment samples (Figure 12). Though a trend may be seen these differences were not statistically significant when analysed using ANNOVAR (pre-treatment vs on-treatment p-value=0.08 and pre-treatment vs post-treatment p-value=0.19). These findings might be influenced by individual patient responses to therapy. Agostini and colleagues (2011) reported that total cfDNA concentrations were significantly decreased from pre- to post-CRT plasma samples in patients with LARC who responded well to therapy (TRG1-2). Similarly, Park and colleagues (2018) reported that total cfDNA in patients with hepatocellular carcinoma did not alter significantly from plasma samples before and after the administration of radiotherapy in patients who responded well to therapy. However, patients who did not respond or responded poorly displayed a significant increase in total

cfDNA concentration after radiotherapy. Therefore, these data may contain two or more different signatures based on clinical outcomes that we cannot currently distinguish.

We did not observe any differences in total cfDNA concentrations according to T/N/M staging (Table 18). This agreed with respective ANNOVAR and two-sample t-tests which showed no significant differences using any of the stages in univariate analyses. A difference was seen using ESMO group staging (Glynne-Jones et al. 2017) in these patients between total cfDNA from stage III v IV patients (p-value=0.02; Figure 12). Stage IV patients here displayed significantly lower total cfDNA concentrations than stage III patients. However, significance was lost when separating cfDNA samples into their respective treatment stage at time of collection. This finding is surprising as previous reports have shown the opposite; a positive correlation between tumour stage and cfDNA concentration in patients with hepatocellular carcinoma (Park et al. 2018).

The difference observed in cfDNA concentrations between patients with stage III and IV disease contradicts previous findings by Boysen and colleagues (2017). This study reported that total cfDNA concentrations positively correlated with AJCC grouped stage in patients with LARC, whereas we observed a decrease from stage III-IV groups according to ESMO guidelines. This same study also reported a significant difference in total cfDNA in patients with LARC who had cancer spread to their lymph nodes and those who did not. In contrast, our study showed no significant differences.

Overall, we were able to detect and quantify total cfDNA from all plasma samples except one using the Invitrogen Qubit. CfDNA concentrations during therapy were generally but not significantly higher than pre- and post-treatment samples using the Qubit. We also observed a significant difference between total cfDNA concentrations between patients with stage III and IV disease, though significance was lost when samples were grouped into pre-, on- or post-treatment. Though general patterns of cfDNA change were observed, no meaningful analysis can be drawn from these results due to our small sample size and general limitations associated with cfDNA analysis.

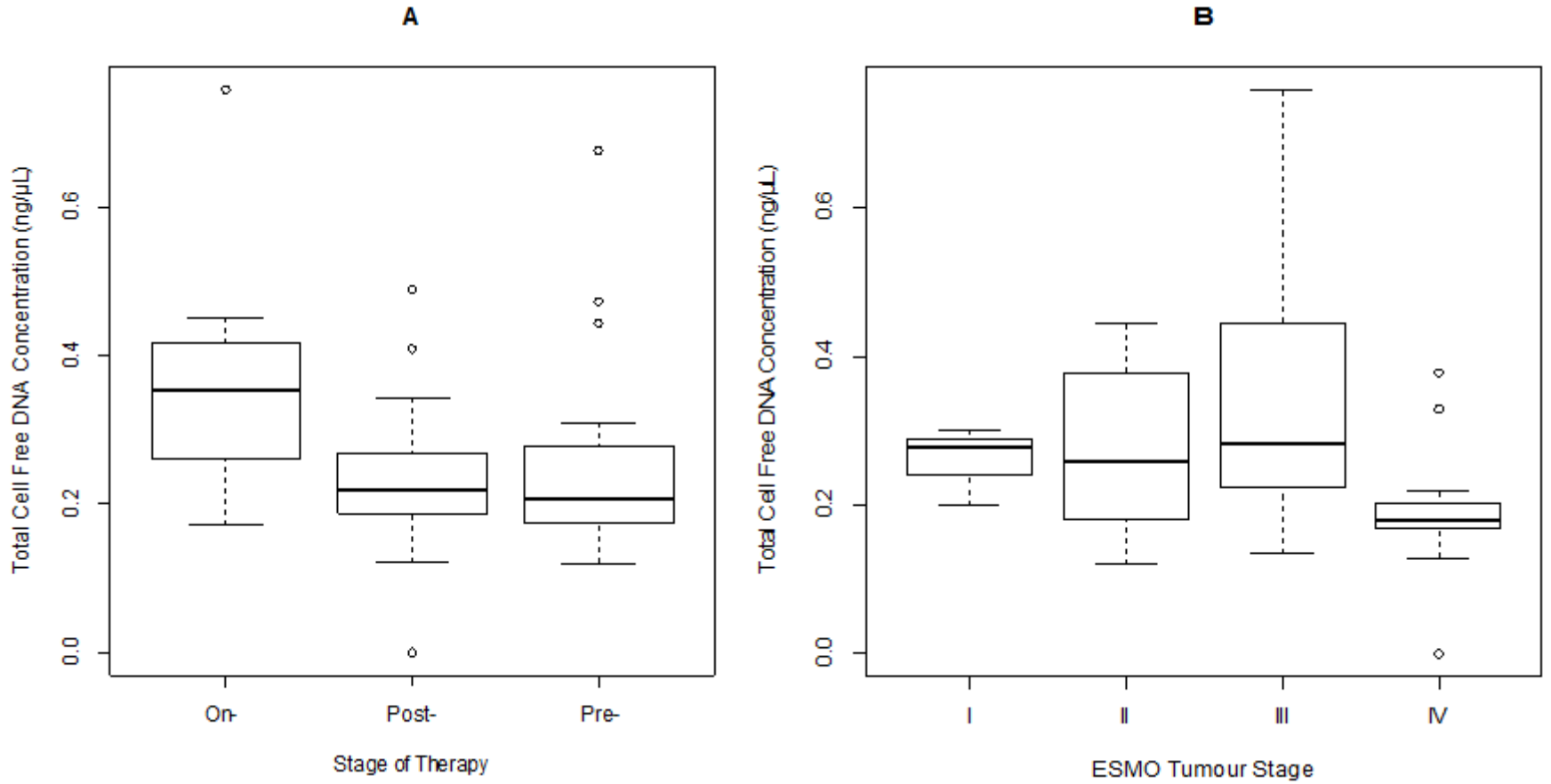

Figure 12: Total cell free DNA concentration and distribution across A- When the plasma samples were collected (Pre-, On- or Post-Radiotherapy) and B- Tumour Stage According to the European Society of Medical Oncology

3.2.4 Patient Clinical Outcomes

3.2.4.1 General Patient Outcomes

Clinical outcomes for all 16 patients became available after the completion of circulating biomarker analysis described above. This data included patient initial patient response to therapy and whether patient experienced distant and/or local relapse within two years of the initiation of therapy. Survival was not included due to the limited follow-up time available for the majority of patients (median: 25.9 months; range: 5-40 months). Two patients (WCB1368 and WCB1473) were excluded from clinical outcome analyses due to having received atypical treatment.

Initial patient response to neoadjuvant therapy was recorded as Tumour Regression Grades (TRG) as reported on CaNISC. TRG systems were developed to quantitatively analyse tumour response according to growth or shrinkage after the administration of neoadjuvant therapy in patients with solid tumours, such as oesophageal (Mandard et al. 1994) or rectal cancer (Dworak et al. 1997). TRG has traditionally been quantified at the time of surgery (pathological TRG; pTRG). pTRG has several limitations (Sclafani et al. 2017). Quantifying a patient's pTRG requires the patient to undergo surgery, which not all patients can, after the administration of neoadjuvant therapy. pTRG has also been reported to have a low inter-user variability (Lindebjerg et al. 2011). Furthermore, patients may be analysed using different TRG systems between different treatment centres, which can create further inconsistencies when deciding whether to administer post-operative therapy.

As a result of advancements in imaging technology, MRI-based Tumour Regression Grading (mrTRG) has recently been investigated as a method to predict patient pTRG after completing neoadjuvant therapy (Patel et al. 2011). Therefore, where patients cannot or do not want to undergo surgery, mrTRG has been suggested to be used in place of pTRG. However, studies have demonstrated that the concordance between these two methods is low (Sclafani et al. 2017), thus mrTRG is not yet reliable enough to be an accurate predictor of pTRG or cCR. As imaging methods continue to develop and improve, mrTRG may become a more reliable predictor of pTRG and long-term outcomes. At present, these two factors do not concord well in patients with LARC.

In this study, TRG principally refers to pTRG, except where such data was unavailable in which case mrTRG was used instead. Although we are aware of the limited correlation between the two response values, there was no other available metric which was more comparable to

pTRG, thus mrTRG was applied only where necessary. Patients who achieved a complete response to therapy (TRG1) may have been gauged either pathologically (pCR) or clinically (cCR). Patient clinical characteristics and outcomes are summarised in Table 20 and Table 21.

Table 20: Treatment outcomes for patients with LARC consented by the Wales Cancer Bank

Patient ID	Disease Stage	Radiotherapy (Gy in Doses)	TRG	Patient Experienced Local and/or Distant Relapse
WCB1262	II	45 in 25	1	None
WCB1263	III	45 in 25	3	None
WCB1268	IV	45 in 25	3	Distant only
WCB1295	III	45 in 25	1	None
WCB1419	III	25 in 5	4	Distant only
WCB1441	II	45 in 25	4	Distant only
WCB1476	III	45 in 25	2	None
WCB1477	II	45 in 25	4	Local only
WCB1478	III	25 in 5	4	n/a
WCB1479	IV	45 in 25	3	Distant only
WCB1493	II	45 in 25	1	Local only
WCB1602	III	45 in 25	1	None
WCB1603	II	45 in 25	3	Local and distant
WCB1604	III	45 in 25	2	Distant only

n/a

Table 21: Treatment strategies and outcomes for all patients with locally advanced rectal cancer

Variables	N	(%)
Radiotherapy Regimen		
Short Course (One Week)	2	14.3
Long Course (Five Weeks)	12	85.7
Surgery Performed		
Yes	9	64.3
No	5	35.7
TRG		
TRG 1	3	21.4
TRG 2	2	14.3
TRG 3	5	35.7
TRG 4	4	28.6
Disease Recurrence within Two Years†		
Yes	9	69.2
No	4	30.8

† - Recurrence data for one patient (WCB1478) at two years was unavailable.

Overall, 9/14 patients (64.3%) received surgery after treatment with radiotherapy. Patient WCB1419 was not fit enough for surgery whilst three patients (WCB1295, WCB1493 and WCB1602) underwent active monitoring as they achieved cCR. In contrast, patient WCB1268 was found to have distant metastases before surgery could take place.

Four patients had neoadjuvant chemotherapy in addition to or before the administration of radiotherapy (WCB1268, WCB1477, WCB1479 and WCB1603) and two patients had detectable metastases at diagnosis (WCB1268 and WCB1479).

Following the completion of treatment, 9/13 (69.2%) patients experienced disease recurrence within two years of follow-up, of which five and six patients experienced local and distant recurrence respectively (including two patients who experienced both distant and local disease recurrence). Median follow-up for these patients was 64.8 months (range: 11.7-96.9 months).

From our cohort, three patients (21.4%) achieved a complete response (TRG 1). This in line with studies which have reported a complete response in 12.5-20.6% of patients with LARC undergoing pre-operative CRT (Maas et al. 2010; Akiyoshi et al. 2012). In total, five patients (35.7%) displayed a good response (TRG 1/2) to neoadjuvant therapy, including the three who achieved cCR, whilst nine patients (64.3%) displayed a poor response.

Objective: Do longitudinal ctDNA levels correlate with clinical outcomes in patients with LARC?

Patient response to therapy was separated in a binary fashion (TRG 1/2 and TRG 3/4 for good and poor responders respectively). In addition, patients were also separated out into those who achieved a complete (TRG 1) vs incomplete response (TRG 2-4) and the recurrence of local and/or distant disease within the first two years of follow-up (see Table 22, Table 23 and Table 24 respectively). We analysed recurrence in patients after two years as studies have reported that disease recurrence within this time period was indicative of long-term patient outcomes (Sargent et al, 2011). This data was acquired with the ambition of comparing circulating biomarker concentrations to each of these clinical responses to identify any clinically significant associations.

Although we had a total of 14 patients with LARC for circulating biomarker analysis, not all patients had plasma collected at all timepoints, nor did all patients undergo ctDNA analysis. Therefore, some patients are missing from certain sections of the analysis, which is illustrated by the variable values in the N column.

Table 22: A comparison of longitudinal ctDNA and cfDNA levels across patients with good and poor responses to neoadjuvant therapy

	Good Response (TRG 1/2)		Poor Response (TRG 3/4)		
Variables	**N**	**Median (Range)**	**N**	**Median (Range)**	**p-value**
Pre-Treatment Levels					
CtDNA VAF (%)	4	0 (0.0 – 0.0)	7	0 (0.0 – 3.5)	0.3
CfDNA Concentration (ng/µL)	6	0.42 (0.22 – 0.45)	9	0.35 (0.22 – 0.76)	0.79
On-Treatment Levels					
CtDNA VAF (%)	2	0 n/a	5	0 n/a	n/a
CfDNA Concentration (ng/µL)	5	0.41 (0.10 – 0.43)	5	0.19 (0.19 – 0.21)	*0.08*
Post-Treatment Levels					
CtDNA VAF (%)	3	0 n/a	6	0 (0.0 – 2.8)	0.52
CfDNA Concentration (ng/µL)	4	0.36 (0.16 – 0.68)	8	0.2 (0.12 – 0.47)	0.14

Table 23: A comparison of longitudinal ctDNA and cfDNA levels across patients with complete and incomplete responses to neoadjuvant therapy

	Complete Response (TRG 1)		Incomplete Response (TRG 2-4)		
Variables	**N**	**Median (Range)**	**N**	**Median (Range)**	**P-Value**
Pre-Treatment Levels					
ctDNA VAF (%)	2	0 (0.0 – 0.0)	9	0 (0.0 – 3.5)	0.52
cfDNA Concentration (ng/µL)	4	0.42 (0.34 – 0.43)	10	0.345 (0.22 – 0.76)	0.78
On-Treatment Levels					
CtDNA VAF (%)	2	0 n/a	5	0 n/a	n/a
CfDNA Concentration (ng/µL)	4	0.42 (0.34 – 0.43)	6	0.19 (0.10 – 0.21)	<0.01
Post-Treatment Levels					
CtDNA VAF (%)	3	0 n/a	6	0 (0.0 – 2.8)	n/a
CfDNA Concentration (ng/µL)	2	0.56 (0.44 – 0.68)	9	0.2 (0.12 – 0.47)	<0.01

Table 24: A comparison longitudinal ctDNa and cfDNA levels between different patient responses to neoadjuvant therapy. Any recurrence refers to patients who experience either local or distant relapse.

	Local Relapse			Distant Relapse			Any Relapse		
Variable	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p- Value**	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p- Value**	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p- Value**
Pre-Treatment Levels									
CtDNA VAF (%)	0 (0.0 – 0.6)	0 (0.0 – 3.5)	0.74	0 (0.0 – 0.0)	0 (0.0 – 3.5)	0.26	0 (0.0 – 0.0)	0 (0.0 – 3.5)	0.50
CfDNA Concentration (ng/µL)	0.36 (0.26 – 0.76)	0.26 (0.22 – 0.41)	0.36	0.42 (0.32 – 0.45)	0.35 (0.22 – 0.76)	0.8	0.43 (0.34 – 0.45)	0.36 (0.22 – 0.76)	0.76
On-Treatment Levels									
CfDNA Concentration (ng/µL)	0.21 (0.19 – 0.38)	0.23 (0.19 – 0.41)	0.85	0.34 (0.19 – 0.41)	0.19 (0.19 – 0.21)	*0.10*	0.34 (0.19 – 0.38)	0.21 (0.19 – 0.43)	0.44
Post-Treatment Levels									
CtDNA VAF	0 (0.0 – 0.0)	0 (0.0 – 2.8)	0.48	0 (0.0 – 0.0)	0 (0 – 2.8)	0.36	0 (0.0 – 0.0)	0 (0.0 – 2.8)	0.60
CfDNA Concentration (ng/µL)	0.26 (0.16 – 0.47)	0.21 (0.12 – 0.68)	0.92	0.27 (0.22 – 0.68)	0.18 (0.16 – 0.47)	0.19	0.26 (0.24 – 0.27)	0.2 (0.12 – 0.68)	0.76

Here we observed significantly higher total cfDNA concentrations in patients who achieved a complete response to treatment both during and after therapy (Table 23). Although a suggestive significance (p=0.08) for patients who had a good response was also seen (Table 22), this was not reflected in the local and/or distant recurrence analysis (Table 24). This remained consistent when patients with pre-treatment metastases were removed from analysis. In contrast, Zitt and colleagues (2008) have previously reported that patients with LARC who responded well to pCRT saw a decrease in total cfDNA concentrations from pre- to post-treatment samples.

Studies have also reported that high cfDNA concentrations at baseline were indicative of an increased risk of recurrence and poor DFS in patients with LARC (Schou et al. 2018). This contrasts with our finding that cfDNA concentrations were higher during and after therapy in patients who achieved a complete response to treatment. We did not see any correlations between tumour stage and pre-treatment cfDNA concentrations, which differs from prior reports (Boysen et al. 2017).

We saw no associations between baseline cfDNA concentrations and initial patient response to therapy nor the likelihood a patient would relapse. This also remained consistent when patients with pre-treatment metastases were removed from analyses. In contrast, Schou and colleagues (2018) reported that high pre-treatment cfDNA was associated with a decreased likelihood of achieving pCR and poor DFS in patients with LARC. This is in agreement with a study from Vivancos and colleagues (2018) who also demonstrated that high pre-treatment cfDNA was associated with an increased risk of recurrence within five years in patients with LARC. However, our data does not support this finding. Our findings may contradict the literature due to several factors, including the reproducibility or sensitivity of the Qubit, the potential impact of pre-analytical handling of samples or the fact that cfDNA is not a tumour specific biomarker.

Whenever ctDNA was detected at baseline, patients would have responded poorly in both (100%) cases. Both patients (WCB1295 and WCB1478) had detectable pre-treatment metastases and lymph node involvement respectively. Although this might suggest that detectable baseline ctDNA is a poor prognostic indicator, this is also likely to be a result of the small sample size and number of positive events. Thus, no meaningful findings can be reported at this time. No patients had detectable ctDNA either during or after the administration of neoadjuvant treatment. This may have been influenced the detectable ctDNA levels. Due to

the low numbers of patients with detected ctDNA at any time, comparisons between different response rates cannot be performed statistically.

3.3 Discussion

3.3.1 Conclusions

Objective: Can ctDNA be detected in the circulation of patients with LARC before the initiation of therapy using NGS?

We detected pre-treatment ctDNA in three patients with LARC using NGS (Table 17) achieving a sensitivity and concordance of 30% and 37.5% respectively using NGS. Previous studies in patients with LARC have shown variable sensitivities when identifying pre-treatment ctDNA using NGS (Tie et al. 2018). We postulate that the decreased levels we observed here are the result of numerous factors addressed below.

Lower concordance rates have been reported in locally advanced cancer patients in comparison to patients with metastatic CRC (Grasselli et al. 2017; Hsu et al. 2018). This is likely a result of the low tumour burden or absence/early stages of metastatic lesions in these patients. Carpinetti and colleagues (2015) reported that ctDNA was difficult to detect in patients with early stage tumours (T1-T2 stage). Siravegna and colleagues (2015) reported that their patients who displayed discordance between tumour tissue and cfDNA all had low tumour burden, an important factor to consider within our patient cohort. A similar finding was reported by Bettegowda and colleagues (2014), where significant differences were observed in the frequency of ctDNA detection between patients with locally advanced and metastatic disease across a range of solid tumours (including patients with CRC).

The sensitivity and concordance of this study are lower than anticipated. Comparisons to other studies in the literature are difficult, as we only found two studies directly investigating patients with LARC which varied widely in results (Sclafani et al. 2018; Tie et al. 2018). Of these, only one used NGS to detect ctDNA (Tie et al. 2018). Although, we intended to compare our findings to studies with metastatic patients, the large differences in patient characteristics did not allow a fair and direct comparison.

Our reduced sensitivity and concordance may be due to numerous factors. A low plasma volume (1-2mL) providing a low cfDNA concentration, created difficulties in variant detection, as the NGS required a larger number of amplification cycles which can cause amplification bias, for either wild-type or mutant DNA.

Furthermore, the NGS system we applied had a relatively large amount of 'background noise' making variant detection <1.0% difficult to discern from artefacts. Therefore, true variants in this range may have been interpreted incorrectly.

Our sensitivity and concordance may also be influenced by sample deterioration and improper pre-analytical handling. CfDNA is an unstable biomarker which can fragment easily if not handled properly or undergoes numerous freeze thaws during storage (El Messaoudi et al. 2013). Furthermore, as the pre-analytical plasma processing, storage and transportation was performed by the WCB, issues may have occurred here which damaged sample quality and concentration. This may have further detrimental effects on downstream analysis.

Another factor which can influence ctDNA detection are the biological characteristics of the patients themselves. The findings that ctDNA is less frequently detected and found at lower levels in patients with LARC than patients with metastatic disease suggests that tumours shed less DNA in these patients. This could be because activity and growth are lower in early-stage tumours, or that tumours may not yet be in close proximity to blood vessels. Perhaps the tumour has not grown to the size appropriate required for necrosis or other forms of cell death to release abundant ctDNA into the circulation for reliable detection. Thus, these patients may not benefit from circulating biomarker analysis.

Overall, we confidently detected ctDNA in pre-treatment samples from patients with LARC using NGS. However, our observed sensitivity for detecting ctDNA was below expectations based on the literature, which may be due to several technical or biological factors.

Objective: *Can ctDNA be reliably detected in sequential plasma samples before, during or after therapy in patients with LARC using ddPCR?*

We observed lower frequencies of ctDNA using ddPCR in comparison to NGS. Our findings were also low in comparison to previous studies using ddPCR (Sclafani et al. 2018) or NGS (Tie et al. 2018) who reported sensitivities of 34% and 76% respectively for detecting ctDNA in patients with LARC. The use of NGS to screen for numerous variants rather than being limited to one by ddPCR may therefore provide some inherent advantages to ctDNA detection. Our findings may also be influenced by the sensitivity of ddPCR. As ddPCR does not amplify target DNA in the same fashion as NGS, analysis is much more limited by sample input. Therefore, detecting low VAFs in low DNA concentrations can be difficult, as shown in the Materials and Methods Chapter.

For a number of these patients, we had no access to tumour tissue for variant detection and instead relied upon variants identified during clinical analysis. Furthermore, for the tissue we did sequence we did so with a targeted sequencing panel, rather than whole exosome or genome sequencing, thus limiting the number of variants we could analyse. This could have also assisted the identification of variants in patients where we did not detect any.

Another important factor which could influence ctDNA detection is cfDNA concentration and plasma volumes. Tie and colleagues (2018) applied 10mL of patient plasma for ctDNA analysis, whereas our study and that by Sclafani and colleagues (2018) only investigated ctDNA in 1-2mL and 2mL of plasma respectively. The use of greater volumes of plasma for cfDNA extraction can benefit ctDNA analysis by providing greater concentrations of cfDNA for analysis whilst increasing the likelihood of low levels of ctDNA being extracted. The increased concentration of cfDNA can improve the confident detection of variants with low VAFs (see Section 2.3.2). The large difference in plasma volume applied here may therefore contribute to the difference in sensitivities observed between these studies.

Pre-analytical sample handling is another key factor which may have contributed to our limited findings. As samples were collected and processed externally, we cannot be certain whether all samples were processed within an appropriate time interval. An increased time period between collection and processing can cause the lysis of white blood cells, which consequently shed DNA into plasma. This can appear as cfDNA during analysis and may mask the presence of ctDNA.

We were only able to perform ctDNA or cfDNA analysis in a very small number of patients with LARC. With small numbers and small positive incidents of ctDNA detection, it is not possible to identify changes which are clinically significant.

The clinical characteristics of our patients are also important to consider when investigating ctDNA in cancer patients. Beije and colleagues (2016) have reported ctDNA had a higher concordance with metastatic sites (55%) than matching primary tumours (39%) in patients with metastatic CRC. This suggests that ctDNA in these patients is largely derived from or associated with metastatic lesions. Our study did include three patients with known metastatic lesions at diagnosis, of which one (WCB1473) had detectable pre-treatment ctDNA. For patient WCB1473 the biopsy was taken from the metastatic lesion, correlating with this finding from Beije and colleagues (2016). Our results may agree with these findings as both patients with detected ctDNA levels had either metastases (M1) or invasion of local lymph nodes (N1) with a

high tumour burden (T4) (Bettegowda et al. 2014). However, this is merely a qualitative observation due to our low ctDNA detection rate. Furthermore, we did not detect ctDNA in patient WCB1479 who also had metastases detected at diagnosis. This might be because the metastatic lesion was not sequenced in this patient, and concordance is higher between metastatic lesions and ctDNA. Otherwise, any number of limitations discussed here could be the cause of not detecting ctDNA in this patient.

From the literature we expected to see detectable levels of ctDNA in pre-treatment plasma, which would have mostly decreased after the initiation of radiotherapy. We then hoped to see changes in ctDNA, after the completion of radiotherapy, which would potentially correlate to patient outcomes and response. However, due to our lack of ctDNA positive patients, we were unable to see any such association.

We also anticipated an initial peak in ctDNA release after initiating radiotherapy, which could have been indicative of sudden tumour cell death. We did not see any such peak, which may be due to this phenomenon not occurring, or issues with plasma timepoints. In order to see such an event, plasma samples might have to be collected more frequently or soon after the administration of radiotherapy in these patients. We are currently unaware of the optimal timings for blood collection to detect the shedding of ctDNA into the circulation and measure circulating biomarkers in relation to radiotherapy. Circulating biomarker analysis throughout radiotherapy is ideal. However, we need to identify when to collect plasma after the administration of radiotherapy to allow optimal circulating biomarker analysis. This requires further investigation and standardisation in the future to facilitate predicting patient outcomes and response.

Using ddPCR, we were unable to reliably detect ctDNA in sequential plasma samples from patients with LARC which is likely to be a consequence of numerous biological and/or technical factors.

Objective: Can we detect and measure quantitative changes in total cfDNA before, during and after treatment in patients with LARC?

We were able to extract and detect cfDNA within all patient samples except one (patient WCB1479 week 5). We were also able to measure quantitative changes in total cfDNA throughout pCRT using 1mL of patient plasma. However, the method used to quantify total cfDNA in these patients has been shown to possess limitations when compared to qPCR (Nakayama et al. 2016). Alternatively, other studies have shown that the Qubit correlated well

with qPCR when quantifying with cfDNA (Ponti et al. 2018). Therefore, further studies are required to analyse the efficacy of using the Qubit for total cfDNA quantification. As a result, we do not yet fully understand whether the small changes in total cfDNA levels are clinically significant.

We observed a general increase in total cfDNA concentration from pre- to on-treatment samples followed by a decrease from on- to post-treatment samples. However, these differences were not statistically significant. In contrast to our findings, Panditharatna and colleagues (2018) reported a decrease in cfDNA concentrations after the initiation of therapy, suggesting that this was indicative of tumour down-staging or decreased tumour growth and less DNA being shed by the tumour and overall decreased total cfDNA concentrations. The increase seen in our study may represent elevated tumour- and surrounding non-tumour cell death as a result of treatment. The subsequent decrease may be a result of the cessation of therapy and lower rates of cell death. This may also be simply the result of natural biological variation in total cfDNA. This hypothesis is based on the total cfDNA levels of patient WCB1603, who displayed drastic changes before the initiation of therapy.(Zitt et al. 2008; Agostini et al. 2011; Park et al. 2018).

Differences may also exist between ours and various other studies in the literature due to several other factors. Where the studies by Zitt (2008), Agostini (2011) and ourselves focusses on changes in patients with LARC, the study by Park and colleagues (2018) investigates patients with hepatocellular carcinoma. Albeit that all four studies focus on changes occurring as a result of radiotherapy. The pre-treatment total cfDNA we observe here is higher than that reported by Zitt and colleagues (0.001-0.017ng/ µL from 125-240µL of plasma) and Agostini and colleagues (0.009-0.27ng/µL from 500µL of plasma) in patients with rectal cancer. Even when normalised, our findings are generally higher than these two studies. This difference might be due to the techniques used to analyse and quantify total cfDNA.

In each study, a different commercial kit was used to extract total cfDNA, although all were based on Qiagen's QIAamp service. Therefore, some natural variation is likely to occur between each of the three methods. Furthermore, where we used the Qubit, the other two studies applied qPCR of reference genes to quantify total cfDNA (Zitt et al. 2008; Agostini et al. 2011). The use of the Qubit is simple but potentially influenced by contamination. The use of qPCR or ddPCR for total cfDNA concentration can be more accurate and reliable than fluorescent based methods like the Qubit. However, as copy number changes can frequently occur in cancer, the number of genes and genes in question for quantification will need

standardisation. Once standardised, qPCR and ddPCR provide distinct advantages over the Qubit.

We did observe a significant difference in total cfDNA concentration between patients with stage III to IV, where stage IV had significantly lower total cfDNA concentrations. This significance was lost when separating cfDNA sample according to treatment stage at time of collection (pre-, on- or post-treatment). Otherwise, we did not observe any significant differences in total cfDNA concentration based on patient TNM or stage grouping.

This contradicts findings from Boysen and colleagues (2017) who reported significant differences between patients with lymph node positive and negative disease as well as a positive correlation between total cfDNA and the stage of patients according the American Joint Committee on Cancer (AJCC). Another study by Schou and colleagues (2018) found no significant differences according to patient AJCC stage either. In this case, Schou and colleagues quantified total cfDNA directly from plasma using a fluorometric method whereas Boysen and colleagues quantified reference genes using ddPCR. The different nature of these methods, where ddPCR detects specifically amplifiable levels of cfDNA whereas the fluorometric assays are less specific could be the cause of differences observed here. Fluorometric methods can theoretically quantify cfDNA which is fragmented beyond the point where amplification is reliable. However, these methods can be influenced by the presence of contaminants. Both studies also investigated larger numbers of patients with LARC (Schou and Boysen analysing 123 and 74 patients with LARC respectively). Therefore, the detection of significant differences in tumour stage may require a larger cohort to reliably observe.

Furthermore, total cfDNA levels can be influenced by other physiological factors, because it can be derived from both tumour and non-tumour tissue. Such factors can include radiation-induced toxicities, exercise, severe infection or other causes of inflammation (Vittori et al. 2019). Significant changes in total cfDNA can also be caused by improper pre-analytical sample handling causing white blood cell lysis and DNA to shed into plasma (Pritchard et al. 2012). Zwirner and colleagues (2018) also reported that cfDNA concentrations were raised for up to three days after radiotherapy. Thus, any concentrations drawn during or shortly after radiotherapy must consider this potential confounding factor. Additionally, the presence of radiation-induced toxicity or infection can also cause significant increases in concentrations, which must be considered for future analyses (Zwirner et al. 2018).

As mentioned previously, our study is limited by small sample size. This can create difficulties in achieving statistical significance and is likely to contribute to our limited findings.

Overall, we were able to detect and quantify total cfDNA in most of our patient cohort. The significantly higher concentrations observed in total cfDNA samples collected during therapy may be indicative of increased tumour and/or non-tumour cell death after the initiation of therapy.

3.3.2 Clinical Implications

Objective Do longitudinal ctDNA levels correlate with clinical outcomes in patients with LARC?

Due to our low frequency of ctDNA detection, comparisons to clinical outcomes to find significant associations was not possible. However, we looked at individual patients and their clinical outcomes. Patient WCB1478 had detectable pre-treatment ctDNA which subsequently depleted before the collection of plasma at week 4 respectively. This patient had a poor response to therapy (TRG4) which contradicts the literature. Hsu and colleages (2018) have previously reported that a decrease in ctDNA levels from pre- to post-treatment were indicative of tumour shrinkage in patients with metastsatic CRC. Similarly, Paditharatna and colleagues (2018) reported that a significant decrease in ctDNA after therapy corelated with tumour shrinkage in patients with gliomas. . The lack of detectable ctDNA in patient WCB1478, may be a consequence of non-biological reasons which are discussed further below or the relatively short time period between treatment completion and sample collection (~2 weeks).

Patient WCB1493 relapsed after the shortest time period of (1.2 months) but had no detectable ctDNA at any time using our methods. This was consistent with pre-treatment clinical factors, such as no detectable lymph node involvement or metastatic lesions. Although we saw no statistical significance, both patients with detectable pre-treatment ctDNA had either detectable metastases or lymph node involvement before the initiation of therapy.

Carpinetti and colleagues (2015) illustrated that ctDNA levels were low at pre-treatment, with initial good responders having a further decrease. However, our low overall ctDNA positivity makes it difficult to compare our findings at this time. This may be, in part, because we applied ddPCR to detect one variant in ctDNA, but changes could be occurring which are missed using this method. With the small sample sizes and low number of patients with detectable ctDNA levels at this time, comparisons are impossible.

Other studies have suggested that a post-therapeutic increase in ctDNA levels could be due to the presence and/or growth of minimal residual disease and/or micro-metastases (Tie et al. 2018). Carpinetti and colleagues (2015) reported that post-therapeutic increases in ctDNA were associated with disease progression whilst Lecomte and colleagues (2002) established that detectable ctDNA at time of surgery predicted poorer overall survival in CRC patients. This is in agreement with numerous reports which suggested that post-therapeutic increases in ctDNA correlated with decreased overall survival and time to recurrence (Diehl et al. 2005; Bettegowda et al. 2014; Provencio et al. 2017). Finally, in renal cell carcinoma, increase in

ctDNA during follow-up correlated with the emergence of brain metastases and was able to detect progression earlier than current clinical imaging methods (Yamamoto et al. 2019b). These reports suggest that detectable levels of ctDNA after the completion of therapy may indicate that treatment did not down-stage tumours to a significant extent. However, due to our lack of ctDNA positive samples, especially post treatment samples, we were unable to determine whether these findings are consistent in this cohort.

Studies have suggested that early changes in ctDNA are less significant regarding clinical outcomes than post-treatment ctDNA levels (Tie et al. 2018). Long and/or short course chemoradiotherapy with total mesorectal excision does provide benefits regarding local recurrence through tumour down-staging but not necessarily regarding overall survival in patients with LARC (Kapiteijn et al. 2001b). This is also highlighted by the fact that, lymph-node down-staging has been reported to be the most significant predictor of patient long-term survival (Willis 2007). This is likely to be due to the specific targeting of primary tumour with radiotherapy, which does not directly impact pre-existing metastatic lesions. Although pre-treatment ctDNA has been reported to correlate with stage and the presence of metastases (Diehl et al. 2005), detectable levels of ctDNA post-therapy appear to provide greater insight into the presence of metastatic lesions. The presence of metastases may also be unknown to the clinical team after the completion of treatment and may therefore influence further treatment-decision making.

Detectable ctDNA may simply be indicative of more advanced disease and/or the presence of metastatic lesions which may not have been identified previously. This finding is in agreement with a previous study which reported that increased baseline ctDNA was predictive of poor survival across patients with various cancer types (Bettegowda et al. 2014). As previously stated, this finding is merely observational and not statistically significant.

Patients with no detectable ctDNA may also be false negatives. CtDNA may have been called negative falsely, due to other causes such as the incorrect choice of variant for ctDNA analysis, pre-analytical handling or limited assay sensitivity as mentioned previously. Alternatively, this may be a consequence of the inherent tumour biology of our non-metastatic patients, who may shed less ctDNA into the circulation than metastatic counterparts.

The role of ctDNA in the clinic is rapidly developing, particularly with regards to targeted therapies and sensitivity- or resistance-associated tumour variants. With regards to patients with LARC undergoing pCRT, the clinical potential of ctDNA is still unravelling. The detection of

ctDNA after the completion of pCRT is a potentially useful tool as it may suggest the effectiveness of therapy. If so, these patients may have an increased risk of local or distant recurrence and therapeutic decisions can be made earlier.

Objective: Do longitudinal cfDNA levels correlate with clinical outcomes in patients with LARC?

CfDNA concentrations have reportedly been able to predict the response of patients with LARC to pCRT. Here we observed that high sequential (on- and post-treatment) cfDNA concentrations were significantly associated with a complete response to therapy. The increased concentrations of cfDNA during treatment may be indicative of increased rates of tumour cell death and shedding of cfDNA into the circulation (Jahr et al. 2001). This effect may still be seen for a short time after the completion of treatment in post-treatment plasma samples. We did not identify any association between cfDNA concentration and disease recurrence.

This conflicts with previous reports from Zitt and colleagues (2008), who reported that patients who responded well to pCRT saw a decrease in total cfDNA concentration at the end of treatment, and those who did not respond well saw an increase. Similarly, .Schou and colleagues (2018) found that high cfDNA concentrations at baseline were predictive of poor recurrence- and disease-free survival, whilst Boysen and colleagues (2017) showed that high pre-surgical cfDNA concentration correlated with risk of recurrence. This is also in agreement with reports from Park and colleagues (2018) who demonstrated in hepatocellular carcinoma that high pre- and post-CRT cfDNA concentrations were indicative of more advanced disease and larger tumours. These reports suggest that high post-treatment cfDNA concentrations could suggest that tumour tissue remains in one form or another.

We observed in our cohort that patient WCB1419 retained consistently high total cfDNA concentrations from pre-to post-treatment. According to these reports, this may be indicative of poor treatment response and a poor predictive marker of outcomes in this patient. This is in agreement with our findings where this patient had a poor response to therapy (TRG4) and experienced distant relapse within two years of follow-up.

We also observed a relatively sharp increase in total cfDNA post therapy in patient WCB1603. According to previous reports, this would suggest ineffective treatment and possible tumour progression in this patient. These predictions concords with the fact that this patient experienced local and disease recurrence within two years. Although promising, these are only two examples and significant findings should not be based on these results.

For total cfDNA, reports have suggested that levels positively correlate with tumour stage (Boysen et al. 2017). Though we did not observe this ourselves, the utility of total cfDNA for this purpose at pre-treatment can help more clearly define tumour stage, which is currently the only available method which influences treatment decision-making.

There may be several reasons that our data may conflict with the literature. Firstly, our study was comprised of a relatively small number of patients with LARC (n=14) whereas the studies by Schou (2018) and Zitt (2008) analysed cfDNA in 123 and 26 patients with LARC respectively . Due to our small sample numbers, our significant findings may also be skewed by a small number of outliers. Even when using medians instead of means as an average, small sample numbers can allow outliers to influence the average cfDNA concentration during statistical analysis. CfDNA concentrations may also be detrimentally impacted by pre-analytical sample handling or an inefficient cfDNA extraction method. Finally, cfDNA is not specifically tumour-derived, and thus, may be influenced by the shedding of non-tumour tissue via other physiological processes, such as inflammation, infections and radiotherapy (Zwirner et al. 2018; Vittori et al. 2019).

This information could theoretically be used to influence treatment-making decisions where high-risk patients can undergo post-operative neoadjuvant therapy. Alternatively, low-risk patients could avoid unnecessary and toxic procedures with such information.

3.3.3 Future Studies

There are numerous ways with which this study can be altered to improve the reliability and consistency of the results. The analysis of more patients would potentially allow us to detect more ctDNA positive patients and interrogate resulting patterns. This would also help provide greater statistical and clinical significance when comparing ctDNA levels to patient outcomes.

Having tumour tissue to sequence for patients with longitudinal plasma samples could aid this project by allowing the analysis of more variants for ctDNA identification. The availability of whole tumours, multiple regions of a single tumour or larger tumour samples for sequencing would provide greater confidence in the variants detected and potentially help to resolve some issues surrounding intra-tumour heterogeneity.

The utility of whole exome or genome sequencing for tumour biopsies alongside germline DNA would have limited the number of patients without detectable tumour-specific variants. This would then increase the number of patients in which ctDNA can be analysed, detected and potential patterns observed.

For longitudinal analysis, the ability to sequence sequential plasma samples could potentially help overcome similar obstacles surrounding inter and intra-tumour heterogeneity. This would allow the analysis of a multitude of variants which could represent separate tumour clones.

Finally, the collection and processing of plasma samples in-house could ensure better sample quality by limiting sample transport and the risk number of freeze-thaws.

4 Analysis and Clinical Utility of Tumour Heterogeneity and Circulating Tumour DNA in Patients with Locally Advanced Rectal Cancer

4.1 Introduction

In Chapter 3, we were able to validate our analytical methods for ctDNA analysis as part of a pilot study in patients with LARC. In this chapter, we aim to expand upon these findings in a larger cohort of patients with LARC undergoing pCRT. In addition, we intend to investigate the ability to detect clonal changes in the primary tumour tissue of patients with LARC using a targeted NGS panel. As part of this expansion, patient samples were acquired from the ARISTOTLE clinical trial.

4.1.1 Study Objectives

Using plasma and matching pre- and post- treatment tissue in patients from ARISTOTLE, we aim to expand upon our findings in the previous chapter. In this chapter, the objectives are:

- How effective is our targeted NGS panel at detecting clinically significant variants in patients with LARC?
    - Can we identify molecular changes within tumour tissue using our targeted sequencing panel in patients with LARC?
    - What molecular changes occur in relation to CRT in patients with LARC?
- Can ctDNA be reliably detected in sequential plasma samples before, during or after neoadjuvant therapy in another cohort of patients with LARC using ddPCR?
    - Is ctDNA released into the circulation in relation to neoadjuvant therapy in patients with LARC?
- Can we detect and measure quantitative changes in total cfDNA before, during and after radiotherapy in another cohort of patients with LARC?
    - Is cfDNA released into the circulation in relation to neoadjuvant therapy in patients with LARC?

4.2 Results

We initially received permission from Professor Nick West for the acquisition of tumour tissue and plasma samples for 24 patients from the ARISTOTLE clinical trial for molecular analysis. These 24 patients were chosen because each had all four plasma timepoints, tumour biopsy and surgical resection tissue available for molecular analysis. Of these 24 patients, five achieved a complete pathological response pCR to neoadjuvant therapy and were excluded from molecular analysis due to having limited amounts of tumour tissue available. In total, paired tumour tissue and sequential plasma samples were available for molecular analysis in 19 patients with LARC from the ARISTOTLE clinical trial.

All 19 patients had five weeks of radiotherapy (45Gy/25 doses) alongside concurrent Capecitabine with or without Irinotecan. Plasma samples were collected at weeks 0 (pre-treatment), week 1 (first week of therapy), week 5 (final week of therapy) and week 10 (post-treatment). Patients underwent surgery 10 weeks after the completion of pCRT. All tissue samples were collected and processed at each patient's respective trial site before being transported to the AWMGS for molecular analysis later.

Plasma samples were collected and handled differently depending upon when the patient was consented during the trial, see Chapter 2 for further details.

4.2.1 Next Generation Sequencing of tumour tissue from patients with LARC

Objective: How effective is our targeted NGS panel at detecting clinically significant variants in patients with LARC?

In order to identify tumour-specific variants in 19 patients from the ARISTOTLE clinical trial, we sequenced 38 FFPE tumour samples from each patient, including matching tumour biopsies and surgical resections. Samples were sequenced using the Ion Torrent CHPv2 as described previously in the Materials and Methods Chapter.

Only the tumour biopsy from patient ARI-166 failed NGS analysis, therefore 37/38 (97.4%) tumour tissue samples were successfully sequenced.

Variants were called, filtered and annotated as previously described in the Materials and Methods Chapter. Briefly, variants were filtered according to VAF, coverage and directional bias. Our limit of detection for variants in tumour tissue in this study was set at 5% VAF based on findings previously described in the Materials and Methods. Thereafter, non-coding

variants were removed, and all remaining variants were checked manually in the IGV software to ensure they were genuine and not sequencing artefacts.

Remaining variants were annotated and filtered according to guidelines suggested by the ASCO and CAP (Li et al. 2017a). Briefly, variants were filtered according to; the reported minor allele frequency in healthy control populations, whether the variant was previously detected in patients with rectal cancer, the impact on protein function, and any current clinical association with rectal cancer. All variants detected at known clinically relevant hotspots in *KRAS* (p.G12, p.G13 and p.Q61), *NRAS* (p.G12, p.G13 and p.Q61), *BRAF* (p.V600E) and *PIK3CA* (p.E542, p.E545 and p.H1047) were automatically included due to their relevance in clinical guidelines from the National Institute for Health and Care Excellence (NICE).

Guidelines from ASCO/CAP suggest categorising variants into separate (Tiers I-IV) to define their level of clinical significance (Li et al. 2017a). The categories include variants with strong clinical significance (Tier I), potential clinical significance (Tier II), unknown significance (Tier III) and benign or likely benign (Tier IV).

Applying these criteria, we identified 34 strong or likely clinically significant (Tier I/II) variants in either pre- or post-treatment tissue from 18 patients with LARC (see Table 25).

Table 25: Summary of all 34 Tier I/II variants detected in 19 patients wih LARC

Patient ID	Genomic Location	Gene ID	Protein Change	VAF in Biopsy (%)	VAF in Resection (%)	Variant Tier
ARI-166	Chr12:25398285C>A	*KRAS*	p.G12C	15	16.8	I
ARI-182	Chr12:25398285C>T	*KRAS*	p.G12S	45.1	1.2	I
ARI-202	Chr12:25398284C>T	*KRAS*	p.G12D	28.3	47	I
ARI-202	Chr17:7578388C>G	*TP53*	p.R181P	35.6	67.6	II
ARI-239	Chr12:25398285C>A	*KRAS*	p.G12C	55.1	37.6	I
ARI-239	Chr17:7577094G>A	*TP53*	p.R282W	54.7	33.1	II
ARI-295	Chr17:7578437G>A	*TP53*	p.Q165X	23	31.8	II
ARI-297	Chr1:115256528T>A	*NRAS*	p.Q61H	41.1	20.2	I
ARI-306	Chr17:7577141C>A	*TP53*	p.G266V	18.9	7.6	II
ARI-316	Chr5:112175390C>T	*APC*	p.Q1367*	27.9	23.3	II
ARI-316	Chr12:25398285C>T	*KRAS*	p.G12S	63	69.4	I
ARI-316	Chr17:7578455C>G	*TP53*	p.A159P	32.4	31.3	II
ARI-341	Chr5:112175426G>T	*APC*	p.E1379*	25	17.5	II
ARI-341	Chr12:25398284C>T	*KRAS*	p.G12D	23.4	23.2	I
ARI-341	Chr12:25398284C>A	*KRAS*	p.G12V	12	0	I
ARI-341	Chr17:7578212G>A	*TP53*	p.R213*	30.5	31	II

Patient ID	Genomic Location	Gene ID	Protein Change	VAF in Biopsy (%)	VAF in Resection (%)	Variant Tier
ARI-346	Chr12:25398284C>A	*KRAS*	p.G12V	6.5	7.5	I
ARI-346	Chr3:178936091G>A	*PIK3CA*	p.E545K	6.9	8	II
ARI-366	Chr5:112175216G>T	*APC*	p.E1309*	21.6	31	II
ARI-366	Chr12:25398284C>A	*KRAS*	p.G12V	15.1	24.8	I
ARI-373	Chr5:112175390C>T	*APC*	p.Q1367*	32	16	I
ARI-373	Chr17:7577120C>T	*TP53*	p.R273H	30	23	II
ARI-378	Chr12:25398284C>A	*KRAS*	p.G12V	44	26.5	I
ARI-378	Chr17:7577556C>T	*TP53*	p.C242Y	47.5	29.9	II
ARI-400	Chr12:25380275T>A	*KRAS*	p.Q61H	22.2	28.8	I
ARI-403	Chr5:112175328C>A	*APC*	p.S1346*	43	58	II
ARI-403	Chr12:25398284C>A	*KRAS*	p.G12V	16.8	24.6	I
ARI-408	Chr12:25398285C>A	*KRAS*	p.G12C	21.8	26.7	I
ARI-413	Chr12:25398284C>A	*KRAS*	p.G12V	29.4	20.5	I
ARI-413	Chr17:7578263G>A	*TP53*	p.R196*	23.3	20.8	II
ARI-437	Chr12:25398281C>T	*KRAS*	p.G13D	10	27.2	I
ARI-437	Chr17:7578388C>G	*TP53*	p.R181P	15.3	37	II

*A nonsense mutation occurred at this location, causing the emergence of a stop codon

Variant Tier – Variant tier of pathogenicity based on guidelines from ASCO/CAP (Li et al. 2017a)

4.2.1.1 Genomic landscape of LARC

In either the tumour biopsy or surgical resection, at least one pathogenic variant was detected in 18/19 (94.7%) patients with LARC. Variants in *KRAS, TP53,* and *APC* were detected in 14 (73.7%), ten (52.6%) and five (26.3%) patients respectively. A single variant in *PIK3CA* and *NRAS* were each detected once in two separate patients (5.3%; Table 26). In this study, all variants detected in *KRAS* and *NRAS* were cross-checked using ddPCR in both pre- and post-treatment tissue. Variants in *TP53, PIK3CA* and *APC*, were cross-checked using ddPCR in 6/14 (42.8%), 1/2 (50%) and 1/8 (12.5%) variants respectively. Not all variants were cross-checked in the same manner due to financial limitations. All variants which were cross-checked were found to be positive.

We were able to detect *TP53* variants in ten patients, of which three were nonsense and seven were missense variants. These variants were detected within the DNA binding domain of p53 (Saha et al. 2015).

In *KRAS,* we were able to detect 15 variants in 14 patients, all of which of were missense and located at p.G12, p.G13 (14/15; 93.3%) or p.Q61 (1/15; 6.7%). Two separate *KRAS* variants (p.G12D and p.G12V) were detected in the tumour biopsy of patient ARI-341. However only p.G12D was detected in the matching resected tumour tissue from this patient.

In *APC*, we were able to detect five pathogenic variants in five patients, all of which were nonsense. These variants were in the mutation cluster region (MCR) of *APC* (Albuquerque et al. 2002).

In *PIK3CA*, we were able to detect one variant in one patient, which was missense (p.E545K). We were only able to detect one single missense variant in *NRAS* (p.Q61H) in tumour tissue from one patient.

The mutational frequencies of these genes are summarised and compared to larger rectal cancer studies from cBioPortal and TCGA Network (Network 2012) in Table 26.

Table 26: Frequency of Variants Detected Within Genes of Interest Within our Cohort in Comparison to Other Selected Cohorts

	Frequency of Variant Detection (%)		
Gene	Our Study	cBioportal	TCGA Network
KRAS	73.7	40.5	42.5
TP53	52.6	68.1	52.4
APC	26.3	70.4	75.9
PIK3CA	5.3	19.2	20.3
NRAS	5.3	4.1	9.0

TCGA – The Cancer Genome Atlas Network (2012)

A total of 831 patients with rectal cancer were included in the cBioportal-based analysis; including 619 from Giannakis and colleagues (2016), 74 from Seshagiri and colleagues (2012) and 138 from Brannon and colleagues (2014).

We observed higher than expected frequencies of variants in *KRAS* according to cBioportal and TCGA Network, whereas variants in *APC* and *PIK3CA* occurred less frequently than anticipated. This is likely to be because our targeted NGS panel sequences mutation hotspots rather than entire genes. Therefore, any variants located outside of such hotspots may be missed during sequencing. The frequency of variant detection in *TP53* and *NRAS* were similar to previous reports.

Variants in *KRAS* were more frequent than we observed previously in patients with LARC consented by the Wales Cancer Bank (WCB) in Chapter 3; which was 52.4% respectively. This may have been impacted by the removal of patients who achieved pCR, creating some selection bias.

Overall, we were able to detect pathogenic variants in 18/19 (94.7%) patients with LARC using our targeted NGS panel. This includes pathogenic variants in a wide range of genes, including *KRAS, TP53, APC, PIK3CA* and *NRAS.* These findings suggest that our targeted NGS panel has been able to effectively detect pathogenic variants in patients with LARC.

4.2.1.2 Molecular Changes Occurring as a Result of Therapy

Objective: Can we identify molecular changes within tumour tissue using our targeted sequencing panel in patients with LARC?

4.2.1.2.1 Calculating Tumour Heterogeneity Scores

A tumour heterogeneity score is a representation of the proportion of a tumour which contains a tumour-specific variant (Normanno et al. 2015). This could provide further information regarding individual variants and whether each was located across the entirety of a tumour sample or in smaller tumour-subclones. In this study, we applied heterogeneity scores to represent individual tumour subclones. Therefore, changes identified over time were hypothesised to represent changes in subclonal tumour architecture because of neoadjuvant therapy.

Heterogeneity scores were calculated from VAFs, which were quantified by ddPCR or NGS data (see Section 2.2.14.3.2), and then normalised according to tumour cellularity in patient tissue as described previously (Normanno et al. 2015).

The tumour heterogeneity scores of detected variants were compared between paired tumour biopsies and resections in our patient cohort (Table 27). Tumour heterogeneity scores can be used to predict whether a single variant is present in a sub-clone or a larger proportion of tumour tissue. A relatively high heterogeneity score might suggest that a specific variant was detected in larger amounts of tumour tissue in comparison to a lower heterogeneity score. Heterogeneity scores of ~100 suggest that a variant is harboured across the entire tumour biopsy or resection. Scores >100 may suggest the presence of copy number alteration in said genes (Li et al. 2017b) and scores <100 may suggest that only a fraction for tumour cells harbour said variant. Very low scores, (e.g. <33) might also suggest this variant is present within a small tumour sub-clone (Normanno et al. 2015).

Table 27: Tumour heterogeneity scores of all detected variants detected in tumour biopsies and resections in patients with locally advanced rectal cancer from the Aristotle clinical trial

			Tumour Biopsy			Surgical Resection			
Patient ID	**Gene**	**Variant**	**VAF (%)**	**Tumour (%)**	**HS**	**VAF (%)**	**Tumour (%)**	**HS**	**Difference In HS**
ARI-166	*KRAS*	p.G12C†	15	40	75	16.8	50	67.2	-7.8
ARI-182	*KRAS*	p.G12S†	45.1	60	150.3	1.2	30	8	-142.3
ARI-202	*KRAS*	p.G12D†	28.3	50	113.2	47	60	156.7	43.5
	TP53	p.R181P	35.6		142.4	67.6		225.3	82.9
ARI-239	*KRAS*	p.G12C†	55.1	40	275.5	37.6	40	188	-87.5
	TP53	p.R282W	54.7		273.5	33.1		165.5	-108
ARI-295	*TP53*	p.Q165X†	23	50	92	31.8	40	159	67
ARI-297	*NRAS*	p.Q61H†	41.1	70	117.4	20.2	50	80.8	-36.6
ARI-306	*TP53*	p.G266V†	18.9	20	189	7.6	30	50.7	-138.3
ARI-316	*APC*	p.Q1367X	27.9	40	139.5	23.3	60	77.7	-61.8
	KRAS	p.G12S†	63		315	69.4		231.3	-83.7
	TP53	p.A159P	32.4		162	31.3		104.3	-57.7
ARI-341	*APC*	p.E1379X†	25	60	83.3	17.5	50	70	-13.3
	TP53	p.R213X	30.5		101.7	31		124	22.3
	KRAS	p.G12D†	23.4		78	23.2		92.8	14.8
	KRAS	p.G12V	12		40	0		0	-40
ARI-346	*PIK3CA*	p.E545K†	6.9	20	69	8	50	32	-37
	KRAS	p.G12V†	6.5		65	7.5		30	-35
ARI-366	*APC*	p.E1309X	21.6	50	86.4	31	60	103.3	16.9
	KRAS	p.G12V†	15.1		60.4	24.8		82.7	22.3
ARI-373	*APC*	p.Q1367†	32	30	213.3	16	40	80	-133.3
	FBXW7	p.R465C	47.2		314.7	22		110	-204.7
	TP53	p.C176Y	12.8		85.3	8		40	-45.3
	TP53	p.R273H†	30		200	23		115	-85
ARI-378	*KRAS*	p.G12V†	44	50	176	26.5	40	132.5	-43.5
	TP53	p.C242Y†	47.5		190	29.9		149.5	-40.5
ARI-400	*KRAS*	p.Q61H†	22.2	30	148	28.8	50	115.2	-32.8
ARI-403	*APC*	p.S1346X	43	30	286.7	58	50	232	-54.7
	KRAS	p.G12C†	16.8		112	24.6		98.4	-13.6
ARI-408	*KRAS*	p.G12C†	21.8	30	145.3	26.7	40	133.5	-11.8
ARI-413	*KRAS*	p.G12V†	29.4	30	196	20.5	40	102.5	-93.5
	TP53	p.R196X	23.3		155.3	20.8		104	-51.3
ARI-437	*KRAS*	p.G13D†	10	30	66.7	27.2	40	136	69.3
	TP53	p.R181P	15.3		102	37		185	+83.0

VAF-Variant Allelic Frequency

HS – Heterogeneity Score

†Variants were cross-checked and VAFs were quantified using ddPCR

Of our detected variants, 33 (97.1%) were shared between matching patient biopsies and resections, whereas only one (2.9%) variant was detected solely in a biopsy and no variants were detected in resections alone (Table 28). Thus, there was no evidence of emerging novel tumour subclones in these patients that we could detect. This suggests that there were no significant changes in the detected variants between pre- and post-treatment tumour tissue.

Table 28: Variant Distribution Between Pre- and Post-Treatment Tumour Tissue

Variant Location	Number of Variants	%
Pre-Treatment Biopsy Only	1	2.9
Surgical Resection Only	0	0
Shared	33	97.1
Total	34	100

In agreement with our findings, Lee and colleagues (2017) reported similar mutations rates and composition between tumour tissue before and after tyrosine kinase inhibitor (TKI) therapy in patients with lung cancer. This differs from previous reports that 26-65%, 58% and 67% of variants overlapped between primary tumour tissue and metastatic lesions before and after treatment in patients with CRC (Harada et al. 2019), ovarian cancer (Lambrechts et al. 2015) and glioblastoma (Kim et al. 2015) respectively. This of course may not be directly comparable to our findings as it focusses on changes occurring in metastatic lesions rather than within the primary tumour. Furthermore, these studies, including our own, all investigated patients with all different tumour types undergoing various modalities of treatment.

Some changes were observed which were of potential interest in the context of clonal changes in this patient cohort. In this analysis, we were able to detect evidence of clonal changes occurring within patient tumours. Evidence of said clonal changes may be seen in patient ARI-341 who demonstrated a complete loss of one variant (*KRAS* p.G12V) from pre- to post-treatment tissue. This may represent the complete elimination of a tumour subclone which

harboured this variant. However, the remaining three variants detected within this patient's tumour tissue showed relatively little change, suggesting limited clonal changes may have occurred as a result of neoadjuvant therapy in these tumour subclone(s) that these variants represent. Whether these variants represent the same subclone or individual clones is also unknown.

We also observed what appeared to be a significant decrease in tumour heterogeneity score from pre- to post-treatment tissue in the single variant (*KRAS* p.G12S) detected in patient ARI-182. Unfortunately, for this patient there were no other pathogenic variants detected to monitor clonal changes. Therefore, we cannot determine whether another individual or group of subclones developed in place of the loss of the subclone represented by this variant. Similarly, ARI-306 displayed a large reduction in the only detectable variant (*TP53* p.G266V) between the tumour biopsy and surgical resection, with no other variant appearing in its place.

We proceeded to identify changes in mutated genes and associated heterogeneity scores between the tumour biopsies and post-treatment surgical resections in patients with LARC using our targeted NGS panel (Table 27). For variants which were cross-checked with ddPCR, the VAF quantified by ddPCR was used when calculating respective heterogeneity scores.

From the variants detected in *KRAS,* we identified a large range of heterogeneity scores ranging from small tumour subclones (44.0) to potential copy number changes (275.5) (median: 114, mean: 126) in patient biopsies. The same was also observed in tumour resections 6.7-235.6 (median: 90, mean: 97). These differences were not statistically significant (p=0.65). In patients with metastatic CRC, Normanno and colleagues (2015) reported that *KRAS* and *NRAS* variants had average heterogeneity scores of 87.1 and 102.8 respectively, whilst Li and colleagues (2017b) found similarly high median heterogeneity scores of 120 and 125 respectively in pre-treatment tumour tissue . Dienstmann and colleagues (2017a) also reported average heterogeneity scores of 112 and 98 for *KRAS* and *NRAS* respectively in FFPE tumour tissue . Overall, these studies have reported that variants in *KRAS* and *NRAS* are harboured in a large proportion of tumour cells with similar averages to those observed in our study.

We saw no significant differences between the average heterogeneity scores of variants between the biopsy and resection in *TP53* (152.8 and 114 respectively) and *APC* (114.6 and 117.5 respectively). Similarly, Dienstmann and colleagues (2017a) reported average heterogeneity scores of 132 and 100 of variants in *TP53* and *APC* respectively in FFPE tumour

tissue. Our findings contradict previous reports from Sakai and colleagues (2014) who reported that variants in *TP53* saw a significant increase in VAF in patient tumour tissue as a result of chemoradiation therapy in patients with LARC. However, this study analysed only the VAF of these variants and not the heterogeneity scores, which complicates a direct comparison. Furthermore, VAFs and heterogeneity scores of tumour suppressor genes, such as *APC* and *TP53* may be more complex to quantify due to the potential presence of loss of heterozygosity (LoH) in these genes.

These results may be influenced by several factors, such as amplification bias from the NGS protocol, the subjectivity of estimating neoplastic cell content and the inaccuracy of macro-dissection which can all impact VAF and heterogeneity score calculations.

Using our targeted NGS panel, we did not see significant molecular changes between tumour biopsies and surgical resections in patients with LARC. We did identify significant molecular changes occurring between pre- and post-treatment tissue in a small number of patients. We hypothesised that these differences may represent changes in subclonal tumour architecture in response to neoadjuvant therapy in this patient cohort.

4.2.2 Sequential circulating tumour DNA analysis using ddPCR

Cell free DNA was extracted from 1mL of plasma from patients in this cohort using the QiaAmp Circulating Nucleic Acids Kit. CtDNA was then analysed by ddPCR as described in the Materials and Methods Chapter. We previously demonstrated that ctDNA analysis using ddPCR had a sensitivity of 1%, but this was highly dependent on the input DNA concentration. Therefore, to provide consistency across patient samples, a minimum of five mutant droplets were required for the confident detection of ctDNA in this study.

We were unable to analyse >1 tumour specific variant in ctDNA in most patients due to limited sample availability. Where possible, variants detected in *KRAS/NRAS* were analysed, as this would also provide consistency and facilitate a potential comparison to previous ctDNA analyses in WCB patients from Chapter 3.

4.2.2.1 Longitudinal Analysis of Circulating Tumour DNA

Objective: Can ctDNA be reliably detected in sequential plasma samples before, during or after neoadjuvant therapy in another cohort of patients with LARC using ddPCR?

All plasma samples were successfully analysed using ddPCR except for two time-points (ARI-341 Week 0 and ARI-373 Week 10). CtDNA was successfully detected in three plasma samples from 3/18 (16.7%) patients with LARC (See Table 29). This sensitivity (16.7%) for the detection

of ctDNA is below our previous findings in patients with LARC consented by the WCB (Chapter 3), in which 25% of patients had ctDNA detected at any time.

In this study, detected ctDNA VAFs varied between 1.0 – 2.7%. Unlike our previous cohort in Chapter 3, no patients had ctDNA detected after the completion of therapy (week 10). We were also able to detect ctDNA in two patients whilst on-treatment (Weeks 1 and 5), whereas none were seen in our previous cohort. This may be seen here because we increased our total sample size and all 18 patients analysed had an incomplete response to therapy, which may have enhanced the likelihood that ctDNA would be observed during therapy.

Here we were able to monitor ctDNA in patients with *KRAS* wild-type tumours, which was not possible previously. This sensitivity for detecting ctDNA is still very low in comparison to previous reports of 76.7% at any time in patients with LARC (Tie et al. 2018). This difference may be a result of differences in patient tumour stage, the lower volumes of plasma available for our study, differences in technology used, the cut-off applied for ctDNA detection or the quality of samples analysed.

Table 29: Variant allele frequency of ctDNA detected in patients from the ARISTOTLE clinical trial

		Variant Allelic Frequency (%)				
Patient ID	**Variant**	**Week 0**	**Week 1**	**Week 5**	**Week 10**	**ctDNA Group**
ARI-373	*TP53* p.R273H	1.0	0	0	F	A
ARI-316	*KRAS* p.G12S	0	1.0	0	0	B
ARI-408	*KRAS* p.G12C	0	0	2.7	0	B
ARI-166	*KRAS* p.G12C	0	0	0	0	C
ARI-182	*KRAS* p.G12S	0	0	0	0	C
ARI-202	*KRAS* p.G12D	0	0	0	0	C
ARI-239	*KRAS* p.G12C	0	0	0	0	C
ARI-295	*TP53* p.Q165X	0	0	0	0	C
ARI-297	*NRAS* p.Q61H	0	0	0	0	C
ARI-306	*TP53* p.G266V	0	0	0	0	C
ARI-341	*KRAS* p.G12D	F	0	0	0	C
ARI-346	*KRAS* p.G12V	0	0	0	0	C
ARI-366	*KRAS* p.G12V	0	0	0	0	C
ARI-378	*KRAS* p.G12V	0	0	0	0	C
ARI-400	*KRAS* p.Q61H	0	0	0	0	C
ARI-403	*KRAS* p.G12C	0	0	0	0	C
ARI-413	*KRAS* p.G12V	0	0	0	0	C
ARI-437	*KRAS* p.G13D	0	0	0	0	C

F – Sample failed ddPCR analysis

Overall, three different patterns of change in ctDNA were observed in sequential plasma samples in our patient cohort. We separated these into Groups A, B and C based on each associated change based on previous findings from Chapter 3.

Within Group A, the ctDNA detected in pre-treatment plasma declined to undetectable levels by the end of therapy. Patients in Group B began with undetectable pre-treatment ctDNA which became detectable during therapy and was then undetectable during follow-up. Finally, patients within Group C had no detectable ctDNA at any time.

The decrease in levels of detectable ctDNA in the circulation of patients with LARC (as seen in Group A) has been previously reported as a result of CRT (Carpinetti et al. 2015). This study hypothesised that this decrease during therapy may be indicative of a good response to therapy.

At this time, we have not identified any studies which investigate the effects of radiotherapy on ctDNA during treatment in patients with LARC (as seen in Group B). In patients with nasopharyngeal carcinoma, one study reported an increase in ctDNA during the first week of radiotherapy (Lo et al. 2000). However, this increase in ctDNA was never compared to clinical outcomes in these patients. This group hypothesised that this change was due to cancer cell death after therapy. This study also reported that a single patient presented a second release of ctDNA during day 14 of therapy, possibly suggesting further cancer cell death at this time. We were unable to identify any studies reporting the release of ctDNA during the last week of radiotherapy in any disease. In this case, such a peak may be representative of cancer cell re-population instead of cell death (Lo et al. 2000). Further studies are required to investigate the clinical implications of these changes.

Most of this patient cohort (15/18) had no ctDNA detected at any time (Group C). This may be due to several technical reasons. For example, the low volume of plasma available would result in low overall total cfDNA concentrations which would detrimentally influence the sensitivity of our ddPCR assay. Furthermore, there may also be issues with sample handling throughout the process. This may also be a consequence of the inherent tumour biology of non-metastatic patients.

Overall, our findings suggest that our methods could not reliably detect ctDNA in this extended cohort of patients with LARC due to our low number of positive events. Where ctDNA was positively detected, we have high confidence in the reliability of these findings.

4.2.3 Sequential Cell Free DNA Changes

4.2.3.1 Variations in Sequential Cell Free DNA Concentrations

Objective: Can we detect and measure quantitative changes in total cfDNA before, during and after radiotherapy in another cohort of patients with LARC?

As discussed in Chapter 3, studies have previously reported that sequential changes in total cfDNA concentrations were associated with patient response to neoadjuvant therapy in patients with LARC (Agostini et al. 2011). Although we cannot compare these data to any patient clinical information, we proceeded to investigate whether we could identify sequential cfDNA changes in these patients; which we intend to compare to our previous findings from Chapter 3.

A total of 74 cfDNA samples were available for molecular analysis in patients with LARC collected from the ARISTOTLE clinical trial. CfDNA was extracted from 1mL of plasma using the Qiagen QIAamp Circulating Nucleic Acids Kit (55114) and quantified using the Invitrogen Qubit High Sensitivity Assay (Q32854) as described in the Materials and Methods Chapter.

For all patients, except ARI-239, there were no obvious patterns of change observed in sequential total cfDNA concentration at this time (Figure 13). For all other patients, total cfDNA varied between 0.1-0.7ng/µL, whereas patient ARI-239 displayed a significantly higher peak of detectable cfDNA at week 5 (2.0 ng/µL). This resembles the pattern of change for cfDNA observed in all samples in Chapter 3, albeit only in this one patient. This may result from cellular changes occurring in the patient's tumour, surrounding non-tumour tissue or systemically as a result of neoadjuvant therapy. Alternatively, this may also be due to inappropriate pre-analytical handling. As this sample was known to be collected in EDTA, it is possible that white blood cell lysis may have occurred before the sample was processed into plasma. Therefore, the leakage of white blood cell DNA into the plasma may have been misconstrued as cfDNA, causing this significantly high peak over all other samples.

Apart from patient ARI-239, our results contrast with previous findings in Chapter 3. This difference may be a consequence of the exclusion of patients who achieved pCR. Previously , Agostini and colleagues (2011) reported that patients with rectal cancer who responded well to CRT saw a significant decrease in cfDNA levels after the initiation of treatment. In patients with hepatocellular carcinoma, patients who responded poorly to therapy displayed a significant increase in total cfDNA concentrations from pre- to post-treatment (Park et al. 2018). We did not see either trend here.

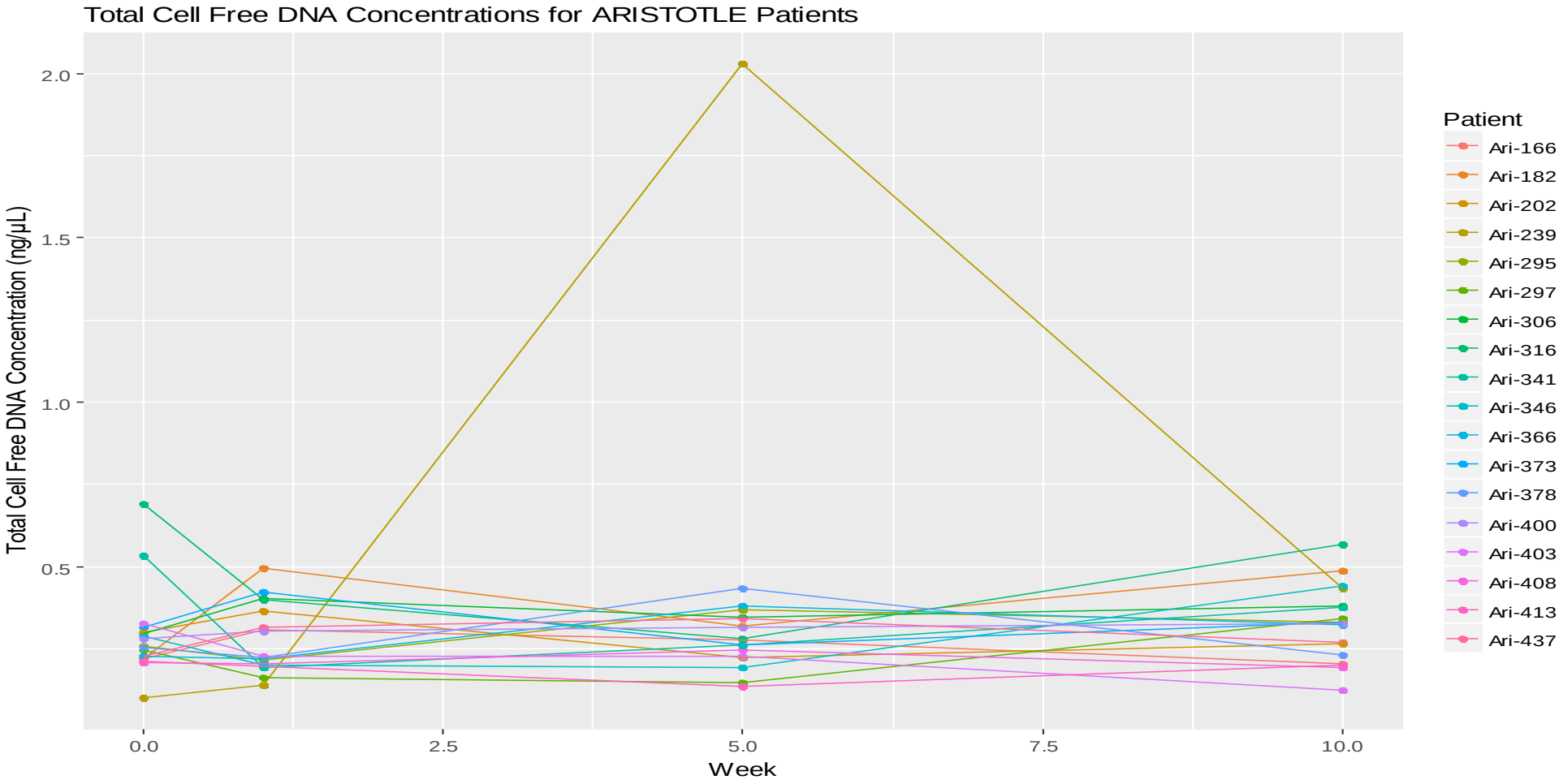

Figure 13: Total Cell Free DNA Concentrations for all ARISTOTLE Patients at Weeks 0, 1, 5 and 10

We proceeded to compare total cfDNA concentrations between whole blood samples which were collected in EDTA to those which were collected in Streck Tubes using an un-paired two-sample T-test (see Appendices Section 9.5). In the first instance, we analysed differences in cfDNA concentration across all timepoints (Figure 14).

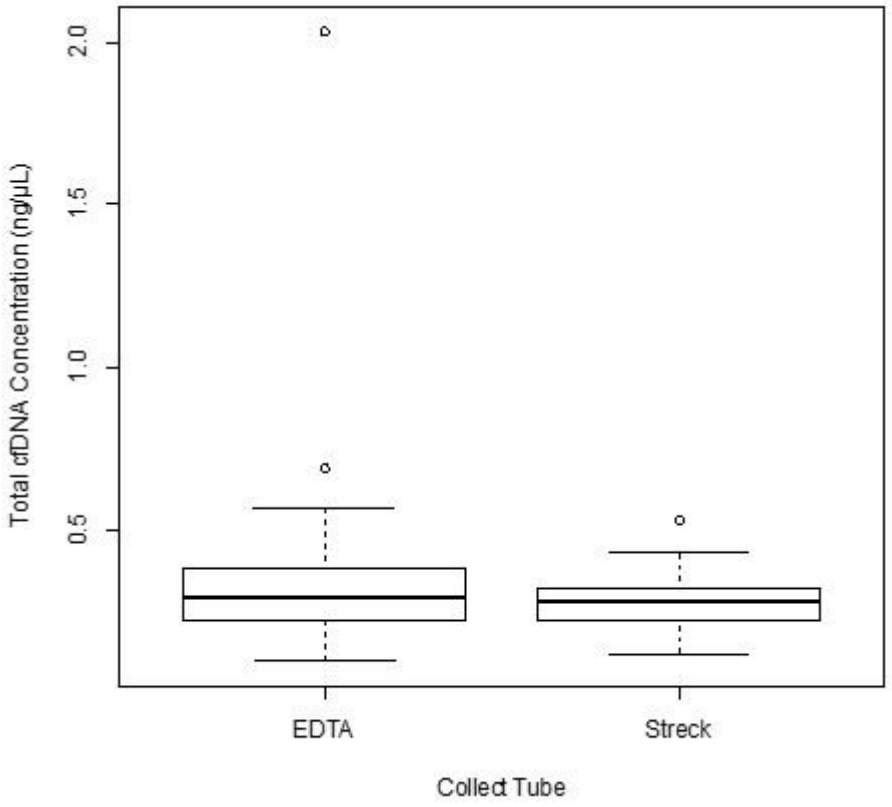

Figure 14: Total cfDNA concentrations between the two collection tubes across all plasma collection timepoints

Across all plasma samples we observed no significant differences in total cfDNA concentrations between the two collection tubes (p=0.12). We proceeded to investigate any differences between the two collection tubes at each individual timepoint (Figure 15)

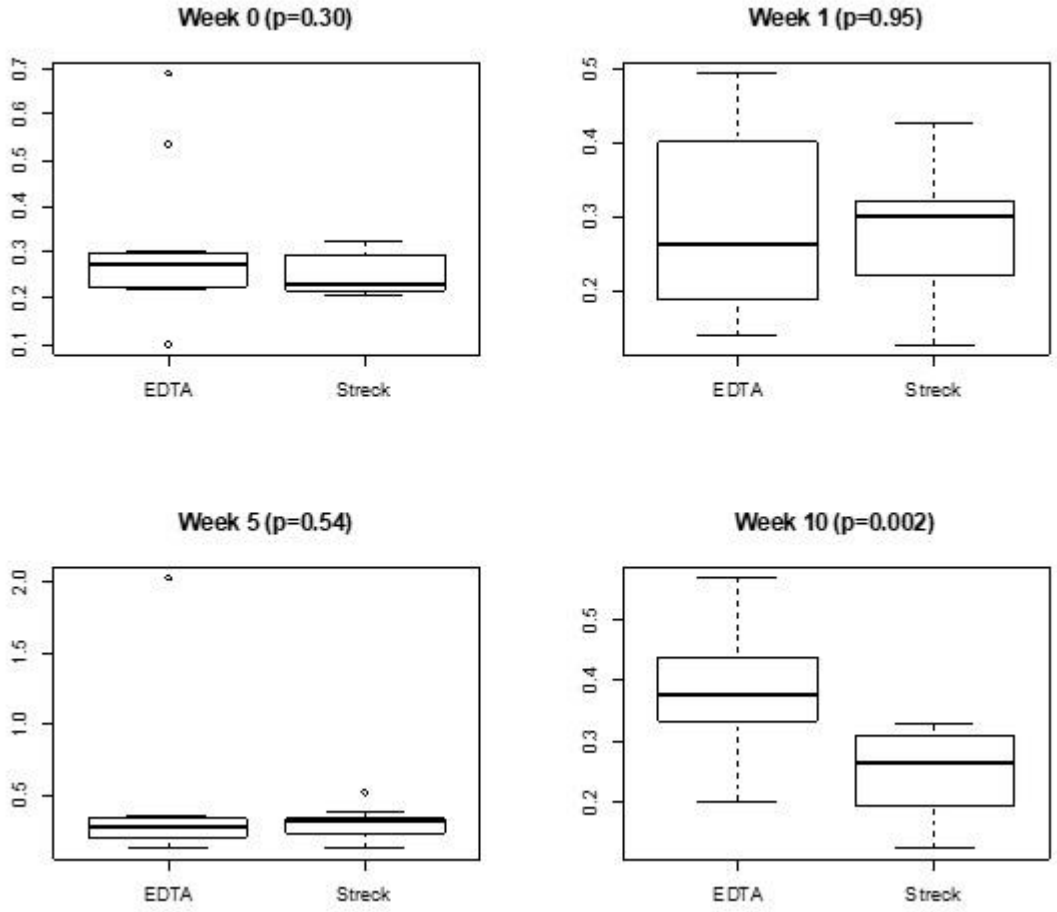

Figure 15: Total cfDNA concentrations between the two collection tubes at each plasma collection timepoint

We observed no significant differences between collections tubes for plasma samples collected at weeks 0, 1 and 5. However, at week 10, we observed that plasma samples collected in EDTA tubes had a significantly higher total cfDNA concentration than those collected in Streck tubes (0.38 vs 0.25 ng/µL; $p<0.01$).

Overall, using the Invitrogen Qubit we did not detect any significant sequential changes in total cfDNA concentrations in our patient cohort. We did detect a significant difference in total cfDNA at week 10 of collection between plasma samples collected in EDTA and Streck tubes, although this may be due to chance rather than a true significant finding.

4.3 Discussion

4.3.1 Conclusions

Objective: How effective is our targeted NGS panel at detecting clinically significant variants in patients with LARC?

Using our targeted NGS panel, we were able to detect pathogenic variants in tumour biopsies or resections from all but one patient with LARC from this cohort. When examining the frequency at which A*PC, PIK3CA* and *KRAS* were mutated, we observed differences between our findings and expected values based on large-scale studies.

We detected a number of nonsense variants in the MCR of *APC* (Albuquerque et al. 2002; Christie et al. 2013). Nonsense and frame-shift variants are frequently detected in the MCR of *APC* as an early genomic 'hit' during colorectal carcinogenesis. These variants have been demonstrated to inhibit the regulatory effect which APC imposes upon β-catenin (Fearon 2011). This provides some confidence in the positive calling of these detected variants.

We may have detected variants in *APC* less frequently than expected for several reasons. Firstly, our targeted panel only covered hotspots in the *APC* gene which revolved around the MCR. However, this would only include <10% of the entire gene (Table 11). Thus, any tumour-specific variants occurring outside of these sequenced regions would have been missed. We may also have missed the presence of large deletions or small indels in our patients due to limitations of the sequencing chemistry which were discussed in the Materials and Methods chapter. Finally, all variants had to be deemed pathogenic to pass through our filtration algorithm. This would not have been the case for cBioPortal or TCGA, where they would simply be confidently detected, and pathogenicity isn't a requirement for variant calling.

In *KRAS,* the majority of our variants were located at hotspots (p.G12, p.G13 and p.Q61) in accordance with previous expectations (Fearon 2011). We saw that *KRAS* was mutated much more frequently than anticipated in our study. This may be a result of patient selection bias. As patients who achieved pCR were excluded from these analyses, our patient mutational profiles may not cover the full array of patient responses to neoadjuvant therapy. This may then skew our findings and may need to be considered during statistical analyses. Additionally, this result may also be impacted by our small sample size in comparison to these previous studies.

With regards to the difference observed in mutation frequency for *PIK3CA,* this is likely to be a consequence of our small sample size, the hotspot analysis of our NGS panel and the low frequency at which this gene is mutated in this population.

The frequency at which we detected variants in *TP53* were found at the approximate expected frequency according to cBioPortal and TCGA (2012). In *TP53* we detected numerous missense and nonsense variants in the DNA binding domain (amino acid 98-322) (Saha et al. 2015). Many of these variants correlated with 'hotspots' at p.R175, p.Y220, p.G245, p.R248, p.R249, p.R273 and p.R282. Disruptions within the DNA binding domain have been predicted to inhibit the ability of p53 to initiate apoptosis, thus providing more support for these variants being genuine and pathogenic (Saha et al. 2015).

There are numerous factors which may have impacted the frequency at which we detected clinically significant variants in this study. Our targeted NGS panel was mostly designed to identify variants in mutational hotspots, therefore not the entire gene is sequenced in most our analysed genes. Consequently, variants outside of hotspot regions may have been missed, resulting in our decreased rate of variant detection for some genes, such as *APC*.

As with most semiconductor sequencers, the CHPv2 has limitations when detecting large copy number variations or small indels. As these types of variants are relatively common in tumour suppressor genes, such as *APC* or *TP53*, this may contribute to a decreased frequency of variant detection.

Furthermore, our variant filtration algorithm may have been more stringent than those applied in the larger studies, as we attempted to replicate clinical analysis in our study. However, clinical laboratories will largely concern themselves with analysing targetable variants, therefore potentially missing variants which are clinically relevant, if not yet targetable. Nonetheless, we are confident that all highlighted variants are clinically significant and not random passenger variants.

Another issue which may have affected our ability to detect variants is sequencing coverage. Although sequencing covering was generally good, all detected variants required ≥250 reads, thus any regions with less total reads or sequencing dropout may have led to true variants being missed.

Overall, our findings suggest that the CHPv2 was effective at detecting pathogenic variants across a range of genes in patients with LARC. This strengthens our initial decision-making

process when choosing this targeted NGS panel, as described in the Materials and Methods Chapter.

Objective: Can we identify molecular changes within tumour tissue using our targeted sequencing panel in patients with LARC?

Using our targeted sequencing panel, we were able to observe changes in tumour heterogeneity scores, which may represent clonal changes in tumour tissue. We could not identify the occurrence of novel variants in tumour resections which might suggest the growth of a tumour sub-clone as a result of therapy; as has been seen in patients being treated by targeted therapies (Gollins et al. 2017).

We observed a significant reduction in tumour heterogeneity score (e.g. ARI-182 or ARI-306) or complete loss of a variant (e.g. ARI-341) from pre- to post-treatment tissue in a small number of patients (Table 27). This may be indicative of clonal shrinkage or loss due to therapy. However, we are unable to verify this association without obtaining patient outcomes.

Previous reports have demonstrated a decrease in the number of detectable variants after stereotactic radiotherapy using an in-house 50-gene panel in a patient with lung cancer (Nakagomi et al. 2017). Findlay and colleagues (2016) used whole exome sequencing and observed that patients with oesophageal cancer who responded well to chemotherapy showed more genetic changes after therapy than patients who responded poorly to therapy. The greater number of changes observed by these studies may be due to the type of sequencing analysis performed in comparison to our study. This may also be a result of the different cancer types and treatment modalities (including time-period of treatment) between the various studies.

Conversely, increases in heterogeneity scores may be indicative of clonal growth within a patient's tumour. In the tumour tissue of patients who displayed a mixture of heterogeneity score changes, we may be able to observe distinct shifts in clonal dominance in response to treatment. Finally, where little or no change is observed, this may suggest genetic stability in the patient tumour.

In selected genes of interest (including *APC, KRAS* and *TP53*) one study reported that variants in these genes remained relatively constant in metastatic lesions after treatment with FOLFOX in patients with CRC (Harada et al. 2019). In contrast, less frequently mutated genes (including *EP300, MED12* and RUNX1) were seen to develop variants after therapy, the clinical significance of which is less understood in the context of CRC (Harada et al. 2019). This finding

may also reflect that variants in *APC, KRAS* and *TP53* are early events in tumour development, and the later events may be more prone to loss during treatment than early stemming variants. Patients with oesophageal cancer also displayed highly variable levels of change in clonal composition after chemotherapy (Findlay et al. 2016; Noorani et al. 2017). Although these studies analysed patients with different diseases, stages and treatment to our patient cohort, this may still contribute to the limited variant and clonal changes we were able to observe here.

In our analysis, surgical resections were collected shortly after therapy, whilst other studies analysed tumour tissue upon disease recurrence. This would provide a longer timeframe for cellular repopulation and clonal changes to occur as a result of therapy, which may contribute to the limited amount of changes we observed.

Additionally, this analysis is also limited by the sensitivity and limit of detection for this targeted panel. Although we have previously demonstrated that variants can be detected at ~1% VAF, we chose the limit of 5% to ensure selected variants were pathogenic and not artefactual. Consequently, genuine pathogenic variants at this VAF may have been missed, as displayed in patient ARI-182, where a variant (*KRAS* p.G12S) was detected at 1.2% and cross-checked using ddPCR. This was only investigated due to its presence in matching tumour biopsy and would have otherwise been missed. Without prior knowledge, identifying variants at this VAF could result in the 'miscalling' of variants.

For this study, likely pathogenic (Tier II) or pathogenic variants (Tier I) were chosen based on guidelines from ASCO and CAP (Li et al. 2017a). Relevant pathogenic variants may have been lost as a result of this stringent filtration process, as they may not directly impact treatment in this patient cohort at this time, or their impact on tumour development is not as well understood as other variants.

The calculation of VAFs and tumour heterogeneity may also be influenced of the presence of LoH. LoH is commonly observed in tumour suppressor genes, such as *APC* and *TP53.* Although LoH is common in these genes in the context of CRC, we have not incorporated any potential impact that LoH may have on VAF quantification. Alternatively, the calculation of VAF may not be entirely accurate due to the potential of PCR amplification bias impacting VAF quantification. As the NGS procedure include multiple rounds of amplification, this is a potential issue. Furthermore, as we only sequenced a small region of tumour tissue, our

calculated VAFs are not necessarily representative of the entire tumour, and we are thus limited by intra-tumour heterogeneity (Burrell and Swanton 2014).

When calculating heterogeneity scores, VAFs were normalised according to neoplastic tumour tissue content, as quantified by Professor Richard Adams from a haematoxylin and eosin stained tissue slide. Since tumour cellularity is a subjective approximation (Normanno et al. 2015) and, therefore, these values may be variable, and our heterogeneity scores may also be somewhat inaccurate.

Due to the subjective nature of macrodissections, the tumour cellularity of macrodissected tumour tissue may vary to that anticipated, by the accidental inclusion of surrounding non-tumour tissue. This may impact heterogeneity scores and any downstream analyses.

Due to the numerous limitations of heterogeneity scores, they may be more appropriately applied as rough approximations rather than definitive values when comparing to tumour subclones. Although we have attempted to directly analyse clonal architecture by using these scores, they were not intended for this purpose. Finally, analysing changes in heterogeneity scores can also include a large amount of subjectivity when deciding which changes were large and which were small.

Using our targeted NGS panel, we were able to detect molecular changes occurring between pre- and post-treatment tissue in patients with LARC. These changes may be a result of the administration of neoadjuvant therapy in these patients. We suspect that said molecular changes may represent changes in clonal architecture may be indicative of tumour response. However, there are distinct limitations with our methodology which must be considered.

Objective: Can ctDNA be reliably detected in sequential plasma samples before, during or after radiotherapy in another cohort of patients with LARC using ddPCR?

In this study, we were able to detect ctDNA in 3/18 (16.7%) patients with LARC, but only in samples collected before or during the administration of neoadjuvant therapy. This sensitivity for ctDNA detection is identical to our previous findings in Chapter 3 (16.7%) but below that of other studies in patients with LARC (Bettegowda et al. 2014; Sclafani et al. 2018; Tie et al. 2018).

Previously, we saw no patients had detectable ctDNA levels during treatment, but some had ctDNA detected after therapy. In contrast here, we observed ctDNA during neoadjuvant treatment but none after the completion of treatment.

We anticipated that the frequency of patients with detectable ctDNA would increase, due to a larger number of variants being available for ctDNA analysis, whereas previously we were limited to variants in *KRAS/NRAS.* This could then have allowed for the inclusion of patients with alternative molecular characteristics, which may have impacted on ctDNA detection rates.

The differences observed in detected ctDNA between the two cohorts may have been caused by several factors. Although both sets of patients had LARC, this cohort was treated with chemotherapy alongside radiotherapy, whereas the previous cohort were largely treated with radiotherapy only. This may impact both the response of patient tumours to treatment and sequential ctDNA changes. Furthermore, as patient samples were collected as part of a clinical trial, the timing and processing of sample collection may be more standardised in comparison to samples collected as part of routine clinics. This may affect sample quality. Finally, the timing of plasma collection after the completion of therapy also differs, with longer timepoints being available in our previous cohort. This allows a greater time frame for molecular changes or tumour recurrence to occur and ctDNA to be detected.

There may be other factors which contributed to the low frequency of ctDNA detected in this study in comparison to findings in the literature. The low volume of plasma (1mL) analysed in this study might have limited the concentrations of ctDNA used for ddPCR analysis, limiting the sensitivity of ddPCR, as has been demonstrated in the Materials and Methods Chapter. Although ddPCR has been reported to be a sensitive assay, the sensitivity of the method is entirely limited as it is based on input DNA concentrations. Furthermore, the locally advanced non-metastatic nature of our patients may not be entirely appropriate for ctDNA analysis due to limited amounts of angiogenesis and consequential DNA shedding into the circulation. Another factor that should be considered is the selection bias associated with these patients. As patients who achieved pCR were excluded from analysis, this may have led to the loss of patients who experienced high amounts of tumour cell death and resulting DNA shedding during neoadjuvant therapy.

There may also be issues with sample collection or the timepoints chosen for analysis, which cannot be commented as they were performed and chosen by the Principle Investigators of the trial but should still be considered. There may have also been issues with how samples were handled during cfDNA extraction and ctDNA analysis which may have limited the efficacy of each phase of analysis. Furthermore, analytical sensitivity may have benefited from the analysis of each sample being performed in triplicate rather than duplicate. This may have

facilitated the confident identification of ctDNA in borderline cases and provided more data for analysis.

An increase in the plasma availability for this patient cohort may help overcome these limitations for future analyses, as would more information or direct involvement in pre-analytical sample handling. As suggested previously in Chapter 3, the processing of samples on-site or delivery of samples over days may contribute to our sensitivity issues and may be addressed by handling on a single site. Perhaps an additional amplification step, such as co-amplification at lower denaturation temperature (COLD) PCR, may help improve detection rates, although this may introduce issues if attempting to quantify ctDNA.

As far as we are aware, this is the first time that ctDNA was demonstrably detected whilst patients were being administered chemoradiotherapy, as studies primarily investigate pre- and post-treatment levels. Therefore, it would be of interest to determine how these patients responded to therapy.

Overall, these findings suggest that these methods are currently unable to reliably detect sequential ctDNA in patients with LARC undergoing neoadjuvant therapy. We have discussed numerous factors which must be addressed to improve the reliability of ctDNA detection. However, these findings may bring into question the clinical efficacy of ctDNA analysis in patients with LARC.

Objective: Can we detect and measure quantitative changes in total cfDNA before, during and after CRT in another cohort of patients with LARC?

We could not decipher any patterns of change in cfDNA concentrations, nor could we identify any associations with any patient molecular characteristics in this patient cohort. This contrasts findings from Chapter 3 where we saw higher cfDNA concentrations in samples collected during therapy.

As previously described in Chapter 3, total cfDNA concentrations are not tumour-specific biomarkers, and can thus be influenced by external factors, such as inflammation, infection, radiation-induced toxicity and exercise (Vittori et al. 2019). This might contribute to such a difference being seen between our two separate cohorts. However, we do not have this information and, therefore, cannot investigate this possibility.

Any differences may also be a result of patient selection bias. Agostini and colleagues (2011) reported that a decrease from pre- to post-treatment cfDNA concentration was observed in

patients with rectal cancer who responded well to radiotherapy. Patients who displayed pCR were excluded from this study, which may have influenced our findings here. There were no methodological changes made from our previous analyses and therefore we are excluding this as a possible cause.

Our findings may also be a consequence of the limitations of the Qubit when quantifying cfDNA. The qubit measures DNA using a fluorescence-based assay, in contrast to amplification-based methods for DNA quantification and analysis, such as the ddPCR. This difference may result in variation in findings, for example a fluorescence-based assay may quantify DNA of any length, whereas PCR-based quantification may only quantify DNA of a minimum length. Fluorescence assays may also be less prone to being influenced by inhibitors of PCR but may also be impacted more by contaminants, such as RNA. We also did not investigate the limit of detection or the reproducibility for the qubit, and thus the impact this may have on quantification is not fully understood and should be investigated in the future.

It is noteworthy that, in patient ARI-239, we observed a significant increase in cfDNA at week 5 in comparison to all other patients and timepoints. This peak may be representative of patient response to treatment. The presence of this peak may also be indicative of improper pre-analytical handling, as this may have caused white cell lysis and the shedding of cellular DNA into plasma. As this concentration is significantly higher than all others observed, this is likely to be the case, although this peak is similar to our findings in Chapter 3. Otherwise, this may be a marker of systemic inflammation and/or radiation- or chemotherapy-induced toxicity (Zwirner et al. 2018). Alternatively, this may be tumour-derived cfDNA which simply did not contain the *KRAS* variants we specifically investigated. Further investigation by NGS would help provide further insight here.

During our analysis, we identified that plasma samples collected in EDTA tubes had a significantly higher total cfDNA concentration at week 10 compared to plasma collected in Streck tubes. This significant difference was not seen at other timepoints. This may be a consequence of how samples were handled and processed at their respective sites, or how samples were shipped to the AWMGs for molecular analysis. If EDTA samples were not processed within a specific period of time (within 2-4 hours), white blood cells may lyse and shed cellular DNA into the plasma (Pritchard et al. 2012). Furthermore, plasma samples shipped on dry ice may have also warmed and/or undergone freeze thaws whilst in transit, impacting sample quality. This may be detected and mistaken for cfDNA in this patient cohort, possibly explaining this finding.

Samples during this analysis were not paired, thus differences may exist in the natural tumour biology between the two patient cohorts. Overall, this outlines the need for consistency in sample handling during clinical trials, as these factors can cause significant differences during molecular analysis.

This difference observed, though statistically significant, may have been a result of chance as it was only observed at week 10 and not at more timepoints. These samples were unpaired, and the EDTA samples may have had significantly higher cfDNA concentrations for separate biological reasons. This may be impacted by other factors such as small sample size, or how patients responded to therapy, a factor we cannot currently account for without clinical outcome data for these patients. This analysis could be improved in the future by acquiring data from factors (such as outcomes to therapy, time taken to transport samples, time before sample processing in EDTA tubes etc.) which may impact cfDNA concentrations and incorporate them into multivariate analysis.

Altogether, we would suggest the ideal method for sample handling in this case, would be collection in Streck tubes before transporting samples at room temperature to a single site (such as the AWMGS) within three calendar days for processing. This will provide a greater consistency in sample extraction, as it would all occur on the same site according to the same protocol. However, how the time taken for transportation may affect this process, should also be considered as part of future analyses

Overall, we were able to detect and measure sequential total cfDNA concentrations in our patients, but we did not identify any significant quantitative changes occurring throughout therapy. We also identified a significantly higher total cfDNA concentrations in plasma samples collected in EDTA tubes compared to Streck tubes at week 10. There may be several factors, both technical and biological, which may have impacted our findings. We were also unable to compare these results to clinical outcomes as this data was unavailable at present.

4.3.2 Clinical Implications

Objective: What molecular changes occur in relation to CRT in patients with LARC?

Variants in *TP53* and/or downstream targets have recently been investigated alongside or in place of p53 expression. Wan and colleagues (Wan et al. 2018a) reported that *TP53* p.T155I was found to be more frequently detected in patients with LARC who responded poorly to CRT. Similarly, Sakai and colleagues (2014) reported that 8/9 patients with rectal cancer who responded poorly to neoadjuvant therapy had detectable *TP53* variants in post-treatment tissue samples. This study also reported that *TP53* variants VAFs increased from pre- to post-treatment tissue as a result of pCRT across all patients. A similar finding was reported in patients with anal squamous cell carcinoma, where variants in *TP53* were only detected in recurrent disease after patients underwent pCRT; but was undetected in treatment-naïve tumour tissue (Cacheux et al. 2016).

These findings suggest that variants in *TP53* may influence tumour response to neoadjuvant therapy in these patients. Furthermore. therapy may cause evolutionary advantage and resulting expansion of mutant *TP53* in cloned tumour tissue. Such changes would have been hypothetically detectable in ctDNA at an earlier stage than in surgically resected tissue.

Variants in *KRAS* were previously demonstrated to correlate with decreased pCR rates in patients with stage II/III CRC (Chow et al. 2016) and rectal adenocarcinoma (Duldulao et al. 2013) undergoing CRT. Whether this is a result of *KRAS* mutant functionality or whether mutant *KRAS* is acting as more of a prognostic marker in these cases, remains unclear.

Studies have recently suggested that there is predictive value of variants in *PIK3CA* or the PI3K-mTOR pathway. In 201 patients with LARC, PI3K pathway mutations were associated with reduced rates of pCR after chemoradiotherapy (Abdul-Jalil et al. 2014) With regards to variants in *NRAS,* due to the rarity of *NRAS* variants in this subgroup of patients, few studies have investigated the potential clinical implications in the context of chemoradiotherapy.

Findlay and colleagues(2016) have previously reported that changes in heterogeneity after neoadjuvant therapy may be indicative of treatment response in patients with oesophageal cancer. The presence of clonal changes might suggest that tumour drivers may be regressing, allowing smaller or more resistant sub-clones to develop.

Overall, reports suggest that levels of heterogeneity and clonal changes may be associated with response to therapy in patients with cancer. Although we cannot compare our findings to clinical outcomes, the molecular changes we observed may possibly be used as a predictive

tool to detect clonal changes in patients with LARC. However, this research is still in its relative infancy, and will require further knowledge on factors or variants predisposing to sensitivity or resistance to neoadjuvant therapy in patients with LARC.

Objective: Is ctDNA released into the circulation in relation to neoadjuvant therapy in patients with LARC?

The potential clinical implications of sequential changes in ctDNA were discussed extensively in Chapter 3. Overall, we were unable to detect any post-treatment ctDNA in any patients using our current methods which may be indicative of patient outcomes. Previously, studies have reported that detectable post-treatment or post-surgical ctDNA was associated with poor response to therapy and clinical outcomes (Tie et al. 2018). As we detected no post-treatment levels at this time, this would suggest these patients may experience good responses to therapy, if not pCR. Otherwise, not detecting post-treatment ctDNA may be due to the limitations of the sensitivity of our assay.

We were able to detect the presence of pre-treatment ctDNA in one patient, which may be indicative of advanced stage or the presence of metastatic/pre-metastatic lesions (Bettegowda et al. 2014). Such information could provide valuable insight and facilitate treatment decision-making for some of these patients. Furthermore, in terms of outcome prediction, pre-treatment samples have been demonstrated to be less useful than post-treatment ctDNA detection (Tie et al. 2018). This may bring into question how useful pre-treatment ctDNA levels are when predicting patient outcomes. However, we could not examine the significance of this as part of this study.

The detection of ctDNA being released during neoadjuvant therapy (weeks 1 and 5) appears to be a novel finding in patients with LARC. These observations match our previous hypothesis that ctDNA may be shed early during radiotherapy, potentially indicating tumour cell death. Similarly, the detection of ctDNA during week 5 of neoadjuvant therapy may be indicative of a delayed response to treatment in these patients, if not tumour-cell repopulation (Lo et al. 2000). However, without clinical outcome data, it is impossible to decipher the clinical implications for ctDNA at present.

Objective: Is cfDNA released into the circulation in relation to neoadjuvant therapy in patients with LARC?

As previously discussed in Chapter 3, sequential changes in cfDNA concentrations have been demonstrated to be associated with patient response to therapy. One study reported that

patients with LARC who demonstrated a decrease in cfDNA concentrations from pre- to post-treatment plasma samples had an improved response to treatment (Zitt et al. 2008). Similarly, high pre-treatment cfDNA levels were also associated with an increased risk of recurrence (Boysen et al. 2017) and poor DFS (Schou et al. 2018) in patients with LARC.

As no distinct patterns were visible, we cannot associate any changes with clinical significance. The general stability of total cfDNA concentrations may be indicative of stable disease in our patient cohort. These findings may be a result of technical factors which had been discussed previously.

Overall, we were unable to discern any patterns of cfDNA secretion in response to neoadjuvant therapy which may be indicative of patient response. Therefore, at present, we are unable to determine the power of sequential total cfDNA concentrations in predicting patient response to therapy.

4.3.3 Future Studies

At present, it is difficult to gauge the implications of our findings in the context of clinical implications, especially without any clinical data or information regarding treatment arms for these patients. Therefore, the acquisition of such information, including both short and long-term outcomes, could allow us to decipher any clinically significant findings. These could then be investigated further in a larger cohort of patients with LARC as part of a prospective study.

In this study, we used the CHPv2 to calculate heterogeneity scores which were analsyed as proxy markers for tumour subclones. From this, one can attempt to infer changes in tumour heterogeneity, which have been previously demonstrated to associate with tumour aggressiveness and response to therapy. Other bioinformatic programs, such as MATH and THI would be more appropriate for such analysis, whilst TMB would be a better tool for looking at tumour burden. However, we did not sequence enough of the genome to provide sufficient power for such analyses. Therefore, we applied tumour heterogeneity scores and quantitative changes as a result of therapy as a surrogate marker for clonal changes to infer changes in tumour heterogeneity, though this is not ideal nor the designed purpose of tumour heterogeneity scores.

It would also be beneficial to expand our small targeted NGS panel to whole exome/genome sequencing or a larger gene panel and to apply this to whole biopsies and/or surgical resections. This would allow analysis of tumour heterogeneity across the tumour, rather than being limited to investigate a small number of variants to analyse clonal changes. The reports that intratumour heterogeneity can be a prognostic indicator could make this an interesting area to investigate further.

Finally, as highlighted by the significant differences between different collection tubes, future analyses would benefit from an optimised and standardised protocol for pre-analytical sample handling and transportation. Although changes may have been made in the trial to address certain issues, this may have impacted circulating biomarker analysis in these patients.

5 Circulating Exosomes and Exosomal Cargo in Patients with Locally Advanced Rectal Cancer

5.1 Introduction

Exosomes are a subset of extracellular vesicles which perform an important role in intercellular communication by transporting molecular cargo to target cells (Valadi et al. 2007; Huang et al. 2012). The molecular cargo can then influence the cellular physiology of the recipient cell. In cancer, exosomes are reportedly co-opted and selectively packaged in order to promote tumour development in nearby and distant cells (Takano et al. 2017).

Exosomes have been detected in a number of bodily fluids, including plasma, serum and urine (Toiyama et al. 2016). The molecular cargo of exosomes has been reported to include proteins (Johnstone et al. 1987), mRNA and microRNA (Valadi et al. 2007). However, the presence of ExoDNA is still widely debated (Kahlert et al. 2014; Jeppesen et al. 2019).

MicroRNAs are small (19-26bp) non-coding RNAs which can supress the expression of a number of genes (Bartel 2004). MicroRNAs have been shown to be dysregulated in solid tumours and are hypothesised to drive tumourigenesis and/or cause treatment resistance (Calin and Croce 2006).

Plasma or serum-derived microRNAs (cell-free or encapsulated into exosomes) have reportedly been able to distinguish between healthy controls and patients with rectal cancer (Jo et al. 2017). Similar studies have also demonstrated the ability of circulating microRNAs to predict response to therapy (D'Angelo et al. 2016). MicroRNAs can be released into the circulation bound to carriers, such as Argonaute proteins, or encapsulated within exosomes (Chim et al. 2008)

The number of published studies investigating exosomes has risen rapidly over recent years (Lotvall et al. 2014). This had led to an increase in the development of commercially available kits with the purpose of facilitating exosome isolation in order to make it faster, simpler and more able to transition into the realm of diagnostics in the future. However, recent studies have demonstrated that different commercial kits result in variable exosome yield and purity leading to differences in downstream analyses of exoRNA signatures (Van Deun et al. 2014; Tang et al. 2017). This highlights the need for methodological standardisation in the analysis of exosomes.

Here, we intend to determine the optimal method for the isolation of exosomes from the plasma of patients with LARC. Isolated exosomes will then be used to validate the presence of and subsequently quantify levels of internal molecular cargo. We will then look to analyse quantitative changes over time of exosome derived molecular cargo.

This will be a proof-of-principle study to determine whether detection, quantification and analysis of exosome-derived molecular cargo is feasible in the circulation of LARC patients before examining these factors in the clinical context.

5.1.1 Study Objectives

In this chapter, the objectives are:

- Which method(s) provide the best yield and purity of exosomes from the plasma of patients with LARC?
- Can we detect, quantify and analyse the molecular cargo of isolated exosomes in patients with LARC?
- Can we detect and measure quantitative changes in the molecular cargo of exosomes before, during and after radiotherapy in patients with LARC?
    - How is exosomal molecular cargo released into the circulation in relation to radiotherapy?
    - Do longitudinal levels of exosomal molecular cargo correlate with clinical outcomes in patients with LARC?

5.2 Results

5.2.1 Exosome Isolation Kit Comparisons

Objectives: Which method(s) provide the best yield and purity of exosomes from the plasma of patients with LARC?

We explored several commercial kits for exosome isolation before choosing three which we then directly compared. The three methods were chosen as each used different technologies to isolate exosomes. These include precipitation by volume excluding polymers (TEI kit), membrane affinity columns (ExoEasy kit) and size exclusion chromatography (ExoSpin columns).

For each method, 1mL of pre-treatment plasma was used for exosome isolation. The size and concentration of isolated vesicles were calculated using the NTA and are illustrated in Figure 16. Sample purity was assessed using a P/P ratio (Table 30) as suggested by Webber and Clayton (2013).

Sample purity is a critical characteristic to consider in the field of exosome analysis. Current methods for examining particle purity involve the investigation of exosome and non-exosome specific markers, as well as electron microscopy to ensure vesicular morphology and size. Electron microscopy, however, is not widely available in a clinical or research setting, and the use of non-exosome markers is neither quantitative nor standardised at this time. Therefore, the P/P ratio has been proposed as a simple and quantitative biomarker for exosome purity analysis (Webber and Clayton 2013).

5.2.1.1 Extracellular Vesicle Purity

Table 30: P/P Ratios and respective exosome purity calculated from protein and particle concentrations according to Webber and Clayton (2012)

Isolation Kit	Particle Concentration (P/mL)	Protein Concentration (µg/mL)	P/P Ratio	Purity
TEI	1.70E+12	6817.789987	2.49E+07	Impure
ExoEasy	5.29E+12	761.8111434	6.94E+09	Low
ExoSpin	2.29E+13	2854.758585	8.02E+09	Low

Overall, we observed stark differences between these three methods with regards to particle size, particle concentration, protein concentration and overall purity.

5.2.1.1.1 Total Exosome Isolation Kit

Eluted microparticles from the TEI kit were deemed to be impure according to the calculated P/P ratios (Webber and Clayton 2013). This finding is in agreement with previous studies from Van Deun and colleagues (2014) and Tang and Colleagues (2017) who reported that the purity of vesicles isolated by the TEI kit were lower in comparison to methods such as sequential ultra-centrifugation. These findings contradict Schageman and colleagues (2013), who reported that vesicles isolated using the TEI serum or cell culture were as pure as exosomes from ultra-centrifugation.

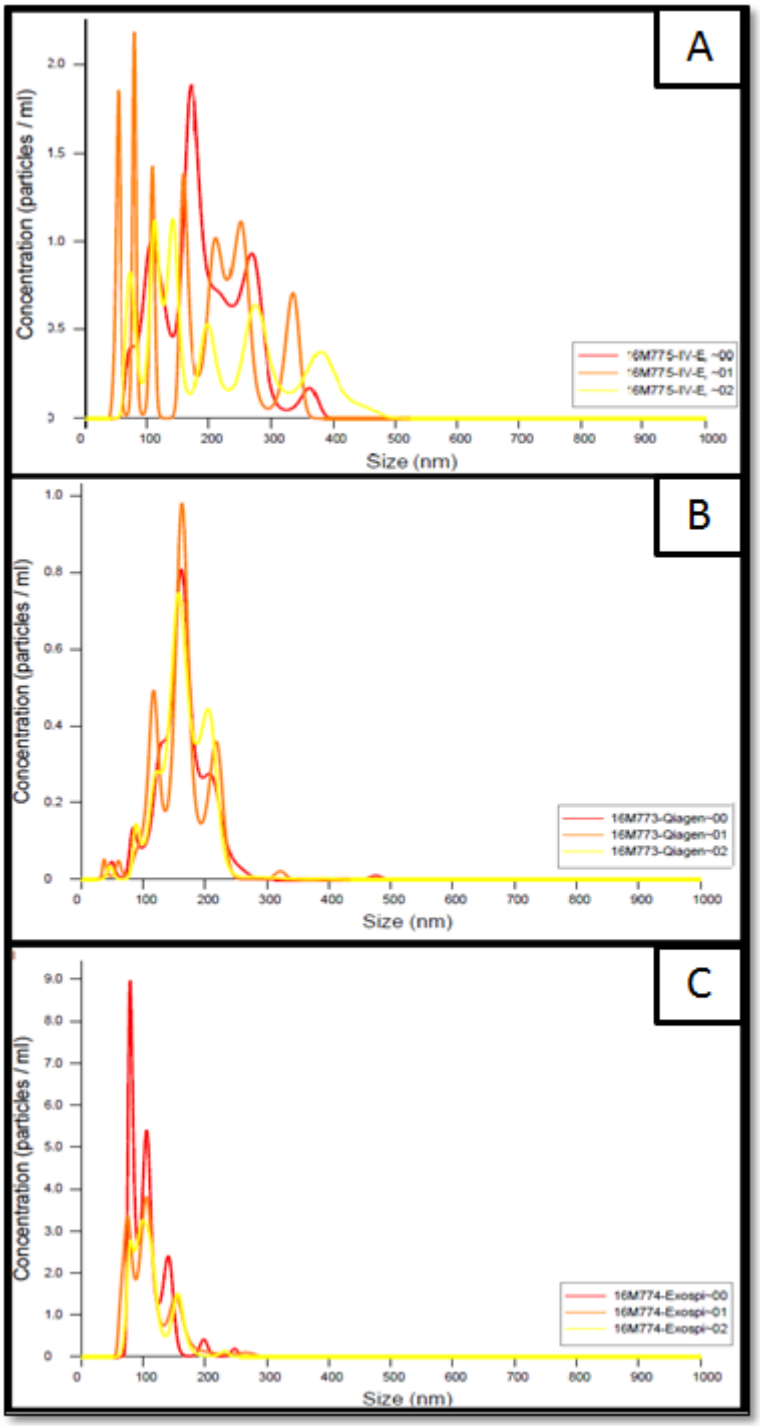

Figure 16: Nanosight Tracking Analysis displaying the concentratuion and size of extracellular vesicles isolated from 1mL of matching patient plasma samples using; A – the Total Exosome Isolation Kit, B–the ExoEasy Kit and C –ExoSpin columns.

The TEI kit appeared to isolate vesicles between 100-400nm in diameter (Figure 16A) with multiple peaks being observed. In contrast, other studies have reported that isolated vesicles were generally <300nm (Schageman et al. 2013; Tang et al. 2017). The highest concentration peak we observed for this method was 2.1 particles/mL.

Tang and colleagues (2017) also reported a large variation in the modal peak size in different serum samples. Schageman and colleagues (2013) reported that the vesicles isolated from the TEI kit were more uniform, with only a single modal peak size being observed. Both of these studies also showed that the TEI kit isolated a greater concentration of exosomes in comparison to sequential ultra-centrifugation (Schageman et al. 2013; Tang et al. 2017) whereas this method provided the lowest exosome yield in our study.

5.2.1.1.2 ExoEasy Kit

Using the ExoEasy kit, we observed a significant decrease in protein concentration in comparison to the TEI kit, however overall purity was still low. This contradicts findings from Enderle and colleagues (2015), who reported that exosomes isolated by the ExoEasy kit were indeed pure.

With this method, we isolated vesicles at 50-300nm in diameter (Figure 16B), with most of the isolated vesicles being ~200nm. This is concordant with previous studies using this method (Enderle et al. 2015; Stranska et al. 2018). These vesicles are slightly larger than expected according to recommended guidelines (Witwer et al. 2013). The yield of exosomes was also greater than the TEI kit. The highest concentration peak we observed for this method was 1.0 particles/mL.

5.2.1.1.3 ExoSpin Columns

ExoSpin columns provided the greatest level of purity and yield of exosomes. This is in agreement with a study from Lobb and colleagues (2015) which demonstrated that ExoSpin columns provided greater exosomal purity than precipitation based methods such as ExoQuick.

Exosomes isolated with ExoSpin columns showed a modal peak at 90-100nm and ranged from 50-200nm (Figure 16C). The highest concentration peak we observed for this method was 9.0 particles/mL. These findings are in accordance with suggestions from the International Society for Extracellular Vesicle (ISEV) (Witwer et al. 2013; Lotvall et al. 2014). We also observed a rise in protein concentration in comparison to the ExoEasy kit, which was also reported in previous studies (Welton et al. 2015).

Overall, our exosome yield and purity data are as expected compared to previous reports. The TEI kit provided the lowest sample purity and the ExoSpin columns provided the highest. The ExoEasy kit appeared to remove the most plasma-derived proteins whereas the ExoSpin kit removed less protein but had vesicles isolated within a narrower and more appropriate size range. However, both methods were limited regarding sample purity.

Due to the superior purity and yield of isolated exosomes, we chose to proceed with ExoSpin columns to analyse the exosomal molecular cargo. In the first instance we chose to investigate exoDNA.

5.2.2 Exosome-Derived DNA Analysis

Objective: Can we detect, quantify and analyse the molecular cargo of isolated exosomes in patients with LARC?

Here we attempted to validate the presence of DNA in circulating exosomes isolated from pre-treatment plasma from patients with LARC. ExoDNA was compared to matching cfDNA with regards to total DNA concentration, tumour-specific variants and DNA fragment length.

For this study, two patients with LARC were consented by the WCB (patients WCB1493 and WCB1262). Both patients had known tumour-specific variants in primary tumour tissue (*KRAS* p.G13D in patient WCB1493 and *KRAS* p.G12D in patient WCB1262).

Exosomes were isolated using ExoSpin columns from 1mL of plasma (Section 2.2.6.3) and DNA was extracted as described in Section 2.2.11.

DNA was quantified using the Qubit and ddPCR system (Table 31).

Table 31: DNA concentrations of cell free DNA and exosome-derived DNA calculated using the qubit and ddPCR

	DNA Concentration (Copies/µL)		**DNA Concentration (ng/µL)**	
Patient ID	ExoDNA	CfDNA	ExoDNA	CfDNA
WCB1493	4.77	28.4	0.15	1.1
WCB1262	5.54	8.16	0.99	1.7

In both patients, exoDNA had a lower concentration than matching cfDNA. This result was expected, as exoDNA would theoretically be captured during cfDNA extraction.

We then attempted to detect and quantify a known-tumour specific variant in each sample type. ddPCR was used to compare VAFs between the two sources for each patient. However, no tumour-specific variants were detected in either patient from cfDNA or exoDNA.

Studies have previously reported that tumour specific variants can be more frequently detected in exoDNA than in cfDNA in patients with pancreatic cancer (Allenson et al. 2017; Bernard et al. 2019). This is in agreement with San Lucas and colleagues' previous reports that a large fraction of exoDNA was tumour-derived (San Lucas et al. 2015) in patients with visceral

cancers. Studies have also reported that median VAF was increased and concordance with tumour tissue was superior in exoDNA compared to ctDNA in patients with metastatic pancreatic cancer (Bernard et al. 2019).

Though our data does not currently support these findings, our studies have numerous differences and limitations which may help explain any discrepancies.

As no significant differences were observed between exoDNA and cfDNA in terms of VAF, we proceeded to compare the length of DNA extracted from each source as described in Section 2.2.9.2.

The length of exoDNA was analysed due to previous reports which have demonstrated that exoDNA is 2-10kb in length (Kahlert et al. 2014; Thakur et al. 2014). Furthermore, longer DNA could allow the analysis of a larger range of variants, expanding from single nucleotide changes to larger scale insertions and deletions. The length of extracted exoDNA was measured using the Agilent Tapsetation 2200 (see Figure 17).

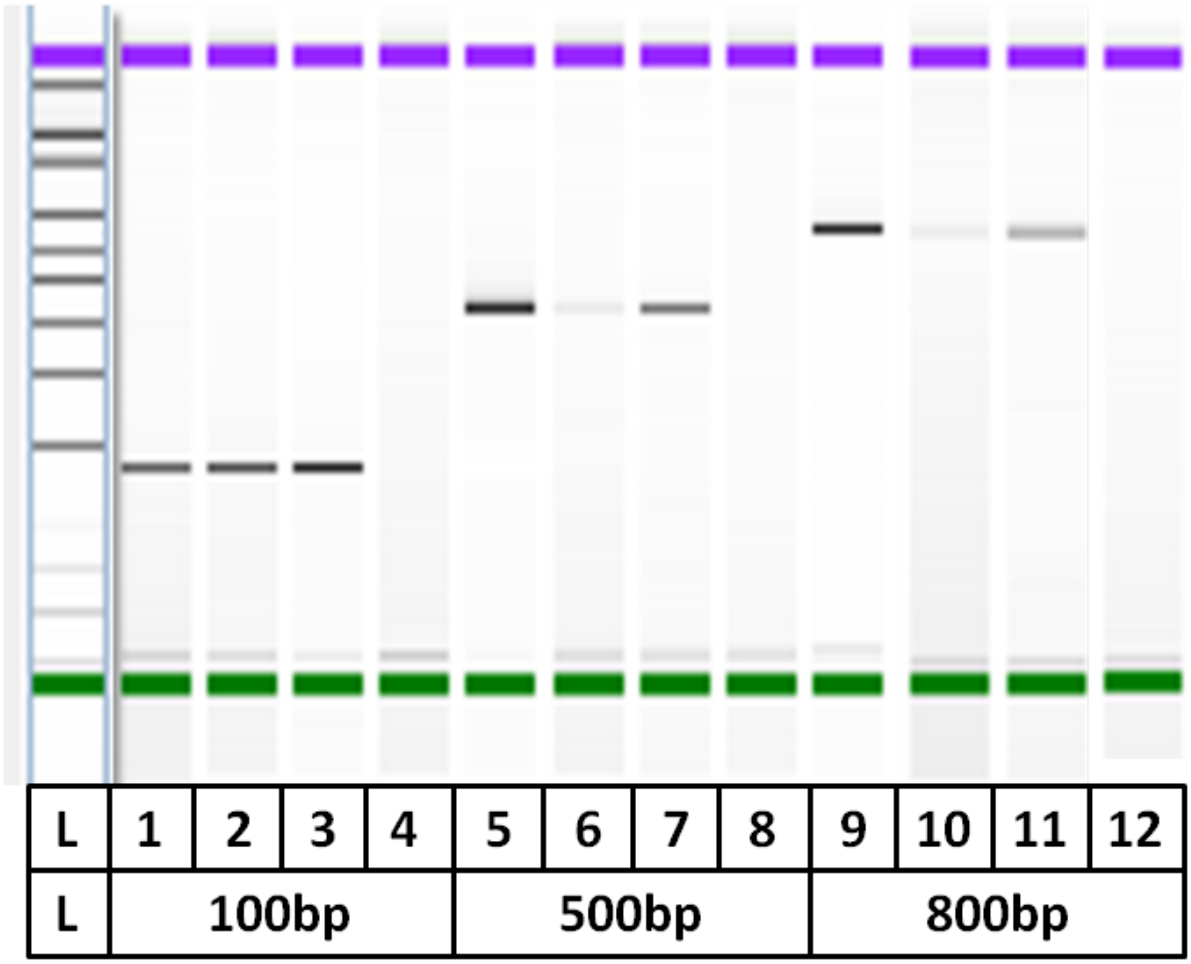

Figure 17: The Gel View of amplified DNA analysed by the BioAnalyser 2100. L – Ladder, samples 1-4 - amplified DNA at 200 base pairs for cellular DNA, exoDNA, cfDNA and No Template Control respectively, 5-8 - amplified DNA at 500 base pairs for cellular DNA, exoDNA, cfDNA and No Template Control respectively, 9-12 - amplified DNA at 800 base pairs for cellular DNA, exoDNA, cfDNA and No Template Control respectively

These results suggest that exoDNA is not longer than matching cfDNA, with very little being observed at ~800bp in comparison to the much stronger band present in cfDNA. This finding contradicts other reports, which have shown exoDNA to be >10kb in length (Kahlert et al. 2014) or up to 2.5kb (Thakur et al. 2014).

On the contrary, our data indicated exoDNA has similar properties to cfDNA, and appears to display none of the theorised benefits of encapsulation in exosomes. We therefore hypothesised that the extracted exoDNA may be located on the external surface of exosomes rather than encapsulated internally. To test this hypothesis, we performed a Proteinase K and DNase I digestion on isolated exosomes before performing DNA extraction. Proteinase K was included to ensure the DNA was not bound to contaminating proteins such as Argonautes which were isolated alongside exosomes. Paired control samples underwent the same temperatures during treatment but were not treated with either enzyme.

We also ensured that exosomes had been successfully lysed by the DNA extraction process by running the lysed sample though the NTA. Here we saw that all exosomes had indeed been lysed by the DNA extraction process (data not shown).

The DNA extracted from paired samples was quantified using the Qubit (Table 32).

Table 32: Total DNA concentrations of DNA extracted from circulating exosomes and matching Proteinase K and DNase I treated circulating exosomes from pre-treatment plasma in patients with LARC

	DNA Concentration (ng/µL)	
Patient ID	**Treated**	**Untreated**
WCB1493	TL	1.3
WCB1262	TL	1.6

TL – DNA was too low to be detected using the Invitrogen Qubit

Here we demonstrated that exoDNA from treated exosomes had been completely lost in comparison to matching undigested samples. This would suggest that the DNA was in fact bound to the external surface of the exosomes. This is in agreement with reports that the DNase I treatment of exosomes eliminated all detectable DNA from exosomes isolated from cell culture media and human plasma samples (Nemeth et al. 2017; Jeppesen et al. 2019).

Furthermore, we saw that the externally bound DNA was largely <1kb in length and shorter than matching cfDNA whereas previous reports have suggested that exoDNA was found up to 2.5kb (Kitai et al. 2017) or >10kb in length (Kahlert et al. 2014) after DNase I digestion.

Overall, our data suggests that the detected exoDNA was bound to contaminating proteins or to the external surface of exosomes, rather than being encapsulated within.

Confident that we would not detect exoDNA in these patients at this time, we proceeded to investigate the potential utility of exoRNA in this cohort.

5.2.3 Exosome-Derived RNA Extraction and Analysis

We have previously developed and optimised an assay for the extraction and analysis of exoRNA from the circulation of patients with LARC (See Appendices 9.7.1.1). Here we aimed to validate the presence of microRNA in circulating exosomes isolated from the plasma of patients with LARC.

Objective: Can we detect, quantify and analyse the molecular cargo of isolated exosomes in patients with LARC?

5.2.3.1 Systematic Review

We performed a systematic review in order to identify circulating microRNAs of interest with respect to patients with CRC or rectal cancer. Pubmed was used to look for the search terms 'Circulating microRNA Rectal Cancer' and 'Circulating MicroRNA Colorectal Cancer'. These two searches provided 12 and 147 hits respectively. Of these studies, six were shared between the two searches, therefore a total of 153 reports were examined as part of this systematic review.

Of the 153 studies, we excluded 37 reviews, 13 studies investigating patients with other cancer types, nine methodological development studies, seven-meta-analyses and two which were not available in English. Overall, we analysed 85 studies investigating circulating microRNAs in patient with colon and/or rectal cancers.

From these studies we identified 23 different microRNAs which were found to be significant in a diagnostic or prognostic manner, each of which was reported in at least three separate studies. The total list of identified microRNAs can be found in the Appendices (Section 9.8). From this list two microRNAs, Mir-31 and -125b, were chosen for investigation in sequential plasma samples throughout radiotherapy in patients with LARC.

Aside from this systematic review, we also chose to investigate Mir-99a*, which had previously been seen to influence DNA damage repair in response to ionizing radiation in prostate, breast and lung cancers (Mueller et al. 2013; Rane et al. 2016; Yin et al. 2018).

5.2.3.2 Assay Validation

To validate our developed assay, sequential plasma samples from three patients (WCB1479, WCB1493 and WCB1263) were used for sequential exosome-derived microRNA analysis. During this validation, microRNAs were analysed without normalisation.

Table 33: Concetration of microRNAs of interest extracted from circulating exosomes in the plasma of patients with LARC at various timepoints around radiotherapy

Patient Sample	Timepoint (Weeks)	Mir-31 (copies/µL)	Mir-99a* (copies/µL)	Mir-125b (copies/µL)
WCB1479	-2	20.8	2.08	225
WCB1479	6	7.31	0.763	153
WCB1479	16	10.1	1.65	287
WCB1493	-2	12.1	0	201
WCB1493	0	20	3.44	403
WCB1493	7	2.32	1.51	111
WCB1263	1	9.74	0.672	88.6
WCB1263	3	2.69	0.468	62.4

Week 1 marks the initiation of therapy for all patients

From this validation cohort, we saw variable concentrations of microRNA between timepoints and patients (Table 33). Only one timepoint had undetectable microRNA (Mir-99a* in patient WCB1493 week -2). Although there was one negative result, we generally saw detectable microRNA levels otherwise, and therefore decided to proceed with longitudinal analysis of the remaining nine patients. For these patients, microRNAs were normalised using RNU6B.

5.2.4 Sequential MicroRNA Analysis

Objective: Can we detect and measure quantitative changes in the molecular cargo of exosomes before, during and after radiotherapy in patients with LARC?

5.2.4.1 Investigating Circulating Exosome-Derived MicroRNA Using ddPCR

We applied our validated assay to quantify normalised levels of selected microRNAs in the circulating exosomes of patients with LARC. This is with the longer-term aim of investigating the ability of exosome-derived microRNAs to prognosticate and/or predict outcomes in patients with LARC in the future.

Nine patients with LARC consented by the WCB were used for longitudinal analysis. Clinical details for these patients can be seen in Table 34.

Table 34: Clinical characteristics of patients with locally advanced rectal cancer consented by the Wales Cancer Bank. ESMO stages grouped as suggested by Glynne-Jones et al. 2017

Patient ID	T	N	M	AJCC Grouped Stage	Radiotherapy (Gy in Doses)	Time Period of Radiotherapy
WCB1295	3	2	0	III	45 in 25	5 Weeks
WCB1368	1	0	0	I	20 in 5	1 Week
WCB1441	4b	0	0	II	45 in 25	5 Weeks
WCB1473	0	0	1	IV	30 in 5	1 Week
WCB1476	3	2	0	III	45 in 25	5 Weeks
WCB1477	3	0	0	II	45 in 25	5 Weeks
WCB1478	4	1	0	III	25 in 5	1 Week
WCB1603	4	0	0	II	45 in 25	5 Weeks
WCB1604	3	1	0	III	45 in 25	5 Weeks

AJCC – Grouped tumour stage in accordance with the American Joint Committee on Cancer

ExoRNA was extracted from sequential plasma samples from each of these patients according to the protocol described Section 2.2.11.1. Extracted RNA was reverse transcribed as described in Section 2.2.12.1.1 and analysed using ddPCR (Section 2.2.15.5).

Overall, all three microRNAs were confidently detected in all patients and timepoints at variable levels (Table 35).

Table 35: Normalised expression levels of microRNAs of interest extracted from circulating exosomes in the plasma of patients with LARC at various timepoints around radiotherapy. Gene expression was normalised using the RNU6B

ID	Week	Patient Status	Mir-31	Mir-99a*	Mir-125b
WCB1295	-3	Pre-Treatment	0.04	0.01	0.57
WCB1295	4	On-Treatment	0.02	0.01	0.44
WCB1368	-2	Pre-Treatment	0.02	0.06	0.24
WCB1368	1	On-Treatment	0.02	0.09	0.37
WCB1441	-2	Pre-Treatment	0.12	0.01	0.36
WCB1441	-1	Pre-Treatment	0.10	0.02	0.54
WCB1441	1	On-Treatment	0.10	0.01	0.51
WCB1473	0	Pre-Treatment	0.06	0.01	0.33
WCB1473	2	On-Treatment	0.02	0.01	0.22
WCB1473	13	Post-Treatment	0.05	0.01	0.30
WCB1476	-4	Pre-Treatment	0.03	0.005	0.19
WCB1476	12	Post-Treatment	0.03	0.02	0.51
WCB1477	-1	Pre-Treatment	0.04	0.01	0.64

ID	Week	Patient Status	Mir-31	Mir-99a*	Mir-125b
WCB1477	12	Post-Treatment	0.04	0.02	0.66
WCB1478	-4	Pre-Treatment	0.02	0.02	0.41
WCB1478	4	On-Treatment	0.02	0.008	0.30
WCB1603	-3	Pre-Treatment	0.05	0.006	0.81
WCB1603	1	On-Treatment	0.04	0.008	0.32
WCB1603	5	On-Treatment	0.03	0.003	0.18
WCB1603	14	Post-Treatment	0.01	0.005	0.23
WCB1604	-8	Pre-Treatment	0.04	0.02	0.34
WCB1604	-3	Pre-Treatment	0.05	0.03	0.63
WCB1604	0	Pre-Treatment	0.04	0.06	0.39
WCB1604	6	Post-Treatment	0.03	0.06	0.39

Week 1 marks the initiation of therapy for all patients

5.2.4.1.1 Mir-31

The reliable detection of Mir-31 across all patients contradicts previous reports which had excluded Mir-31 from sequential analysis, due to very low expression levels when analysed using qPCR in patients with LARC (Azizian et al. 2015).

Levels of Mir-31 did not significantly associate with T/N/M or grouped staging according to the AJCC in our cohort. This contradicts previous findings where Mir-31 expression correlated with tumour stage in the plasma and tissue of patients with CRC (Kanaan et al. 2012; Eslamizadeh et al. 2018).

We could not identify any distinct trends or patterns of change in Mir-31 throughout the course of therapy in our patient cohort. In contrast to this, studies have previously reported a significant decrease in circulating Mir-31 levels after the completion of therapy in rectal cancer (Jo et al. 2017). In agreement with this, Kim and colleagues (2014) reported that Mir-31 levels decreased in irradiated colon cancer cells *in vitro*. Though we may have seen a general decrease in circulating mir-31 in some patient samples, no significant differences were observed.

We did notice that patient WCB1441 had higher levels of Mir-31 than the other patients. This difference was observed against all patients except WCB1473 ($p = 0.06$) and WCB1603 ($p=0.06$) at pre-treatment.

5.2.4.1.2 Mir-99a*

We saw no specific trend of change in Mir-99a* levels from pre- to post-treatment samples. Previously, Mir-99a* has been reported to increase or decrease after therapy in patients with head and neck squamous cell carcinoma (Hou et al. 2015) or ovarian cancer (Yoshimura et al. 2018) respectively. Therefore, there are no consistent findings with which to compare results against on this subject. We also saw no significant associations with clinical pathological factors.

Levels of Mir-99a* were found to be expressed at a greater level in patient WCB1604 in comparison to other patients. This difference reached statistical significance in post-treatment samples against all available patients (See Appendices Section 9.10 for details). We did not see any statistically significant differences between patient WCB1368 and any other patients in this analysis, although levels appear to be very high, more so than patient WCB1604 (Figure 18).

5.2.4.1.3 Mir-125b

No patients displayed significantly high or low levels of Mir-125b in circulating exosomes at this time. No studies appear to have examined correlations between clinical pathological features and Mir-125b expression in patients with LARC or CRC. Previous reports have suggested no association exists in patients with chronic lymphocytic leukaemia (Gagez et al. 2017) and hepatocellular carcinoma (Liu et al. 2016). These contrast with reports that Mir-125b levels, in both tissue and plasma, were associated with clinical pathological features in patients with lung adenocarcinoma (Zhao et al. 2015). Furthermore, levels of Mir-125b have also been demonstrated to negatively correlate with tumour size (Sanchez-Sendra et al. 2018) and the presence of metastases (Huang et al. 2012) in the primary tumours of patients with melanoma, and squamous cell carcinoma respectively. Studies have also shown that low levels of circulating mir-125b were associated with more advanced stage in patients with osteosarcoma (Luo et al. 2016).

Raychaudhuri and colleagues (2017) reported that Mir-125b levels significantly increased after chemotherapy in the circulation of patients with breast cancer. Here we saw no significant changes as a result of radiotherapy in our patients.

Generally, no patterns of change could be easily discerned from total circulating exoRNA levels from these patients. Though concentrations were seen to vary, whether these variations are

significant changes which occurred as a result of therapy, or natural biological variation remains to be seen.

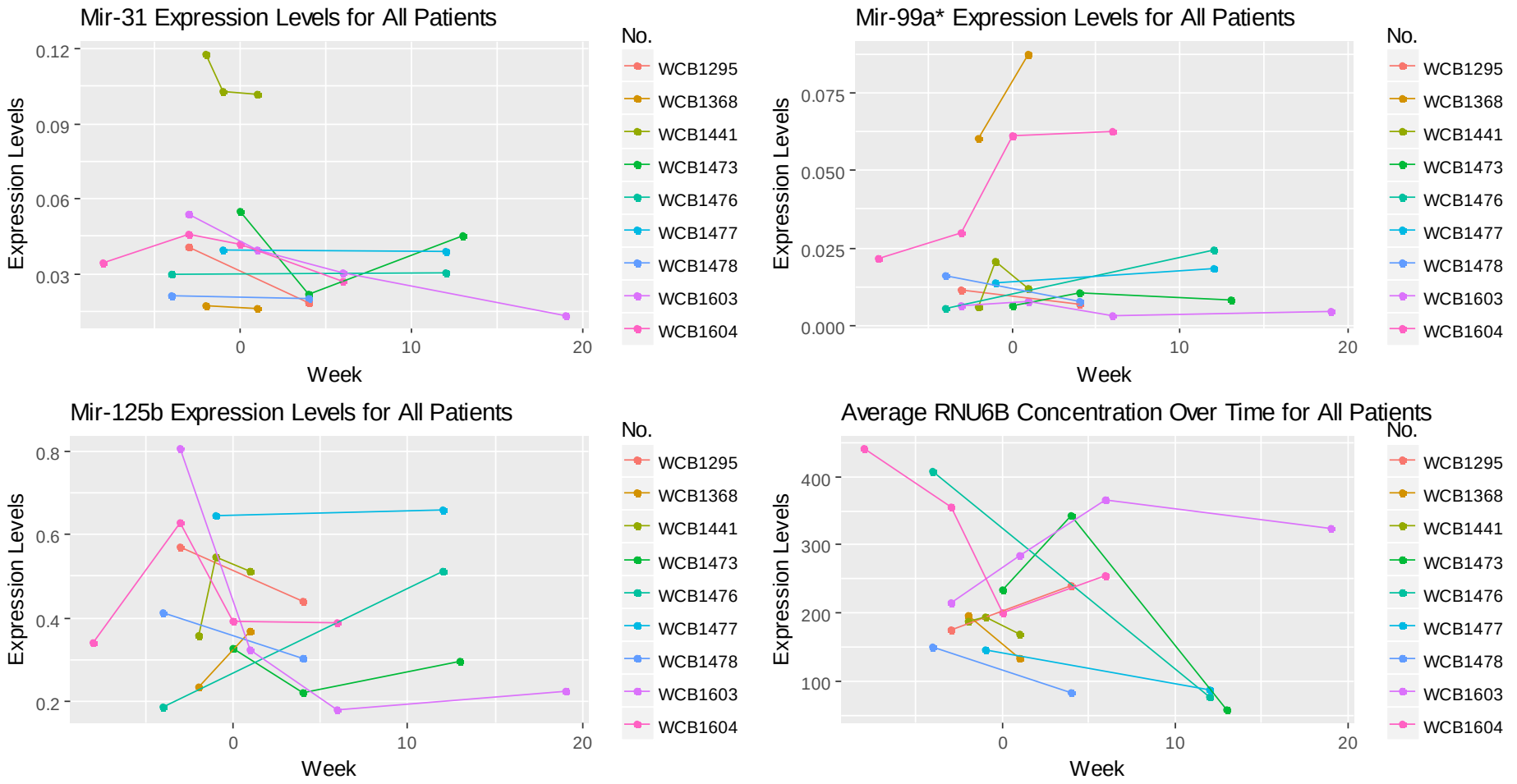

Figure 18: Normalised expression levels of microRNAs of interest and across all available timepoints from all patients. Gene expression was normalised using RNU6B. Average RNU6B here represents total RNA.

5.2.5 Patient Clinical Outcomes

Objective: Do longitudinal levels of exosomal molecular cargo correlate with clinical outcomes in patients with LARC?

Findings from exoRNA analysis were compared to patient clinical outcomes, specifically initial response to therapy and disease recurrence within two years as previously performed in section 3.2.4. As previously discussed in Chapter 3, two patients (WCB1368 and WCB1473) were excluded from clinical outcome analysis due to receiving atypical treatment. The clinical outcomes for the remaining seven patients is illustrated below in Table 36.

Table 36: Treatment outcomes for patients with LARC consented by the Wales Cancer Bank

Patient ID	Disease Stage	Radiotherapy (Gy in Doses)	TRG	Patient Experienced Local and/or Distant Relapse
WCB1295	III	45 in 25	1	None
WCB1441	II	45 in 25	4	Distant only
WCB1476	III	45 in 25	2	None
WCB1477	II	45 in 25	4	Local only
WCB1478	III	25 in 5	4	n/a
WCB1603	II	45 in 25	3	Local and distant
WCB1604	III	45 in 25	2	Distant only

As performed previously, clinical outcomes were separated in a binary fashion, including good vs poor response (TRG1/2 vs TRG 3/4), complete vs incomplete response (TRG1 vs TRG2-4) and whether or not patients experienced local and/or distant relapse in two years of follow-up. These are illustrated below in Table 37, Table 38 and Table 39 respectively.

Analysis of exoRNA levels in relation to patient outcomes was limited, as only seven patients who underwent exoRNA analysis had known outcomes. Furthermore, although all eight had pre-treatment plasma available, only one patient experienced a complete response to neoadjuvant therapy. Therefore, statistical analysis is not possible for the complete vs incomplete response (Table 38).

Table 37: A comparison of longitudinal circulating exoRNA levels across patients with good and poor responses to neoadjuvant therapy

	Good Response (TRG 1/2)		Poor Response (TRG 3/4)		
Variables	**N**	**Median (Range)**	**N**	**Median (Range)**	**p-value**
Pre-Treatment Levels					
Mir-31 Concentration (Copies/µL)	2	0.04 (0.03 – 0.04)	5	0.05 (0.01 – 0.1)	0.44
Mir-99a* Concentration (Copies/µL)	2	0.02 (0.01 – 0.02)	5	0.02 (0.01 – 0.06)	0.40
Mir-125b Concentration (Copies/µL)	2	0.38 (0.19 – 0.57)	5	0.48 (0.36 – 0.81)	0.40
On-Treatment Levels					
Mir-31 (Copies/µL)	1	0.02 n/a	4	0.03 (0.02 – 0.10)	n/a
Mir-99a* (Copies/µL)	1	4 n/a	4	0.01 (0.003 – 0.02)	n/a
Mir-125b (Copies/µL)	1	0.44 n/a	4	0.41 (0.18 – 0.54)	n/a
Post-Treatment Levels					
Mir-31 (Copies/µL)	1	0.03 n/a	4	0.04 (0.01 – 0.05)	n/a
Mir-99a* (Copies/µL)	1	0.02 n/a	4	0.02 (0.01 – 0.06)	n/a
Mir-125b (Copies/µL)	1	0.51 n/a	4	0.45 (0.23 – 0.66)	n/a

Table 38: A comparison of longitudinal circulating exoRNA levels across patients with complete and incomplete responses to neoadjuvant therapy

	Complete Response (TRG 1)		Incomplete Response (TRG 2-4)		
Variables	**N**	**Median (Range)**	**N**	**Median (Range)**	**P-Value**
Pre-Treatment Levels					
Mir-31 Concentration (Copies/µL)	1	0.04 n/a	6	0.04 (0.02 – 0.10)	n/a
Mir-99a* Concentration (Copies/µL)	1	0.01 n/a	6	0.01 (0.005 – 0.06)	n/a
Mir-125b Concentration (Copies/µL)	1	0.57 n/a	6	0.41 (0.19 – 0.81)	n/a
On-Treatment Levels					

	Complete Response (TRG 1)		Incomplete Response (TRG 2-4)		
Mir-31 (Copies/µL)	1	0.02 n/a	4	0.03 (0.02 – 0.10)	n/a
Mir-99a* (Copies/µL)	1	0.01 n/a	4	0.01 (0.003 – 0.02)	n/a
Mir-125b (Copies/µL)	1	0.44 n/a	4	0.41 (0.18 – 0.56)	n/a
Post-Treatment Levels					
Mir-31 (Copies/µL)	0	n/a n/a	5	0.03 (0.01 – 0.05)	n/a
Mir-99a* (Copies/µL)	0	n/a n/a	5	0.02 (0.01 – 0.06)	n/a
Mir-125b (Copies/µL)	0	n/a n/a	5	0.51 (0.23 – 0.66)	n/a

Table 39: A comparison longitudinal levels of circulating exoRNAs between different patient responses to neoadjuvant therapy. Any recurrence refers to patients who experience either local or distant relapse within two years of follow-up.

	Local Relapse			Distant Relapse			Any Relapse		
Variable	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p- Value**	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p- Value**	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p- Value**
Pre-Treatment Levels									
Mir-31 Concentration (Copies/µL)	0.04 (0.03 – 0.10)	0.04 (0.02 – 0.06)	0.65	0.04 (0.03 – 0.04)	0.05 (0.02 – 0.10)	0.16	0.04 (0.03 – 0.04)	0.05 (0.02 – 0.10)	0.28
Mir-99a* Concentration (Copies/µL)	0.01 (0.01 – 0.02)	0.01 (0.01 – 0.06)	0.56	0.01 (0.01 – 0.01)	0.02 (0.01 – 0.06)	0.34	0.01 (0.01 – 0.01)	0.02 (0.01 – 0.06)	0.45
Mir-125b Concentration (Copies/µL)	0.54 (0.19 – 0.57)	0.41 (0.36 – 0.81)	0.51	0.57 (0.19 – 0.64)	0.41 (0.36 – 0.81)	0.74	0.38 (0.19 – 0.57)	0.48 (0.36 – 0.81)	0.37
On-Treatment Levels									
Mir-31 Concentration (Copies/µL)	0.06 (0.02 – 0.10)	0.02 (0.02 – 0.03)	0.48	0.02 (n/a)	0.03 (0.01 – 0.05)	n/a	0.02 (n/a)	0.03 (0.02 – 0.10)	n/a
Mir-99a* Concentration (Copies/µL)	0.01 (0.01 – 0.01)	0.01 (0.01 – 0.06)	0.88	0.01 (n/a)	0.01 (0.003 – 0.02)	n/a	0.01 (n/a)	0.01 (0.003 – 0.02)	n/a
Mir-125b Concentration (Copies/µL)	0.48 (0.44 – 0.51)	0.3 (0.18 – 0.54)	0.6	0.44 (n/a)	0.41 (0.18 – 0.54)	n/a	0.44 (n/a)	0.41 (0.18 – 0.54)	n/a
Post-Treatment Levels									
Mir-31 Concentration (Copies/µL)	0.03 (n/a)	0.04 (0.01 – 0.05)	n/a	0.04 (0.03 – 0.04)	0.03 (0.01 – 0.05)	0.77	0.03 (n/a)	0.04 (0.01 – 0.05)	n/a
Mir-99a* Concentration (Copies/µL)	0.02 (n/a)	0.02 (0.01 – 0.06)	n/a	0.02 (0.02 – 0.02)	0.01 (0.01 – 0.06)	0.27	0.02 (n/a)	0.02 (0.01 – 0.06)	n/a
	0.51	0.45	n/a	0.59	0.39	0.31	0.51	0.45	n/a

	Local Relapse			Distant Relapse			Any Relapse		
Mir-125b Concentration (Copies/µL)	(n/a)	(0.23 – 0.66)		(0.51 – 0.66)	(0.23 – 0.51)		(n/a)	(0.23 – 0.66)	

Previous reports have demonstrated that Mir-99a* was up-regulated in the serum of patients with ovarian cancer (Yoshimura et al. 2018). Other studies have demonstrated that Mir-99a* was significantly down-regulated in the circulation of patients with breast (Yu et al. 2018) and bladder cancer (Feng et al. 2014; Du et al. 2015). Furthermore, Mir-31 has been demonstrated to be down-regulated in patients with rectal cancer (Jo et al. 2017) and CRC (Wang et al. 2014). However, we did not find any studies which have investigated circulating Mir-31 or Mir-99a* levels in relation with response to radiotherapy in patients with LARC at this time.

Increased levels of Mir-125b have been reported in the serum of patients with rectal adenocarcinoma which was associated with poor response to neoadjuvant therapy (D'Angelo et al. 2016). Similarly, in 56 patients with breast cancer, increased levels of Mir-125b were associated with resistance to chemotherapy (Wang et al. 2012). We did not see any such correlations in our cohort.

5.3 Discussion

5.3.1 Conclusions

5.3.1.1 Exosome Isolation Comparison

Objective: Which method(s) provide the best yield and purity of exosomes from the plasma of patients with LARC?

The purity and yield of isolated exosomes are important characteristics to consider due to the implications that these factors can have on down-stream analysis, either in a research or clinical setting (Van Deun et al. 2014). Here we observed considerable variation in the yield and purity of exosomes isolated by each method.

5.3.1.1.1 Total Exosome Isolation Kit

The TEI kit is a precipitation-based method which utilises volumes exclusion polymers, such as polyethylene glycol, to isolate extracellular vesicles. This method provided the lowest yield and exosome purity of the three tested. This was shown by the large range of the size of vesicles isolated and the numerous modal peaks observed. This presence of numerous peaks has been previously reported using this kit (Tang et al. 2017), suggesting that larger vesicles were isolated alongside exosomes. This contradicts studies which have reported that isolated vesicles were uniform when analysed by NTA (Schageman et al. 2013). Differences here may exist due to the nature of the isolation reagent. The large and numerous peaks we observed may have been the result of vesicle aggregation rather than necessarily representing the presence of microparticles of various sizes.

In this study, we determined vesicle purity using a P/P ratio whereas previous studies investigated the presence of exosome-specific markers in accordance with the ISEV (Lotvall et al. 2014). Studies have shown that exosomes isolated by the TEI kit lacked exosome-specific markers, such as CD63 (Van Deun et al. 2014) and CD9 (Tang et al. 2017). These results suggest that other non-exosomal microparticles are co-isolated using this method. In contrast, Schageman and colleagues (2013) demonstrated that exosomes isolated by the TEI kit expressed both CD63 and CD9. However, this study did not examine these markers using fluorescent microscopy as Van Deun and colleagues had, to prove that all isolated vesicles contained these markers (Van Deun et al. 2014).

In agreement with previous studies, our findings have demonstrated that exosomes isolated using the TEI kit generally lacked purity.

5.3.1.1.2 ExoEasy Kit

The ExoEasy kit is a membrane-affinity-based isolation method, meaning exosomes are captured by being bound to the membrane of a spin column. Exosomes isolated using the ExoEasy kit had an improved purity and yield in comparison to those extracted using the TEI kit. The size and range of isolated vesicles were similar to that demonstrated in previous studies (Enderle et al. 2015). However, this kit isolated microparticles >200nm which are larger than typical exosomes and other small extracellular vesicles (Thery et al. 2018).

Enderle and colleagues (2015) previously demonstrated that exosomes isolated using the ExoEasy kit displayed the exosome-specific marker TSG101. However, the ISEV also suggested the inclusion of at least two exosome specific markers as well as one non-exosomal marker (such as calnexin) for purity analysis (Lotvall et al. 2014). A recent study reported that exosomes isolated using this method had very low to undetectable levels of the exosome-specific markers TSG101 and CD81, suggesting low exosome purity (Stranska et al. 2018). This study also reported a high protein concentration with particles >200nm believed to be protein aggregates.

Overall, this approach provided greater yield and purity of exosomes than the TEI kit, but still has limitations regarding vesicle purity to a lesser degree.

5.3.1.1.3 ExoSpin Columns

ExoSpin is based on size exclusion chromatography and separates vesicles and microparticles based on particle size. We demonstrated here that the ExoSpin columns isolated exosomes with the greatest purity and yield of our tested methods. This is unsurprising as size exclusion chromatography-based methods have been demonstrated to provide superior sample purity over precipitation-based methods (Ludwig et al. 2018). Welton and colleagues (2015) demonstrated the purity of exosomes isolated with the ExoSpin columns using P/P ratios as well as CD9, CD63 and CD81 as exosome-specific markers. However, this report also demonstrated that a high level of Apo-B, a lipoprotein marker, was also detected in isolated vesicles using this method (Welton et al. 2015). This too brings into question the purity of the isolated exosomes in this fraction. Though some plasma albumin was detected, isolating exosomes from this fraction was reported to removed >99% of all albumin in patient plasma (Welton et al. 2015).

In this study we used a P/P ratio to examine sample purity. This is not a commonly used method for determining vesicle purity in comparison to the detection of exosome and non-

exosome specific markers, which is currently the gold standard. The detection of such markers, by western blots or, ideally, fluorescent microscopy provides greater clarity regarding vesicle purity and the presence of contaminating microparticles. The findings displayed in this study, though may be a good indicator, will need to be expanded upon, especially with the view to moving into the clinical setting in the future.

Sequential-ultracentrifugation and density gradients were not used or directly compared to the methods used in this study. Although sequential ultra-centrifugation was excluded (due to the potential difficulties transitioning such a method into the clinical setting in the future) this study could have benefitted with its inclusion in a technological comparison.

Although our purity analysis here was limited, we used our results in conjunction with findings from previous studies to select the optimal exosome isolation method in subsequent investigations.

5.3.1.2 Exosome-Derived DNA Analysis

Objective: Can we detect, quantify and analyse the molecular cargo of isolated exosomes in patients with LARC?

We initially detected exoDNA from the circulation of patients with LARC which was subsequently lost after exosomes were treated with Proteinase K and DNase I. This would suggest that it was extracellular DNA bound to contaminating proteins or the external surface of our isolated vesicles. This is consistent with previous reports (Valadi et al. 2007; Nemeth et al. 2017; Jeppesen et al. 2019) but contradicts a number of others (Kahlert et al. 2014; Thakur et al. 2014; Jin et al. 2016; Yang et al. 2017; Vagner et al. 2018).

Upon closer inspection, various reasons were identified for this contrast in findings. Yang and colleagues (2017) did not include a DNase digestion step whilst Jin and colleagues (2016) detected exoDNA in exosomes isolated using ExoQuick. ExoQuick has previously been shown to isolate vesicles of low purity (Van Deun et al. 2014). Therefore, non-exosomal vesicles (such as micro-vesicles or apoptotic bodies) may be the source of DNA in this case. In agreement with this, when Vagner and colleagues (2018) reported the presence of EV-derived DNA, it was encapsulated within larger vesicles (1 - 5.5µm) compared to smaller vesicles (100-400nm).

Nevertheless, studies have still reported the presence of exoDNA after appropriate exosome isolation and DNase treatment. Furthermore, exoDNA has also been reportedly detected during viral infections (Ambrosio et al. 2019) or pregnancy (Sheller-Miller et al. 2017).

Therefore, it appears DNA can be detected within exosomes under certain conditions, which numerous studies have proposed is related to cellular stress.

Studies have previously demonstrated that DNA and nuclear proteins can re-localise to the cytoplasm as a result of oxidative or cellular stress (Montermini et al. 2015; Sheller-Miller et al. 2017). Once in the cytoplasm, DNA can then be packaged into exosomes. This has also been demonstrated to occur to cells undergoing high level of DNA damage (Kitai et al. 2017; Takahashi et al. 2017). Whether cellular stress causes an active or passive secretion of genomic DNA remains to be determined. These studies may suggest that exoDNA appears or rapidly increases as a result of cellular stress and/or DNA damage.

A number of publications have reported the presence of double-stranded DNA in exosomes isolated from the serum of pancreatic cancer patients (Kahlert et al. 2014) and cancer cell lines (Thakur et al. 2014). These studies both identified tumour-specific variants from parental cells in exoDNA. In agreement with this, Lee and colleagues (Lee et al. 2014) demonstrated that exosomes can transfer oncogenic DNA to target cells *in vitro*. However, none of these studies have directly reported the presence of exoDNA in patients with CRC.

Overall, these studies suggest that DNA may be packaged into exosomes under specific conditions. These conditions appear to rely on the presence of cellular stress or DNA damage, causing DNA to relocate from the nucleus into the cytoplasm and be packaged into exosomes. Cellular stress and DNA damage are both associated with hallmarks of cancer (Hanahan and Weinberg 2011), and thus this leak of DNA into the cytoplasm may be a common occurrence in patient tumours and cancer cell lines. This may also be further amplified by patient treatment, especially radiotherapy, which can cause large amounts of both DNA damage and cellular stress in patient tumours.

We may have been unable to detect exoDNA in our patients for several reasons. Here we only analysed DNA from 1mL of plasma, where perhaps more is required to detect exoDNA using our methods in these patients. In addition, we only investigated this in two patients, thus further investigation of a larger number of patients may have provided more information on the subject. Furthermore, these patients were treatment-naïve and exoDNA may have been detectable after the initiation of radiotherapy and the associated DNA damage.

We may have been unable to detect exoDNA due to the sensitivities of the methods applied here for DNA quantification. Nor did we examine the sensitivities of our DNA quantification assays. Thus, our negative detection of exoDNA is limited according to these unknown

sensitivities. Finally, as our patients have locally advanced disease, with a generally low systemic tumour burden, their tumours may not yet be secreting exoDNA. These tumours may have not grown to the capacity and acquired the levels of cellular stress required for exoDNA secretion at this time. Therefore, this may not be the most appropriate patient cohort to investigate this feature with. For now, the question of exoDNA remains unanswered.

5.3.1.3 Circulating Exosome-Derived MicroRNA Analysis in Patients with LARC

Objective: Can we detect, quantify and analyse the molecular cargo of isolated exosomes in patients with LARC?

We were able to successfully develop a method to extract and analyse microRNA from exosomes isolated from the circulation of patients with LARC. From this exoRNA, we were able to detect and quantify the presence of three microRNAs of interest using ddPCR.

Here we showed that Mir-31 levels were reliably detected using the TaqMan MicroRNA RT kit and ddPCR. Previously, Azizian and colleagues (2015) reported that Mir-31 levels were too low to monitor in the serum of patients with LARC using the MiScript II RT kit and qPCR. However, here we showed that our kit allowed the targeted reverse transcription of small genes to increase cDNA levels. This can create separate limitations with regards to amplification bias which will need to be explored further in the future. Furthermore, ddPCR is a more sensitive method for RNA quantification than qPCR (Wang et al. 2019). Therefore, though levels were low, we were able to reliably detect and normalise Mir-31.

We also detected consistently low levels of Mir-99a*. This is not surprising as previous reports have suggested that Mir-99a* was down-regulated in the tumour tissue of patients with head and neck squamous cell carcinoma (Hou et al. 2015) and bladder cancer (Feng et al. 2014). As far as we know, to date, no studies have investigated the role of Mir-99a* in tumour tissue from patients with LARC. Mir-99a* displayed little variation, due to the very low levels detected and small numbers of patients analysed. Therefore, further work would be required to determine whether this microRNA can be used to predict response or clinical outcome in patients with LARC.

Mir-125b levels were relatively high in comparison to Mirs-31 and -99a*. This finding is as expected, due to previous reports that Mir-125b is up-regulated in the circulation of patients with CRC (Liu, 2018; Yamada, 2015). Furthermore, D'Angelo and colleagues (2016) reported that high Mir-125b levels were associated with poor response to pCRT in patients with LARC. There was much higher variation in Mir-125b levels between patients and timepoints in this

cohort in comparison to Mirs-31 and 99a*. Further investigation is required to determine whether these differences are indicative of clinical outcomes or a result of greater natural variation.

Expression levels varied between both microRNAs of interest and between patients. We saw that patients WCB1441 and WCB1604 had significantly higher levels of Mir-31 and Mir-99a* respectively in comparison to other patients. This might be due to up-regulation of these microRNAs within cancer cells or being actively secreted in exosomes. Alternatively, these differences may be a result of natural variation between patients. We also saw higher levels of Mir-99a* in patient WCB1368 but this was not statistically significant, possibly due to small sample numbers.

Expression levels of each microRNA and total RNA were not associated with T/N/M or AJCC staging. Most studies which compare clinical pathological features to microRNA levels generally investigate this in tumour tissue. Any correlations therefore do not necessarily translate into circulating exoRNA, due to the masking presence of non-tumour derived exosomes. This is also emphasised by the low stage and tumour burden associated with our patient cohort. Low tumour burden might suggest that a limited number of exosomes are being secreted into the circulation by the tumour. Therefore, a smaller number of tumour derived exosomes may be detected in the circulation in comparison to more advanced patients. As a result, the presence of any signatures may be made more difficult or impossible to identify. If these patient tumours are not shedding exosomes at a high rate, then the analysis of circulating biomarkers may be meaningless in these patients.

We must also consider the effects of changing exosome isolation methods for exoRNA analysis. The ExoRNEasy kit, an expansion of the ExoEasy kit, was applied for exoRNA extraction due to the larger range of plasma input volume, allowing us to reliably detect exoRNA. However, this method isolated exosomes of lower purity in comparison to ExoSpin columns. Our quantified microRNA expression levels may be influenced by the presence of contaminating vesicles and proteins. Furthermore, due to the low expression levels we observe here in Mirs-31 and 99a*, differences can be difficult to identify, especially in such a small sample cohort. The reverse transcription process used here included targeted amplification of microRNAs of interest, which can create amplification bias and the presence of any correlations or association may therefore be warped or lost. Finally, these microRNAs were chosen based on a review of the literature rather than us performing large scale analyses ourselves in these patients. Therefore, we cannot be entirely certain that these microRNAs are appropriate within our patient cohort.

Objective: Can we detect and measure quantitative changes in the molecular cargo of exosomes before, during and after radiotherapy in patients with LARC?

We did not identify any significant changes over time for any of our microRNAs nor any significant associations to clinical outcomes, which may be due to several reasons. If differences did indeed exist, the small number of patients here may not be a large enough sample size to identify them with any reliability or confidence.

One of the most important factors to consider when evaluating such findings are the patient characteristics of the cohort in question. As previously stated, these patients' tumours may not be secreting exosomes into the circulation at a particularly high rate. Therefore, the microRNAs we have analysed may originate largely from non-tumour tissue, rendering such analyses redundant, especially as there is currently no way to investigate tumour-specific exosomes or microRNA, in contrast to using tumour-specific variants in ctDNA. Therefore, such studies can be limited by the lack of tumour-derived microRNA in such analyses, specifically when patients have locally advanced disease with low tumour burden. As a result, the changes we observe, whether significant or not, may be the result of natural biological variation over time rather than tumour specific differences.

Another reason that we could not identify sequential changes may be the choice of microRNAs. These microRNAs may not be affected by ionizing radiation or be involved in cellular response to radiotherapy in patients with LARC. This could be due to reports that microRNAs have been demonstrated to be differentially expressed and perform various functions between different cancer types. Furthermore, the ExoRNEasy kit may have significant effects on gene expression analysis in these patients through the presence of contaminating vesicles or proteins and their associated microRNA levels.

Changes may also be distorted due to the use of targeted amplification during reverse transcription. This may, in turn, cause amplification bias and potentially influence microRNA levels and level changes between timepoints. However, use of such targeted amplification may be necessary to reliably detect some of these microRNA in the circulating exosomes of patients with LARC at all. Therefore, more knowledge is required regarding the consistency of amplifications in genes of interest using this method.

Although we analysed exoRNA levels, we did not analyse the concentration of circulating exosomes. The inclusion of circulating exosome concentration in tandem with exoRNA data could provide more clinically relevant information regarding a patient tumour. Furthermore, as

we could not quantify microRNA concentrations before performing reverse transcription, we are not aware of how consistent our ddPCR analysis was, due to the potentially variable concentrations of input microRNA.

In this study, RNU6B was used as a reference gene to normalise microRNA expression levels. This gene was chosen due to utility in previous studies but may not necessarily be appropriate for utility in our cohort.

At present, we are unaware of the optimal timings for blood collection to measure circulating biomarkers in relation to radiotherapy. Circulating biomarker analysis throughout radiotherapy is ideal. However, we need to identify when to collect plasma after the administration of radiotherapy, whether it is a few hours after therapy or on which days or weeks would allow optimal circulating biomarker analysis. This requires further investigation and standardisation to facilitate predicting patient outcomes and response.

5.3.2 Clinical Implications

Objective: Do longitudinal levels of exosomal molecular cargo correlate with clinical outcomes in patients with LARC?

5.3.2.1 Mir-31

Mir-31 has previously been associated with having important roles in numerous solid tumours. In patients with LARC, Mir-31 has been seen to be overexpressed in 17 patient tumours and down-regulated in 111 patient plasma in comparison to normal mucosa and healthy control patients respectively (Jo et al. 2017). Similarly, other studies have reported that high Mir-31 levels were associated with decreased response to 5-FU in colon cancer cell lines (Wang et al. 2010) and tumour tissue in patients with LARC (Carames et al. 2016). Although the study by Wang and colleagues (2010) was only performed *in vitro*, without expansion into patient samples, these findings have since been supported in patients by more recent reports (Carames et al. 2016). However, only one of these studies investigated circulating Mir-31 (Jo et al. 2017). Overall, research in this field is limited with regards to patients with rectal or colorectal cancer. Therefore, further investigations are required to ensure these findings are consistent in tissue and in circulating exosomes.

Notably, Mir-31 was chosen for evaluation due to it having been previously reported to be associated with ionizing radiation response. In colon cancer cells, Mir-31-5p was shown to induce radio-sensitivity in MMR proficient cells by targeting *MLH1* (Kim et al. 2014). Similarly, Mir-31 levels were positively associated with improved response and cell death in oesophageal cancer patients and cell-lines respectively by targeting DNA repair genes (Lynam-Lennon et al. 2012). This study also reported that Mir-31 was informatically predicted to target DNA repair genes, including *PARP* and *MLH1*, and that these genes were down-regulated in radiosensitive tumour samples. The targeting of *MLH1* and/or other DNA repair genes by Mir-31 may be an integral mechanism in cellular response to radiotherapy.

Furthermore, Mir-31 has been previously linked to p53 status in serous ovarian cancer, which is also commonly investigated in the context of response to radiotherapy (Creighton et al. 2010). Finally, Mir-31 has also previously been linked to the Ras pathway (Sun et al. 2013a) which has been implicated in response to radiotherapy (Chow et al. 2016). However, the function and expression of levels of Mir-31 has been shown to differ between specific cancer types.

We did not identify any significant differences or associations regarding Mir-31 levels and patient response as part of this analysis. We decided to investigate patient WCB1441 further as they displayed significantly higher levels of Mir-31 than other patients. This patient responded poorly to therapy (TRG 4) and developed distant metastases in the lung after 20.6 months. They also had a large tumour volume (T-stage 4b), which may have influenced Mir-31 levels. However, other patients with lower levels of Mir-31 also experienced a poor treatment response with progression.

Circulating Mir-31 levels may be useful as a prognostic marker in patients with LARC, as studies have shown that levels are indicative of patients' survival to pCRT. However, this will require further analysis to verify.

*5.3.2.2 Mir-99a**

In tumour tissue, Mir-99a* has previously been reported to be down-regulated in patients with LARC who respond poorly to pCRT (Svoboda et al. 2012). We therefore hope to examine here whether Mir-99a* levels in plasma-derived exosomes can be used in a similar manner.

Considering diagnostic potential, Yan and colleagues (2018) reported that Mir-99a* was down-regulated in prostate cancer in comparison to normal prostate tissue. Svoboda and colleagues (2012) reported that Mir-99a* was linked to DNA repair by targeting RAD51C and RAD9B, genes known to be involved in DNA repair. They therefore hypothesised that Mir-99a* would cause a down regulation of DNA repair genes leading to radiosensitivity. This tumour-suppressive nature of Mir-99a* is in agreement with reports from Molina-Pinelo and colleagues (2014) who stated that up-regulated Mir-99a* was associated with improved progression-free and overall survival in patients with advanced colorectal cancers. Mir-99a* was also shown to inhibit cell proliferation, migration and invasion in prostate cancer cell-lines (Arai et al. 2018).

Studies have generally demonstrated that Mir-99a* provides a tumour suppressive function. This is in agreement with reports that Mir-99a* levels are low in the primary tumour in patients with more aggressive disease in prostate cancer (Rane et al. 2016). The role of Mir-99a* and DNA repair in response to radiotherapy makes this an interesting candidate for analysis in the cohort. However, studies have primarily focussed on tumour tissue and cancer cell-lines, therefore expectations for detection, quantification and clinical relevance of Mir-99a* in these patients is currently unknown.

We did not identify any clinically significant associations between circulating Mir-99a* levels and patient clinical outcomes within this study. We hypothesised that patient WCB1604 would respond well to treatment, due to their significantly high levels of detect circulating Mir-99a* in comparison to other patients. However, this patient experienced a poor response (TRG3) to radiotherapy, although had previously received chemotherapy. This patient continued to develop metastases in the lungs after 15.8 months.

5.3.2.3 Mir-125b

Mir-125b has been demonstrated to influence apoptosis through targeting genes like BAK1, Puma, Cyclin C and p53 (Jia et al. 2012; Zhao et al. 2012; Banzhaf-Strathmann and Edbauer 2014). Hypothetically, Mir-125b levels may be indicative of apoptosis and, therefore, may be able to predict or infer response to radiotherapy in patients with LARC. Another study demonstrated that circulating levels increased after treatment in patients with breast cancer (Raychaudhuri et al. 2017). Mir-125b was reported to be down-regulated before the initiation of therapy in the tumour tissue of patients with rectal cancer (D'Angelo et al. 2016) and oral squamous cell carcinoma (Shiiba et al. 2013). Mir-125b was also seen to be up-regulated in the circulation of patients with breast cancer before therapy began (Wang et al. 2012). Consequently, we expected to observe that high Mir-125b was associated with improved response to therapy. However, we found that no patients were seen to have significantly high or low levels of Mir-125b nor any significant changes throughout radiotherapy. Nor were such levels significantly associated with any type of response. This finding is unsurprising due to our small sample numbers.

Although we did not see any initial association with clinical characteristics, these microRNAs investigated in this study may still be able to play an important role in predicting outcome and response to radiotherapy in patients with LARC.

Although the three exoRNAs (Mir-31, -99a* and -125b) were chosen based on a review of the literature, further investigation will be required to show that these were appropriate choices for exoRNA analysis in this cohort of patients. There were also likely to be variable levels of amplification bias between microRNAs and patients during this analysis, potentially masking any significant association. The presence of these exoRNAs being released by non-tumour tissue and physiological processes could also impact our ability to identify any significant data. Ideally, further investigation is required to determine whether concentrations of tissue-derived microRNA and exoRNA correlate or not.

There was also a lack of uniformity for the collection of plasma samples in terms of both stage of treatment and the number of samples collected throughout our cohort. This can create difficulties when attempting to perform a statistical comparison. Finally, only a small number of patients (n=8) underwent exoRNA examination, thus this analysis is underpowered and our lack of significant findings is not unexpected.

The primary objective of this chapter was to develop an assay which would facilitate exoRNA analysis from 1mL of plasma in patients in LARC. Therefore, clinical analysis was a secondary objective and a way to attempt to apply the data. This assay will require further development in the future before attempting further clinical analysis. This will likely include the necessity of further optimisation for both exoRNA extraction and/or ddPCR analysis before ensuring the assay provides regular reproducible findings from separate aliquots of matching plasma samples. Furthermore, the chosen microRNAs will require further investigation to determine whether they are suitable for such analysis, and, if not, the identification of suitable microRNAs from hypothesis-free studies, such as transcriptomic sequencing or microRNA array analysis. Although we achieved our primary objectives, extended investigations will be required before repeating such clinical analyses performed here.

5.3.3 Future Studies

Studies have shown that specific protein markers, like Glypican-1 can potentially distinguish tumour and non-tumour derived exosomes in patients with pancreatic (Melo et al. 2015) and colorectal cancer (Li et al. 2017). Though this field is still in its infancy, utility of such a marker in rectal cancer could greatly improve results and clinical applicability of microRNA expression in the future. This may discount the role of the tumour-microenvironment, including cancer associated fibroblasts, which may be clinically useful. As previously stated, this field is still in an early developmental phase and more research is required.

It would be interesting to investigate whether exoDNA is released and/or detectable after the initiation and/or completion of radiotherapy. This is due to the connection that exoDNA reportedly has with cellular stress and DNA damage. We would also like to investigate this in larger plasma volumes and more advanced patients, in order to identify the extent to which these factors can influence exoDNA secretion and detection.

It would be ideal in the future to perform hypothesis free analytical methods (such as NGS or microarrays) to identify a much larger number of microRNAs in patient tissue and/or circulating exosomes. From this, a small number can then be analysed further. This would also assist in identifying appropriate reference genes for normalisation. This would certainly require further assay development for exosome isolation and RNA extraction, as such methods would have limited sensitivities in comparison to ddPCR applied here.

Having tumour tissue available from these patients for microRNA expression analysis could assist in identifying microRNAs of interest. The use of tumour tissue before and after radiotherapy could also provide valuable insight as to which microRNAs are most heavily involved in providing radio-sensitivity or -resistance to cancer cells. Finally, monitoring these microRNAs in circulating exosomes and comparing them to changes in tumour tissue, could help infer whether microRNAs are released in an active or passive fashion.

It would also be of interest to compare microRNA profiles of isolated circulating exosomes to matching plasma. Findings here could either further enforce the role of circulating exosomes in cancer or assist overcoming numerous technical limitations by removing the need to isolate exosomes.

Finally, in future studies we could attempt to analyse the presence of mRNA in depth in place of microRNA. This could also potentially allow the detection and quantification of tumour-specific variants and allow a direct comparison to ctDNA and cfDNA.

6 Analysis and Clinical Utility of Circulating Biomarkers in Predicting Outcome in Locally Advanced Rectal Cancer Patients

6.1 Introduction

Standard therapy for patients with locally advanced rectal cancer (LARC) is neoadjuvant therapy and surgery with curative intent. Dependent upon tumour stage, up to 20% of patients can achieve pCR (Maas et al. 2010). Patients who achieve pCR or a partial clinical response have improved overall survival OS and DFS (Janjan et al. 2001; Maas et al. 2010). However, the current methods for predicting patient response to therapy remain limited.

Pre-operative CEA is currently used to predict response and/or survival to neoadjuvant therapy in patients with rectal cancer (Colloca et al. 2017). CEA has recently been reported to predict OS and DFS in a systematic review in patients with non-metastatic rectal cancer, suggesting its continued clinical importance for patients with LARC (Colloca et al. 2017).

Several circulating biomarkers may provide benefits over CEA, some of which may have use in routine clinical practice. These include; components of the full blood count and related systemic inflammatory indices (such as NLR, PLR and LMR).

Here we analysed full blood count data for 235 patients with LARC who were treated at the Velindre Cancer Centre. With this cohort, we aimed to acquire greater statistical power to identify significant associations between circulating biomarkers and clinical outcomes in patients with LARC.

We also investigated pre-treatment levels of haemoglobin and CEA to detmerine whether they could be predictive of patient outcomes and response to therapy in 14 patients with LARC consented by the WCB on which ctDNA, cfDNA and exoRNA analyses were previously performed.

6.1.1 Study Objectives

This study intends to determine the predictive power of circulating biomarkers in patients with LARC undergoing neoadjuvant therapy. In this study, our objectives are:

- Can clinical full blood counts and derived immune ratios be used to predict clinical outcomes in response to neoadjuvant therapy in patients with LARC?
- Can baseline CEA or haemoglobin levels be used to predict response to neoadjuvant therapy in patients with LARC?

6.2 Results

6.2.1 Circulating Biomarker Analysis in Patients with LARC

Objective: Can clinical full blood counts and derived immune ratios be used to predict clinical outcomes in response to neoadjuvant therapy in patients with LARC?

Clinical data for these patients included the age at which patients were diagnosed, ranging from 27-92 (median 62). All patients diagnosed with LARC were treated with curative intent. A total of 18 (7.7%) patients received short-course (1 week) therapy and 96 (40.9%) received long-course (5 weeks) treatment. No detailed treatment data was available for the remaining 122 patients (51.5%). Blood was collected at baseline (pre-treatment) and week 5 for each of these patients as part of routine clinical care. Week 5 refers to the fifth week after patients began therapy, which may be during or after the administration of treatment for patients on long- and short-course therapy respectively. From routine blood collections, platelet, lymphocyte, neutrophil, and monocyte concentrations were available at both timepoints. Median follow-up for the whole cohort was 37 months, and 177 patients had a minimum of two years follow-up available for analysis. A summary of available clinical characteristics of this cohort are displayed in Table 40.

Table 40: Clinicopathological characteristics of all 235 patients with locally advanced rectal cancer undergoing circulating biomarker analysis

Clinicopathological Characteristics		
Characteristic	**N**	**%**
Sex		
Male	160	68.1
Female	75	31.9
Clinical T Stage		
T1	1	0.4
T2	32	13.6
T3	174	74
T4	26	10.8
Unknown	2	0.9
Clinical N Stage		
N0	62	26.4
N1	87	37
N2	72	30.6
Unknown	14	7.8
Clinical M Stage		
M0	195	83
M1	15	6.4

Clinicopathological Characteristics		
Characteristic	**N**	**%**
Unknown	25	10.6
Recurrence Within 2 years		
Local	45	19.1
Distant	66	28.1
Any	72	30.6
None	115	48.9
Unknown	48	20.4
Course of Therapy		
Long Course (5 Weeks)	96	40.9
Short Course (1 Week)	17	7.2
Unknown	122	51.9
Response to Therapy		
Complete Response	9	3.8
Incomplete Response	175	74.5
Unknown	56	23.8

We initially investigated whether any components from full blood counts or derived immune ratios could predict how patients with LARC responded to neoadjuvant therapy. Response to therapy was measured using two separate clinical endpoints. At first, patients were separated according to their initial response to therapy (complete vs incomplete). We then compared concentrations of each respective circulating biomarker between the two responses to therapy. We also investigate whether these biomarkers could predict patients who experienced local and/or distant disease recurrence within the first two years of follow-up. We analysed recurrence in patients after two years as studies have reported that disease recurrence within this time period was indicative of long-term outcomes (Sargent et al. 2011).

Complete response was defined as ypT0 and ypN0 in this analysis. We applied ypT0 and ypN0 as definitions of pCR in this part of the study because they refer to the complete absence of primary tumour tissue (ypT0) and lymph node involvement (ypN0) after neoadjuvant therapy. Local recurrence was defined as intra-pelvic recurrence after primary cancer resection (Cai et al. 2014). Therefore, distant disease recurrence arises in lymph nodes or other organs outside of the pelvic region.

The circulating biomarkers analysed within this study included platelet, neutrophil, monocyte and lymphocyte concentrations, as well as PLR, NLR and LMR values. An unpaired two-sample T-Test was performed to identify any differences between the separate responses to therapy in all patients (see Table 41 and Table 42).

Table 41: A comparison of pre-treatment blood cell counts between patients who experienced complete and incomplete response to neoadjuvant therapy in all 235 patients. Change in variable refers to the difference between baseline and week 5 for each respective biomarker.

	Complete Response	Incomplete Response	
Variable	**Median (Range)**	**Median (Range)**	**P-Value**
Pre-Treatment Platelet Count	211 (162 – 308)	275 (88.0 – 601.0)	*0.07*
Week 5 Platelet Count	183 (133 – 221)	227 (64.0 – 515.0)	0.02
Change in Platelet Count	-54 (-98 – 10)	-49 (-326.0 – 230.0)	0.86
Pre-Treatment Lymphocyte Count	1.5 (0.8 – 2.9)	1.7 (0.7 - 3.8)	0.34
Week 5 Lymphocyte Count	0.7 (0.4 – 1.2)	0.7 (0.2 - 3.2)	0.93
Change in Lymphocyte Count	-0.8 (-1.7 – -0.2)	-1.1 (-2.6 – 0.8)	0.32
Pre-Treatment Neutrophil Count	3.7 (2.6 – 7.2)	4.6 (1.9 – 11.0)	0.26
Week 5 Neutrophil Count	3.2 (1.8 – 4.4)	3.8 (1.4 – 15.0)	0.20
Change in Neutrophil Count	-1.6 (-3.1 – 0.7)	-1.1 (-6.4 – 8.4)	0.93
Pre-Treatment Monocyte Count	0.5 (0.3 – 1.1)	0.6 (0.1 – 1.3)	0.54
Week 5 Monocyte Count	0.4 (0.3 – 0.9)	0.5 (0.2 – 1.2)	0.43
Change in Monocytes	0 (-0.21 – 0.20)	0 (-0.6 – 0.5)	0.97
Pre-Treatment PLR	139.5 (92.6 – 340.0)	158 (55.2 – 641.0)	0.56
Week 5 PLR	242.9 (141.0 – 470.0)	335 (69.7 – 1477.0)	0.12
Change in PLR	103.4 (8.6 – 244.0)	172 (-93.9 – 975.8)	0.10
Pre-Treatment NLR	2.6 (1.3 – 5.9)	2.6 (1.0 – 11.1)	0.87
Week 5 NLR	4.6	5.5	0.23

	Complete Response	Incomplete Response	
	(1.6 – 8.2)	(0.9 – 21.4)	
Change in NLR	2.1	2.9	0.23
	(-1.0 – 5.1)	(-5.3 – 18.3)	
Pre-Treatment LMR	0.32	0.3	0.86
	(0.2 – 0.6)	(0.1 – 1.0)	
Week 5 LMR	0.6	0.8	0.25
	(0.3 – 1.3)	(0.2 – 2.7)	
Change in LMR	0.37	0.5	0.19
	(0.02 – 0.7)	(-0.3 – 2.4)	

Table 42: A comparison of sequential platelet counts, lymphocyte counts, neutrophil counts, monocyte counts, PLR, NLR and LMR between patients who experienced local and/or distant recurrence within two years of follow up in comparison to those who did not. All values were analysed as continuous variables, not as high or low based on a designated threshold. Change in variable refers the difference between baseline and week 5 for each respective biomarker. Any recurrence refers to patients who experience either local or distant relapse.

	Local Relapse			**Distant Relapse**			**Any Relapse**		
Variable	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p-Value**	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p-Value**	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p-Value**
Pre-Treatment Platelet Count	277.5 (88.0-494.0)	182 (174.0-540.0)	0.65	281 (126.0-601.0)	276 (88.0-562.0)	0.57	280 (126.0-601.0)	280 (88.0-652.0)	0.81
Week 5 Platelet Count	224 (64.0-465.0)	167.5 (128.0-515.0)	0.11	231.5 (115.0-465.0)	239 (64.0-515.0)	0.37	232 (115.0-465.0)	241 (64.0-515.0)	0.51
Change in Platelet Count	-54 (-326.0-177)	-17 (-245-230.0)	0.3	-56 (-326-166.0)	-38 (-326-230.0)	0.07	-54 (-326-166.0)	-43 (-245-230.0)	0.38
Pre-Treatment Lymphocyte Count	1.8 (0.6-3.8)	1.4 (0.9-3.7)	0.72	1.8 (0.6-3.8)	1.7 (0.9-3.7)	0.62	1.8 (0.6-3.8)	1.7 (0.9-3.7)	0.42
Week 5 Lymphocyte Count	0.6 (0.4-2.8)	0.5 (0.3-3.2)	0.12	0.6 (0.2-2.8)	0.7 (0.3-3.2)	0.13	0.6 (0.2-2.8)	0.7 (0.3-3.2)	0.22
Change in Lymphocyte Count	-1.1 (-2.0-1.6)	-1 (-2.5-0.8)	0.24	-1.1 (-2.9-1.6)	-1 (-2.5-0.8)	0.27	-1.1 (-2.9-1.6)	-0.9 (-2.5-0.8)	0.22
Pre-Treatment Neutrophil Count	4.6 (4.0-11.0)	4.6 (1.9-8.4)	0.46	4.6 (1.9-11.0)	4.6 (2.5-9.6)	0.76	4.6 (1.9-11.0)	4.8 (2.5-9.6)	0.77
Week 5 Neutrophil Count	3.5 (1.4-9.3)	2.9 (2.1-15.0)	0.01	3.5 (1.4-9.3)	4 (1.9-15.0)	0.06	3.5 (1.4-9.3)	4 (1.9-15.0)	*0.06*
Change in Neutrophil Count	-1.2 (-6.1-4.8)	-0.9 (-4.6-8.4)	0.09	-1.2 (-6.1-4.8)	-1.1 (-3.9-8.4)	0.06	-1.2 (-6.1-4.8)	-1.2 (-4.6-8.4)4	0.22

	Local Relapse			Distant Relapse			Any Relapse		
Pre-Treatment Monocyte Count	0.6 (0.2-1.22)	0.5 (0.1-1.1)	0.35	0.6 (0.1-1.22)	0.6 (0.1-1.1)	0.47	0.5 (0.3-1.22)	0.5 (0.1-1.1)	0.5
Week 5 Monocyte Count	0.5 (0.2-1.8)	0.5 (0.3-1.2)	0.16	0.5 (0.2-1.8)	0.5 (0.3-1.2)	0.16	0.5 (0.2-1.8)	0.5 (0.3-1.2)	0.2
Change in Monocyte Count	0 (-0.5-0.6)	-0.2 (-0.7-0.5)	0.85	0 (-0.6-0.6)	-0.1 (-0.4-0.5)	0.79	0 (-0.6-0.6)	-0.1 (-0.7-0.5)	0.95
Pre-Treatment PLR	155 (102.1-641.4)	124.3 (205.8-435.5)	0.8	154.6 (50.0-641.4)	160.7 (55.2-323.3)	0.74	153.2 (50.0-641.4)	161.3 (55.2-535.5)	0.98
Week 5 PLR	337.5 (97.8-1476.7)	342 (70.0-1077.5)	0.25	346 (101.4-1476.7)	333.9 (70.0-1077.5)	0.12	344.3 (101.4-1476.7)	330.5 (70.0-1077.5)	0.12
Change in PLR	172 ??	204 (20.7-543.3)	0.15	179.1 (-139.4-975.8)	133.5 (-93.9-865.8)	0.13	183.7 (-139.4-975.8)	133.3 (-93.9-865.8)	*0.08*
Pre-Treatment NLR	2.5 (1.0-11.1)	2.6 (1.3-6.3)	0.47	2.6 (1.0-11.1)	2.7 (1.9-5.5)	0.93	2.5 (1.0-11.1)	2.8 (1.3-6.3)	0.69
Week 5 NLR	5.4 (1.0-20.0)	5 (0.9-21.4)	0.82	5.5 (1.0-20.0)	5.8 (0.9-21.4)	0.62	5.4 (1.0-20.0)	5.7 (0.9-21.4)	0.62
Change in NLR	3 (-5.3-18.3)	2.3 (-0.9-17.6)	0.99	3 (-5.3-18.3)	2.8 (-2.9-17.6)	0.72	3 (-5.3-18.3)	2.6 (-2.9-17.6)	0.52
Pre-Treatment LMR	3.3 (1.0-12.5)	3.3 (1.3-16.7)	0.17	3.4 (1.0-14.3)	3 (1.3-16.7)	0.25	3.5 (1.0-12.5)	3 (1.3-16.7)	0.24
Week 5 LMR	1.3 (0.4-9.1)	1.1 (0.6-6.25)	0.48	1.25 (0.6-6.25)	1.3 (0.5-6.25)	0.35	1.25 (0.6-6.25)	1.3 (0.5-6.25)	0.34
Change in LMR	2	1.8	0.23	2	2.1	2.1	2	2.1	0.14

	Local Relapse			Distant Relapse			Any Relapse		
	(-0.4-3.0)	(-0.8-5.6)		(-0.4-5.0)	(-0.6-3.0)		(-0.4-5.0)	(-0.6-3.0)	

Where data was available, we also separated patients according to the treatment regimen received (short or long-course). This data can be found in the Appendices (Section 9.11).

From Table 41, we observed that patients who had an incomplete response to therapy had significantly higher platelet count at week 5 (p=0.01). Similarly, patients who had an incomplete response also had higher baseline platelet counts, although this difference was not significant (p=0.07). In agreement with these findings, in patients receiving short-course therapy an increase in platelet count from baseline to week 5 was suggestive of distant disease recurrence (p=0.05) and any disease recurrence (p=0.05) within two years of follow-up. This did not correlate with any significant differences with regards to PLR at any time point or change between timepoints.

This finding has similarities with prior reports that high pre-treatment platelet counts were associated with a decreased rate of pCR, DFS and OS in patients with LARC (Belluco et al. 2018). This study also reported that local recurrence and distant metastases both occurred at increase rates in patients with elevated platelet counts using a cut-off of $300x10^{9/}L$ (Belluco et al. 2018). Although, we saw no association with disease recurrence (Table 42), another study reported that a high pre-treatment platelet count was associated with decreased DFS in patients with rectal cancer (Toiyama et al. 2015). Such studies have primarily investigated pre-treatment platelet levels, thus comparing our findings at week 5 to the literature is difficult.

Although platelet counts were significantly associated with tumour response, we found no significant association between PLR and tumour response or disease recurrence. The literature on the predictive value of PLR in patients with rectal cancer is inconsistent. Li and colleagues (2016) have reported that high pre-operative PLR (>144) was associated with decreased OS and DFS, in patients with non-metastatic rectal cancer, but only in univariate analysis. In contrast, Jung and colleagues (2017) reported that high PLR (>92.88) was associated with improved RFS in both univariate and multivariate analysis. Dudani and colleagues (2019) reported that PLR had no significant association with OS, DFS or pCR in patients with LARC. Therefore, our lack of significant findings when predicting pCR and disease recurrence using this inflammatory ratio was not unexpected.

Here we observed that significantly higher neutrophil count at week 5 was associated with an increased risk of local recurrence (Table 42). As with previous findings, this should be treated with caution, considering the number of factors analysed in this study and the probability of seeing a significant result by chance alone. As far as we are aware, no studies have reported

that neutrophil count alone was associated with response to neoadjuvant therapy or disease recurrence. However, numerous studies have investigated the potential utility of NLRs in patients with LARC.

Although we saw no associations, we would have expected to see that increased NLR was associated with an incomplete response to therapy or an increased likelihood of disease recurrence. Studies have shown that high pre-treatment NLR was associated with poor OS (Ward et al. 2018) and DFS (Braun et al. 2019) in patients with rectal cancer. Similarly, patients with rectal cancer who had higher pre-treatment NLR were significantly less likely to achieve pCR (Jeon et al. 2019) and more likely to experience local relapse (Vallard et al. 2018). In patients with NSCLC treated with an anti-PD-1 antibody, high post-treatment NLR (week 6) was predictive of decreased DFS and OS (Cao et al. 2018). These studies show the potential utility of post-treatment NLR as a predictor of survival in patients with different solid tumours.

Although we saw no direct significance when analysing NLR, there is some agreement between the literature and our association of high neutrophil counts with increased rates of local relapse. However, this assumes that neutrophil count and NLR have similar clinical implications in the context of patients with LARC undergoing neoadjuvant therapy. This may also be underestimating the importance of the lymphocyte count in the clinical utility of NLR.

We found no associations between initial patient response or disease recurrence and LMR in our patient cohort. However, in patients who received long-course therapy, low baseline LMR was suggestive of an increased risk of local disease recurrence ($p=0.05$). Similarly, in patients receiving short-course therapy, high monocyte concentrations at week 5 were also suggestive of an increased risk of local recurrence ($p=0.05$). Previously, Jung and colleagues (2017) reported that high pre-treatment LMR (>6.8) was associated with improved RFS under univariate but not multivariate analysis in patients with rectal cancer. Similarly, Xiao and colleagues (2015) reported that high (>3.78) LMR was associated with improved DFS in patients with rectal cancer. Other studies have also reported that high LMR was associated with improved OS (Zhao et al. 2017; Ward et al. 2018). However, we did not investigate OS in our patient cohort, and thus our findings are not directly comparable. Overall, we saw some potential for LMR as a predictive biomarker in our patients.

Using this clinical dataset, we were able to analyse patient response and disease recurrence in a larger cohort than possible in previous chapters. This analysis has demonstrated that high platelet and neutrophil concentrations at week 5 were associated with significantly improved

response and an increased likelihood of local recurrence. However, other white blood cells and inflammatory indices displayed limited power when predicting these clinical endpoints in patients with LARC.

6.2.2 Pre-Treatment Clinical Characteristics of Patients with LARC

Objective: Can baseline CEA or haemoglobin levels be used to predict response to neoadjuvant therapy in patients with LARC?

6.2.2.1 Baseline Clinicopathological Features

Previously, in Chapters 3 and 5, we analysed concentrations of ctDNA, cfDNA and exoRNA and compared these to patient clinical outcomes, in a small cohort of up to 14 patients with LARC. In this chapter, we have expanded on the previous circulating biomarkers by also investigating the possible predictive power of pre-treatment haemoglobin and CEA levels in the same 14 patients. A reminder of the baseline clinical characteristics and clinical outcomes for these patients and a summary of the pretreatment haemoglobin and CEA levels are described below in Table 43 and Table 44 respectively.

Table 43: Clinicopathological characteristics of patients with locally advanced rectal cancer undergoing circulating biomarker analysis to predict response to radiotherapy plus surgery

Patient ID	Disease Stage	Radiotherapy (Gy in Doses)	Baseline CEA (ng/mL)	Baseline Haemoglobin (g/dl)	TRG	Patient Experienced Local and/or Distant Relapse
WCB1262	II	45 in 25	2	105	1	None
WCB1263	III	45 in 25	4	119	3	None
WCB1268	IV	45 in 25	220	136	3	Distant only
WCB1295	III	45 in 25	2	137	1	None
WCB1419	III	25 in 5	2	142	4	Distant only
WCB1441	II	45 in 25	1	131	4	Distant only
WCB1476	III	45 in 25	0.5	153	2	None
WCB1477	II	45 in 25	1	90	4	Local only
WCB1478	III	25 in 5	41	127	4	n/a
WCB1479	IV	45 in 25	10	139	3	Distant only
WCB1493	II	45 in 25	2	131	1	Local only
WCB1602	III	45 in 25	5	160	1	None
WCB1603	II	45 in 25	12	109	3	Local and distant
WCB1604	III	45 in 25	6	117	2	Distant only

AJCC– Tumour staging according to the American Joint Committee on Cancer (Glynne-Jones et al. 2017)

Table 44: Pre-treatment blood cell counts of all patients with locally advanced rectal cancer undergoing pre-operative radiotherapy

Variable	N	Median	Range
Haemoglobin (10^9/L)	14	131	90-160
CEA (10^9/L)	14	3	0.5-220

As in previous chapters, patient outcomes were separated in a binary fashion into good vs poor response (TRG1/2 vs TRG3/4), complete vs incomplete response (TRG1 vs TRG2-4) and whether patients experienced local and/or distant relapse within two years of follow-up). A statistical comparison of CEA levels to each of these outcomes is performed and described below in Table 45, Table 46 and Table 47.

6.2.3 Circulating Biomarkers for Predicting Response to Therapy in Patients with LARC

Table 45: A comparison of pre-treatment haemoglobin and CEA levels across patients with good and poor responses to neoadjuvant therapy

	Good Response (TRG 1/2)		Poor Response (TRG 3/4)		
Variables	**N**	**Median (Range)**	**N**	**Median (Range)**	**p-value**
Pre-Treatment Levels					
Haemoglobin (10^9/L)	5	138 (105 – 160)	9	127 (90 – 142)	0.12
CEA (10^9/L)	5	2.0 (0.5 – 10.0)	9	4.0 (1.0 – 220.0)	0.35

n/a – ranges or p-values could not be calculated due to limited number of samples

Table 46: A comparison on pre-treatment haemoglobin and CEA levels across patients with complete and incomplete responses to therapy

	Complete Response (TRG 1)		Incomplete Response (TRG 2-4)		
Variables	N	Median	N	Median	P-Value
Pre-Treatment Levels					
Haemoglobin (10^9/L)	4	134 105-160	10	130 90-142	0.55
CEA (10^9/L)	4	2 2.0-5.0	10	4 0.5-220	0.47

n/a – ranges or p-values could not be calculated due to limited number of samples

Table 47: A comparison circulating pre-treatment haemoglobin and CEA levels between different patient responses to neoadjuvant therapy. Any recurrence refers to patients who experience either local or distant relapse.

	Local Relapse			Distant Relapse			Any Relapse		
Variable	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p-Value**	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p-Value**	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p- Value**
Pre-Treatment Levels									
Haemoglobin (10^9/L)	136.5 (105 – 153)	123.5 (90 – 160)	0.35	131 (90 – 160)	131 (109 – 142)	0.9	128 (105 – 153)	131 (90 – 160)	1.0
CEA (10^9/L)	3.0 (0.5 – 220.0)	2.0 (1.0 – 12.0)	0.63	2.0 (0.5 – 5.0)	10.0 (1.0-220.0)	0.17	2.0 (0.5-5.0)	6.0 (1.0-220.0)	0.39

n/a – ranges or p-values could not be calculated due to limited number of samples

This analysis demonstrated that pre-treatment levels of CEA were not significantly associated with response to radiotherapy in our cohort of patients with LARC. This is in agreement with a systematic review which reported that CEA levels had a low sensitivity and positive predictive value when detecting recurrences in patients with CRC (Sorensen et al. 2016). However, reports have also demonstrated that pre-treatment CEA levels were significantly higher in patients with LARC who did not experience a complete response to neoadjuvant therapy (Yang et al. 2019).

We observed no significant associations between haemoglobin levels at any time and patient response to therapy. In contrast, studies have recently demonstrated that high pre-treatment levels of haemoglobin were associated with an improved response rate to neoadjuvant therapy (Formica et al. 2018) and OS (Franco et al. 2018) in patients with LARC and anal cancer respectively. Similarly, Khan and colleagues (2013) reported that low haemoglobin levels were associated with an increased risk of local recurrence in a study of 463 patients with LARC. However, Clarke and colleagues (2017) found only a suggestive association between pre-treatment haemoglobin levels and response to therapy in patients with LARC.

6.3 Discussion

6.3.1 Conclusion

In this study, we looked at how numerous circulating biomarkers could predict patients who achieved a complete response to therapy or patients who experienced disease recurrence within two years of follow-up.

Objective: Can clinical full blood counts and derived immune ratios be used to predict clinical outcomes in response to neoadjuvant therapy in patients with LARC?

We found that increased platelet and neutrophil counts at week 5 were associated with incomplete response and an increased likelihood of local relapse after neoadjuvant therapy (Table 41). This finding remained consistent when patients were separated according to the course of therapy administered.

Our findings agree with prior reports that high platelet counts are indicative of poor prognosis in patients with rectal cancer. However, these studies generally analysed pre-treatment concentrations, whereas we observed a significant association at week 5 and not pre-treatment. This significance was not reflected in PLR values at any time.

High platelet counts have previously been suggested to drive cancer progression through inducing metastatic dissemination through angiogenesis (Wojtukiewicz et al. 2017), and by protecting tumour cells from immune elimination (Gay and Felding-Habermann 2011). Elevated platelet counts have also been identified in patients with rectal cancer, which was found to be associated with poor prognosis and tumour aggression in these patients (Belluco et al. 2018).

High pre-treatment neutrophil counts were found to be a significant predictor for local recurrence within two years of follow-up in our patient cohort (Table 42). However, the literature is limited at this time regarding neutrophil counts only, and largely refers to pre-treatment NLRs. Nevertheless, this finding is still in relative agreement with studies which have shown that high baseline NLR was predictive of poor DFS in patients with rectal cancer (Braun et al. 2019).

Although neutrophil concentration was predictive of disease recurrence, NLR was not found to be significantly associated with any clinical outcome measured in this study. NLRs, among other systemic inflammatory indices, have recently emerged as a circulating biomarker of interest due to our increasing knowledge of the role of the immune system in cancer

development. Neutrophils have recently been recognised as having a pro-tumour influence via their capacity to induce inflammatory and angiogenic pathways, including IL-6, IL-8 and Vascular Endothelial Growth Factor (Terzic et al. 2010). In contrast, lymphocytes are hypothesised to have an anti-tumour effect, due to the beneficial presence that TILs have (Pine et al. 2015).

However, as previously stated, we saw no significant associations between NLR and any of our clinical endpoints. This finding may differ from the literature for several reasons. The cut-offs of high and low NLR varied from 1.7-5.0 in the literature, which can cause inconsistencies when comparing studies and attempting to draw conclusions. Furthermore, NLR values may have been influenced by the administration of neoadjuvant therapy. Kitayama and colleagues (2010) reported that lymphocyte numbers decreased during radiotherapy administration, whereas the levels of neutrophils specifically, were relatively unaffected (Kitayama et al. 2010). Although we would assume that this would impact all week 5 NLR values, the variability of this effect between patients and therapies has not yet been determined. Therefore, the inflammatory ratios measured at week 5 may have been influenced by neoadjuvant therapy and a longer follow-up may be beneficial before NLRs are examined again to allow normalisation of lymphocyte counts. This may also suggest why we observed a significant difference using the neutrophil count alone.

Any association that neutrophil counts have with clinical outcomes may be due to the role that neutrophils may perform in facilitating distant metastasis in patients with cancer (Wculek and Malanchi 2015). Potentially, neutrophil counts could be used to identify high-risk patients who may require additional post-surgical adjuvant therapy, closer follow-up or even avoiding the unnecessary toxicity of surgery if a complete response can be predicted and then confirmed pre-surgery. This data would require validation in a much larger cohort and might then impact how patients are treated or managed during follow-up. Neutrophil counts may also be impacted by external, non-tumour factors, such as bacterial infection (Kitayama et al. 2010).

We did not identify any significant associations between pre-treatment inflammatory ratios and patient outcomes to neoadjuvant therapy in our main patient cohort. When patients were divided according to treatment received, patients on long-course chemoradiotherapy, low baseline LMR was suggestive of an increased risk of local disease recurrence. Similarly, high monocytes at week 5 were also suggestive of an increased likelihood of local disease recurrence. As previously mentioned, there are some conflicting results in the literature and no significance was seen in our main cohort. Thus, this finding should not necessarily be taken

at face value and further analysis will be necessary to confirm the predictive power of LMR in this cohort.

In recent years, PLRs and LMRs have been investigated either alongside or in place of other inflammatory markers (such as NLR) for clinical utility in patients with rectal cancer. This has been coupled with an increased understanding of the role of platelets and monocytes in cancer progression.

As previously stated, the literature regarding PLR in patients with rectal cancers are inconsistent. This has created difficulties when assessing whether our findings agree with previous studies. This might be the result of numerous factors. Before inflammatory ratios could be applied in the clinical setting, more uniformity and standardisation will be necessary. For example, although considerable research has been performed regarding NLR, there is currently no standard cut-off value to distinguish between high and low values in patients with LARC. Previous studies have reported a wide range of cut-offs ranging from 1.8-5. Most of these studies chose their respective cut-off values based upon Receiver Operating Characteristic (ROC) Curve and/or Area Under Curve (AUC) analysis. With such a wide range being applied for inflammatory marker analyses, proper standardisation is required before clinical utility or testing can be considered. This factor remains consistent in both PLR and LMR studies.

Other non-cancer factors may also affect immune-derived ratios, such as the impact of neoadjuvant therapy or infection. The consistency and reproducibility of PLR, NLR or LMR values over hours or days is also unknown and may require further investigation.

As the full blood count data was obtained as part of routine clinical management, we can have reasonable confidence in the accuracy and reproducibility of each circulating biomarker. Furthermore, the size and breadth of data available within the large dataset allows us to perform more powerful statistical analysis than we could in previous chapters. Nevertheless, this analysis will still have inherent limitations which must be considered. No multivariate analysis or statistical corrections were performed. Therefore, this analysis did not incorporate additional relevant factors, such as whether patients received concurrent chemotherapy, types of surgery or any post-surgical interventions or the number of factors analysed in total.

Finally, this dataset would have benefited from the inclusion of other circulating biomarkers of interest which are included in standard clinical analysis, such as baseline and sequential haemoglobin and/or CEA.

Our lack of significant findings in immune-derived ratios suggest that these inflammatory indices may not be entirely appropriate for patients with LARC. This has been noted previously, where studies have demonstrated that ctDNA is more readily detectable in patients with metastatic disease, rather than locally advanced cancer. As mentioned previously, monocytes, platelets and neutrophils can all play a key role in facilitating the development of tumour metastases. Therefore, these circulating biomarkers may be more clinically significant in patients with metastatic tumours, than those analysed as part of this study.

This may be a consequence of the tumour biology and relatively early stage of these patients as neutrophils, platelets and monocytes have all been implicated in facilitating tumour metastasis. Therefore, the non-metastatic nature of this patient cohort may limit the effectiveness of these circulating biomarkers. However, there may be important scope for detecting patients which are more likely to develop or have already developed undetected micro-metastatic lesions. As we did not investigate this clinical outcome, we cannot comment on this.

Furthermore, using systemic inflammatory indices relies on the assumption that lymphocytes only fulfil an anti-tumour function. However, studies have illuminated the role of $CD4^+$ regulatory T-cells which dampen the immune function of $CD8^+$ cytotoxic T-lymphocytes (Ramsay 2013). Therefore, such ratios might be limited by the inclusion of $CD4^+$ and $CD8^+$ T-cells in the lymphocyte count.

As this study progresses, we begin to suspect that this subgroup of patients, may not shed tumour-derived biomarkers into the circulation at a readily detectable rate. This could contribute to our lack of significant findings in this study. However, we have evidence to suggest that high platelet and neutrophil counts at week 5 may be predictors of incomplete response to therapy and local disease recurrence respectively in patients with LARC.

Objective: Can baseline CEA or haemoglobin levels be used to predict response to neoadjuvant therapy in patients with LARC?

We expected to see that high baseline CEA concentrations were associated with poorer response to neoadjuvant therapy, as this has been previously reported in patients with rectal cancer. Huh and colleagues (2018) reported that post-treatment CEA levels could be significantly elevated in patients with rectal cancer who had a poor response to CRT. This study also demonstrated that elevated post-treatment CEA levels were indicative of decreased 5-

year OS. Similarly, Jeong and colleagues (2016) reported that high post-operative CEA was predictive of distant recurrence and was significantly associated with shortened RFS.

We observed no such association at this time. This may be because this patient cohort is only locally advanced and, therefore, may secrete less CEA. Alternatively, this may suggest that CEA may have limitations as a circulating biomarker for this cohort. Finally, our lack of significant findings may be a result of low sample numbers limiting statistical analysis.

Overall, we were unable to identify any statistically significant associations between pre-treatment haemoglobin levels and patient response to neoadjuvant therapy. However, there were different findings reported on the potential predictive value of baseline haemoglobin in patients with rectal cancer (Khan et al. 2013; Clarke et al. 2017). This may have been a consequence of variable cut-offs defining high and low levels (12-14 g/dL). Furthermore, clinical endpoints, though similar, were not identical between studies, and patient treatment will have some variation between different treatment centres (including our own). Overall, these confounding factors may have contributed to our lack of significant associations between haemoglobin levels and patient response to therapy.

Both CEA and haemoglobin analyses were limited as they only included pre-treatment samples from a small number of patients. This study would benefit from increasing this to include sequential analysis in a larger cohort, such as that investigated in Section 6.2.1. However, these two circulating biomarkers will be relatively reliable as they would have been collected to high standards as part of routine clinical management. Therefore, there are limited concerns regarding pre-analytical handling or sample quality.

As the remaining circulating biomarkers provided limited predictive value for patient response to therapy, we again question whether our chosen circulating biomarkers are appropriate for clinical analysis in patients with LARC. Alternatively, these findings may be attributed to limitations with the assays applied, such as choice of methods, statistical analysis or technological sensitivity.

6.3.2 Future Studies

This study would benefit from the inclusion of pre-treatment and post-treatment CEA and haemoglobin concentrations in the large dataset analysed in Section 6.2.1. This would allow further exploration into these routine clinical biomarkers in line with that previously performed for NLR etc.

This study would also benefit from longer follow-up (>3 years) to analyse longer-term outcomes, such as PFS and OS. The circulating biomarkers may have greater predictive power for longer-term outcomes than initial response.

As stated in previous chapters, greater volumes of plasma for cfDNA, ctDNA and exoRNA analysis could help improve the sensitivity of circulating biomarker analyses. A larger number of patients would also help to identify any statistically significant associations with patient outcomes. Expanding the number of microRNAs in exoRNA analysis alongside numerous reference genes may also help overcome some of the limitations highlighted previously for this biomarker.

7 General Discussion and Future Research

7.1 Summary of Key Findings

Recent advancements in surgical techniques and the administration of pre-operative radiotherapy have improved local control in patients with locally advanced rectal cancer (LARC) (Peeters et al. 2007). Despite this improvement, many patients will still experience distant disease recurrence and disease-related mortality (Maas et al. 2010).

At this time, the only clinical factors which can predict long-term clinical outcomes are achieving tumour-downstaging or pCR to therapy (Maas et al. 2010). There are no clinical or molecular markers that can predict how patients will respond to pre-operative chemoradiotherapy pCRT.

As part of this thesis, our investigation has focussed on the potential utility of circulating biomarkers in predicting and/or monitoring response to neoadjuvant therapy in patients with LARC. Research into circulating biomarkers has often focussed upon patients with metastatic disease, because these patients have more advanced disease or metastases and because outcomes for these patients are generally very poor. This increases the likelihood of systemic and circulating biomarkers being detected and clinically relevant. These patients also have targeted treatments available for them, such as cetuximab, and thus have known resistance associated variants to detect and monitor in ctDNA (Reinert et al. 2015).

In the first instance, our aim was to select a clinically relevant and economically efficient NGS panel for the detection of tumour-specific variants in both tumour tissue and ctDNA. Our choice of panel appeared to be appropriate, given the high rate of detection for clinically significant variants in tumour tissue from patients with LARC. However, this did not translate into ctDNA due to several limitations. We also determined a limit of detection threshold for both NGS and ddPCR to be ~1%, although ddPCR was demonstrated to be dependent on input DNA concentration.

We proceeded to determine whether ctDNA and cfDNA could be reliably detected in patients with LARC. Overall, we were able to detect ctDNA at any time using ddPCR in 2/12 (16.7%) and 3/18 (16.7%) patients in two separate cohort from Chapters 3 and 4 respectively. In Chapter 3, our cohort consisted of 16 patients with LARC who were consented at the Velindre Cancer Centre by the WCB. Of these 16 patients, 12 had *KRAS* variants previously detected which

could be used for ctDNA analysis in longitudinal plasma samples. This sensitivity for detecting ctDNA was lower than expected according to previous reports.

We also observed higher total cfDNA concentrations from plasma samples collected during compared to those collected before or after. We hypothesised that this was a result of elevated tumour and non-tumour cell death, causing increased levels of cfDNA release into the circulation. However, this finding was not consistent when analysed in a similar patient cohort in Chapter 4.

In chapter 4, we expanded our ctDNA analysis further using samples from a cohort of patients with LARC acquired from the ARISTOTLE clinical trial. We also aimed to investigate the ability of our targeted NGS panel to detect variants and analyse clonal changes in tumour tissue occurring as a result of neoadjuvant therapy. We successfully detected tumour-specific variants in rectal tumour tissue from 18/19 (94.7%) patients with LARC.

Using this NGS data, we calculated heterogeneity scores based upon tumour cellularity and calculated VAFs as previously described by Normanno and colleagues (2015). We used these variant heterogeneity scores as a proxy marker for individual tumour clones and were also able to observe some clonal changes between pre- and post-treatment tumour tissue. We speculated that observed clonal changes in tumours after neoadjuvant therapy may be indicative of patient response.

Continuing from chapters 3 and 4, we continued our investigation to include exoRNA as a potential circulating biomarker which, unlike ctDNA, is still in its relative translational infancy. Therefore, in chapter 5, we needed to develop a technique for the extraction and analysis of exoRNA in patients with LARC. For this chapter, we used plasma samples from the same patients used previously in chapter 3, who were consented by the WCB. We successfully developed and optimised an analytical process for the extraction and analysis of exoRNA from this patient cohort. This included choosing the optimal technology for exosome isolation, a key component to determine as this can impact sample purity during downstream analysis. This also required the optimisation of exoRNA extraction and appropriate reverse transcription, to ensure resulting microRNAs were reliably detected during ddPCR analysis.

Three microRNAs of interest (Mir-31, -99a* and -125b) and a reference gene (RNU6B) were chosen for sequential analysis based on a review of the literature. We were able to confidently detect each of these microRNAs and the reference gene at variable levels in sequential plasma samples from our patients using ddPCR. We also investigated the more controversial subject of

exosome-derived DNA (exoDNA). Although we did not detect any exoDNA in our patients, there are further avenues to explore and the presence of exoDNA may not be as straight-forward as exoRNA.

As part of Chapter 6, we investigated the utility of components of clinical full blood counts and related inflammatory ratios to predict patient clinical outcomes in 235 patients with LARC. These included neutrophil, lymphocyte, platelet, and monocyte counts, as well as NLR, PLR and LMR. We demonstrated that patients who achieved a complete response to therapy had significantly lower platelet counts at week 5 of treatment, particularly in patients who received long-course neoadjuvant therapy. Furthermore, a high neutrophil count at week 5 was also significantly associated with disease recurrence in these patients.

In addition, we also compared CEA and haemoglobin concentrations alongside previously quantified circulating biomarkers from chapters 3 (ctDNA and cfDNA) and 5 (exoRNA) to patient clinical outcomes in our prior 14 patient cohort. We observed that patients who had a complete response to treatment had significantly higher cfDNA concentrations both during and after therapy. We did not identify any other significant associations between circulating biomarkers and patient outcomes.

7.1.1 Limitations of our Findings

Although we were able to make some significant achievements throughout this thesis, our findings have inherent limitations. With regards to our ctDNA analysis in Chapter 3 (using patients from the WCB), our detection rate was low compared to expectations based on the literature (Tie et al. 2018). We have speculated that this may be a result of low plasma volumes leading to low cfDNA concentrations. This can then detrimentally impact the sensitivity of ddPCR and NGS analysis as demonstrated in the Materials and Methods, suggesting a cause for our low ctDNA detection rate.

When comparing these findings to patient clinical outcomes, we have no significant associations between ctDNA levels and patient response to therapy. This was unsurprising due to the low number of ctDNA positive timepoints identified in this analysis. However, = we did observe that patients who achieved a complete response to therapy had significantly higher total cfDNA concentrations. This may have been due to increased DNA shedding as a result of greater levels of tumour cell death after therapy.

Our detection rate for ctDNA in Chapter 4 was also still below expectations. Furthermore, although we identified some interesting findings in variant heterogeneity scores, we did not

have access to any clinical information from the ARISTOTLE trial with which we could compare our findings, these results will become available in late 2020. Therefore, at this time, we cannot discern any clinically significant results until such data becomes available.

In Chapter 5, although we were able to confidently detect our genes of interest in exoRNA, we did not observe any significant associations between sequential exoRNA levels or sequential changes therein to any clinical characteristics in these patients. This may be a consequence of our choice of microRNAs and reference gene, the non-metastatic status of our patients or the relatively small size of our cohort. When comparing longitudinal exoRNA levels to clinical ooutcomes, we found no significant associations in this cohort of patients.

In Chapter 6, we identified platelet and neutrophil counts as significant predictors of response to therapy and disease recurrence within two years of follow-up in 235 patients with LARC. There was a possibility that this significance was seen by chance due to the large number of factors analysed. Furthermore, each factor was analysed in univariate, therefore significance may have been lost during multivariate analysis. Each value here was analysed as a continuous variable and not as high or low based on a designated threshold. This was because we did not identify a common threshold from the literature, nor from AUC/ROC analysis. Furthermore, we only investigated these factors in relation to the recurrence of the disease and not directly against PFS or OS.

7.2 Application of our Findings for Future Clinical Implementation

7.2.1 Molecular Characterisation of Tumour Tissue in Patients with Locally Advanced Rectal Cancer

The detection of tumour-specific variants is critically important across various solid tumour types. In patients with metastatic CRC, the detection of variants in RAS are associated with resistance to anti-EGFR monoclonal antibodies, such as cetuximab or panitumumab (Misale et al. 2012). There are currently no molecular biomarkers which can predict patient response to chemo- or radiotherapy, which is the current standard treatment for patients with LARC (Glynne-Jones et al. 2017).

In our cohort, variants were frequently detected in both *KRAS* and *TP53*, both of which have been previously hypothesised to impact patient response to neoadjuvant therapy in patients with LARC. However, we could not compare these molecular findings to clinical outcomes in this cohort as no clinical data were available (see chapter 4).

Once molecular markers which can predict response to neoadjuvant therapy have been identified, it will be vital to ensure that current methods (NGS and ddPCR) can be applied as part of a reliable and robust clinical assay.

We have demonstrated that the 50-gene CHPv2 can reliably detect variants in a wide range of genes (including *TP53*) and hotspot mutations in genes such as *KRAS*. Whether these variants are relevant in the context of neoadjuvant therapy remains to be confirmed. Furthermore, this panel focusses on mutational hotspots in genes of interest in CRC, capturing most clinically relevant variants. However, important variants outside of these hotspots or within other genes not included in this panel may be missed during sequencing. These may include genes involved in p53- or non-p53-mediated apoptosis. Such variants could influence response to neoadjuvant therapy in this patient cohort.

Predicting response to neoadjuvant therapy may not be possible using tumour-specific variants alone. The CRC Subtyping Committee (CRCSC) have recently developed a novel subtyping system referred to as the Consensus Molecular Subtypes (CMS) which are separated into groups 1-4 (Guinney et al. 2015). These subtypes are based on the gene expression profiling of >4000 patients with CRC. Similarly, Bertotti and colleagues (2016) have also developed a novel subtyping system referred to as ColoRectal cancer Intrinsic Subtypes (CRIS), which is also based upon gene expression profiling in patients with CRC. The recent emergence of these novel subtyping systems has increased our understanding in how variable expression profiles can be between different CRC subgroups and, in some cases, which genes or pathways are frequently altered. Therefore, using these subtyping systems, or similar gene expression profiling methods, we may be able to identify expression alterations in genes or pathways which can predict response to neoadjuvant therapy. Thus, the presence of specific variants alone may not be significant enough to achieve this.

Although these studies can provide valuable information regarding patients with colon cancer, their potential value for patients with rectal cancer specifically may be more limited. For the ~2,500 patients with known primary tumours sites, only ~15% were located in the rectum in the cohort analysed by the CRCSC (Guinney et al. 2015). As part of this analysis, the CRCSC initially compared RNA sequencing data between colon and rectal tumours which was previously published by TCGA Network. This study reported no significant differences in the RNA sequencing data between the two primary sites, suggesting that these subtypes may be applied to patients with rectal cancer. Further testing in a larger number of patients with rectal cancer, is required to verify this.

In contrast, the proportion of patients who had primary rectal tumours in the study from Bertotti and colleagues (Bertotti et al. 2016) was not disclosed. Therefore, the clinical value of this subtyping system in patients with rectal cancer will require further validation in the future.

Although our findings in relation to the literature may provide some interesting avenues for future investigations, the analytical context of our study must be considered. In this thesis, we had access to both pre-treatment tumour biopsies and post-treatment surgical resections for molecular analysis. Whether this is feasible in the clinical setting will remain to be seen. The utility of both forms of tissue (pre- and post-treatment) may reveal more information regarding the molecular changes occurring in the tumour as a result of therapy. This combined information may prove to be more valuable than the molecular data gathered from either pre- or post-treatment tissue in isolation.

A major concern for the molecular characterisation of LARC tumour is that each tissue sample will only represent a snapshot of the tumour as a whole and is not necessarily representative of the entire tumour. Nor is this information representative of distant metastases which can arise despite the implementation of neoadjuvant therapy. Therefore, in the clinical setting, the detrimental impact of both inter- and intra-tumour heterogeneity should be considered in tandem with genetic testing as part of treatment-decision making.

7.2.2 Sequential Analysis of Circulating Biomarkers in Patients with LARC

Of the circulating biomarkers investigated within this thesis, ctDNA is the only biomarker currently being used in clinical practise to monitor patient response to therapy. In comparison, the clinical potential of exosomes is still in its relative infancy but with growing interest. Although studies have shown that exosomes are generally increased in patients with cancer, there is considerable inconsistency regarding the molecular cargo of circulating exosomes between studies, particularly regarding microRNA (Silva et al. 2012). This inconsistency also exists in studies investigating circulating microRNAs in patients with cancer and is not specific to exosome-derived microRNAs alone. With studies having demonstrated that most circulating microRNAs are exosome-derived, these two issues are certainly linked.

The inconsistency of microRNAs of interest reported between studies is a cause for concern regarding future clinical implementation. Although we chose three microRNAs of interest which were previously reported in patients with LARC, there was a wide array from which to choose in the literature and studies frequently contradicted one another. These inconsistencies may exist for several clinical or technical reasons. The clinical variability may

arise from many factors, including the different clinical endpoints used during analysis, variable treatment being applied and inconsistent plasma sampling timepoints between studies. With regards to technical variables, these include methods for exosome isolation, RNA extraction and microRNA analysis and expression normalisation. The combinations of these factors were certain to create inconsistencies in the literature. Furthermore, the analysis of circulating biomarkers such as ctDNA and circulating exosomes can also be influenced by the individual patient's tumour, specifically how active the tumour is and how much material the tumour sheds into each patients' circulation. This is another major factor which can detriment circulating biomarker analysis in non-metastatic disease.

Although white cell counts are currently used in clinical practise, their combination as systemic inflammatory indices is not yet regularly applied. Studies tend to agree that systemic inflammatory indices (such as NLR, PLR and LMR) are useful prognostic indicators but cut-offs remain highly variable in the literature and require standardisation. As white cell counts are widely used, such markers will be relatively easy to integrate into routine clinical practice. Furthermore, as the methods for white cell counting have been used for an extensive period, both the technical and analytical validity of these assays has already been assessed.

CtDNA is currently being used to experimentally to detect emerging treatment resistance in patients with metastatic CRC undergoing anti-EGFR monoclonal antibody therapy. The presence or quantity of ctDNA (regarding *KRAS* or related variants) has been used to identify tumour resistance earlier than current methods allow. However, ctDNA is yet to be applied in the context of neoadjuvant therapy, likely because there are no resistance-associated variants to specifically analyse. The presence of detectable ctDNA in post-surgical plasma samples in patients undergoing neoadjuvant therapy with surgery was demonstrated to be a poor prognostic marker for disease-free and overall survival (Tie et al. 2018). This is consistent with our findings in Chapter 6, where patients with detectable ctDNA after chemoradiotherapy experienced a faster time to recurrence.

7.2.3 Technical Evaluation

NGS platforms have the potential to screen large regions of the genome in order to identify variants or expression profiles of interest. However, prior results have demonstrated that this technology can be prone to generating sequencing artefacts. Furthermore, the bioinformatic analysis of NGS data can be highly variable between different studies and research groups, both for variant calling and the removal of such artefacts. The calling of variant pathogenicity is also variable between studies. Although freely available online predictive tools are often used

for this aspect, different tools can provide variable results. Guidelines to standardise variant pathogenicity calling have been suggested by the ACMG-AMP which combat this issue. This lack of consensus in the literature regarding both sequencing technologies and bioinformatic analyses are hurdles which must be overcome as this becomes more common practise in the clinical setting.

ddPCR has been frequently applied for the detection of low-level variants in both ctDNA and tumour tissue. Here we demonstrated that the sensitivity of variant detection by ddPCR appeared to be dependent on input DNA concentrations, with more DNA allowing improved sensitivity for mutant detection. In the materials and methods chapter of this thesis (Chapter 2), we reported that variants at 1% VAF could be detected using 10ng of DNA, but this sensitivity was reduced when using 1ng of DNA. This is critical for the analysis of ctDNA from patients with low volume disease such as LARC. In this thesis, cfDNA concentrations from patients with LARC were very low, which caused numerous sensitivity and detection issues during ddPCR analysis. Here we applied a cut-off of five molecules and 1% VAF for confident variant detection in ctDNA. This cut-off value was variable in the literature and, therefore, requires standardisation, both in the literature and in the clinical setting. As ddPCR sensitivity is highly variable and dependant on input DNA, variable cut-offs for confident detection may be necessary between different input DNA concentrations or total DNA molecules detected during analysis. As this field continues to grow rapidly, such factors will have to be addressed.

For exosome isolation and exoRNA extraction and analysis, a much greater range of commercial off-the-shelf kits used in the literature. This creates difficulties with the analysis or review of other studies, the selection of methods for research and the standardisation of the process with clinical implementation in mind. Exosome isolation kits have been shown, both in this thesis and in the wider literature, to isolate extracellular vesicles of variable size and purity (Van Deun et al. 2014; Tang et al. 2017). This brings into question whether these studies are truly analysing the same microparticles. Although we did not investigate this as part of this thesis, studies have also demonstrated that RNA extraction methods can extract different RNA species preferentially, which influences down-stream molecular analysis (Tang et al. 2017). The same possibly applies to different off-the-shelf reverse transcription kits. These factors could have compromised the microRNA concentrations and gene expression levels detected by ddPCR in this thesis.

Overall, different commercial kits are available for each step between exosome isolation and microRNA reverse transcription, each of which can cause variability and impact analytical

findings. Much further work will be required to identify why differences exist between certain methods and which methods would be most appropriate for future clinical implementation. This will require large retrospective patient studies once the process has been standardised.

7.3 Guidelines for the Clinical Implementation of Genetic Testing

The transfer of genetic tests from research to clinical practise is defined as stage three translational research (Unim et al. 2019). In this context, translational research is defined as research which 'attempts to move evidence-based guidelines into health practice' (Unim et al. 2019). This can be influenced by factors inherent to the research, the delivery of healthcare and external commercial factors. Consequently, models or frameworks have recently been introduced with the purpose of facilitating the translation of genetic tests from research into the clinic.

Most of these models were designed for genetic tests which identify germ-line variants which can predict or predispose a patient to a congenital disorder. In contrast with the research performed in this thesis, either specific models must be generated (of which we could find none at this time) or the current available models must be adapted for purpose. This is because these models include the concept of germ-line variant screening in a population, which is not directly applicable to variants for predicting response to neoadjuvant therapy in patients with LARC.

The most frequently referenced model is the Analytic validity, Clinical validity, Clinical utility and ethical, legal and social implications (referred to as the Rapid-ACCE model) (Gudgeon et al. 2007). Other frameworks have also included these factors in their respective frameworks (Pitini et al. 2018).

The Rapid-ACCE model is a list of 44 questions aimed at creating a comprehensive review of genetic testing (Gudgeon et al. 2007). These questions aim to address analytical and clinical validity of a genetic test, as well as the clinical utility and ethical, legal and social implications of the test.

Analytical validity is defined as the ability of a test to accurately and reliably measure the genotype in question (Pitini et al. 2018). This would also include the sensitivity and specificity of an assay, i.e. how often is a test positive when the variant is present, and negative when absent. The Rapid-ACCE also questions the quality control and precision of said assay as well as robustness across multiple laboratories (Gudgeon et al. 2007). In the context of patients with

cancer, proving an assay is precise, accurate, robust and reliable will be the first essential condition for analytical validity.

Clinical validity is defined as the ability of a test or assay to accurately and reliably detect or predict a clinical condition (Pitini et al. 2018). Clinical validity also refers to the sensitivity and specificity of an assay. In this case, how often does a positive variant represent the presence of a disorder, and how often is a negative result representative of disease absence. The Rapid-ACCE also questions the positive (PPV) and negative predictive values (NPV) of the assay in this setting. In the context of patients with LARC, we are attempting to identify patients who are more or less likely to respond to neoadjuvant therapy (Gudgeon et al. 2007). Therefore, clinical validity require verification as to what proportion of patients who test positive respond or are resistant to neoadjuvant therapy.

Clinical utility compares the risks and benefits of testing and provides clinical usefulness for the integration of the test/assay (Pitini et al. 2018). This factor relates to the clinical implications of a positive or negative result on patient care and the availability of treatment in either case. This is also an economic evaluation, comparing the cost and consequences of various tests. The Rapid-ACCE also questions what other tests are available for this condition, whether quality assurance measures are in place for patients who receive a positive result, availability of methods for long-term monitoring and the economic benefits resulting from the result patient testing (Gudgeon et al. 2007). In the context of patients with LARC, the benefits of identifying patients who would not respond to therapy could improve quality of life by avoiding unnecessary exposure to cytotoxic agents. Alternatively, it may suggest that a more aggressive approach to treatment would be more appropriate. In either case it would suggest a change in treatment strategy, although which alternative strategies would provide better outcomes remains to be seen.

Ethical, legal and social implications are less appropriate in the context of non-hereditary genetic tests as this refers to the presence of any impediments (such as stigmatisms or discrimination) resulting from a positive diagnosis (Pitini et al. 2018). The Rapid-ACCE also questions whether effective safeguards are properly in-place to avoid such impediments causing issues with patients in the future (Gudgeon et al. 2007).

Although there are models and frameworks in place to appropriately facilitate the clinical implementation of genetic tests, our findings are far from this phase of research. A larger number of studies, with larger patient cohorts collecting ‘big data’ will be required to

thoroughly investigate the molecular characteristics of LARC tumours and how these influence patient response to therapy and/or long-term clinical outcomes such as survival. Any findings will then have to be validated in larger patient cohorts as, possibly as part of a retrospective study.

With regards to exoRNA analysis, our study thus far has largely been proof-of-principle to verify that such biomarkers can be analysed in these patients. These findings will require further investigation to identify clinically significant signatures or sequential changes in patients with LARC before the test can be validated using a larger cohort.

7.4 Concluding Remarks and Future Directions

At present, there are no biomarkers available to predict response to neoadjuvant therapy in patients with LARC, and the only current marker for survival in these patients is response to therapy. The ability to predict which patients will respond to treatment before the initiation of therapy will help limit patients who undergo the unnecessary toxicity of chemoradiotherapy and potentially optimise the selection of patients for non-surgical approaches.

There is evidence to suggest that variants in *TP53, KRAS* and/or *PIK3CA* may impact response to neoadjuvant therapy in patients with LARC (Duldulao et al. 2013; Sakai et al. 2014; Chow et al. 2016). However, these could simply be general prognostic factors. With the collection of pre- and post-treatment tissue in large scale retrospective studies, such as ARISTOTLE, we can begin to perform deeper molecular analysis to identify any variants or signatures of interest in this context.

More research needs to be targeted towards this group of patients who have been underrepresented in the literature. Once a variant or molecular signature has been determined which can predict response to radiation-based therapy, huge strides can be made in treatment-stratification for these patients, rather than the one-size-fits-all method currently applied.

Once a tissue-based marker has been identified, resistance-associated variants or signatures can be analysed in circulating biomarkers, either via ctDNA or exoRNA respectively. Studies have previously demonstrated that ctDNA can identify the presence of minimal residual disease and/or new tumour growth more swiftly than current methods allow. Even without the knowledge of resistance associated signatures, known tumour-specific variants can still allow the monitoring of tumour activity/growth after therapy. Greater knowledge about

tumour resistance signatures can allow an expanded repertoire for variants to monitor in ctDNA after surgery.

A greater understanding of exosome isolation and exoRNA extraction methods and their respective strengths and weaknesses will be required before strides can be taken towards clinical implementation. Although we have shown, in this thesis, that exoRNA can be analysed under clinical conditions, much more basic research is required to standardise, optimise and further validate the most appropriate methods for such analysis. Once this has been accomplished, translational work can then proceed to identify exoRNA or exoDNA signatures which can be used to predict outcomes or response to neoadjuvant therapy in patients with LARC.

Each of these assays will require both technical and clinical validation in the future, to ensure the analytical results are accurate, robust and sensitive enough to be applied for clinical utility in non-metastatic patients. The sensitivity, specificity, positive- and negative predictive values of each assay would then be ascertained to ensure findings were clinically relevant and economically viable. Large scale retrospective studies will be required to achieve this.

Our findings have suggested that patients with LARC might receive a limited benefit from the clinical implementation of circulating biomarkers, due to the non-metastatic nature of their disease. CtDNA was only detected in patients who had distant metastases before the initiation of treatment, which is consistent with previous reports. If most of these patients demonstrated limited shedding of molecular components into the circulation, this would contribute to the small number of significant findings achieved in this thesis. This would also detrimentally affect the clinical relevance of circulating biomarkers in this population. However, circulating biomarkers may provide valuable information on the presence of undetected micro-metastases or lymph-node invasion which can then benefit treatment decision-making. For the remaining patients, there is a possibility that tumours do not shed enough material to be detected with our current methods and at present, may not benefit from the clinical implementation of circulating biomarker analysis.

Although our findings here are mainly proof-of-concept, we have demonstrated that, in a subgroup of patients with LARC, ctDNA can be detected and monitored in sequential plasma samples. We have demonstrated that we can isolate circulating exosomes and extract and analyse exoRNA under restrictive conditions in these same patients. More work is still required

before considering clinical implementation, from technical standardisation to retrospective translational patient studies.

8 References

Abdul-Jalil, K. I. et al. 2014. The frequencies and clinical implications of mutations in 33 kinase-related genes in locally advanced rectal cancer: a pilot study. *Ann Surg Oncol* 21(8), pp. 2642-2649. doi: 10.1245/s10434-014-3658-x

Abraha, I. et al. 2018. Preoperative radiotherapy and curative surgery for the management of localised rectal carcinoma. *Cochrane Database Syst Rev* 10, p. Cd002102. doi: 10.1002/14651858.CD002102.pub3

Agostini, M. et al. 2011. Circulating cell-free DNA: a promising marker of pathologic tumor response in rectal cancer patients receiving preoperative chemoradiotherapy. *Ann Surg Oncol* 18(9), pp. 2461-2468. doi: 10.1245/s10434-011-1638-y

Akiyoshi, T. et al. 2012. Predicting the response to preoperative radiation or chemoradiation by a microarray analysis of the gene expression profiles in rectal cancer. *Surg Today* 42(8), pp. 713-719. doi: 10.1007/s00595-012-0223-8

Albuquerque, C. et al. 2002. The 'just-right' signaling model: APC somatic mutations are selected based on a specific level of activation of the beta-catenin signaling cascade. *Hum Mol Genet* 11(13), pp. 1549-1560.

Alderdice, M. et al. 2017. Natural killer-like signature observed post therapy in locally advanced rectal cancer is a determinant of pathological response and improved survival. *Mod Pathol* 30(9), pp. 1287-1298. doi: 10.1038/modpathol.2017.47

Allenson, K. et al. 2017. High prevalence of mutant KRAS in circulating exosome-derived DNA from early-stage pancreatic cancer patients. *Ann Oncol* 28(4), pp. 741-747. doi: 10.1093/annonc/mdx004

Allgauer, M. et al. 2018. Implementing tumor mutational burden (TMB) analysis in routine diagnostics-a primer for molecular pathologists and clinicians. *Transl Lung Cancer Res* 7(6), pp. 703-715. doi: 10.21037/tlcr.2018.08.14

Aloisio, M. et al. 2016. A technical application of quantitative next generation sequencing for chimerism evaluation. *Mol Med Rep* 14(4), pp. 2967-2974. doi: 10.3892/mmr.2016.5593

Ambrosio, M. R. et al. 2019. Putative Role of Circulating Human Papillomavirus DNA in the Development of Primary Squamous Cell Carcinoma of the Middle Rectum: A Case Report. *Front Oncol* 9, p. 93. doi: 10.3389/fonc.2019.00093

Ameres, S. L. and Zamore, P. D. 2013. Diversifying microRNA sequence and function. *Nat Rev Mol Cell Biol* 14(8), pp. 475-488. doi: 10.1038/nrm3611

Arai, T. et al. 2018. Regulation of NCAPG by miR-99a-3p (passenger strand) inhibits cancer cell aggressiveness and is involved in CRPC. *Cancer Med* 7(5), pp. 1988-2002. doi: 10.1002/cam4.1455

Azizian, A. et al. 2015. Preoperative Prediction of Lymph Node Status by Circulating Mir-18b and Mir-20a During Chemoradiotherapy in Patients with Rectal Cancer. *World J Surg* 39(9), pp. 2329-2335. doi: 10.1007/s00268-015-3083-8

Banzhaf-Strathmann, J. and Edbauer, D. 2014. Good guy or bad guy: the opposing roles of microRNA 125b in cancer. *Cell Commun Signal* 12, p. 30. doi: 10.1186/1478-811x-12-30

Bartel, D. P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116(2), pp. 281-297.

Beije, N. et al. 2016. Somatic mutation detection using various targeted detection assays in paired samples of circulating tumor DNA, primary tumor and metastases from patients undergoing resection of colorectal liver metastases. *Mol Oncol* 10(10), pp. 1575-1584. doi: 10.1016/j.molonc.2016.10.001

Belluco, C. et al. 2018. Elevated platelet count is a negative predictive and prognostic marker in locally advanced rectal cancer undergoing neoadjuvant chemoradiation: a retrospective multi-institutional study on 965 patients. *BMC Cancer* 18(1), p. 1094. doi: 10.1186/s12885-018-5022-1

Bentley, D. R. et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456(7218), pp. 53-59. doi: 10.1038/nature07517

Bernard, V. et al. 2019. Circulating Nucleic Acids Are Associated With Outcomes of Patients With Pancreatic Cancer. *Gastroenterology* 156(1), pp. 108-118.e104. doi: 10.1053/j.gastro.2018.09.022

Bertini, I. et al. 2012. Metabolomic NMR fingerprinting to identify and predict survival of patients with metastatic colorectal cancer. *Cancer Res* 72(1), pp. 356-364. doi: 10.1158/0008-5472.can-11-1543

Bertotti, A. et al. 2016. Abstract 107: Unsupervised analysis of cancer-cell intrinsic transcriptional traits defines a new classification system for colorectal cancer with improved predictive and prognostic value. doi: 10.1158/1538-7445.AM2016-107

Bettegowda, C. et al. 2014. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* 6(224), p. 224ra224. doi: 10.1126/scitranslmed.3007094

Bork, U. et al. 2015. Circulating tumour cells and outcome in non-metastatic colorectal cancer: a prospective study. *Br J Cancer* 112(8), pp. 1306-1313. doi: 10.1038/bjc.2015.88

Bottarelli, L. et al. 2018. Potential predictive biomarkers in locally advanced rectal cancer treated with preoperative chemo-radiotherapy. *Acta Biomed* 89(9-s), pp. 102-106. doi: 10.23750/abm.v89i9-S.7881

Boysen, A. K. et al. 2017. Cell-free DNA levels and correlation to stage and outcome following treatment of locally advanced rectal cancer. *Tumour Biol* 39(11), p. 1010428317730976. doi: 10.1177/1010428317730976

Brannon, A. R. et al. 2014. Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions. *Genome Biol* 15(8), p. 454. doi: 10.1186/s13059-014-0454-7

Braun, L. H. et al. 2019. Neutrophil-to-Lymphocyte Ratio in Rectal Cancer-Novel Biomarker of Tumor Immunogenicity During Radiotherapy or Confounding Variable? *Int J Mol Sci* 20(10), doi: 10.3390/ijms20102448

Buijsen, J. et al. 2014. Blood biomarkers are helpful in the prediction of response to chemoradiation in rectal cancer: a prospective, hypothesis driven study on patients with locally advanced rectal cancer. *Radiother Oncol* 111(2), pp. 237-242. doi: 10.1016/j.radonc.2014.03.006

Burrell, R. A. and Swanton, C. 2014. Tumour heterogeneity and the evolution of polyclonal drug resistance. *Mol Oncol* 8(6), pp. 1095-1111. doi: 10.1016/j.molonc.2014.06.005

Cacheux, W. et al. 2016. Mutational analysis of anal cancers demonstrates frequent PIK3CA mutations associated with poor outcome after salvage abdominoperineal resection. *Br J Cancer* 114(12), pp. 1387-1394. doi: 10.1038/bjc.2016.144

Cai, Y. et al. 2014. Prognostic factors associated with locally recurrent rectal cancer following primary surgery (Review).*Oncol Lett*. Vol. 7. pp. 10-16.

Calin, G. A. and Croce, C. M. 2006. MicroRNA signatures in human cancers. *Nat Rev Cancer* 6(11), pp. 857-866. doi: 10.1038/nrc1997

Calin, G. A. et al. 2002. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* 99(24), pp. 15524-15529. doi: 10.1073/pnas.242606799

Cao, D. et al. 2018. A reliable and feasible way to predict the benefits of Nivolumab in patients with non-small cell lung cancer: a pooled analysis of 14 retrospective studies. *Oncoimmunology* 7(11), doi: 10.1080/2162402x.2018.1507262

Cao, L. et al. 2017. Advances in digital polymerase chain reaction (dPCR) and its emerging biomedical applications. *Biosens Bioelectron* 90, pp. 459-474. doi: 10.1016/j.bios.2016.09.082

Carames, C. et al. 2016. MicroRNA-31 Emerges as a Predictive Biomarker of Pathological Response and Outcome in Locally Advanced Rectal Cancer. *Int J Mol Sci* 17(6), doi: 10.3390/ijms17060878

Carpinetti, P. et al. 2015. The use of personalized biomarkers and liquid biopsies to monitor treatment response and disease recurrence in locally advanced rectal cancer after neoadjuvant chemoradiation. *Oncotarget* 6(35), pp. 38360-38371. doi: 10.18632/oncotarget.5256

Castillo, J. et al. 2017. Surfaceome profiling enables isolation of cancer-specific exosomal cargo in liquid biopsies from pancreatic cancer patients. *Ann Oncol*, doi: 10.1093/annonc/mdx542

Cayrefourcq, L. et al. 2015. Establishment and characterization of a cell line from human circulating colon cancer cells. *Cancer Res* 75(5), pp. 892-901. doi: 10.1158/0008-5472.can-14-2613

Cedermark, B. et al. 1997. Improved survival with preoperative radiotherapy in resectable rectal cancer. *N Engl J Med* 336(14), pp. 980-987. doi: 10.1056/nejm199704033361402

Ceelen, W. et al. 2009. Preoperative chemoradiation versus radiation alone for stage II and III resectable rectal cancer: a systematic review and meta-analysis. *Int J Cancer* 124(12), pp. 2966-2972. doi: 10.1002/ijc.24247

Cekaite, L. et al. 2015. MicroRNAs as growth regulators, their function and biomarker status in colorectal cancer. *Oncotarget*, doi: 10.18632/oncotarget.6390

Chang, P. Y. et al. 2016. MicroRNA-223 and microRNA-92a in stool and plasma samples act as complementary biomarkers to increase colorectal cancer detection. *Oncotarget*, doi: 10.18632/oncotarget.7119

Chang, Y. S. et al. 2018. Correlation of genomic alterations between tumor tissue and circulating tumor DNA by next-generation sequencing. *J Cancer Res Clin Oncol* 144(11), pp. 2167-2175. doi: 10.1007/s00432-018-2747-9

Chen, M. B. et al. 2012. P53 status as a predictive biomarker for patients receiving neoadjuvant radiation-based treatment: a meta-analysis in rectal cancer. *PLoS One* 7(9), p. e45388. doi: 10.1371/journal.pone.0045388

Chen, W. Y. et al. 2015. The potential of plasma miRNAs for diagnosis and risk estimation of colorectal cancer. *Int J Clin Exp Pathol* 8(6), pp. 7092-7101.

Chim, S. S. et al. 2008. Detection and characterization of placental microRNAs in maternal plasma. *Clin Chem* 54(3), pp. 482-490. doi: 10.1373/clinchem.2007.097972

Chomczynski, P. 1993. A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotechniques* 15(3), pp. 532-534, 536-537.

Chow, O. S. et al. 2016. KRAS and Combined KRAS/TP53 Mutations in Locally Advanced Rectal Cancer are Independently Associated with Decreased Response to Neoadjuvant Therapy. *Ann Surg Oncol* 23(8), pp. 2548-2555. doi: 10.1245/s10434-016-5205-4

Christie, M. et al. 2013. Different APC genotypes in proximal and distal sporadic colorectal cancers suggest distinct WNT/beta-catenin signalling thresholds for tumourigenesis. *Oncogene* 32(39), pp. 4675-4682. doi: 10.1038/onc.2012.486

Clarke, T. L. et al. 2017. Predicting response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer with serum biomarkers. *Ann R Coll Surg Engl* 99(5), pp. 373-377. doi: 10.1308/rcsann.2017.0030

Cohen, R. et al. 2019. Immunotherapy and metastatic colorectal cancers with microsatellite instability or mismatch repair deficiency. *Bull Cancer* 106(2), pp. 137-142. doi: 10.1016/j.bulcan.2018.09.004

Colloca, G. et al. 2017. Pre-treatment carcinoembryonic antigen and outcome of patients with rectal cancer receiving neo-adjuvant chemo-radiation and surgical resection: a systematic review and meta-analysis. *Med Oncol* 34(10), p. 177. doi: 10.1007/s12032-017-1037-8

Colombino, M. et al. 2002. Prevalence and prognostic role of microsatellite instability in patients with rectal carcinoma. *Ann Oncol* 13(9), pp. 1447-1453. doi: 10.1093/annonc/mdf240

Conde-Muino, R. et al. 2015. Predictive Biomarkers to Chemoradiation in Locally Advanced Rectal Cancer. *Biomed Res Int* 2015, p. 921435. doi: 10.1155/2015/921435

Condeelis, J. and Pollard, J. W. 2006. Macrophages: obligate partners for tumor cell migration, invasion, and metastasis. *Cell* 124(2), pp. 263-266. doi: 10.1016/j.cell.2006.01.007

Costa-Silva, B. et al. 2015. Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nat Cell Biol* 17(6), pp. 816-826. doi: 10.1038/ncb3169

Crane, C. H. et al. 2003. The addition of continuous infusion 5-FU to preoperative radiation therapy increases tumor response, leading to increased sphincter preservation in locally advanced rectal cancer. *Int J Radiat Oncol Biol Phys* 57(1), pp. 84-89. doi: 10.1016/s0360-3016(03)00532-7

Creighton, C. J. et al. 2010. Molecular profiling uncovers a p53-associated role for microRNA-31 in inhibiting the proliferation of serous ovarian carcinomas and other cancers. *Cancer Res* 70(5), pp. 1906-1915. doi: 10.1158/0008-5472.can-09-3875

CRUK. 2016. *Bowel Cancer Statistics*. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer: Available at: [Accessed 08/09/2019].

D'Angelo, E. et al. 2016. Serum miR-125b is a non-invasive predictive biomarker of the pre-operative chemoradiotherapy responsiveness in patients with rectal adenocarcinoma. *Oncotarget* 7(19), pp. 28647-28657. doi: 10.18632/oncotarget.8725

De Caluwe, L. et al. 2013. Preoperative chemoradiation versus radiation alone for stage II and III resectable rectal cancer. *Cochrane Database Syst Rev* (2), p. Cd006041. doi: 10.1002/14651858.CD006041.pub3

de Leng, W. W. et al. 2016. Targeted Next Generation Sequencing as a Reliable Diagnostic Assay for the Detection of Somatic Mutations in Tumours Using Minimal DNA Amounts from Formalin Fixed Paraffin Embedded Material. *PLoS One* 11(2), p. e0149405. doi: 10.1371/journal.pone.0149405

Den Oudsten, B. L. et al. 2012. Higher prevalence of sexual dysfunction in colon and rectal cancer survivors compared with the normative population: a population-based study. *Eur J Cancer* 48(17), pp. 3161-3170. doi: 10.1016/j.ejca.2012.04.004

Deng, Y. X. et al. 2017. Lymphocyte-to-monocyte ratio before chemoradiotherapy represents a prognostic predictor for locally advanced rectal cancer. *Onco Targets Ther* 10, pp. 5575-5583. doi: 10.2147/ott.s146697

Diehl, F. et al. 2005. Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci U S A* 102(45), pp. 16368-16373. doi: 10.1073/pnas.0507904102

Diehl, F. et al. 2008. Circulating mutant DNA to assess tumor dynamics. *Nat Med* 14(9), pp. 985-990. doi: 10.1038/nm.1789

Dienstmann, R. et al. 2017a. Analysis of mutant allele fractions in driver genes in colorectal cancer - biological and clinical insights. *Mol Oncol* 11(9), pp. 1263-1272. doi: 10.1002/1878-0261.12099

Dienstmann, R. et al. 2017b. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat Rev Cancer* 17(2), pp. 79-92. doi: 10.1038/nrc.2016.126

Du, C. et al. 2013. Prognostic value of microsatellite instability in sporadic locally advanced rectal cancer following neoadjuvant radiotherapy. *Histopathology* 62(5), pp. 723-730. doi: 10.1111/his.12069

Dudani, S. et al. 2019. Neutrophil-to-lymphocyte and platelet-to-lymphocyte ratios as predictive and prognostic markers in patients with locally advanced rectal cancer treated with neoadjuvant chemoradiation. *BMC Cancer* 19(1), p. 664. doi: 10.1186/s12885-019-5892-x

Duldulao, M. P. et al. 2013. Mutations in specific codons of the KRAS oncogene are associated with variable resistance to neoadjuvant chemoradiation therapy in patients with rectal adenocarcinoma. *Ann Surg Oncol* 20(7), pp. 2166-2171. doi: 10.1245/s10434-013-2910-0

Dworak, O. et al. 1997. Pathological features of rectal cancer after preoperative radiochemotherapy. *Int J Colorectal Dis* 12(1), pp. 19-23.

EdgeBio. 2019. *Variant Calling on Ion Torrent Data | EdgeBio*. Available at: https://www.edgebio.com/variant-calling-ion-torrent-data [Accessed 29/08/2019].

El Messaoudi, S. et al. 2013. Circulating cell free DNA: Preanalytical considerations. *Clin Chim Acta* 424, pp. 222-230. doi: 10.1016/j.cca.2013.05.022

Enderle, D. et al. 2015. Characterization of RNA from Exosomes and Other Extracellular Vesicles Isolated by a Novel Spin Column-Based Method. *PLoS One* 10(8), p. e0136133. doi: 10.1371/journal.pone.0136133

Erlandsson, J. et al. 2017. Optimal fractionation of preoperative radiotherapy and timing to surgery for rectal cancer (Stockholm III): a multicentre, randomised, non-blinded, phase 3, non-inferiority trial. *Lancet Oncol* 18(3), pp. 336-346. doi: 10.1016/s1470-2045(17)30086-4

Erlandsson, J. et al. 2019. Tumour regression after radiotherapy for rectal cancer - Results from the randomised Stockholm III trial. *Radiother Oncol* 135, pp. 178-186. doi: 10.1016/j.radonc.2019.03.016

Eslamizadeh, S. et al. 2018. The Role of MicroRNA Signature as Diagnostic Biomarkers in Different Clinical Stages of Colorectal Cancer. *Cell J* 20(2), pp. 220-230. doi: 10.22074/cellj.2018.5366

Fahoum, I. et al. 2018. Characterization of Factors Affecting the Detection Limit of EGFR p.T790M in Circulating Tumor DNA. *Technol Cancer Res Treat* 17, p. 1533033818793653. doi: 10.1177/1533033818793653

FDA, O. o. t. C. 2016. Press Announcements - FDA approves first blood test to detect gene mutation associated with non-small cell lung cancer.

Fearon, E. R. 2011. Molecular genetics of colorectal cancer. *Annu Rev Pathol* 6, pp. 479-507. doi: 10.1146/annurev-pathol-011110-130235

Feng, Y. et al. 2014. microRNA-99a acts as a tumor suppressor and is down-regulated in bladder cancer. *BMC Urol* 14, p. 50. doi: 10.1186/1471-2490-14-50

Ferrucci, P. F. et al. 2015. Baseline neutrophil-to-lymphocyte ratio is associated with outcome of ipilimumab-treated metastatic melanoma patients. *Br J Cancer* 112(12), pp. 1904-1910. doi: 10.1038/bjc.2015.180

Findlay, J. M. et al. 2016. Differential clonal evolution in oesophageal cancers in response to neo-adjuvant chemotherapy. *Nat Commun* 7, p. 11111. doi: 10.1038/ncomms11111

Fokas, E. et al. 2014. Tumor regression grading after preoperative chemoradiotherapy for locally advanced rectal carcinoma revisited: updated results of the CAO/ARO/AIO-94 trial. *J Clin Oncol* 32(15), pp. 1554-1562. doi: 10.1200/jco.2013.54.3769

Formica, V. et al. 2018. Hemoglobin level and XRCC1 polymorphisms to select patients with locally advanced rectal cancer candidate for neoadjuvant chemoradiotherapy with concurrent capecitabine and a platinum salt. *Med Oncol* 35(6), p. 83. doi: 10.1007/s12032-018-1141-4

Franco, P. et al. 2018. The prognostic role of hemoglobin levels in patients undergoing concurrent chemo-radiation for anal cancer. *Radiat Oncol* 13(1), p. 83. doi: 10.1186/s13014-018-1035-9

Friedman, C. F. and Postow, M. A. 2016. Emerging Tissue and Blood-Based Biomarkers that may Predict Response to Immune Checkpoint Inhibition. *Curr Oncol Rep* 18(4), p. 21. doi: 10.1007/s11912-016-0509-x

Fujiyoshi, K. et al. 2017. Metastatic Pattern of Stage IV Colorectal Cancer with High-Frequency Microsatellite Instability as a Prognostic Factor. *Anticancer Res* 37(1), pp. 239-247. doi: 10.21873/anticanres.11313

Gaedcke, J. et al. 2010. KRAS and BRAF mutations in patients with rectal cancer treated with preoperative chemoradiotherapy. *Radiother Oncol* 94(1), pp. 76-81. doi: 10.1016/j.radonc.2009.10.001

Gagez, A. L. et al. 2017. miR-125b and miR-532-3p predict the efficiency of rituximab-mediated lymphodepletion in chronic lymphocytic leukemia patients. A French Innovative Leukemia Organization study. *Haematologica* 102(4), pp. 746-754. doi: 10.3324/haematol.2016.153189

Gao, J. et al. 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6(269), p. pl1. doi: 10.1126/scisignal.2004088

Garrido-Cardenas, J. A. et al. 2017. DNA Sequencing Sensors: An Overview. *Sensors (Basel)* 17(3), doi: 10.3390/s17030588

Gasch, C. et al. 2013. Heterogeneity of epidermal growth factor receptor status and mutations of KRAS/PIK3CA in circulating tumor cells of patients with colorectal cancer. *Clin Chem* 59(1), pp. 252-260. doi: 10.1373/clinchem.2012.188557

Gay, L. J. and Felding-Habermann, B. 2011. Contribution of platelets to tumour metastasis. *Nat Rev Cancer* 11(2), pp. 123-134. doi: 10.1038/nrc3004

Germano, G. et al. 2017. Parallel Evaluation of Circulating Tumor DNA and Circulating Tumor Cells in Metastatic Colorectal Cancer. *Clin Colorectal Cancer*, doi: 10.1016/j.clcc.2017.10.017

Giannakis, M. et al. 2016. Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Rep* 15(4), pp. 857-865. doi: 10.1016/j.celrep.2016.03.075

Glynne-Jones, R. et al. 2017. Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 28(suppl_4), pp. iv22-iv40. doi: 10.1093/annonc/mdx224

Gollins, S. et al. 2017. Preoperative chemoradiation with capecitabine, irinotecan and cetuximab in rectal cancer: significance of pre-treatment and post-resection RAS mutations. *Br J Cancer* 117(9), pp. 1286-1294. doi: 10.1038/bjc.2017.294

Gomez-Millan, J. et al. 2017. The impact of body mass index and nuclear beta-catenin on survival in locally advanced rectal cancer treated with preoperative radiochemotherapy. *J Surg Oncol* 115(3), pp. 301-306. doi: 10.1002/jso.24494

Gorges, T. M. et al. 2016. Improved Detection of Circulating Tumor Cells in Metastatic Colorectal Cancer by the Combination of the CellSearch(R) System and the AdnaTest(R). *PLoS One* 11(5), p. e0155126. doi: 10.1371/journal.pone.0155126

Goto, S. et al. 1997. Organizing and computing metabolic pathway data in terms of binary relations. *Pac Symp Biocomput*, pp. 175-186.

Grasselli, J. et al. 2017. Concordance of blood- and tumor-based detection of RAS mutations to guide anti-EGFR therapy in metastatic colorectal cancer. *Ann Oncol* 28(6), pp. 1294-1301. doi: 10.1093/annonc/mdx112

Greystoke, A. et al. 2015. Development of a circulating miRNA assay to monitor tumor burden: From mouse to man. *Mol Oncol*, doi: 10.1016/j.molonc.2015.10.004

Grillet, F. et al. 2017. Circulating tumour cells from patients with colorectal cancer have cancer stem cell hallmarks in ex vivo culture. *Gut* 66(10), pp. 1802-1810. doi: 10.1136/gutjnl-2016-311447

Gudgeon, J. M. et al. 2007. Rapid ACCE: experience with a rapid and structured approach for evaluating gene-based testing. *Genet Med* 9(7), pp. 473-478. doi: 10.1097GIM.0b013e3180a6e9ef

Guinney, J. et al. 2015. The consensus molecular subtypes of colorectal cancer. *Nat Med* 21(11), pp. 1350-1356. doi: 10.1038/nm.3967

Gut, M. et al. 1999. One-tube fluorogenic reverse transcription-polymerase chain reaction for the quantitation of feline coronaviruses. *J Virol Methods* 77(1), pp. 37-46. doi: 10.1016/s0166-0934(98)00129-3

Hanahan, D. and Weinberg, R. A. 2011. Hallmarks of cancer: the next generation. *Cell* 144(5), pp. 646-674. doi: 10.1016/j.cell.2011.02.013

Hao, Y. X. et al. 2017. KRAS and BRAF mutations in serum exosomes from patients with colorectal cancer in a Chinese population. *Oncol Lett* 13(5), pp. 3608-3616. doi: 10.3892/ol.2017.5889

Harada, K. et al. 2019. Comparative sequence analysis of patient-matched primary colorectal cancer, metastatic, and recurrent metastatic tumors after adjuvant FOLFOX chemotherapy. *BMC Cancer* 19(1), p. 255. doi: 10.1186/s12885-019-5479-6

Hardingham, J. E. et al. 2015. Detection and Clinical Significance of Circulating Tumor Cells in Colorectal Cancer--20 Years of Progress. *Mol Med* 21 Suppl 1, pp. S25-31. doi: 10.2119/molmed.2015.00149

Hasan, S. et al. 2018. Microsatellite Instability (MSI) as an Independent Predictor of Pathologic Complete Response (PCR) in Locally Advanced Rectal Cancer: A National Cancer Database (NCDB) Analysis. *Ann Surg*, doi: 10.1097/sla.0000000000003051

Hindson, B. J. et al. 2011. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem* 83(22), pp. 8604-8610. doi: 10.1021/ac202028g

Hong, S. P. et al. 2012. The differential impact of microsatellite instability as a marker of prognosis and tumour response between colon cancer and rectal cancer. *Eur J Cancer* 48(8), pp. 1235-1243. doi: 10.1016/j.ejca.2011.10.005

Horizon. 2019. *Multiplex I cfDNA Reference Standard Set*. Available at: https://horizondiscovery.com/en/products/diagnostic-reference-standards/PIFs/Multiplex-I-cfDNA-Reference-Standard-Set [Accessed: 25/11/2019].

Hou, B. et al. 2015. Circulating microRNAs as novel prognosis biomarkers for head and neck squamous cell carcinoma. *Cancer Biol Ther* 16(7), pp. 1042-1046. doi: 10.1080/15384047.2015.1045692

Hrebien, S. et al. 2016. Reproducibility of Digital PCR Assays for Circulating Tumor DNA Analysis in Advanced Breast Cancer. *PLoS One* 11(10), p. e0165023. doi: 10.1371/journal.pone.0165023

Hsu, H. C. et al. 2018. Targeted Sequencing of Circulating Tumor DNA to Monitor Genetic Variants and Therapeutic Response in Metastatic Colorectal Cancer. *Mol Cancer Ther* 17(10), pp. 2238-2247. doi: 10.1158/1535-7163.mct-17-1306

Hsu, H. C. et al. 2016. Mutations of KRAS/NRAS/BRAF predict cetuximab resistance in metastatic colorectal cancer patients. *Oncotarget* 7(16), pp. 22257-22270. doi: 10.18632/oncotarget.8076

Hu, J. L. et al. 2019. CAFs secreted exosomes promote metastasis and chemotherapy resistance by enhancing cell stemness and epithelial-mesenchymal transition in colorectal cancer. *Mol Cancer* 18(1), p. 91. doi: 10.1186/s12943-019-1019-x

Huang, A. et al. 2016. Detecting Circulating Tumor DNA in Hepatocellular Carcinoma Patients Using Droplet Digital PCR Is Feasible and Reflects Intratumoral Heterogeneity. *J Cancer* 7(13), pp. 1907-1914. doi: 10.7150/jca.15823

Huang, L. et al. 2012. Downregulation of six microRNAs is associated with advanced stage, lymph node metastasis and poor prognosis in small cell carcinoma of the cervix. *PLoS One* 7(3), p. e33762. doi: 10.1371/journal.pone.0033762

Huang, X. et al. 2014. Relationship between circulating tumor cells and tumor response in colorectal cancer patients treated with chemotherapy: a meta-analysis. *BMC Cancer* 14, p. 976. doi: 10.1186/1471-2407-14-976

Huang, X. et al. 2015. Meta-analysis of the prognostic value of circulating tumor cells detected with the CellSearch System in colorectal cancer. *BMC Cancer* 15, p. 202. doi: 10.1186/s12885-015-1218-9

Huang, Y. et al. 2019. Local environment in biopsy better predict the pathological response to neoadjuvant chemoradiotherapy in rectal cancer. *Biosci Rep* 39(3), doi: 10.1042/bsr20190003

Huang, Z. and Feng, Y. 2017. Exosomes Derived From Hypoxic Colorectal Cancer Cells Promote Angiogenesis Through Wnt4-Induced beta-Catenin Signaling in Endothelial Cells. *Oncol Res* 25(5), pp. 651-661. doi: 10.3727/096504016x14752792816791

Huang, Z. et al. 2010. Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. *Int J Cancer* 127(1), pp. 118-126. doi: 10.1002/ijc.25007

Hubbard, T. et al. 2002. The Ensembl genome database project. *Nucleic Acids Res* 30(1), pp. 38-41. doi: 10.1093/nar/30.1.38

Huggett, J. F. and Whale, A. 2013. Digital PCR as a novel technology and its potential implications for molecular diagnostics. *Clinical chemistry* 59(12), pp. 1691-1693.

Huh, J. W. et al. 2016. Mismatch Repair Gene Expression as a Predictor of Tumor Responses in Patients With Rectal Cancer Treated With Preoperative Chemoradiation. *Medicine (Baltimore)* 95(3), p. e2582. doi: 10.1097/md.0000000000002582

Ignatiadis, M. et al. 2015. Circulating Tumor Cells and Circulating Tumor DNA: Challenges and Opportunities on the Path to Clinical Utility. *Clin Cancer Res* 21(21), pp. 4786-4800. doi: 10.1158/1078-0432.ccr-14-1190

Isella, C. et al. 2017. Selective analysis of cancer-cell intrinsic transcriptional traits defines novel clinically relevant subtypes of colorectal cancer. *Nat Commun* 8, p. 15107. doi: 10.1038/ncomms15107

Ivanov, M. et al. 2017. Towards standardization of next-generation sequencing of FFPE samples for clinical oncology: intrinsic obstacles and possible solutions. *J Transl Med* 15(1), p. 22. doi: 10.1186/s12967-017-1125-8

Jahr, S. et al. 2001. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res* 61(4), pp. 1659-1665.

Jakob, C. et al. 2008. Predictive value of Ki67 and p53 in locally advanced rectal cancer: correlation with thymidylate synthase and histopathological tumor regression after neoadjuvant 5-FU-based chemoradiotherapy. *World J Gastroenterol* 14(7), pp. 1060-1066. doi: 10.3748/wjg.14.1060

Janjan, N. A. et al. 2001. Improved overall survival among responders to preoperative chemoradiation for locally advanced rectal cancer. *Am J Clin Oncol* 24(2), pp. 107-112.

Janni, W. et al. 2016. Pooled Analysis of the Prognostic Relevance of Circulating Tumor Cells in Primary Breast Cancer. *Clin Cancer Res*, doi: 10.1158/1078-0432.ccr-15-1603

Jennings, L. J. et al. 2017. Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *J Mol Diagn* 19(3), pp. 341-365. doi: 10.1016/j.jmoldx.2017.01.011

Jeon, B. H. et al. 2019. Neutrophil to Lymphocyte Ratio: A Predictive Marker for Treatment Outcomes in Patients With Rectal Cancer Who Underwent Neoadjuvant Chemoradiation Followed by Surgery. *Ann Coloproctol* 35(2), pp. 100-106. doi: 10.3393/ac.2018.10.01

Jeppesen, D. K. et al. 2019. Reassessment of Exosome Composition. *Cell* 177(2), pp. 428-445.e418. doi: 10.1016/j.cell.2019.02.029

Jia, H. Y. et al. 2012. MicroRNA-125b functions as a tumor suppressor in hepatocellular carcinoma cells. *Int J Mol Sci* 13(7), pp. 8762-8774. doi: 10.3390/ijms13078762

Jia, S. et al. 2017. Clinical and biological significance of circulating tumor cells, circulating tumor DNA, and exosomes as biomarkers in colorectal cancer. *Oncotarget*, doi: 10.18632/oncotarget.17184

Jin, Y. et al. 2016. DNA in serum extracellular vesicles is stable under different storage conditions. *BMC Cancer* 16(1), p. 753. doi: 10.1186/s12885-016-2783-2

Jo, P. et al. 2017. Changes of Microrna Levels in Plasma of Patients with Rectal Cancer during Chemoradiotherapy. *Int J Mol Sci* 18(6), doi: 10.3390/ijms18061140

Johnstone, R. M. et al. 1987. Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes). *J Biol Chem* 262(19), pp. 9412-9420.

Jung, S. W. et al. 2017. Association of immunologic markers from complete blood counts with the response to preoperative chemoradiotherapy and prognosis in locally advanced rectal cancer. *Oncotarget* 8(35), pp. 59757-59765. doi: 10.18632/oncotarget.15760

Kahlert, C. et al. 2014. Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer. *J Biol Chem* 289(7), pp. 3869-3875. doi: 10.1074/jbc.C113.532267

Kanaan, Z. et al. 2012. Plasma miR-21: a potential diagnostic marker of colorectal cancer. *Ann Surg* 256(3), pp. 544-551. doi: 10.1097/SLA.0b013e318265bd6f

Kapiteijn, E. et al. 2001. Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer. *N Engl J Med* 345(9), pp. 638-646. doi: 10.1056/NEJMoa010580

Kent, W. J. et al. 2002. The human genome browser at UCSC. *Genome Res* 12(6), pp. 996-1006. doi: 10.1101/gr.229102

Khan, A. A. et al. 2013. Association between pretreatment haemoglobin levels and morphometric characteristics of the tumour, response to neoadjuvant treatment and long-term outcomes in patients with locally advanced rectal cancers. *Colorectal Dis* 15(10), pp. 1232-1237. doi: 10.1111/codi.12307

Kidess-Sigal, E. et al. 2016. Enumeration and targeted analysis of KRAS, BRAF and PIK3CA mutations in CTCs captured by a label-free platform: Comparison to ctDNA and tissue in metastatic colorectal cancer. *Oncotarget* 7(51), pp. 85349-85364. doi: 10.18632/oncotarget.13350

Kim, S. B. et al. 2014. Inhibition of microRNA-31-5p protects human colonic epithelial cells against ionizing radiation. *Life Sci Space Res (Amst)* 1, pp. 67-73. doi: 10.1016/j.lssr.2014.02.001

Kim, T. G. et al. 2018. Baseline neutrophil-lymphocyte ratio and platelet-lymphocyte ratio in rectal cancer patients following neoadjuvant chemoradiotherapy. *Tumori*, p. 300891618792476. doi: 10.1177/0300891618792476

Kitai, Y. et al. 2017. DNA-Containing Exosomes Derived from Cancer Cells Treated with Topotecan Activate a STING-Dependent Pathway and Reinforce Antitumor Immunity. *J Immunol* 198(4), pp. 1649-1659. doi: 10.4049/jimmunol.1601694

Kitayama, J. et al. 2010. Circulating lymphocyte number has a positive association with tumor response in neoadjuvant chemoradiotherapy for advanced rectal cancer. *Radiat Oncol* 5, p. 47. doi: 10.1186/1748-717x-5-47

Ko, J. et al. 2015. Detection and isolation of circulating exosomes and microvesicles for cancer monitoring and diagnostics using micro-/nano-based devices. *Analyst*, doi: 10.1039/c5an01610j

Kobayashi, H. et al. 2007. Absence of cyclooxygenase-2 protein expression is a predictor of tumor regression in rectal cancer treated with preoperative short-term chemoradiotherapy. *Dis Colon Rectum* 50(9), pp. 1354-1362. doi: 10.1007/s10350-006-0881-y

Kong, J. C. et al. 2019. Prognostic Impact of Tumor-Infiltrating Lymphocytes in Primary and Metastatic Colorectal Cancer: A Systematic Review and Meta-analysis. *Dis Colon Rectum* 62(4), pp. 498-508. doi: 10.1097/dcr.0000000000001332

Kreutz, J. E. et al. 2011. Theoretical design and analysis of multivolume digital assays with wide dynamic range validated experimentally with microfluidic digital PCR. *Anal Chem* 83(21), pp. 8158-8168. doi: 10.1021/ac201658s

Lallo, A. et al. 2017. Circulating tumor cells and CDX models as a tool for preclinical drug development.*Transl Lung Cancer Res*. Vol. 6. pp. 397-408.

Lambrechts, D. et al. 2015. Evaluation of efficacy and safety markers in a phase II study of metastatic colorectal cancer treated with aflibercept in the first-line setting. *Br J Cancer* 113(7), pp. 1027-1034. doi: 10.1038/bjc.2015.329

Lecomte, T. et al. 2002. Detection of free-circulating tumor-associated DNA in plasma of colorectal cancer patients and its association with prognosis. *Int J Cancer* 100(5), pp. 542-548. doi: 10.1002/ijc.10526

Lee, C. K. et al. 2017. Next-generation sequencing reveals novel resistance mechanisms and molecular heterogeneity in EGFR-mutant non-small cell lung cancer with acquired resistance to EGFR-TKIs. *Lung Cancer* 113, pp. 106-114. doi: 10.1016/j.lungcan.2017.09.005

Lee, J. W. et al. 2015. KRAS Mutation Status Is Not a Predictor for Tumor Response and Survival in Rectal Cancer Patients Who Received Preoperative Radiotherapy With 5-Fluoropyrimidine Followed by Curative Surgery. *Medicine (Baltimore)* 94(31), p. e1284. doi: 10.1097/md.0000000000001284

Lee, T. H. et al. 2014. Oncogenic ras-driven cancer cell vesiculation leads to emission of double-stranded DNA capable of interacting with target cells. *Biochem Biophys Res Commun* 451(2), pp. 295-301. doi: 10.1016/j.bbrc.2014.07.109

Li, H. et al. 2016. Preoperative neutrophil-to-lymphocyte ratio is a more valuable prognostic factor than platelet-to-lymphocyte ratio for nonmetastatic rectal cancer. *Int Immunopharmacol* 40, pp. 327-331. doi: 10.1016/j.intimp.2016.09.014

Li, M. M. et al. 2017a. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J Mol Diagn* 19(1), pp. 4-23. doi: 10.1016/j.jmoldx.2016.10.002

Li, W. et al. 2017b. Major challenges related to tumor biological characteristics in accurate mutation detection of colorectal cancer by next-generation sequencing. *Cancer Lett* 410, pp. 92-99. doi: 10.1016/j.canlet.2017.09.014

Lim, S. H. et al. 2014. Effect of neoadjuvant chemoradiation on tumor-infiltrating/associated lymphocytes in locally advanced rectal cancers. *Anticancer Res* 34(11), pp. 6505-6513.

Lin, M. T. et al. 2014. Clinical validation of KRAS, BRAF, and EGFR mutation detection using next-generation sequencing. *Am J Clin Pathol* 141(6), pp. 856-866. doi: 10.1309/ajcpmwgwgo34egod

Lindebjerg, J. et al. 2011. Factors influencing reproducibility of tumour regression grading after high-dose chemoradiation of locally advanced rectal cancer. *Histopathology* 59(1), pp. 18-21. doi: 10.1111/j.1365-2559.2011.03888.x

Lindner, K. et al. 2015. Circulating microRNAs: emerging biomarkers for diagnosis and prognosis in patients with gastrointestinal cancers. *Clin Sci (Lond)* 128(1), pp. 1-15. doi: 10.1042/cs20140089

Liu, L. et al. 2012. Comparison of Next-Generation Sequencing Systems. *J Biomed Biotechnol* 2012, doi: 10.1155/2012/251364

Liu, M. et al. 2016. A Preoperative Measurement of Serum MicroRNA-125b May Predict the Presence of Microvascular Invasion in Hepatocellular Carcinomas Patients. *Transl Oncol* 9(3), pp. 167-172. doi: 10.1016/j.tranon.2016.03.002

Lo, Y. M. et al. 2000. Kinetics of plasma Epstein-Barr virus DNA during radiation therapy for nasopharyngeal carcinoma. *Cancer Res* 60(9), pp. 2351-2355.

Lobb, R. J. et al. 2015. Optimized exosome isolation protocol for cell culture supernatant and human plasma. *J Extracell Vesicles* 4, p. 27031. doi: 10.3402/jev.v4.27031

Lotvall, J. et al. 2014. Minimal experimental requirements for definition of extracellular vesicles and their functions: a position statement from the International Society for Extracellular Vesicles. *J Extracell Vesicles* 3, p. 26913. doi: 10.3402/jev.v3.26913

Ludwig, A. K. et al. 2018. Precipitation with polyethylene glycol followed by washing and pelleting by ultracentrifugation enriches extracellular vesicles from tissue culture supernatants in small and large scales. *J Extracell Vesicles* 7(1), p. 1528109. doi: 10.1080/20013078.2018.1528109

Luo, Z. et al. 2016. Association of circulating miR-125b and survival in patients with osteosarcoma-A single center experience. *J Bone Oncol* 5(4), pp. 167-172. doi: 10.1016/j.jbo.2016.06.002

Lynam-Lennon, N. et al. 2012. MicroRNA-31 modulates tumour sensitivity to radiation in oesophageal adenocarcinoma. *J Mol Med (Berl)* 90(12), pp. 1449-1458. doi: 10.1007/s00109-012-0924-x

Ma, L. et al. 2007. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* 449(7163), pp. 682-688. doi: 10.1038/nature06174

Maas, M. et al. 2010. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *Lancet Oncol* 11(9), pp. 835-844. doi: 10.1016/s1470-2045(10)70172-8

Mandard, A. M. et al. 1994. Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer* 73(11), pp. 2680-2686.

Margalit, O. et al. 2019. Locally advanced rectal adenocarcinoma: Are preoperative short and long course radiotherapy truly equivalent? *Mol Clin Oncol* 10(5), pp. 555-559. doi: 10.3892/mco.2019.1825

Marijnen, C. A. 2015. Organ preservation in rectal cancer: have all questions been answered? *Lancet Oncol* 16(1), pp. e13-22. doi: 10.1016/s1470-2045(14)70398-5

Martellucci, J. et al. 2015. Role of KRAS mutation as predictor of pathologic response after neoadjuvant chemoradiation therapy for rectal cancer. *Updates Surg* 67(1), pp. 47-53. doi: 10.1007/s13304-015-0281-8

Matsumura, T. et al. 2015. Exosomal microRNA in serum is a novel biomarker of recurrence in human colorectal cancer. *Br J Cancer* 113(2), pp. 275-281. doi: 10.1038/bjc.2015.201

Matsutani, S. et al. 2018. Significance of tumor-infiltrating lymphocytes before and after neoadjuvant therapy for rectal cancer. *Cancer Sci* 109(4), pp. 966-979. doi: 10.1111/cas.13542

McCall, C. M. et al. 2014. False positives in multiplex PCR-based next-generation sequencing have unique signatures. *J Mol Diagn* 16(5), pp. 541-549. doi: 10.1016/j.jmoldx.2014.06.001

Mehrotra, M. et al. 2018. Detection of somatic mutations in cell-free DNA in plasma and correlation with overall survival in patients with solid tumors. *Oncotarget* 9(12), pp. 10259-10271. doi: 10.18632/oncotarget.21982

Melo, S. A. et al. 2015. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. *Nature* 523(7559), pp. 177-182. doi: 10.1038/nature14581

Miki, H. et al. 2018. Pretreatment Serum Carbohydrate Antigen 19-9 Concentration Is a Predictor of Survival of Patients Who Have Undergone Curative Resection of Stage IV Rectal Cancer. *Dig Surg* 35(5), pp. 389-396. doi: 10.1159/000480247

Mirjolet, C. et al. 2018. Tumor lymphocyte immune response to preoperative radiotherapy in locally advanced rectal cancer: The LYMPHOREC study. *Oncoimmunology* 7(3), p. e1396402. doi: 10.1080/2162402x.2017.1396402

Misale, S. et al. 2012. Emergence of KRAS mutations and acquired resistance to anti EGFR therapy in colorectal cancer. *Nature* 486(7404), pp. 532-536. doi: 10.1038/nature11156

Misyura, M. et al. 2016. Comparison of Next-Generation Sequencing Panels and Platforms for Detection and Verification of Somatic Tumor Variants for Clinical Diagnostics. *J Mol Diagn* 18(6), pp. 842-850. doi: 10.1016/j.jmoldx.2016.06.004

Molina-Pinelo, S. et al. 2014. MiR-107 and miR-99a-3p predict chemotherapy response in patients with advanced colorectal cancer. *BMC Cancer* 14, p. 656. doi: 10.1186/1471-2407-14-656

Molinari, C. et al. 2015. Biomarkers and Molecular Imaging as Predictors of Response to Neoadjuvant Chemoradiotherapy in Patients With Locally Advanced Rectal Cancer. *Clin Colorectal Cancer* 14(4), pp. 227-238. doi: 10.1016/j.clcc.2015.05.014

Montermini, L. et al. 2015. Inhibition of oncogenic epidermal growth factor receptor kinase triggers release of exosome-like extracellular vesicles and impacts their phosphoprotein and DNA content. *J Biol Chem* 290(40), pp. 24534-24546. doi: 10.1074/jbc.M115.679217

Mostert, B. et al. 2013. KRAS and BRAF mutation status in circulating colorectal tumor cells and their correlation with primary and metastatic tumor tissue. *Int J Cancer* 133(1), pp. 130-141. doi: 10.1002/ijc.27987

Mueller, A. C. et al. 2013. The miR-99 family regulates the DNA damage response through its target SNF2H. *Oncogene* 32(9), pp. 1164-1172. doi: 10.1038/onc.2012.131

Nakagomi, T. et al. 2017. Elucidation of radiation-resistant clones by a serial study of intratumor heterogeneity before and after stereotactic radiotherapy in lung cancer.*J Thorac Dis*. Vol. 9. China, pp. E598-e604.

Nakayama, Y. et al. 2016. Pitfalls of DNA Quantification Using DNA-Binding Fluorescent Dyes and Suggested Solutions. *PLoS One* 11(3), p. e0150528. doi: 10.1371/journal.pone.0150528

Nemeth, A. et al. 2017. Antibiotic-induced release of small extracellular vesicles (exosomes) with surface-associated DNA. *Sci Rep* 7(1), p. 8202. doi: 10.1038/s41598-017-08392-1

Network, C. G. A. 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487(7407), pp. 330-337. doi: 10.1038/nature11252

Nicholson, B. D. et al. 2015. Blood CEA levels for detecting recurrent colorectal cancer. *Cochrane Database Syst Rev* 12, p. Cd011134. doi: 10.1002/14651858.CD011134.pub2

Noorani, A. et al. 2017. A comparative analysis of whole genome sequencing of esophageal adenocarcinoma pre- and post-chemotherapy. *Genome Res* 27(6), pp. 902-912. doi: 10.1101/gr.214296.116

Normanno, N. et al. 2015. Heterogeneity of KRAS, NRAS, BRAF and PIK3CA mutations in metastatic colorectal cancer and potential effects on therapy in the CAPRI GOIM trial. *Ann Oncol* 26(8), pp. 1710-1714. doi: 10.1093/annonc/mdv176

Ogata-Kawata, H. et al. 2014. Circulating exosomal microRNAs as biomarkers of colon cancer. *PLoS One* 9(4), p. e92921. doi: 10.1371/journal.pone.0092921

Ostwal, V. et al. 2019. Low prevalence of deficient mismatch repair (dMMR) protein in locally advanced rectal cancers (LARC) and treatment outcomes. *J Gastrointest Oncol* 10(1), pp. 19-29. doi: 10.21037/jgo.2018.10.01

Overman, M. J. et al. 2017. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. *Lancet Oncol* 18(9), pp. 1182-1191. doi: 10.1016/s1470-2045(17)30422-9

Panditharatna, E. et al. 2018. Clinically Relevant and Minimally Invasive Tumor Surveillance of Pediatric Diffuse Midline Gliomas Using Patient-Derived Liquid Biopsy. *Clin Cancer Res*, doi: 10.1158/1078-0432.ccr-18-1345

Pantel, K. et al. 2012. Circulating epithelial cells in patients with benign colon diseases. *Clin Chem* 58(5), pp. 936-940. doi: 10.1373/clinchem.2011.175570

Park, S. et al. 2018. Plasma Cell-Free DNA as a Predictive Marker after Radiotherapy for Hepatocellular Carcinoma. *Yonsei Med J* 59(4), pp. 470-479. doi: 10.3349/ymj.2018.59.4.470

Patel, U. B. et al. 2011. Magnetic resonance imaging-detected tumor response for locally advanced rectal cancer predicts survival outcomes: MERCURY experience. *J Clin Oncol* 29(28), pp. 3753-3760. doi: 10.1200/jco.2011.34.9068

Peeters, K. C. et al. 2007. The TME trial after a median follow-up of 6 years: increased local control but no survival benefit in irradiated patients with resectable rectal carcinoma. *Ann Surg* 246(5), pp. 693-701. doi: 10.1097/01.sla.0000257358.56863.ce

Pender, A. et al. 2015. Efficient Genotyping of KRAS Mutant Non-Small Cell Lung Cancer Using a Multiplexed Droplet Digital PCR Approach. *PLoS One* 10(9), p. e0139074. doi: 10.1371/journal.pone.0139074

Peng, H. et al. 2018. Analysis of Clinical characteristics to predict pathologic complete response for patients with locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy. *J Cancer* 9(15), pp. 2687-2692. doi: 10.7150/jca.25493

Peng, H. et al. 2016. Predictive value of APAF-1 and COX-2 expression in pathologic complete response to neoadjuvant chemoradiotherapy for patients with locally advanced rectal adenocarcinoma. *Oncotarget* 7(23), pp. 35233-35240. doi: 10.18632/oncotarget.9125

Pentheroudakis, G. and Twelves, C. 2002. Capecitabine (Xeloda): from the laboratory to the patient's home. *Clin Colorectal Cancer* 2(1), pp. 16-23. doi: 10.3816/CCC.2002.n.007

Pine, J. K. et al. 2015. Systemic neutrophil-to-lymphocyte ratio in colorectal cancer: the relationship to patient survival, tumour biology and local lymphocytic response to tumour.*Br J Cancer*. Vol. 113. pp. 204-211.

Pitini, E. et al. 2018. How is genetic testing evaluated? A systematic review of the literature. *Eur J Hum Genet* 26(5), pp. 605-615. doi: 10.1038/s41431-018-0095-5

Ponti, G. et al. 2018. The value of fluorimetry (Qubit) and spectrophotometry (NanoDrop) in the quantification of cell-free DNA (cfDNA) in malignant melanoma and prostate cancer patients. *Clin Chim Acta* 479, pp. 14-19. doi: 10.1016/j.cca.2018.01.007

Portale, G. et al. 2018. Platelet-to-Lymphocyte Ratio and Neutrophil-to-Lymphocyte Ratio Are Not Prognostic Biomarkers in Rectal Cancer Patients with Curative Resection. *J Gastrointest Surg* 22(9), pp. 1611-1618. doi: 10.1007/s11605-018-3781-2

Potapov, V. and Ong, J. L. 2017. Examining Sources of Error in PCR by Single-Molecule Sequencing. *PLoS One* 12(1), p. e0169774. doi: 10.1371/journal.pone.0169774

Primrose, J. N. et al. 2014. Effect of 3 to 5 years of scheduled CEA and CT follow-up to detect recurrence of colorectal cancer: the FACS randomized clinical trial. *Jama* 311(3), pp. 263-270. doi: 10.1001/jama.2013.285718

Pritchard, C. C. et al. 2012. Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. *Cancer Prev Res (Phila)* 5(3), pp. 492-497. doi: 10.1158/1940-6207.capr-11-0370

Provencio, M. et al. 2017. Dynamic circulating tumor DNA quantificaton for the individualization of non-small-cell lung cancer patients treatment. *Oncotarget* 8(36), pp. 60291-60298. doi: 10.18632/oncotarget.20016

Pugh, S. et al. 2017. Association between miR-31-3p expression and cetuximab efficacy in patients with KRAS wild-type metastatic colorectal cancer: a post-hoc analysis of the New EPOC trial. *Oncotarget* 8(55), pp. 93856-93866. doi: 10.18632/oncotarget.21291

Quail, M. A. et al. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13, p. 341. doi: 10.1186/1471-2164-13-341

Rachiglio, A. M. et al. 2016. Limits and potential of targeted sequencing analysis of liquid biopsy in patients with lung and colon carcinoma. *Oncotarget* 7(41), pp. 66595-66605. doi: 10.18632/oncotarget.10704

Rai, A. et al. 2019. Exosomes Derived from Human Primary and Metastatic Colorectal Cancer Cells Contribute to Functional Heterogeneity of Activated Fibroblasts by Reprogramming Their Proteome. *Proteomics* 19(8), p. e1800148. doi: 10.1002/pmic.201800148

Ramsay, A. G. 2013. Immune checkpoint blockade immunotherapy to activate anti-tumour T-cell immunity. *Br J Haematol* 162(3), pp. 313-325. doi: 10.1111/bjh.12380

Rane, J. K. et al. 2016. Inhibition of the glucocorticoid receptor results in an enhanced miR-99a/100-mediated radiation response in stem-like cells from human prostate cancers. *Oncotarget* 7(32), pp. 51965-51980. doi: 10.18632/oncotarget.10207

Raychaudhuri, M. et al. 2017. MicroRNAs miR-7 and miR-340 predict response to neoadjuvant chemotherapy in breast cancer. *Breast Cancer Res Treat* 162(3), pp. 511-521. doi: 10.1007/s10549-017-4132-9

Reid, A. L. et al. 2015. Detection of BRAF-V600E and V600K in melanoma circulating tumour cells by droplet digital PCR. *Clin Biochem* 48(15), pp. 999-1002. doi: 10.1016/j.clinbiochem.2014.12.007

Reinert, T. et al. 2015. Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery. *Gut*, doi: 10.1136/gutjnl-2014-308859

Reinert, T. et al. 2016. Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery. *Gut* 65(4), pp. 625-634. doi: 10.1136/gutjnl-2014-308859

Renehan, A. G. et al. 2016. Watch-and-wait approach versus surgical resection after chemoradiotherapy for patients with rectal cancer (the OnCoRe project): a propensity-score matched cohort analysis. *Lancet Oncol* 17(2), pp. 174-183. doi: 10.1016/s1470-2045(15)00467-2

Research, C. f. D. E. a. 2018a. Approved Drugs - FDA grants accelerated approval to pembrolizumab for first tissue/site agnostic indication.

Research, C. f. D. E. a. 2018b. Approved Drugs - FDA grants nivolumab accelerated approval for MSI-H or dMMR colorectal cancer.

Riva, F. et al. 2017. Patient-Specific Circulating Tumor DNA Detection during Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer. *Clin Chem* 63(3), pp. 691-699. doi: 10.1373/clinchem.2016.262337

Rodel, C. et al. 2005. Prognostic significance of tumor regression after preoperative chemoradiotherapy for rectal cancer. *J Clin Oncol* 23(34), pp. 8688-8696. doi: 10.1200/jco.2005.02.1329

Rothberg, J. M. et al. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475(7356), pp. 348-352. doi: 10.1038/nature10242

Ruiz-Lopez, L. et al. 2018. The role of exosomes on colorectal cancer: A review. *J Gastroenterol Hepatol* 33(4), pp. 792-799. doi: 10.1111/jgh.14049

Russo, M. et al. 2015. Tumor heterogeneity and lesion-specific response to targeted therapy in colorectal cancer. *Cancer Discov*, doi: 10.1158/2159-8290.cd-15-1283

Saha, T. et al. 2015. Structural and sequential context of p53: A review of experimental and theoretical evidence. *Prog Biophys Mol Biol* 117(2-3), pp. 250-263. doi: 10.1016/j.pbiomolbio.2014.12.002

Saito, G. et al. 2018. Relations of Changes in Serum Carcinoembryonic Antigen Levels before and after Neoadjuvant Chemoradiotherapy and after Surgery to Histologic Response and Outcomes in Patients with Locally Advanced Rectal Cancer. *Oncology* 94(3), pp. 167-175. doi: 10.1159/000485511

Sakai, K. et al. 2014. Chemoradiation provides a physiological selective pressure that increases the expansion of aberrant TP53 tumor variants in residual rectal cancerous regions. *Oncotarget* 5(20), pp. 9641-9649. doi: 10.18632/oncotarget.2438

San Lucas, F. A. et al. 2015. Minimally invasive genomic and transcriptomic profiling of visceral cancers by next-generation sequencing of circulating exosomes. *Ann Oncol*, doi: 10.1093/annonc/mdv604

Sanchez-Sendra, B. et al. 2018. Downregulation of intratumoral expression of miR-205, miR-200c and miR-125b in primary human cutaneous melanomas predicts shorter survival. *Sci Rep* 8(1), p. 17076. doi: 10.1038/s41598-018-35317-3

Sargent, D. et al. 2011. Two or Three Year Disease Free Survival (DFS) as a Primary Endpoint in Stage III Adjuvant Colon Cancer Trials with fluoropyrimidines with or without Oxaliplatin or Irinotecan: Data from 12,676 patients from MOSAIC, X-ACT, PETACC-3, C-06, C-07, and C89803. *Eur J Cancer* 47(7), pp. 990-996. doi: 10.1016/j.ejca.2010.12.015

Sastre, J. et al. 2012. Circulating tumor cell count is a prognostic factor in metastatic colorectal cancer patients receiving first-line chemotherapy plus bevacizumab: a Spanish Cooperative Group for the Treatment of Digestive Tumors study. *Oncologist* 17(7), pp. 947-955. doi: 10.1634/theoncologist.2012-0048

Sauer, R. et al. 2004. Preoperative versus postoperative chemoradiotherapy for rectal cancer. *N Engl J Med* 351(17), pp. 1731-1740. doi: 10.1056/NEJMoa040694

Schageman, J. et al. 2013. The Complete Exosome Workflow Solution: From Isolation to Characterization of RNA Cargo. *Biomed Res Int* 2013, doi: 10.1155/2013/253957

Schou, J. V. et al. 2018. Circulating cell-free DNA as predictor of treatment failure after neoadjuvant chemo-radiotherapy before surgery in patients with locally advanced rectal cancer. *Ann Oncol* 29(3), pp. 610-615. doi: 10.1093/annonc/mdx778

Sclafani, F. et al. 2017. Comparison between MRI and pathology in the assessment of tumour regression grade in rectal cancer. *Br J Cancer* 117(10), pp. 1478-1485. doi: 10.1038/bjc.2017.320

Sclafani, F. et al. 2018. KRAS and BRAF mutations in circulating tumour DNA from locally advanced rectal cancer. *Sci Rep* 8(1), p. 1445. doi: 10.1038/s41598-018-19212-5

Sclafani, F. et al. 2019. Analysis of KRAS, NRAS, BRAF, PIK3CA and TP53 mutations in a large prospective series of locally advanced rectal cancer patients. *Int J Cancer*, doi: 10.1002/ijc.32507

Seshagiri, S. et al. 2012. Recurrent R-spondin fusions in colon cancer. *Nature* 488(7413), pp. 660-664. doi: 10.1038/nature11282

Sheller-Miller, S. et al. 2017. Damage-Associated molecular pattern markers HMGB1 and cell-Free fetal telomere fragments in oxidative-Stressed amnion epithelial cell-Derived exosomes. *J Reprod Immunol* 123, pp. 3-11. doi: 10.1016/j.jri.2017.08.003

Shelton, J. W. et al. 2013. In vitro and in vivo enhancement of chemoradiation using the oral PARP inhibitor ABT-888 in colorectal cancer cells. *Int J Radiat Oncol Biol Phys* 86(3), pp. 469-476. doi: 10.1016/j.ijrobp.2013.02.015

Shen, Z. et al. 2015. Prospective controlled study of the safety and oncological outcomes of ELAPE procure with definitive anatomic landmarks versus conventional APE for lower rectal cancer. *Eur J Surg Oncol* 41(4), pp. 472-477. doi: 10.1016/j.ejso.2015.01.017

Shiiba, M. et al. 2013. MicroRNA-125b regulates proliferation and radioresistance of oral squamous cell carcinoma. *Br J Cancer* 108(9), pp. 1817-1821. doi: 10.1038/bjc.2013.175

Shinto, E. et al. 2014. CD8+ and FOXP3+ tumor-infiltrating T cells before and after chemoradiotherapy for rectal cancer. *Ann Surg Oncol* 21 Suppl 3, pp. S414-421. doi: 10.1245/s10434-014-3584-y

Silva, J. et al. 2012. Analysis of exosome release and its prognostic value in human colorectal cancer. *Genes Chromosomes Cancer* 51(4), pp. 409-418.

Siravegna, G. et al. 2015. Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat Med* 21(7), pp. 795-801. doi: 10.1038/nm.3870

Slater, A. et al. 2016. Transanal endoscopic microsurgery for rectal cancer. *Clin Radiol* 71(2), pp. e121-129. doi: 10.1016/j.crad.2015.10.022

Smith, F. M. et al. 2012. The surgical significance of residual mucosal abnormalities in rectal cancer following neoadjuvant chemoradiotherapy. *Br J Surg* 99(7), pp. 993-1001. doi: 10.1002/bjs.8700

Smith, J. J. and Garcia-Aguilar, J. 2015. Advances and challenges in treatment of locally advanced rectal cancer. *J Clin Oncol* 33(16), pp. 1797-1808. doi: 10.1200/jco.2014.60.1054

Song, J. et al. 2018. Predictive value of carcinoembryonic antigen and carbohydrate antigen 19-9 related to downstaging to stage 0-I after neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Cancer Manag Res* 10, pp. 3101-3108. doi: 10.2147/cmar.s166417

Sorensen, C. G. et al. 2016. The diagnostic accuracy of carcinoembryonic antigen to detect colorectal cancer recurrence - A systematic review. *Int J Surg* 25, pp. 134-144. doi: 10.1016/j.ijsu.2015.11.065

Stiksma, J. et al. 2014. CA 19-9 as a marker in addition to CEA to monitor colorectal cancer. *Clin Colorectal Cancer* 13(4), pp. 239-244. doi: 10.1016/j.clcc.2014.09.004

Stranska, R. et al. 2018. Comparison of membrane affinity-based method with size-exclusion chromatography for isolation of exosome-like vesicles from human plasma. *J Transl Med* 16(1), p. 1. doi: 10.1186/s12967-017-1374-6

Sun, D. et al. 2013a. MicroRNA-31 activates the RAS pathway and functions as an oncogenic MicroRNA in human colorectal cancer by repressing RAS p21 GTPase activating protein 1 (RASA1). *J Biol Chem* 288(13), pp. 9508-9518. doi: 10.1074/jbc.M112.367763

Sun, W. et al. 2013b. The advantage of circulating tumor cells over serum carcinoembryonic antigen for predicting treatment responses in rectal cancer. *Future Oncol* 9(10), pp. 1489-1500. doi: 10.2217/fon.13.91

Sun, W. et al. 2016. Circulating tumor cells: A promising marker of predicting tumor response in rectal cancer patients receiving neoadjuvant chemo-radiation therapy. *Oncotarget* 7(43), pp. 69507-69517. doi: 10.18632/oncotarget.10875

Sun, W. et al. 2014. The role of plasma cell-free DNA detection in predicting preoperative chemoradiotherapy response in rectal cancer patients. *Oncol Rep* 31(3), pp. 1466-1472. doi: 10.3892/or.2013.2949

Sun, X. et al. 2018. Monitoring colorectal cancer following surgery using plasma circulating tumor DNA.*Oncol Lett*. Vol. 15. pp. 4365-4375.

Sun, Y. et al. 2019. Pathological complete response may underestimate distant metastasis in locally advanced rectal cancer following neoadjuvant chemoradiotherapy and radical surgery: Incidence, metastatic pattern, and risk factors. *Eur J Surg Oncol* 45(7), pp. 1225-1231. doi: 10.1016/j.ejso.2019.03.005

Surinova, S. et al. 2015. Prediction of colorectal cancer diagnosis based on circulating plasma proteins. *EMBO Mol Med* 7(9), pp. 1166-1178. doi: 10.15252/emmm.201404873

Svoboda, M. et al. 2012. MicroRNA expression profile associated with response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer patients. *Radiat Oncol* 7, p. 195. doi: 10.1186/1748-717x-7-195

Sysmex-Inostics. 2016. *OncoBEAMÂ® RAS CRC test has been granted CE Mark approval - sysmex inostics*. Privacy Policy. Available at: http://www.sysmex-inostics.com/news-and-events/press-releases/new-liquid-biopsy-ras-testing-for-metastatic-colorectal-cancer-patients-now-available-for-clinical-practices-546.html [Accessed 05/09/2019].

Takahashi, A. et al. 2017. Exosomes maintain cellular homeostasis by excreting harmful DNA from cells. *Nat Commun* 8, p. 15287. doi: 10.1038/ncomms15287

Takano, Y. et al. 2017. Circulating exosomal microRNA-203 is associated with metastasis possibly via inducing tumor-associated macrophages in colorectal cancer. *Oncotarget* 8(45), pp. 78598-78613. doi: 10.18632/oncotarget.20009

Tan, K. et al. 2018. Longitudinal monitoring reveals dynamic changes in circulating tumor cells (CTCs) and CTC-associated miRNAs in response to chemotherapy in metastatic colorectal cancer patients. *Cancer Lett* 423, pp. 1-8. doi: 10.1016/j.canlet.2018.02.039

Tang, Y. T. et al. 2017. Comparison of isolation methods of exosomes and exosomal RNA from cell culture medium and serum. *Int J Mol Med* 40(3), pp. 834-844. doi: 10.3892/ijmm.2017.3080

TCGA, N. 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487(7407), pp. 330-337. doi: 10.1038/nature11252

Teng, F. et al. 2015. Tumor infiltrating lymphocytes (TILs) before and after neoadjuvant chemoradiotherapy and its clinical utility for rectal cancer. *Am J Cancer Res* 5(6), pp. 2064-2074.

Terzi, C. et al. 2008. Survivin, p53, and Ki-67 as predictors of histopathologic response in locally advanced rectal cancer treated with preoperative chemoradiotherapy. *Int J Colorectal Dis* 23(1), pp. 37-45. doi: 10.1007/s00384-007-0376-x

Terzic, J. et al. 2010. Inflammation and colon cancer. *Gastroenterology* 138(6), pp. 2101-2114.e2105. doi: 10.1053/j.gastro.2010.01.058

Thakur, B. K. et al. 2014. Double-stranded DNA in exosomes: a novel biomarker in cancer detection.*Cell Res*. Vol. 24. England, pp. 766-769.

Theil, G. et al. 2016. The Use of a New CellCollector to Isolate Circulating Tumor Cells from the Blood of Patients with Different Stages of Prostate Cancer and Clinical Outcomes - A Proof-of-Concept Study.*PLoS One*. Vol. 11.

Thierry, A. R. et al. 2010. Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. *Nucleic Acids Res* 38(18), pp. 6159-6175. doi: 10.1093/nar/gkq421

Tie, J. et al. 2018. Serial circulating tumour DNA analysis during multimodality treatment of locally advanced rectal cancer: a prospective biomarker study. *Gut*, doi: 10.1136/gutjnl-2017-315852

Tie, J. et al. 2019. Serial circulating tumour DNA analysis during multimodality treatment of locally advanced rectal cancer: a prospective biomarker study. *Gut* 68(4), pp. 663-671. doi: 10.1136/gutjnl-2017-315852

Tie, J. et al. 2015. Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer. *Ann Oncol* 26(8), pp. 1715-1722. doi: 10.1093/annonc/mdv177

Toiyama, Y. et al. 2015. Elevated platelet count as predictor of recurrence in rectal cancer patients undergoing preoperative chemoradiotherapy followed by surgery. *Int Surg* 100(2), pp. 199-207. doi: 10.9738/intsurg-d-13-00178.1

Toiyama, Y. et al. 2013. Serum miR-21 as a diagnostic and prognostic biomarker in colorectal cancer. *J Natl Cancer Inst* 105(12), pp. 849-859. doi: 10.1093/jnci/djt101

Toiyama, Y. et al. 2016. Circulating cell-free microRNAs as biomarkers for colorectal cancer. *Surg Today* 46(1), pp. 13-24. doi: 10.1007/s00595-015-1138-y

Trakarnsanga, A. et al. 2012. Treatment of locally advanced rectal cancer: controversies and questions. *World J Gastroenterol* 18(39), pp. 5521-5532. doi: 10.3748/wjg.v18.i39.5521

Tsai, W. S. et al. 2016. Circulating Tumor Cell Count Correlates with Colorectal Neoplasm Progression and Is a Prognostic Marker for Distant Metastasis in Non-Metastatic Patients. *Sci Rep* 6, p. 24517. doi: 10.1038/srep24517

Tsongalis, G. J. et al. 2014. Routine use of the Ion Torrent AmpliSeq Cancer Hotspot Panel for identification of clinically actionable somatic mutations. *Clin Chem Lab Med* 52(5), pp. 707-714. doi: 10.1515/cclm-2013-0883

Unim, B. et al. 2019. Current Genetic Service Delivery Models for the Provision of Genetic Testing in Europe: A Systematic Review of the Literature. *Front Genet* 10, p. 552. doi: 10.3389/fgene.2019.00552

Untergasser, A. et al. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* 35(Web Server issue), pp. W71-74. doi: 10.1093/nar/gkm306

Vagner, T. et al. 2018. Large extracellular vesicles carry most of the tumour DNA circulating in prostate cancer patient plasma. *J Extracell Vesicles* 7(1), p. 1505403. doi: 10.1080/20013078.2018.1505403

Valadi, H. et al. 2007. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol* 9(6), pp. 654-659. doi: 10.1038/ncb1596

Vallard, A. et al. 2018. Outcomes prediction in pre-operative radiotherapy locally advanced rectal cancer: leucocyte assessment as immune biomarker. *Oncotarget* 9(32), pp. 22368-22382. doi: 10.18632/oncotarget.25023

Van Deun, J. et al. 2014. The impact of disparate isolation methods for extracellular vesicles on downstream RNA profiling. *J Extracell Vesicles* 3, doi: 10.3402/jev.v3.24858

van Gijn, W. et al. 2011. Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer: 12-year follow-up of the multicentre, randomised controlled TME trial. *Lancet Oncol* 12(6), pp. 575-582. doi: 10.1016/s1470-2045(11)70097-3

Vanni, I. et al. 2015. Next-Generation Sequencing Workflow for NSCLC Critical Samples Using a Targeted Sequencing Approach by Ion Torrent PGM Platform. *Int J Mol Sci* 16(12), pp. 28765-28782. doi: 10.3390/ijms161226129

Vittori, L. N. et al. 2019. Circulating Cell-Free DNA in Physical Activities. *Methods Mol Biol* 1909, pp. 183-197. doi: 10.1007/978-1-4939-8973-7_14

Vivancos, A. et al. 2018. Circulating cell-free DNA as predictor of treatment failure after neoadjuvant chemoradiotherapy before surgery in patients with locally advanced rectal cancer: is it ready for primetime? *Ann Oncol* 29(3), pp. 532-534. doi: 10.1093/annonc/mdy043

Vlassov, A. V. et al. 2012. Exosomes: current knowledge of their composition, biological functions, and diagnostic and therapeutic potentials. *Biochim Biophys Acta* 1820(7), pp. 940-948. doi: 10.1016/j.bbagen.2012.03.017

Voelkerding, K. V. et al. 2009. Next-generation sequencing: from basic research to diagnostics. *Clin Chem* 55(4), pp. 641-658. doi: 10.1373/clinchem.2008.112789

Vogelstein, B. and Kinzler, K. W. 1999. Digital PCR. *Proc Natl Acad Sci U S A* 96(16), pp. 9236-9241.

Wan, J. F. et al. 2018a. Aneuploidy of chromosome 8 and mutation of circulating tumor cells predict pathologic complete response in the treatment of locally advanced rectal cancer. *Oncol Lett* 16(2), pp. 1863-1868. doi: 10.3892/ol.2018.8831

Wan, Z. et al. 2018b. Exosome-mediated cell-cell communication in tumor progression. *Am J Cancer Res* 8(9), pp. 1661-1673.

Wang, C. et al. 2019. Droplet digital PCR improves urinary exosomal miRNA detection compared to real-time PCR. *Clin Biochem* 67, pp. 54-59. doi: 10.1016/j.clinbiochem.2019.03.008

Wang, C. J. et al. 2010. Suppression of microRNA-31 increases sensitivity to 5-FU at an early stage, and affects cell migration and invasion in HCT-116 colon cancer cells.*BMC Cancer*. Vol. 10. p. 616.

Wang, H. et al. 2012. Circulating MiR-125b as a marker predicting chemoresistance in breast cancer. *PLoS One* 7(4), p. e34210. doi: 10.1371/journal.pone.0034210

Wang, J. et al. 2014. Identification of a circulating microRNA signature for colorectal cancer detection. *PLoS One* 9(4), p. e87451. doi: 10.1371/journal.pone.0087451

Wang, W. et al. 2017. Characterization of the release and biological significance of cell-free DNA from breast cancer cell lines. *Oncotarget* 8(26), pp. 43180-43191. doi: 10.18632/oncotarget.17858

Wang, X. et al. 2018. Preoperative short-course radiotherapy and long-course radiochemotherapy for locally advanced rectal cancer: Meta-analysis with trial sequential analysis of long-term survival data. *PLoS One* 13(7), p. e0200142. doi: 10.1371/journal.pone.0200142

Ward, W. H. et al. 2018. Predictive Value of Leukocyte- and Platelet-Derived Ratios in Rectal Adenocarcinoma. *J Surg Res* 232, pp. 275-282. doi: 10.1016/j.jss.2018.06.060

Wculek, S. K. and Malanchi, I. 2015. Neutrophils support lung colonization of metastasis-initiating breast cancer cells. *Nature* 528(7582), pp. 413-417. doi: 10.1038/nature16140

Webber, J. and Clayton, A. 2013. How pure are your vesicles? *J Extracell Vesicles* 2, doi: 10.3402/jev.v2i0.19861

Weber, M. et al. 2003. A real-time polymerase chain reaction assay for quantification of allele ratios and correction of amplification bias. *Anal Biochem* 320(2), pp. 252-258. doi: 10.1016/s0003-2697(03)00396-8

Welton, J. L. et al. 2015. Ready-made chromatography columns for extracellular vesicle isolation from plasma. *J Extracell Vesicles* 4, p. 27269. doi: 10.3402/jev.v4.27269

Wennerberg, E. et al. 2017. Immune recognition of irradiated cancer cells. *Immunol Rev* 280(1), pp. 220-230. doi: 10.1111/imr.12568

Wilkins, A. et al. 2019. Radiotherapy enhances responses of lung cancer to CTLA-4 blockade.*J Immunother Cancer*. Vol. 7. England, p. 64.

Willis, J. E. 2007. The pathologist's role in rectal cancer patient assessments. *Clin Colon Rectal Surg* 20(3), pp. 158-166. doi: 10.1055/s-2007-984860

Witwer, K. W. et al. 2013. Standardization of sample collection, isolation and analysis methods in extracellular vesicle research. *J Extracell Vesicles* 2, doi: 10.3402/jev.v2i0.20360

Wojtukiewicz, M. Z. et al. 2017. Platelets and cancer angiogenesis nexus. *Cancer Metastasis Rev* 36(2), pp. 249-262. doi: 10.1007/s10555-017-9673-1

Wood, G. et al. 2017. Derived neutrophil to lymphocyte ratio as a prognostic factor in patients with advanced colorectal cancer according to RAS and BRAF status: a post-hoc analysis of the MRC COIN study. *Anticancer Drugs* 28(5), pp. 546-550. doi: 10.1097/cad.0000000000000488

Xiao, W. W. et al. 2015. A Low Lymphocyte-to-Monocyte Ratio Predicts Unfavorable Prognosis in Pathological T3N0 Rectal Cancer Patients Following Total Mesorectal Excision. *J Cancer* 6(7), pp. 616-622. doi: 10.7150/jca.11727

Xu, Y. and Her, C. 2015. Inhibition of Topoisomerase (DNA) I (TOP1): DNA Damage Repair and Anticancer Therapy. *Biomolecules* 5(3), pp. 1652-1670. doi: 10.3390/biom5031652

Yamamoto, A. et al. 2019a. Clinical Implications of Pretreatment: Lymphocyte-to-Monocyte Ratio in Patients With Rectal Cancer Receiving Preoperative Chemoradiotherapy. *Dis Colon Rectum* 62(2), pp. 171-180. doi: 10.1097/dcr.0000000000001245

Yamamoto, Y. et al. 2019b. Clinical significance of the mutational landscape and fragmentation of circulating tumor DNA in renal cell carcinoma. *Cancer Sci* 110(2), pp. 617-628. doi: 10.1111/cas.13906

Yan, H. B. et al. 2018. Expression of microRNA-99a-3p in Prostate Cancer Based on Bioinformatics Data and Meta-Analysis of a Literature Review of 965 Cases. *Med Sci Monit* 24, pp. 4807-4822. doi: 10.12659/msm.908057

Yang, J. et al. 2019. Analyses of predictive factors for pathological complete remission in neoadjuvant therapy for locally advanced rectal cancer. *J buon* 24(1), pp. 77-83.

Yang, N. et al. 2018. The characteristics of ctDNA reveal the high complexity in matching the corresponding tumor tissues. *BMC Cancer* 18(1), p. 319. doi: 10.1186/s12885-018-4199-7

Yang, S. et al. 2017. Detection of mutant KRAS and TP53 DNA in circulating exosomes from healthy individuals and patients with pancreatic cancer. *Cancer Biol Ther* 18(3), pp. 158-165. doi: 10.1080/15384047.2017.1281499

Yang, X. et al. 2016. Quantification of mutant alleles in circulating tumor DNA can predict survival in lung cancer. *Oncotarget* 7(15), pp. 20810-20824. doi: 10.18632/oncotarget.8021

Yasuda, K. et al. 2011. Density of CD4(+) and CD8(+) T lymphocytes in biopsy samples can be a predictor of pathological response to chemoradiotherapy (CRT) for rectal cancer. *Radiat Oncol* 6, p. 49. doi: 10.1186/1748-717x-6-49

Yin, H. et al. 2018. MiR-99a Enhances the Radiation Sensitivity of Non-Small Cell Lung Cancer by Targeting mTOR. *Cell Physiol Biochem* 46(2), pp. 471-481. doi: 10.1159/000488615

Yoshimura, A. et al. 2018. Exosomal miR-99a-5p is elevated in sera of ovarian cancer patients and promotes cancer cell invasion by increasing fibronectin and vitronectin expression in neighboring peritoneal mesothelial cells. *BMC Cancer* 18(1), p. 1065. doi: 10.1186/s12885-018-4974-5

Yu, J. I. et al. 2016. Prognostic significance of survivin in rectal cancer patients treated with surgery and postoperative concurrent chemo-radiation therapy. *Oncotarget* 7(38), pp. 62676-62686. doi: 10.18632/oncotarget.10445

Zeitoun, G. et al. 2019. The Immunoscore in the Clinical Practice of Patients with Colon and Rectal Cancers. *Chirurgia (Bucur)* 114(2), pp. 152-161. doi: 10.21614/chirurgia.114.2.152

Zhang, F. et al. 2017. Metabolomics for biomarker discovery in the diagnosis, prognosis, survival and recurrence of colorectal cancer: a systematic review. *Oncotarget* 8(21), pp. 35460-35472. doi: 10.18632/oncotarget.16727

Zhang, L. N. et al. 2015. Elevated CA19-9 as the Most Significant Prognostic Factor in Locally Advanced Rectal Cancer Following Neoadjuvant Chemoradiotherapy. *Medicine (Baltimore)* 94(45), p. e1793. doi: 10.1097/md.0000000000001793

Zhao, A. et al. 2012. MicroRNA-125b induces cancer cell apoptosis through suppression of Bcl-2 expression. *J Genet Genomics* 39(1), pp. 29-35. doi: 10.1016/j.jgg.2011.12.003

Zhao, J. et al. 2017. Clinical and Prognostic Significance of Pathological and Inflammatory Markers in Mucinous Rectal Cancer Patients Receiving Neoadjuvant Chemoradiotherapy and Curative Surgery. *Med Sci Monit* 23, pp. 4826-4833. doi: 10.12659/msm.904116

Zhao, Q. et al. 2015. Circulating miRNAs is a potential marker for gefitinib sensitivity and correlation with EGFR mutational status in human lung cancers. *Am J Cancer Res* 5(5), pp. 1692-1705.

Zheng, G. et al. 2014. Serum microRNA panel as biomarkers for early diagnosis of colorectal adenocarcinoma. *Br J Cancer* 111(10), pp. 1985-1992. doi: 10.1038/bjc.2014.489

Zhu, J. et al. 2015. Targeted serum metabolite profiling and sequential metabolite ratio analysis for colorectal cancer progression monitoring. *Anal Bioanal Chem* 407(26), pp. 7857-7863. doi: 10.1007/s00216-015-8984-8

Zhu, M. et al. 2017. A panel of microRNA signature in serum for colorectal cancer diagnosis. *Oncotarget* 8(10), pp. 17081-17091. doi: 10.18632/oncotarget.15059

Zitt, M. et al. 2008. Circulating cell-free DNA in plasma of locally advanced rectal cancer patients undergoing preoperative chemoradiation: a potential diagnostic tool for therapy monitoring. *Dis Markers* 25(3), pp. 159-165.

Zwirner, K. et al. 2018. Circulating cell-free DNA: A potential biomarker to differentiate inflammation and infection during radiochemotherapy. *Radiother Oncol*, doi: 10.1016/j.radonc.2018.07.016

9 Appendices

9.1 Application for Access to Samples to the Wales Cancer Bank

APPLICATION FOR ACCESS TO SAMPLES (v2.8 June 2015)

I. This application (v2.8) is intended for the use and processing of samples utilised by the laboratory and/or personnel that fall under the supervision of the Principal Investigator listed in the application. Any transfer of samples or aliquots to personnel or laboratories that are not under the supervision of the indicated PI requires the following:
An explanation of the need to transfer the materials and benefit to the investigator's research
A copy of the enclosed Wales Cancer Bank agreement page signed by the collaborator

The WCB does not supply samples to banks solely for distribution to third party researchers; those researchers should be encouraged to apply to the WCB directly.

The information requested in these forms is necessary in order to document correctly your request for tissue and other services and to ensure that the WCB operates within the guidelines of the Human Tissue Authority. When submitting a written request for supply of material:

Please print neatly or type.

Patient identity is confidential. Samples will be coded and supplied with a minimum data set. The cost recovery and/or processing fee per sample will vary according to the type of sample requested.

The Wales Cancer Bank is authorised by the REC for Wales to release samples to researchers. Researchers receiving samples from WCB are NOT required to have approval from NRES for the use of these samples as samples will be provided anonymously with only the minimum data set. However, researchers must be able to satisfy the External Review Panel of the WCB that the project they submit is both ethically and scientifically valid. IF researchers are already in possession of NRES approval for their projects, a copy of the NRES letter should be supplied with the application. Researchers are advised that it is their responsibility to ensure that they comply with the Human Tissue Act or other appropriate laws that cover the use of human material in research. An HTA licence is NOT needed to store tissue sourced from WCB for an approved project that is subject to a signed Material Transfer Agreement.

If the research has been peer reviewed by a recognised funding body as part of a grant application, please include the grant application and approval letter. If the sample collection and use of samples is clearly detailed (to the satisfaction of WCB) in the approved application, WCB will not send it for further external review.

Transfer of samples from WCB to researchers will be by Courier. Researchers are required to cover the cost of transport of their samples and supply appropriate customs declarations if appropriate.

Please email the completed application form to wcbresearchapplications@cf.ac.uk and send hard copy of the signed Material Transfer Agreement (final page) to:

Wales Cancer Bank
Cardiff University
2nd Floor, Room 2LB4 64
A Block

Main Building, University Hospital of Wales
Heath Park
Cardiff
CF14 4XN

For additional information please contact the Wales Cancer Bank Secretariat on +44 (0)2920 743243

II. INVESTIGATOR DATA

A. Principal Investigator Adams

Adams	Richard A	MD
Last Name	*First Name*	*Middle Initial*

Degree

Investigator's Title	Dr
Address	**Institute of Cancer & Genetics** Cardiff University School of Medicine Velindre Hospital Cardiff
Post code	
Phone / Fax	029 2031 6206
Email	Richard.Adams@wales.nhs.uk

Contact Person (if different from above)

Name	Panayiotis Georgiades
Contact number	07792838208
Email	georgiadespa@cardiff.ac.uk

B. Shipping Address (*if different from above*):

Institute of Medical Genetics Building

Cardiff University

Heath Park Campus

Post code

C. Invoice information. Is a purchase order required for shipment of specimens to your institution?

Yes No [X] Yes, please supply purchase order when project has been approved.

Invoices will be sent to the shipping address listed in section B. If you would like the original invoice to be provided by post to another location (eg. your finance department), please enter that address below. A shipping list will be included with the samples, please complete and fax back to WCB to acknowledge receipt.

Person to whom invoice should be addressed (*if different from above*):

Rachel Butler

Invoice Address (*if different from above*):

Institute of Medical Genetics, UHW. CF14 4XW

Post code

Courier services are provided by either DHL (Europe and Japan) or Federal Express (USA). Please state if you require another Courier and provide the appropriate customer number below.

Preferred Courier _____Customer Number _____

III. RESEARCH INFORMATION

Specimens will be provided to all investigators, based either in Academia or Industry

A. Please indicate the source of funds for your proposed project. If this is Institutional Funding, please enclose a letter from your Head of Department indicating that funds and premises are available to complete your project.

Funding Source	Period of Support
Richard Adams PhD research account	January 2016 to January 2018

B. Please provide a short lay summary (max 200 words) of the intended research. Please note, this information will be used in WCB reports to funders, NRES, in the public Annual Report and on the WCB website so only include information that is not commercially sensitive. This summary will be reviewed by a lay representative.

Over the past decade, the advancement of sequencing technologies has greatly improved our understanding of the mutational spectrum of colorectal cancer (CRC). We now have greater knowledge of the location of 'hotspots' where mutations are likely to occur, which can affect treatment and prognostis. We can apply this knowledge by interrogating circulating biomarkers such as circulating tumour DNA (ctDNA), exosomal nucleic acids and tumour educated platelets (TEPs) and proteins in patient blood. Exosomes are cell-derived vesicles that are present in many biological fluids. TEPs are platelets, which have been shown to sequester cell-free nucleic acids.

ctDNA levels within plasma have been investigated as a sensitive method of monitoring patients and predict relapse and response to treatment in certain cancer types. Circulating exosomes originate from tumour and/or wild type cells, providing nucleic acids with more structural stability and preventing degradation. We plan to use exosome, platelet and protein isolation techniques with sequencing to interrogate and compare all biomarkers in CRCs at diagnosis and throughout therapy.

These studies may provide clinicians with a non-invasive method with which to assess patient response and monitor progress. This has potential implications in choice of treatment and enables clinicians to halt toxic treatments that appear ineffective.

C. Please provide the **title** and a short research summary (2-4 pages of A4) of the proposed research on the samples you are requesting from the WCB (*use additional pages where necessary*). Sufficient information should be provided to enable the External Review Panel to determine the scientific validity of your study. Please fully justify the number and type of samples requested and address ALL the headings below.

If your project has already been successfully peer reviewed as part of a grant application, you may supply a copy of the scientific part of that grant application and confirmation of the grant award. If details of the sample collection and usage are clearly shown in the application (to the satisfaction of WCB), no further information is required in this form. Please go to the Material Transfer Agreement.

Title:

Temporal changes in circulating biomarkers in patients undergoing chemo or radiotherapy in colorectal cancers

ii. Introduction (including an overview of the state of the art in your proposed project area):

Problems regularly arising in molecular diagnostics by sequencing tumour biopsies include;

Inadequate tissue being available from the diagnostic specimen
No evidence of invasive cancer in diagnostic specimen
No tumour sample being available for analysis due to the location of tumour blocks being unknown
For patients whose CRC was first diagnosed before progression to metastatic disease, the use of archival pre-metastatic tumour sample for RAS testing instead of a recent sample, to avoid re-biopsy.
Heterogeneity of tumours such that a single biopsy is not indicative of the overall tumour load
The time taken to locate and retrieve tumour samples, which can cause delays in RAS testing turnaround times that prohibit timely treatment of eligible patients with anti-EGFR therapy
The cost of retrieval and transport of tumour samples, which places a burden on the pathology laboratories

Research in recent years has had particular emphasis on the use of liquid biopsies to aid both in monitoring disease and in determining treatment options. It has recently become routine clinical practice, to sequence for mutations in the tumour biopsy, in genes such as *KRAS* and *NRAS* in CRCs (Misale et al. 2014). Data from such analyses, however, can be limited as they provide a single snapshot in time of a particular region of the tumour. Deep sequencing of different circulating biomarkers, such as ctDNA and exosomal nucleic acids, has been hypothesised to provide an unbiased overview of the genetic landscape of the tumour, and any metastases, at the time of the biopsy being taken (Diaz et al. 2012). It should also be noted that other circulating biomarkers, such as circulating tumour cells and cell free microRNA have also been analysed as circulating biomarkers.

Cell-Free Tumour DNA (ctDNA)

Recent studies suggest that a complex array of genetic alterations exist in all human cancers. Genetic alterations, which include gene deletions, gene amplifications, point mutations, and chromosomal rearrangements, play a major role in the development and progression of cancers and can be utilised as unique identifiers that distinguish tumour cells from their wild type counterparts. The exclusive nature of tumour-defining genetic alterations makes them attractive systemic biomarkers with a theoretical specificity of 100%, when detectable.

As cells turn over, nucleic acids are released into the systemic circulation. In patients with a malignancy, a fraction of the circulating nucleic acids (CNAs) are tumour-derived (Diehl et al. 2005). Distinguishing the fraction of tumour-derived nucleic acids in plasma from normal nucleic acids has been accomplished by screening for genetic alterations. The most amenable

and best studied genetic alterations for analysis from the CNA pool are point mutations. Unlike deletions and amplifications, point mutations can easily be differentiated from the complementary normal nucleic acids in plasma. In addition, point mutations are often found in clusters and in a large percentage of patients with a given tumour type.

Deep sequencing studies into both hereditary and somatic CRCs have provided a large amount of information about the molecular alterations critical for the development of CRC. This knowledge allows us to perform targeted sequencing on a panel of 'hotspot' genes where mutations are likely to occur, as well as give a potential prognostic indicator in the case of KRAS mutations (Diaz et al. 2012). These mutations are detected in the blood and can be used to monitor levels of disease and detect the emergence of resistant clones.

Exosomal Nucleic Acids

In recent years, there has been an increasing interest in investigating circulating exosomes as biomarkers in a large number of cancers. Exosomes are small microvesicles (30-120nm in diameter), which have a role in cell-cell communication (Raposo and Stoorvogel 2013). They perform these roles by budding off from their parent cell to later fuse with a target cell and deposit its cargo, which includes, DNA proteins, mRNA and microRNAs. It has thus been hypothesized that the contents of tumour exosomes can give an indication as to the genetic profile of their parent tumour cells by sequencing the nucleic acids stored. Nucleic acids stored in exosomes have also found to be more stable (Kahlert et al. 2014) allowing larger fragments to be interrogated for larger-scale genetic changes, chromosomal aberrations or expression levels.

Exosomes are present in the circulation of both healthy and cancer patients (Vlassov et al. 2012) but have been reported to be at increased levels in cancer patients and during cancer progression (Ko et al. 2015). Studies have since looked into using the genetic contents of circulating exosomes as biomarkers in a variety of cancers including ovarian (Taylor and Gercel-Taylor 2008), lung (Rabinowits et al. 2009), prostate (Hessvik et al. 2012) and, more recently, colorectal (Matsumura et al. 2015), with a large emphasis looking at exosomal microRNA (miRNA). These works have generally been looking to define miRNA profiles for early detection, monitoring and prediction of prognosis and recurrence in these cancers, though much work still needs to be done validate these studies for clinical utility.

Tumour Educated Platelets

Platelets have very recently emerged in the literature as a potential biomarker and source of tumour RNA (Nilsson et al. 2015). Studies have shown that platelets can carry tumour nucleic acids in circulation, the methods of which are not entirely understood. It has been demonstrated that platelets can accumulate tumour derived nucleic acids through microvesicle-dependent mechanisms, e.g. exosomes in circulation (Nilsson et al. 2015). These have been referred to as tumour educated platelets (TEPs). TEPs have been hypothesized to act through microvesicle-independent mechanisms, such as by sequester circulating free or protein bound nucleic acids in circulation (Best et al. 2015; Nilsson et al. 2011). Though this was shown to be the case some time ago (Nilsson et al. 2011), it has only very recently been attempted as a biomarker in certain cancer types.

Proteins: as an additional component of this pilot we will aim to see if proteins are effectively stabilised in Streck tubes, as this will inform future optimisation of collections.

Here we propose a translational research study to define the utility of exosomes and TEPs to improve our understanding of cancer progression and prognosis and to offer insights in to the evolution and heterogeneity of colorectal cancer.

ii. Aim(s) (please clearly state the aim(s) of your project):

Primary project objective

To evaluate, in patients with newly diagnosed and/or advanced CRC, the feasibility of extracting exosomes from plasma, the best anticoagulants in which to store patient samples and whether sequencing results from exosomes correlate with that of ctDNA and the primary tumour in FFPE. These samples will also be used to determine and optimize the best method with which to isolate exosomes from patient plasma and the best technique with which to validate the presence and concentration of exosomes. The same investigations will then be performed looking at optimizing isolation and nucleic acid extraction of tumour derived nucleic acids in TEPs. These will then be compared to circulating exosomes, cell free nucleic acids and the FFPE tumour biopsy.

Secondary project objectives

To assess the concentration and fragment size of nucleic acids extracted from circulating exosomes and tumour derived nucleic acids from platelets. Samples may also be used to determine whether mutations or genetic variants detected in exosomes and platelets correlate with those found in the original tumour biopsy and compared wild-type tissue located in the FFPE block and/or buffy coat.

We shall also aim to assess the impact of timing of treatment and of diurnal variation on exosome and/or platelet secretion to the circulation.

Other objectives

Samples may also be used for quality control and other assay development purposes.

iv. Experience of group and/or company carrying out analysis (please provide information to indicate that your research group has experience in the techniques you intend to use, either by use of preliminary data from other work carried out in your group or by providing references to publications from your group/company that are relevant to this application):

Recent work has been performed in this group using the Proton Torrent (Life Technologies) and droplet digital PCR (Bio-Rad) in order to look at levels of ctDNA in three different lung cancer subtypes in 30 patients sequentially throughout chemo/radiotherapeutic treatment. This work was performed by Dr D. Nelmes and supported by WCB. The samples for this project will thus be expanding on this study in lung cancers and focussing on CRCs. Collaborations with Dr A. Clayton's group, who have experience in extracting and quantitating exosomes in both plasma (Welton et al., 2015) and analysing exosomes functionally in cell-lines (Chowdhury et al., 2015), will allow us to optimize our techniques and compare and contrast ctDNA with exosomal nucleic acids as potential biomarkers of CRC. Though no work has been performed using platelets by our group, platelet extraction and storage are routinely used in diagnostics, and nucleic acid extraction and analysis from these samples will be similar to that of exosomes.

References

Chowdhury, R. et al. 2015. Cancer exosomes trigger mesenchymal stem cell differentiation into pro-angiogenic and pro-invasive myofibroblasts. *Oncotarget* 6(2), pp. 715-731

Welton, J. L. et al. 2015. Ready-made chromatography columns for extracellular vesicle isolation from plasma. *J Extracell Vesicles* 4, p. 27269.

v. If hypothesis generation is the specific purpose of your application, what do you envisage its application in the clinical setting will be?

Please give as much detail as possible on target identification, validation etc

We hypothesise that one may be able to track minimal residual disease, detect mutations of interest and potentially subgroup CRC patients based on DNA sequencing and mRNA expression data from exosomes and/or TEPs.

vi. Methods (please detail the methods you intend to use, indicating controls and the experimental design you will use where relevant include statistical information):

Sample storage

Samples collected will be spun down stored as plasma at -80°C until exosomes can be isolated and quantified and nucleic acids can be extracted. We shall explore collection both in Streck tubes and in EDTA tubes following standardized local protocols. All samples will be stored with their unique anonymised WCB identifier and no other details. In the case of TEP studies, platelets may be isolated and lysed within 24 hours and stored at -80°C in lysis buffer until RNA extraction is performed.

Exosomal isolation

Several different methods of exosomal isolation will be assessed during this pilot, including ultracentrifugation, ultracentrifugation with a sucrose gradient and kits manufactured for the purpose such as Exoquick (Systembio), PureExo (101bio), Exo-Spin (CellGS) or ExoEasy Maxi/Midi kit (Qiagen).

Exosomal Validation

In order to determine the presence of exosomes and their concentration from the sample taken, several methods may be applied. Use of electron microscopy can be applied to visualise the microvesicles and determine their size and shape. Use of Nanosight has a similar function in a real-time scenario. Exosomes can be further validated by their surface markers (CD63, CD9 and/or CD81) through flow cytometry or western blotting.

Assessment of Nucleic Acid Content and Concentration

The concentration of nucleic acids extracted from exosomes can be assessed by using the Bioanalyzer 2100 to determine both concentration and fragment length of the nucleic acids. This can then be validated using the Bio-Rad droplet digital PCR.

Panel Sequencing

Exosomal nucleic acids will likely be sequenced and compared to results from FFPE tumour tissue using next generation sequencing based techniques. This will involve the application of a 50 gene cancer 'hotspot' panel and the Proton Torrent both supplied by Life Technologies or whole exome/transcriptome sequencing on the HiSeq Supplied by Illumina. Mutations detected in the FFPE will be interrogated in exosomal nucleic acids to ensure all results match and/or correlate.

Data collection and management

Patients identified and recruited to this research, will be consented by WCB. On consent, they are assigned a unique, anonymised WCB research identifier. Clinical/research samples collected from each patient will be labelled with their unique WCB identifier. For each patient, paired clinical data, e.g. scan results will also be labelled with the unique WCB identifier and anonymised.

Data storage, relating to clinical information or results from lab work on clinical research samples, will be done using the anonymised WCB numbers, on secured Cardiff University computer and server.

Data analysis

Samples will be processed using only the anonymised WCB identifier, and results generated to be analysed will be done using the same.

vii. How many samples do you need to do your study and explain why you need these numbers? This is a finite resource so therefore it is important for us to understand the number and type of samples required. NB. Larger sample requests will require a more detailed explanation.

We would initially like to intake 20 patient samples, being baseline plasma and matching FFPE tumour biopsy. This will be in order to optimize certain aspects of the project, such as the anticoagulant for the blood samples the exosome and platelet isolation techniques and the nucleic acid extraction techniques.

Once protocols have been optimized, we would like to request samples from approximately 40 patients (20 undergoing chemotherapy and 20 undergoing radiotherapy), with one baseline and three sequential blood samples from easch patient. These patients can be a mixture of early and advanced/metastatic CRC, with preferably more emphasis on advanced (15:5 in each). This would allow us to directly assess the ability extract appropriate samples in a quality controlled fashion and also to compare the effects of timing of samples in the therapeutic strategy and their relation to outcome.

IV. SPECIMENS REQUESTED *(Please refer to the biosample search facility on www.walescancerbank.com to ascertain the type of tissue or tissue products available The*

WCB will undertake bespoke collection of material if required. If you require a bespoke collection please supply details of your request on a separate sheet.)

Please specify **exactly** what you require e.g. 20 samples of RNA extracted from ER positive invasive ductal carcinoma of the breast RIN>7

Please list sample requirements and format (RNA, serum etc). Ensure any age, pathological sub-types are clearly indicated

Sample Information Required: *(Anatomic site of tissue, pathological diagnosis, patient age, sex will be provided for all samples.)* **Additional patient information may be available, but you must request it in this application and justify its necessity for your research. It may be possible to provide some samples with details of treatment and outcome – although this may not be possible for all samples.**

For single sample patients; two plasma samples collected in both Streck DNA (10ml) and EDTA tubes (10ml), the latter spun and frozen within 1 hour.
For multi-sample patients; three sequential plasma samples of patients undergoing chemo and/or radiotherapy for colorectal cancer collected in Streck tubes (10-20ml) and/or EDTA tubes (10-20ml); additional to a baseline sample at diagnosis
Corresponding FFPE tissue (4-5 slides) from tumour samples at diagnosis for each patient where KRAS/NRAS or general mutational status is unavailable. Where possible, these slides would preferably contain both tumour and normal adjacent tissue.
The patients can either be consented by Dr Adams or WCB.
This is a prospective study and all samples must be taken fresh and not be historical

AGREEMENT FOR USE OF TISSUE

The recipient/investigator agrees that the tissues provided by the Wales Cancer Bank (WCB) will be used only for the purposes specified in this application. The recipient agrees not to attempt to obtain information identifying the individuals providing tissues to the WCB. The recipient agrees that it shall not sell any portion of the tissues provided by the WCB, or products directly extracted from these tissues (e.g. protein, mRNA or DNA). The recipient also agrees that they shall not transfer tissue (or any portion thereof) supplied by the WCB to third parties without the prior written permission of the WCB. Any subsequent transfer that may be made to other parties, with prior agreement from WCB, will require signature of this agreement between the final recipients of the material and the WCB.

The recipient understands that while the WCB attempts to avoid providing tissues that are contaminated with highly infectious agents such as hepatitis and HIV, all tissues should be handled as if potentially infectious. The individuals who have supplied tissue to the WCB have not agreed to have clinical tests performed on this tissue (e.g. for the presence of infective agents such as hepatitis), therefore, the recipient agrees not to perform such tests on the tissues supplied by the WCB**.** The recipient acknowledges that the institution where the tissue will be used follows Human Tissue Authority or appropriate local regulations if outside England, Wales and Northern Ireland, for handling human specimens and will instruct their staff to abide by those rules. The recipient further agrees to assume all responsibility for informing and training personnel in the dangers and procedures for safe handling of human tissues.

Tissues are provided as a service to the research community without warranty of merchantability or fitness for a particular purpose or any other warranty, express or implied. The WCB accepts no responsibility for any injury (including death) damages or loss that may arise either directly or indirectly from their use.

The recipient agrees to acknowledge the contributions of the Wales Cancer Bank in all publications resulting from the use of these tissues. Recommended wording to the methods or acknowledgement section is as follows: *"Tissue samples were obtained from the Wales Cancer Bank which is funded by the Welsh Government and Cancer Research Wales. Other investigators may have received specimens from the same subjects."*

The institution agrees to assume all risks and responsibility in connection with the receipt, handling, storage and use of tissues from the Wales Cancer Bank. It further agrees to indemnify and hold harmless the Wales Cancer bank and the Welsh Government from any claims costs, damages or expenses resulting from the use of the tissues provided by the WCB. The undersigned warrant that they have authority to execute this agreement on behalf of the recipient institution.
BY MY SIGNATURE I AGREE TO THE TERMS SET FORTH IN THE ABOVE AGREEMENT

Richard Adams	Institute of Cancer & Genetics	Velindre Cancer Centre
Typed Name of Principal Investigator	**Institution**	**Division or Department**

02/02/2016		
Signature of Principal Investigator	**Date**	**Institutional signatory (if applicable)**

UPON RECEIPT OF THESE SIGNED UNDERSTANDINGS AND THE INFORMATION REQUESTED ABOVE, THE WALES CANCER BANK WILL CONSIDER THIS REQUEST. Specific questions about your application should be directed to Dr Alison Parry-Jones, Project Manager, Wales Cancer Bank, Cardiff University, A2 corridor, 2 Floor, Main Building, University Hospital of Wales, Heath Park, Cardiff CF14 4XN. Tel: +44 (0)2920 743243, Fax: +44 (0)2920 744309, Email: parry-jonesa@cardiff.ac.uk

Further Amendments

Query plasma preparation methodology and whether appropriate for cfDNA, exosomes and platelet DNA – standard methods for cfDNA isolation recommend that include 2 centrifugation steps.

Two centrifugation steps will be performed for any sample tubes which will be analysed for ctDNA only (shown in modified section below). For platelets and (potentially) exosomes, these will also undergo two-step centrifugation of different speeds to those for ctDNA. We will first produce platelet-rich plasma with a slow spin followed by a faster spin to pellet platelets and leave platelet-poor-plasma. Isolation of exosomes from platelet-poor-plasma will be tested for efficacy.

Plasma Isolation

For cell-free nucleic acid and exosome isolation, samples will be spun twice, at 1000rpm for 10min, 2ml plasma aliquots made and spun again at 4000rpm for 10 min to isolate plasma. For platelet isolation, blood will be spun down at 150rcf for 10min to produce platelet-rich plasma, aliquots of which will be spun again at 1000rpm to pellet the platelets and produce platelet-

poor plasma. This platelet poor plasma will also be examined for potential exosome isolation and cell-free nucleic acid utility.

Applicant is requesting plasma to be frozen at -80oC within 1 hour of blood sample collection – is this suitable for the subsequent preparations of exosomes and platelet DNA.

We will perform studies to determine whether freezing has any effect on the RNA content of exosomes or platelets. Platelets will, in fact, by lysed in a lysis buffer before being frozen, as suggested by other researchers in the field (see modified section below). If we find that freezing of lysed cells or plasma affects RNA quality at any point, we will alter protocols to process samples fully through to RNA extraction before freezing.

Sample storage

Streck tube samples will be processed into plasma within 72 hours of collection, being left at room temperature in-between. EDTA samples will be processed and platelets and platelet-poor-plasma isolated within 24 hours of collection, being left at room temperature until processing. Variation in times and temperatures before processing will be analysed for future clinical utility. Plasma will be stored at -20°C to -80°C before exosome and/or ctDNA isolation, depending on manufacturer's instructions and time-frame before extraction. If isolation or extraction immediately follows plasma centrifugation, samples will be kept on ice. This is a preliminary plan and may be altered if found certain conditions affect exosome quality or quantity.

All samples will be stored with their unique anonymised WCB identifier and no other details. In the case of TEP studies, platelets will be isolated and lysed within 24 hours and stored at -80°C in lysis buffer until RNA extraction is performed

Mention in hypothesis that will include mRNA expression data from exosomes and platelet nucleic acids – this is not evident in any other part of the proposal. Are they going to measure mRNA expression?

Yes we will be sequencing and quantifying mRNA expression data from exosomes and platelets.

Amended Lay Summary

Recent improvements of DNA sequencing technology have improved our knowledge colorectal cancer (CRC) genetics. CRC is a very complex disease, associated with several genetic changes and pathways. We will use biological markers to detect important genetic changes in the blood and personalise patient therapy. These biomarkers include ctDNA, exosomes and platelets.

Each of these is a different molecule which can be found in the blood. We can examine each of them by analysing their protein, DNA or RNA sequence and quantities. RNA is an intermediate molecule between DNA and protein, and is often measured to calculate DNA expression levels.

We will isolate proteins and sequence DNA or RNA from each of these sources. We will investigate and compare ctDNA, platelets and exosomes as biomarkers in CRC. Samples will be taken at diagnosis and throughout therapy, to see which is most effective at disease monitoring and sub-grouping.

These studies may provide clinicians with a non-invasive method to assess patient response and monitor progress. The use of blood-based testing to monitor patients can potentially allow regular testing. This

testing can also include monitoring disease burden and choosing therapies which would benefit patients most. This can also enable clinicians to halt ineffective toxic treatments.

9.2 Clinical Factors Acquire for Patients from the Wales Cancer Bank

Table 48: Patient clinical factors for the small group of 16 patients consented by the WCB

Patient Clinical Factors
cT Stage
cN Stage
cM Stage
CEMVI+/-
ypT
ypN
EMVI
R0/R1/R2
CEA concentration
Tumour location: High/Mid/low
Baseline Neutrophil Count
Baseline Lymphocyte Count
Baseline Monocyte Count
Baseline Platelet Count
Baseline Haemoglobin Concentration
White Cell Count
Lymphocyte Count at Week 5
Neutrophil Count at Week 5
***KRAS* Status**
***NRAS* Status**
***BRAF* Status**
Surgery
Follow-Up
Radiotherapy (Gy)
mrTRG
ypTRG
Censored Alive
Censored Disease Free
Additional Notes

9.3 Cancer Genes using cBioPortal

Table 49: List of cancer relevant genes frequently mutated in patients with rectal cancer from cBioportal

Gene	Frequency Mutated (%)	Gene	Frequency Mutated (%)
APC	78.6	LRRK2	6.4
TP53	71.4	CREBBP	5.7
KRAS	37.1	EPHA5	5.7
FAT4	18.6	EPHB1	5.7
FBXW7	17.9	FAT1	5.7

LRP1B	16.4	PRKDC	5.7
PIK3CA	15.0	PTEN	5.7
SMAD4	14.3	PTPN13	5.7
ATM	10.0	ROS1	5.7
NRAS	10.0	ARID1A	5.7
TCF7L2	9.3	TRRAP	5.7
ERBB4	8.6	PCLO	5.7
AMER1	8.6	TET1	5.7
RELN	7.9	BCL9L	5.7
PDE4DIP	7.9	BCL9	5.0
RNF213	7.9	CBLB	5.0
CARD11	7.9	CTNNB1	5.0
BRCA2	7.1	DNMT3B	5.0
RUNX1T1	7.1	KDR	5.0
ROBO1	7.1	SMAD2	5.0
PTPRT	7.1	NF1	5.0
GRIN2A	6.4	PIK3CG	5.0
MKI67	6.4	RANBP2	5.0
PCBP1	6.4	TRIP11	5.0
PIK3R1	6.4	KMT2B	5.0
PTPRC	6.4	MED12	5.0
SMARCA1	6.4	MGA	5.0
SOX9	6.4	SETD2	5.0
ZNF521	6.4	PREX2	5.0
EP400	6.4		

9.4 cfDNA concentrations for patients from the WCB

Table 50: Total cfDNA concentrations and collection tube used for each patient and plasma timepoint

Sample ID	Week	cfDNA Concentration (ng/uL)
WCB1203	0	0.741
WCB1251	0	0.432
WCB1260	0	0.324
WCB1262	0	0.342
WCB1262	1	0.282
WCB1263	1	0.439
WCB1263	3	0.194
WCB1263	9	0.238
WCB1263	23	0.308
WCB1266	0	0.301
WCB1268	0	0.36
WCB1268	2	0.185
WCB1269	0	0.482
WCB1295	0	0.433

WCB1295	4	0.344
WCB1367	0	0.939
WCB1419	0	0.758
WCB1419	11	0.472
WCB1441	0	0.378
WCB1441	0	0.259
WCB1441	1	0.212
WCB1473	0	0.328
WCB1473	2	0.183
WCB1473	4	0.175
WCB1473	13	0.218
WCB1476	0	0.451
WCB1476	12	0.271
WCB1477	0	0.315
WCB1477	12	0.215
WCB1478	0	0.283
WCB1478	4	0.195
WCB1479	0	0.378
WCB1479	5	TL
WCB1479	6	0.128
WCB1479	8	0.162
WCB1479	10	0.175
WCB1493	0	0.416
WCB1493	1	0.144
WCB1493	3	0.257
WCB1493	4	0.408
WCB1493	6	0.444
WCB1602	0	0.406
WCB1602	1	0.238
WCB1602	2	0.488
WCB1602	5	0.229
WCB1602	10	0.676
WCB1603	0	0.377
WCB1603	1	0.122
WCB1603	5	0.189
WCB1603	14	0.119
WCB1604	0	0.443
WCB1604	0	0.197
WCB1604	0	0.223
WCB1604	6	0.134

9.5 Blood collection tubes and cfDNA concentrations for patients from ARISTOTLE

Table 51: Total cfDNA concentrations and collection tube used for each patient and plasma timepoint

Patient ID.	Week of Sample Collection	CfDNA Concentration	Collection Tube
ARI-166	Week 0	0.22	EDTA
ARI-166	Week 1	0.31	EDTA
ARI-166	Week 10	0.20	EDTA
ARI-166	Week 5	0.28	EDTA
ARI-182	Week 0	0.22	EDTA
ARI-182	Week 1	0.49	EDTA
ARI-182	Week 10	0.49	EDTA
ARI-182	Week 5	0.32	EDTA
ARI-202	Week 0	0.3	EDTA
ARI-202	Week 1	0.37	EDTA
ARI-202	Week 10	0.26	EDTA
ARI-202	Week 5	0.22	EDTA
ARI-239	Week 0	0.10	EDTA
ARI-239	Week 1	0.14	EDTA
ARI-239	Week 10	0.43	EDTA
ARI-239	Week 5	2.03	EDTA
ARI-295	Week 0	0.26	EDTA
ARI-295	Week 1	0.22	EDTA
ARI-295	Week 10	0.33	EDTA
ARI-295	Week 5	0.37	EDTA
ARI-297(a)	Week 0	0.26	EDTA
ARI-297(a)	Week 1	0.18	EDTA
ARI-297(a)	Week 10	0.35	EDTA
ARI-297(a)	Week 5	0.13	EDTA
ARI-297(b)	Week 0	0.23	EDTA
ARI-297(b)	Week 1	0.14	EDTA
ARI-297(b)	Week 10	0.34	EDTA
ARI-297(b)	Week 5	0.16	EDTA
ARI-306	Week 0	0.30	EDTA
ARI-306	Week 1	0.40	EDTA
ARI-306	Week 10	0.38	EDTA
ARI-306	Week 5	0.34	EDTA
ARI-316	Week 0	0.69	EDTA
ARI-316	Week 1	0.40	EDTA
ARI-316	Week 10	0.57	EDTA
ARI-316	Week 5	0.28	EDTA
ARI-341	Week 0	0.53	EDTA
ARI-341	Week 1	0.19	EDTA
ARI-341	Week 10	0.38	EDTA
ARI-341	Week 5	0.26	EDTA

ARI-346	Week 0	0.29	EDTA
ARI-346	Week 1	0.20	EDTA
ARI-346	Week 10	0.44	EDTA
ARI-346	Week 5	0.19	EDTA
ARI-366	Week 0	0.23	Streck
ARI-366	Week 1	0.22	Streck
ARI-366	Week 10	0.32	Streck
ARI-366	Week 5	0.38	Streck
ARI-373	Week 0	0.31	Streck
ARI-373	Week 1	0.42	Streck
ARI-373	Week 5	0.34	Streck
ARI-378(a)	Week 0	0.22	Streck
ARI-378(a)	Week 1	0.13	Streck
ARI-378(a)	Week 10	0.15	Streck
ARI-378(a)	Week 5	0.53	Streck
ARI-378(b)	Week 0	0.30	Streck
ARI-378(b)	Week 1	0.32	Streck
ARI-378(b)	Week 10	0.31	Streck
ARI-378(b)	Week 5	0.34	Streck
ARI-400	Week 0	0.28	Streck
ARI-400	Week 1	0.30	Streck
ARI-400	Week 10	0.33	Streck
ARI-400	Week 5	0.32	Streck
ARI-403	Week 0	0.33	Streck
ARI-403	Week 1	0.23	Streck
ARI-403	Week 10	0.12	Streck
ARI-403	Week 5	0.23	Streck
ARI-408	Week 0	0.21	Streck
ARI-408	Week 1	0.20	Streck
ARI-408	Week 10	0.19	Streck
ARI-408	Week 5	0.24	Streck
ARI-412	Week 1	0.43	Streck
ARI-412	Week 10	0.28	Streck
ARI-412	Week 5	0.31	Streck
ARI-413	Week 0	0.21	Streck
ARI-413	Week 10	0.20	Streck
ARI-413	Week 5	0.13	Streck
ARI-437	Week 0	0.23	Streck
ARI-437	Week 1	0.31	Streck
ARI-437	Week 10	0.27	Streck
ARI-437	Week 5	0.34	Streck

9.6 BCA Assay Set Up

Table 52: Concentrations for the BCA assay standard curve

BSA Concentration (ng/ml)
0
1.953125
3.90625
7.8125
15.625
31.25
62.5
125
250
500
1000
2000

BSA – Bovine Serum Albumin

9.7 ExoRNEasy Protocol Changes

9.7.1.1 Assay Development and Optimisation

Due to the consistency of its reported overexpression in CRCs, we used Mir-21 during the development of exosome-derived microRNA extraction and analysis. Simultaneously, we looked to detect levels of mRNA using an mRNA specific *KRAS* primer/probe set. The *KRAS* gene was chosen for validation as the majority of these WCB patients had *KRAS* mutant tumours. Levels of mutant *KRAS* could later be compared between ctDNA and exoRNA in the circulation of these patients.

The Mir-21 probe coupled with the Qiagen MiScript RT Kit II could not distinguish between DNA and microRNA. Therefore, exosomes had to be treated with Proteinase K, DNase I and Rnase A during assay development. Throughout methodological optimisation, contaminating DNA levels were examined in parallel as described in Section 2.2.15.

Exosomes were isolated from 1mL of plasma from patients with LARC using ExoSpin columns. These exosomes were treated and then applied directly to ExoRNEasy kit in place of patient plasma. Exosomes were bound to a membrane affinity column where they were lysed for RNA extraction using Qiazol.

We also spiked non-human RNA (*C. Elegans*) into the assay to verify RNA extraction efficiency without interfering with results.

Using 1mL of plasma collected from patient with LARC, no microRNA or mRNA was observed whilst the spike-in control *C. elegans* mir-39 (cel-mir-39) was detected by qPCR. The protocol was adapted to elute exosomes into an elution buffer (400µL of buffer XE; See Section 2.2.6.2) before lysis, which improved spike-in detectable yield but had no bearing on exoRNA.

We hypothesised that 1mL of plasma was not enough for reliable exoRNA detection in these patients at this time. We chose to revert to adding pre-filtered (0.22µm) patient plasma directly into the Qiagen ExoRNEasy Maxi Kit (an extension of the ExoEasy kits). This allowed the application of 1-4mL of patient plasma.

The purity and yield of exosomes would be directly comparable to that expected of the ExoEasy kit, so we did not repeat the comparisons of exosome quality. We compared input volumes of 2mL and 3mL of patient plasma to determine how much sample is required for confident and reliable RNA detection.

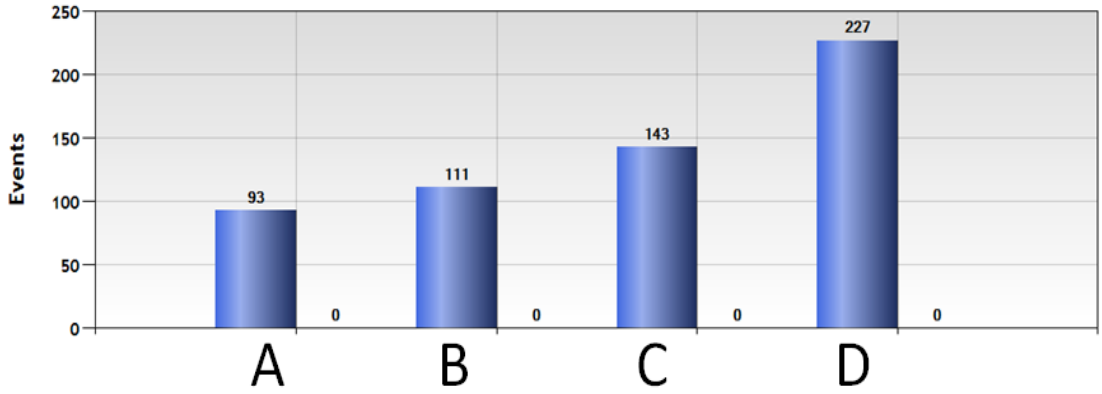

Figure 19: Levels of Mir-21 Detected in 2mL and 3mL of Patient Plasma using Droplet Digital PCR. Exosome Derived RNA was Extracted the Optimised ExoRNEasy Kit. A & B – 2mL Patient Plasma was used for ExoRNA Extraction, C&D 3mL of Patient Plasma was used for ExoRNA Extractions.

Mir-21 was successfully detected in 2mL and 3mL of plasma. The use of 3mL increased the confidence and reliability of sample detection. As the use of >1mL plasma was not compatible with the ExoSpin kit, pre-filtered plasma was added directly to ExoRNEasy Maxi kit during further microRNA analyses. We could not detect mRNA in any of these sample volumes and therefore chose to proceed with microRNA analysis alone.

We expected to see ~50 copies of Mir-21 in exoRNA from 1mL of plasma, however we were unable to see any (data not shown). This may be because of the small volume and reverse transcription process. There may have been no Mir-21 extracted from 1mL of plasma, or at least not enough to allow reliable reverse transcription. The levels we observe in 2mL and 3mL may also be deceiving, as they underwent two hours of reverse transcription, possible increasing the detected levels cDNA further. We, therefore, hypothesised that Mir-21 was simply undetectable or too low for reliable detection from 1mL of plasma in patients with LARC at this time.

To avoid the potential of DNA contamination influencing results, a new reverse transcription kit (TaqMan MicroRNA RT Kit) was tested. As a result, Mir-21 analysis was discontinued, due to incompatibility of our Mir-21 probe with the Taqman MicroRNA RT Kit.

A total of 10ng of genomic DNA underwent reverse transcription to examine the impact DNA contamination would have on reverse transcription and ddPCR. No cDNA was detected after reverse transcription and ddPCR analysis of genomic DNA (data not shown). This change of reverse transcription chemistry removed any issues regarding DNA contamination influencing microRNA quantification.

With a developed assay ready, we proceeded to investigate our microRNAs of interest in circulating exosomes from patients with LARC.

9.8 MicroRNA Literature Review

Table 53: A representation of the which circulating microRNAs have been investigated in patients with colon and/or rectal cancer and the number of times analysed. Review was performed in January 2018

MicroRNA	N
Let-7a	3
Let-7e	1
let-7g	2

Mir-1	1
Mir-10	2
Mir-100	1
Mir-103	1
Mir-106a	4
Mir-106b	1
Mir-10a	1
Mir-10b	1
Mir-122	1
Mir-1229	1
Mir-1246	3
Mir-1254	1
Mir-125a	3
Mir-125b	4
Mir-126	1
Mir-129	1
Mir-1290	1
Mir-130	1
Mir-130b	1
Mir-133a	2
Mir-134	1
Mir-135b	4
Mir-139-3p	1
Mir-141	5
Mir-142	2
Mir-143	3
Mir-145	4
Mir-146	1
Mir-148a	1
Mir-149	1
Mir-150	4
Mir-152	1
Mir-155	1
Mir-15b	2
Mir-16	3
Mir-16-2	1
Mir-16-5p	1
Mir-17	7
Mir-18	1
Mir-181	1
Mir-181b	2
Mir-181d	1
Mir-182	1
Mir-1826	1
Mir-187	1
Mir-188	1
Mir-18a	5

Mir-18b	2
Mir-191	1
Mir-192	1
Mir-193	2
Mir-193a	2
Mir-194	1
Mir-195	1
Mir-196	1
Mir-196a	1
Mir-198a	1
Mir-199a	1
Mir-19a	4
Mir-19b	2
Mir-200	2
Mir-200b	1
Mir-200c	1
Mir-203	5
Mir-20a	10
Mir-21	21
Mir-210	2
Mir-2110	1
Mir-21-5p	1
Mir-216a	1
Mir-22*	1
Mir-221	3
Mir-222	4
Mir-223	5
Mir-224	2
Mir-23a	4
Mir-23b	3
Mir-24	3
Mir-241	1
Mir-26a	1
Mir-27	1
Mir-27a	2
Mir-27b	3
Mir-29	1
Mir-29a	10
Mir-29b	2
Mir-29c	2
Mir-30a	1
Mir-30b	2
Mir-30c	2
Mir-31	7
Mir-3156	1
Mir-320a	4
Mir-326	1

Mir-328	1
Mir-331	1
Mir-335	1
Mir-338	1
Mir-342	1
Mir-345	1
Mir-34a	3
Mir-372	2
Mir-376a	1
Mir-376c	1
Mir-378	1
Mir-38a	1
Mir-423	1
Mir-425	2
Mir-4669	1
Mir-483	1
Mir-484	1
Mir-486	1
Mir-495	1
Mir-532	1
Mir-572	1
Mir-576	1
Mir-579	1
Mir-601	1
Mir-652	1
Mir-720	2
Mir-760	1
Mir-777	1
Mir-92	5
Mir-92a	12
Mir-95	1
Mir-96	1
Mir-96-5p	1
RNU6B	1
U2	1

N – Number of articles which each analysed

9.9 Circulating Mir-31 Analysis

Table 54: ANNOVAR for Mir-31 concentrations across each patient in this study. Significant findings are highlighted in bold, and suggestive significance in italics

Patient Comparison	Difference	Lower Limit	Upper Limit	Adjusted p-Value
WCB1441-WCB1368	0.09	0.05	0.13	<0.01
WCB1603-WCB1441	-0.07	-0.11	-0.04	<0.01
WCB1478-WCB1441	-0.09	-0.12	-0.05	<0.01
WCB1604-WCB1441	-0.07	-0.10	-0.04	<0.01
WCB1441-WCB1295	0.08	0.04	0.12	<0.01

WCB1476-WCB1441	-0.08	-0.12	-0.04	<0.01
WCB1473-WCB1441	-0.07	-0.10	-0.03	<0.01
WCB1477-WCB1441	-0.07	-0.11	-0.03	<0.01
WCB1473-WCB1368	0.02	-0.01	0.06	0.44
WCB1604-WCB1368	0.02	-0.02	0.06	0.55
WCB1477-WCB1368	0.02	-0.02	0.06	0.61
WCB1478-WCB1473	-0.02	-0.06	0.02	0.65
WCB1603-WCB1368	0.02	-0.02	0.05	0.73
WCB1604-WCB1478	0.02	-0.02	0.05	0.77
WCB1478-WCB1477	-0.02	-0.06	0.02	0.80
WCB1603-WCB1478	0.01	-0.02	0.05	0.91
WCB1476-WCB1368	0.01	-0.03	0.05	0.96
WCB1368-WCB1295	-0.01	-0.05	0.03	0.97
WCB1473-WCB1295	0.01	-0.03	0.05	0.97
WCB1476-WCB1473	-0.01	-0.05	0.03	0.98
WCB1477-WCB1295	0.01	-0.03	0.05	0.99
WCB1477-WCB1476	0.01	-0.03	0.05	0.99
WCB1604-WCB1295	0.01	-0.03	0.04	1.00
WCB1478-WCB1476	-0.01	-0.05	0.03	1.00
WCB1604-WCB1476	0.01	-0.03	0.04	1.00
WCB1478-WCB1295	-0.01	-0.05	0.03	1.00
WCB1603-WCB1473	-0.01	-0.04	0.03	1.00
WCB1603-WCB1477	-0.01	-0.04	0.03	1.00
WCB1603-WCB1295	0.00	-0.03	0.04	1.00
WCB1603-WCB1476	0.00	-0.03	0.04	1.00
WCB1604-WCB1603	0.00	-0.03	0.03	1.00
WCB1604-WCB1473	0.00	-0.04	0.03	1.00
WCB1478-WCB1368	0.00	-0.04	0.05	1.00
WCB1604-WCB1477	0.00	-0.04	0.03	1.00
WCB1476-WCB1295	0.00	-0.04	0.04	1.00
WCB1477-WCB1473	0.00	-0.04	0.04	1.00
WCB1441-All other patients	0.07	-0.09	-0.06	<0.01

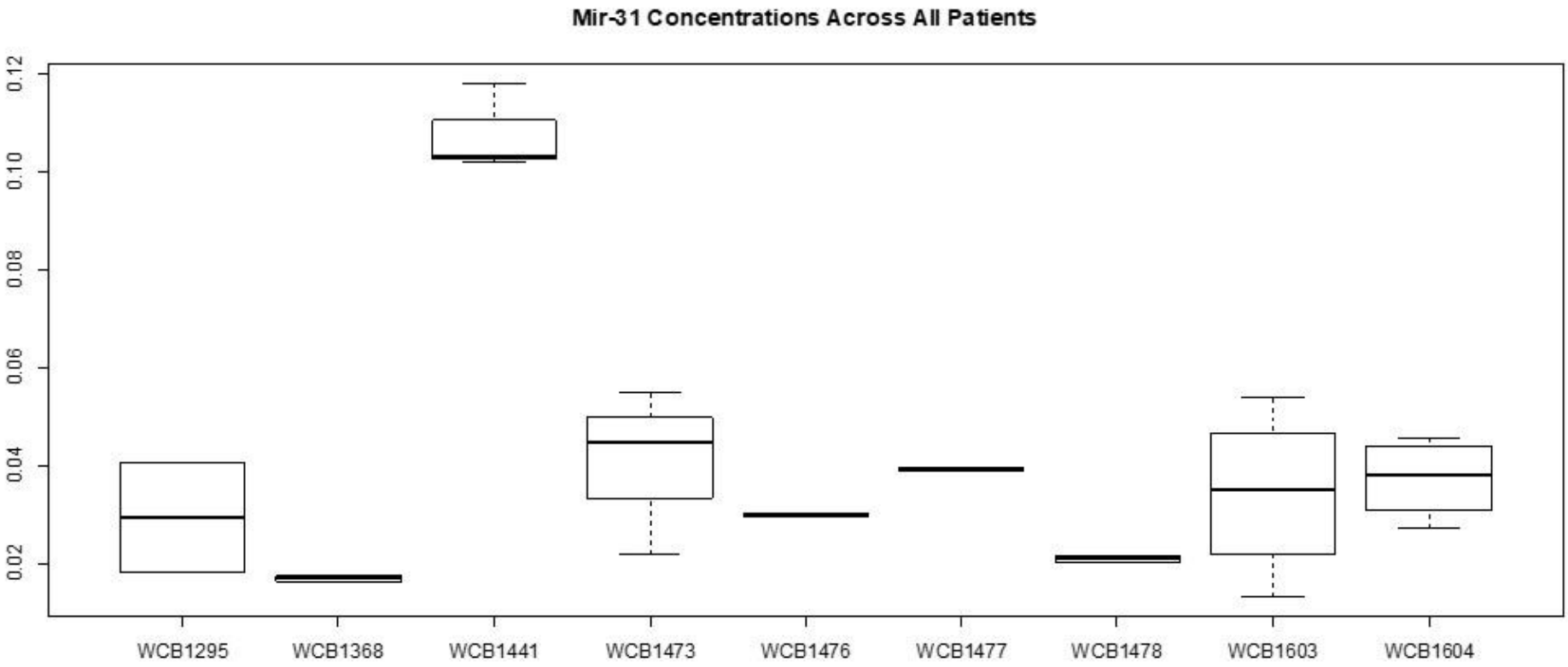

Figure 20: Boxplot representation of Mir-31 concentrations across all nine patients analysed

9.10 Circulating Mir-99a* Analysis

Table 55: ANNOVAR for Mir-99a concentrations across each patient in this study. Significant findings are highlighted in bold, and suggestive significance in italics*

Patient Comparison	Difference	Lower Limit	Upper Limit	Adjusted p-Value
WCB1603-WCB1368	-0.07	-0.11	-0.03	<0.01
WCB1473-WCB1368	-0.07	-0.10	-0.03	<0.01
WCB1441-WCB1368	-0.06	-0.10	-0.02	<0.01
WCB1368-WCB1295	0.06	0.02	0.11	<0.01
WCB1478-WCB1368	-0.06	-0.10	-0.02	<0.01
WCB1476-WCB1368	-0.06	-0.10	-0.02	<0.01
WCB1477-WCB1368	-0.06	-0.10	-0.02	<0.01
WCB1604-WCB1603	0.04	0.01	0.07	<0.01
WCB1604-WCB1473	0.04	0.00	0.07	0.03
WCB1604-WCB1441	0.03	0.00	0.06	*0.06*
WCB1604-WCB1295	0.03	0.00	0.07	*0.07*
WCB1604-WCB1478	0.03	0.00	0.07	0.11
WCB1604-WCB1368	-0.03	-0.07	0.01	0.15
WCB1604-WCB1476	0.03	-0.01	0.07	0.19
WCB1604-WCB1477	0.03	-0.01	0.06	0.23
WCB1603-WCB1477	-0.01	-0.05	0.03	0.98
WCB1603-WCB1476	-0.01	-0.05	0.03	0.99
WCB1603-WCB1441	-0.01	-0.04	0.03	1.00
WCB1477-WCB1473	0.01	-0.03	0.05	1.00
WCB1603-WCB1478	-0.01	-0.04	0.03	1.00
WCB1476-WCB1473	0.01	-0.03	0.05	1.00
WCB1477-WCB1295	0.01	-0.04	0.05	1.00
WCB1476-WCB1295	0.01	-0.04	0.05	1.00
WCB1473-WCB1441	0.00	-0.04	0.03	1.00
WCB1603-WCB1295	0.00	-0.04	0.03	1.00
WCB1478-WCB1477	0.00	-0.05	0.04	1.00
WCB1603-WCB1473	0.00	-0.04	0.03	1.00
WCB1441-WCB1295	0.00	-0.04	0.04	1.00
WCB1477-WCB1441	0.00	-0.04	0.04	1.00
WCB1478-WCB1473	0.00	-0.04	0.04	1.00
WCB1478-WCB1476	0.00	-0.05	0.04	1.00
WCB1478-WCB1295	0.00	-0.04	0.05	1.00
WCB1476-WCB1441	0.00	-0.04	0.04	1.00
WCB1473-WCB1295	0.00	-0.04	0.04	1.00
WCB1478-WCB1441	0.00	-0.04	0.04	1.00
WCB1477-WCB1476	0.00	-0.04	0.04	1.00
WCB1368-All other patients	0.06	-0.19	0.08	0.13
WCB1604-All other patients	0.03	-0.06	0	*0.08*

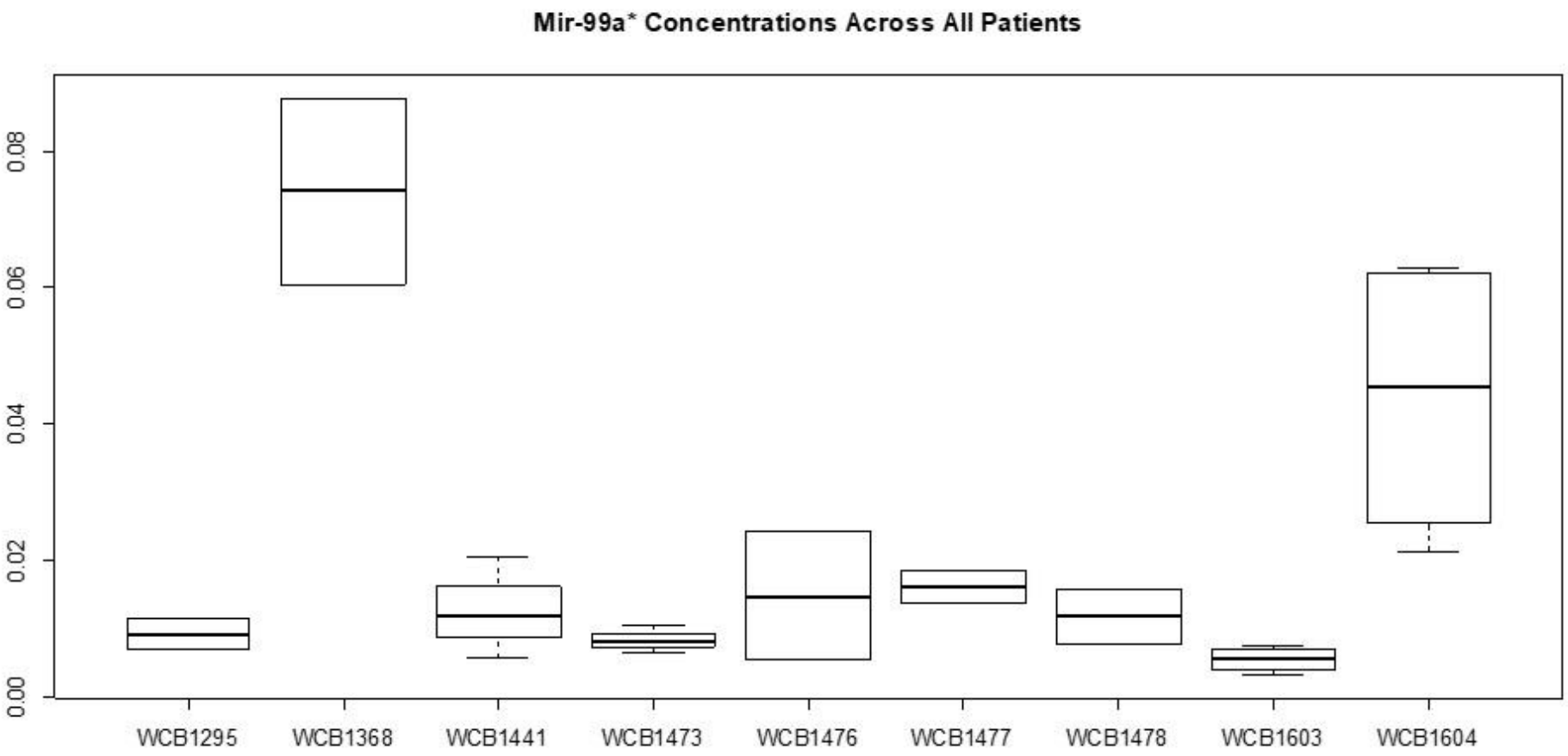

Figure 21: Boxplot representation of Mir-99a concentrations across all nine patients analysed.*

9.11 Local and/or Distant Relapse rates Within an Extended WCB Cohort

Table 56: A comparison of sequential platelet counts, lymphocyte counts, neutrophil counts, monocyte counts, PLR, NLR and LMR between patients undergoing short-course therapy who experienced local and/or distant recurrence within two years of follow up in comparison to those who did not. All values were analysed as continuous variables, not as high or low based on a designated threshold. Change in variable refers the difference between baseline and week 5 for each respective biomarker. Any recurrence refers to patients who experience either local or distant relapse

	Local Relapse			**Distant Relapse**			**Any Relapse**		
Variable	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p-Value**	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p-Value**	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p-Value**
Baseline Platelet Count	282 (143.0 - 465.0)	217 (188.0 - 386.0)	0.62	275 (143.0 - 465.0)	253 (188.0 - 386.0)	0.68	305 (143.0 - 465.0)	217 (188.0 - 386.0)	0.38
Week 5 Platelet Count	221 (141.0 - 424.0)	377 (223.0 - 459.0)	0.37	221 (141.0 - 344.0)	295 (209.0 - 459.0)	0.19	221 (141.0 - 344.0)	295 (209.0 - 459.0)	0.19
Change in Platelet Count	-22 (-245 - 135)	6 -11 - 73)	0.47	-41 (-245 - -2)	21 (-11 - 135)	*0.05*	-41 (-245 - -2)	21 (-11 - 135)	*0.05*
Baseline Lymphocyte Count	1.9 (1.4 - 3.4)	2.3 (1.0 - 3.7)	0.62	1.8 (1.4 - 3.4)	2.2 (1.0 - 3.7)	0.52	1.8 (1.4 - 3.4)	2.3 (1.0 - 3.7)	0.43
Week 5 Lymphocyte Count	1.1 (0.8 - 1.8)	2.8 (0.7 - 3.2)	*0.07*	1.1 (0.8 - 1.8)	1.3 (0.7 - 3.2)	2.7	1.1 (0.8 - 1.8)	1.3 (0.7 - 3.2)	0.28
Change in Lymphocyte Count	-0.7 (-1.5 - -0.2)	-0.8 (-1.0 - 0.8)	0.36	-0.4 (-1.3 - -0.2)	-0.8 (-1.5 - 0.8)	0.91	-0.4 (-1.3 - -0.2)	-0.8 (-1.5 - 0.8)	0.91
Baseline Neutrophil Count	5.5 (2.5 - 7.2)	5.9 (4.3 - 6.6)	0.86	5.9 (2.5 - 7.2)	4.7 (4.2 - 6.6)	0.46	5.9 (2.5 - 7.2)	4.8 (4.2 - 6.6)	0.6
Week 5 Neutrophil Count	3.7 (2.5 - 8.0)	3.7 (2.5 - 8.0)	0.19	3.7 (2.6 - 8.0)	3.8 (2.5 - 15.0)	0.59	3.7 (2.5 - 8.0)	3.8 (2.5 - 15.0)	0.59

Change in Neutrophil Count	-1.6	-1.3	0.17	-1.6	-1.3	0.39	-1.6	-1.3	0.39
	(-3.0 - 1.6)	(-1.3 - 8.4)		(-3.0 - 1.6)	(-1.7 - 8.4)		(-3.0 - 1.6)	(-1.7 - 8.4)	
Baseline Monocyte Count	0.7	0.7	0.98	0.7	0.7	0.84	0.6	0.7	0.80
	(0.4 - 1.2)	(0.4 - 0.9)		(0.4 - 1.2)	(0.5 - 0.9)		(0.4 - 1.2)	(0.4 - 0.9)	
Week 5 Monocyte Count	0.5	0.9	*0.07*	0.4	0.5	0.18	0.4	0.5	0.18
	(0.3 - 0.6)	(0.6 - 1.2)		(0.3 - 0.6)	(0.5 - 1.2)		(0.3 - 0.6)	(0.5 - 1.2)	
Change in Monocyte Count	-0.2	-0.1	0.23	-0.2	-0.1	0.66	-0.2	-0.1	0.66
	(-0.4 - 0.1)	(-0.2 - 0.3)		(-0.4 - 0.1)	(-0.3 - 0.3)		(-0.4 - 0.1)	(-0.3 - 0.3)	
Baseline PLR	145.7	104	0.62	156.3	104.8	0.51	156.9	104	0.18
	(95.3 - 191.6)	(81.7 - 198.0)		(81.7 - 191.6)	(90.4 - 198.0)		(95.3 - 191.6)	(81.7 - 198.0)	
Week 5 PLR	196.6	173	0.92	196.6	190	0.48	196.6	190	0.48
	(121.1 - 333.9)	(69.7 - 421.4)		(128.2 - 210.5)	(69.7 - 421.4)		(128.2 - 210.5)	(69.7 - 421.4)	
Change in PLR	32.8	51.6	0.68	23.9	91.1	0.10	23.9	91.1	0.10
	(-2.9 - 228.4)	(-20.7 - 241.1)		(-2.9 - 54.2)	(-20.7 - 241.4)		(-2.9 - 54.2)	(-20.7 - 241.4)	
Baseline NLR	2.6	2.6	0.89	2.7	2.2	0.60	2.8	2.5	0.51
	(1.5 - 4.3)	(1.7 - 3.9)		(1.8 - 4.3)	(1.5 - 4.5)		(1.8 - 4.3)	(1.5 - 4.5)	
Week 5 NLR	3.3	1.8	0.31	3.3	1.9	0.62	3.3	1.9	0.62
	(1.9 - 7.3)	(0.9 - 21.4)		(2.1 - 7.3)	(0.9 - 21.4)		(2.1 - 7.3)	(0.9 - 21.4)	
Change in NLR	0.5	0	0.21	0.5	0.4	0.42	0.5	0.4	0.42
	(-1.0 - 3.0)	(-0.9 - 17.5)		(-1.0 - 3.0)	(-0.9 - 17.5)		(-1.0 - 3.0)	(-0.9 - 17.5)	
Baseline LMR	3.3	3.3	0.96	3	3.2	0.95	2.9	3.2	0.79
	(2.5 - 5.0)	(2.0 - 5.0)		(2.5 - 5.0)	(2.0 - 5.0)		(2.5 - 5.0)	(2.0 - 5.0)	
Week 5 LMR	2.5	3.3	0.37	2.7	2.4	0.50	2.7	2.4	0.50
	(1.7 - 5.0)	(0.6 - 5.0)		(1.7 - 5.0)	(0.6 - 5.0)		(1.7 - 5.0)	(0.6 - 5.0)	

Change in LMR	0	0	0.28	0	0	0.40	0	0	0.40
	(0.0 - 5.0)	(-0.8 - 10.0)		(0.0 - 5.0)	(-0.8 - 10.0)		(0.0 - 5.0)	(-0.8 - 10.0)	

Table 57; A comparison of sequential platelet counts, lymphocyte counts, neutrophil counts, monocyte counts, PLR, NLR and LMR between patients undergoing long-course therapy who experienced local and/or distant recurrence within two years of follow up in comparison to those who did not. All values were analysed as continuous variables, not as high or low based on a designated threshold. Change in variable refers the difference between baseline and week 5 for each respective biomarker. Any recurrence refers to patients who experience either local or distant relapse

	Local Relapse			**Distant Relapse**			**Any Relapse**		
Variable	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p-Value**	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p-Value**	**No Relapse; Median (Range)**	**Relapse; Median (Range)**	**p-Value**
Baseline Platelet Count	287	300	0.31	287	320	0.27	282	307	0.21
	(126.0 - 601.0)	(192.0 - 540.0)		(126.0 - 601.0)	(182.0 - 426.0)		(126.0 - 601.0)	(182.0 - 562.0)	
Week 5 Platelet Count	233	244	0.25	233	244	0.11	233	244	0.16
	(126.0 - 452.0)	(168.0 - 294.0)		(126.0 - 452.0)	(163.0 - 426.0)		(126.0 - 452.0)	(163.0 - 426.0)	
Change in Platelet Count	-57	-72	0.85	-58	-56	0.92	-56.6	-72	*0.49*
	(-245.0 - 102.0)	(-245.0 - 6.0)		(-326.0 - 82.0)	(-236.0 - 8.0)		(-326.0 - 82.0)	(-245.0 - 102.0)	
Baseline Lymphocyte Count	1.8	1.8	0.54	1.8	1.9	0.79	1.7	1.9	0.86
	(0.6 - 3.8)	(0.9 - 3.5)		(0.6 - 3.8)	(0.9 - 3.5)		(0.6 - 3.8)	(0.9 - 3.5)	
Week 5 Lymphocyte Count	0.6	0.7	0.27	0.6	0.7	0.11	0.6	0.7	0.1
	(0.3 - 2.6)	(0.4 - 0.9)		(0.3 - 1.2)	(0.4 -1.5)		(0.3 - 1.2)	(0.4 - 1.5)	
Change in Lymphocyte Count	-1.1	-0.9	0.21	-1.1	-1.1	0.61	-1.1	-1.1	0.53
	(-2.9 - -0.2)	(-2.5 - -0.3)		(-2.9 - -0.2)	(-2.5 - -0.3)		(-2.9 - -0.2)	(-2.5 - -0.3)	
	4.6	51	0.78	4.7	4.8	0.86	4.6	5.1	0.37

Baseline Neutrophil Count	(2.5 - 10.3)	(4.0 - 5.8)		(2.5 - 10.3)	(2.6 - 5.9)		(2.5 - 10.3)	(2.6 - 9.6)	
Week 5 Neutrophil Count	3.4	3.7	0.45	3.4	3.8	0.25	3.4	3.8	0.25
	(1.4 - 9.3)	(2.8 - 4.4)		(1.4 - 9.3)	(1.9 - 4.9)		(1.4 - 9.3)	(1.9 - 7.4)	
Change in Neutrophil Count	-1.1	-1.2	0.71	-1.1	-1.2	0.44	-1.1	-1.3	0.95
	(-6.1 - 2.8)	(-4.5 - 1.7)		(-6.2 - 2.8)	(-3.9 - 1.7)		(-6.1 - 2.8)	(-4.6 - 1.7)	
Baseline Monocyte Count	0.5	0.6	0.14	0.5	0.6	0.2	0.5	0.6	*0.05*
	(0.3 - 1.0)	(0.3 - 1.1)		(0.3 - 1.0)	(0.1 - 1.1)		(0.3 - 1.0)	(0.1 - 1.1)	
Week 5 Monocyte Count	0.5	0.5	0.24	0.5	0.5	0.2	0.4	0.5	0.11
	(0.3 - 1.2)	(0.2 - 0.7)		(0.3 - 0.9)	(0.3 – 0.8)		(0.3 - 1.2)	(0.3 - 0.9)	
Change in Monocyte Count	0	-0.1	0.61	0	-0.1	0.88	0	-0.1	0.54
	(-0.3 - 0.4)	(-0.5 - 0.1)		(-0.5 - 0.4)	(-0.4 - 0.2)		(-0.3 - 0.4)	(-0.4 - 0.5)	
Baseline PLR	154.6	161.3	0.41	154.6	161.3	0.84	154.3	161.3	0.53
	(50.0 - 471.7)	(102.1 - 435.5)		(50.0 - 471.7)	(78.3 - 25-.6)		(50.0 - 471.7)	(78.3 - 435.5)	
Week 5 PLR	396.7	338	0.32	396.7	352.5	0.33	398.3	338	0.27
	(140.9 - 1180.0)	(251.3 - 473.3)		(140.9 - 1180)	(163.6 - 497.5)		(140.9 - 1180)	(163.6 - 497.5)	
Change in PLR	220.4	159.6	0.12	209.3	178.1	0.21	217.6	159.6	0.11
	(8.6 - 801.4)	(-44.5 - 287.5)		(8.6 - 801.4)	(-44.5 - 295.5)		(8.6 - 801.4)	(-44.5 - 295.5)	
Baseline NLR	2.4	3	0.66	2.6	2.7	0.65	2.4	2.8	0.91
	(1.1 - 8.1)	(2.1 - 3.9)		(1.1 - 8.1)	(1.3 - 3.9)		(1.1 - 8.1)	(1.3 - 3.9)	
Week 5 NLR	5.3	5.8	0.41	5.4	5.3	0.35	5.4	5.3	0.32
	(1.6 - 16.0)	(3.8 - 7.5)		(1.6 - 16.0)	(2.8 - 7.5)		(1.6 - 16.0)	(2.8 - 7.5)	
Change in NLR	3	2.2	0.28	3	2.5	0.44	3.1	2.2	0.27
	(-5.3 - 12.8)	(01.3 - 4.5)		(-5.3 - 12.8)	(-1.3 - 5.9)		(-5.3 - 12.8)	(-1.3 - 5.9)	

Baseline LMR	3.3	2.5	*0.06*	3.3	3.3	0.36	3.3	2.9	0.14
	(1.4 - 10.0)	(1.0 - 5.0)		(1.4 - 10.0)	(2.0 - 10.0)		(1.4 - 10.0)	(1.25 - 10.0)	
Week 5 LMR	0.8	0.8	0.63	0.8	0.8	0.42	0.8	0.83	0.48
	(0.4 - 3.3)	(1.4 - 2.0)		(0.4 - 3.3)	(0.8 - 3.3)		(0.4 - 3.3)	(0.8 - 3.3)	
Change in LMR	0.5	0.475	0.18	0.5	0.5	0.19	0.5	0.5	0.15
	(-0.6 - 10)	(0 - 1.4)		(-0.6 - 10.0)	(0 - 1.4)		(-0.6 - 10.0)	(0 - 1.4)	