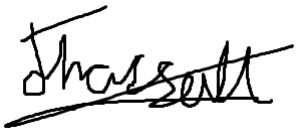
INVESTIGATING RADIATION RESPONSIVENESS IN RECTAL CANCER

James C. Hassall

JULY 30, 2020
144 MAIN STREET, SWITHLAND, LEICESTERSHIRE, LE12 8TJ

Declaration

This report entitled 'Investigating Radiation Responsiveness in Rectal Cancer' has been composed by James Hassall, and I confirm that the work presented in this thesis is my own, unless otherwise mentioned. No part of this thesis has been submitted for any degree, diploma or any other type of qualification at any other institution.



James Hassall

4th February 2021

Acknowledgements

I thank my project supervisor, Professor Ilyas for his continual and reliable support throughout most of my academic life. Professor Ilyas has taught me invaluable lessons on how best to conduct my research and the important aspects to focus on from both a clinical and scientific point of view. These lessons have allowed me to logically structure and analyse my research with the patient always in mind.

I am grateful for the expertise and discipline passed on by my former mentor and colleague, Dr. Fadhil. Without his strict attitude towards laboratory practice and deep knowledge of PCR, I would have achieved far less.

I thank my funders, the Pathological Society of Great Britain and Ireland for allowing me to conduct my research.

I thank my parents and partner (Victoria Sutcliffe) for supporting me throughout the process. Their support has made the process far easier and I cannot thank them enough.

Finally, I would like to thank all the members of the Pathology Research group from 2014 onwards for their help and support.

Dedication

I dedicate this thesis to my parents as they have supported me tremendously throughout my whole education. Words will never be enough for me to show how grateful I am.

Thesis related publications and patents

Thesis related patents (approval pending)

1. Hassall, J., Ebili, H., Ilyas, M and The University of Nottingham. (2019). Detection of nucleic acid sequences. Pub. No.: WO/2019/145734. International application No.: PCT/GB2019/050221.

Thesis related publications

1. Susanti, S., Fadhil W., Murtaza, S., **Hassall, JC.**, Ebili HO., Oniscu, A. and Ilyas, M. (2019). Positive association of PIK3CA mutation with KRAS mutation but not BRAF mutation in colorectal cancer suggests co-selection is gene specific but not pathway specific. Journal of Clinical Pathology, Mar;72(3):263-264. doi: 10.1136/jclinpath-2018-205483.
2. Henry O. Ebili., **James C. Hassall.**, Wakkas Fadhil., Hersh Ham-Karim., Abutaleb Asiri., Teresa P. Raposo,. Ayodeji Johnson Agboola. and Mohammad Ilyas. "Squirrel" Primer-Based PCR Assay for Direct and Targeted Sanger Sequencing of Short Genomic Segments. (2017). Journal of Biomolecular Techniques, 28(3): 97–110. doi: 10.7171/jbt.17-2803-001.
3. Hersh Abdul Ham-Karim., Henry Okuchukwu Ebili., Wakkas Fadhil., Abutaleb Asiri., **James Hassall.** And

Mohammad Ilyas. COLD-HRM: a combination of methods to infer the nature of somatic mutations. (2017). Advances in Cytology & Pathology, 2(2): 54-61.

4. Ebili, HO., **Hassall, J.**, Asiri, A., Ham-Karim, H., Fadhil, W., Agboola, AJ. and Ilyas, M. (2017). QMC-PCRx: a novel method for rapid mutation detection. Journal of Clinical Pathology, 70(8):702-711. doi: 10.1136/jclinpath-2016-204264.

Presentations and conference communications

1. **Hassall, JC.**, and Ilyas, M. (2018, June). An Exquisitely Sensitive Method for Mutation Detection. Presentation at Maastricht Pathology 2018, Pathological Society, Maastricht, NL.
2. **Hassall, JC.**, and Ilyas, M. (2019, July). The Use of HOT_ARMS PCR in Liquid Biopsies for the Management of Patients With Colorectal Cancer. Presentation at Leeds Pathology 2019, Pathological Society, Leeds, UK.

List of abbreviations

1. ARMS = Amplification-Refractory Mutation detection System
2. ATM = Ataxia Telangiectasia Mutated protein
3. ATR = Ataxia Telangiectasia and Rad3 related protein
4. BCSP = Bowel Cancer Screening Program
5. BNA = Bridged Nucleic Acid
6. Bp = base pairs
7. CC = Colon cancer
8. cCR = Complete Clinical Response
9. cDNA = Copy DNA
10. CEA = carcinoembryonic antigen
11. cfDNA = circulating free DNA
12. cfNA = circulating nucleic acid
13. CIMP = CpG island methylator phenotype
14. CIN = Chromosomal Instability
15. CMS = Consensus Molecular Subtypes
16. CNAK = Qiagen Circulating Nucleic Acid Kit
17. COLD-PCR = CO-amplification at Lower Denaturation temperature – Polymerase Chain Reaction
18. CRC = Colorectal Cancer
19. CRT = Chemoradiotherapy
20. CT = Computed tomography
21. Ct = Cycle threshold

22. CV = Coefficient of variation
23. ctDNA = circulating-tumour DNA
24. ddPCR = digital-droplet Polymerase Chain Reaction
25. DDR = DNA damage response
26. DFS = Disease-Free Survival
27. dsDNA = double-stranded DNA
28. E-Ice COLD PCR = Enhanced Improved and Complete Enrichment Co-amplification at Lower Denaturation temperature Polymerase Chain Reaction
29. ESMO = European Society for Medical Oncology
30. FAP = Familial Adenomatous Polyposis
31. FDA = Food and Drug Administration
32. FIT = Fecal Immunohistochemical test
33. FFPE = Formalin-Fixed Paraffin-Embedded
34. gFOBT = Guaic-Based Fecal Occult Blood Test
35. Gy = Gray
36. HMM = Histogenic Molecular Mapping
37. HOT_ARMS PCR: Highly Optimised annealing Temperature_Amplification Refractory Mutation System Polymerase Chain Reaction
38. HOT_PCR: Highly Optimised annealing Temperature_Molecular Adaptation for Maximum Annealing Specificity Polymerase Chain Reaction

39. HOT_PI PCR: Highly Optimised annealing Temperature_Probe Inhibited Polymerase Chain Reaction
40. HR = Hazard Ratio
41. HRM/HRMA = High Resolution Melting Analysis
42. IDL = Insertion and Deletion Loops
43. IHC = Immunohistochemistry
44. Indel = Insertion and deletion
45. LET = Linear energy transfer
46. LNA = Locked Nucleic Acid
47. LOD = Limit Of Detection
48. M = Metastasis status
49. MAF = Mutant allele frequency
50. MIC = Magnetic Induction Cycler
51. MMR = Mismatch Repair
52. MRI = Magnetic Resonance Imaging
53. mRNA = messenger RNA
54. MSI = Microsatellite Instability
55. MSI-H = Microsatellite Instability – High
56. MSI-L = Microsatellite Instability – Low
57. MSS = Microsatellite Stable
58. Mv = Mega volt
59. N = Node status
60. N_CaRT: Nottingham_Cancer Recurrence Test
61. NGS = Next-Generation Sequencing

62. Nt = nucleotides
63. NTC = No Template Control
64. OS = Overall Survival
65. PET-CT = Positron Emission Tomography – Computed Tomography
66. pCR = Pathological Complete Response
67. PCR = Polymerase Chain Reaction
68. PPC = Chance of amplicon formation
69. QC = Quality Control
70. QA = Quality assessment
71. qPCR = real-time PCR
72. RC = Rectal cancer
73. RFS = Recurrence-Free Survival
74. ROS = Reactive Oxygen Species
75. RPPA = Reverse Phase Protein Array
76. RT-qPCR = Reverse Transcription real-time PCR
77. SCNA = somatic copy number alterations
78. SNV = Single Nucleotide Variant
79. ssDNA = single-stranded DNA
80. T = Tumour depth
81. Ta = primer Annealing Temperature
82. Tc = cold-pcr Critical Temperature
83. Tm = primer Melting Temperature
84. TRG = Tumour Regression Grade

85. UK = United Kingdom
86. WT = Wild-Type
87. 5-FU = 5-Fluorourac

Abstract

Currently rectal cancer patients undergoing radiotherapy are subject to a 'watch-and-wait' scenario whereby the clinician and patient are not certain if the tumour is responding to the radiotherapy until completion. Real-time monitoring of response to radiotherapy may be possible via liquid biopsy testing and it is hoped that this will inevitably improve outcome by streamlined patient stratification. Liquid biopsies are a promising method to interrogate the molecular characteristics of the tumour and determine presence of the tumour in real-time by blood test. This is possible when dying cells release DNA into the blood stream (cell-free DNA). This also occurs in tumour cells (circulating tumour DNA) at a higher rate due to their instability, producing regular injections of DNA into the blood stream. Each tumour contains a plethora of mutations which can be identified in cell free DNA by an appropriate mutation detection methodology. Cell-free DNA is highly fragmented (130-170bp), low in mutant allele frequency (as low as 0.01%) and low in concentration (1-100ng/ml of blood). Therefore, these characteristics present a major hurdle for mutation detection methods to determine presence of circulating tumour DNA.

This study has developed two novel methods that are suitable for cell-free DNA mutation detection as other available

methods are currently flawed in a number of ways: either they are too expensive for serial analysis of liquid biopsy (which is required for treatment-response monitoring), not sensitive enough, not scalable enough or too complicated for clinical integration. The two methods 'Highly Optimised annealing Temperature_Amlification Refractory Mutation System PCR (HOT_ARMS PCR)' and Highly Optimised annealing Temperature_Probe Inhibited PCR (HOT_PI PCR)' are a combination of methods for liquid biopsy interrogation. HOT_ARMS is an extremely simple, robust, inexpensive, and exquisitely sensitive test that uses mutation-specific primers for detection of any kind of mutation which results in a sequence change. HOT_ARMS improves the previously published ARMS system by increasing the specificity of the PCR via primer modification to raise the annealing temperature (Ta) to >65°C. The increased kinetic energy of the primers reduces non-specific 3' base-pairing and the modification increases selectivity for the mutant amplicon. This simple adaptation of ARMS PCR with a working annealing of 71°C generates an unrivalled all-round PCR-based method of detecting low frequency mutant alleles. It is a single stage closed-tube test which does not require expensive equipment and, because it is not reliant on probes, it is easy to set up and requires little optimisation for most mutations. The speed

of the test means it could theoretically be established in the hospital out-patient and even the primary care setting.

HOT_PI PCR is an extremely simple, robust, inexpensive, and efficient method of mutation enrichment on up to 6 nucleotides for any sequence change. This can result in 60-fold enrichment of mutant alleles, allowing methods with insufficient limits of detection the ability to test liquid biopsies. High-resolution melting analysis was deployed and could detect 0.1% MAF (a major improvement from 6% MAF). HOT_PI also uses a working annealing temperature of 71°C meaning that it can be combined with HOT_ARMS PCR for a streamlined approach to screen for the presence of mutations in pre-treatment liquid biopsies. This will determine the mutational profile of the tumour for treatment-response monitoring; allowing for quantitative singleplex HOT_ARMS assays to be deployed and ultimately save precious FFPE tissue for histology testing. HOT_PI is a single stage closed-tube test which is easy to set up and does not require expensive equipment. The speed of the test means it could work alongside HOT_ARMS PCR and reduce the number of required tests.

Rectal cancer patients undergoing radiotherapy currently vary in response. Twenty percent of patients will completely

respond to radiotherapy and up to 60% will achieve tumour downstaging. This leaves a large proportion of patients receiving treatment unnecessarily. It may be possible to create a predictive test through biomarker analysis by comparing biopsy specimens from patients with known response status. This study interrogated redox-homeostasis genes in 10 biopsy samples known to respond to radiotherapy (Mandard 1-2) and 10 samples known to not respond to radiotherapy (Mandard 3-5) by reverse-transcription real-time PCR using custom designed assays. Redox-homeostasis has previously been found to be predictive of radiotherapy response in breast and lung cancer.

List of content

List of figures

Appendices

List of tables

Appendices

1. Introduction: Literature review

1. Introduction: literature review

A: Rectal cancer

i. Anatomy

The rectum is part of the large intestine; otherwise known as the large bowel. The rectum is located on the left side of the large intestine where it begins at the rectosigmoid junction and ends at the anorectal ring[1] as shown in figure 1 below.

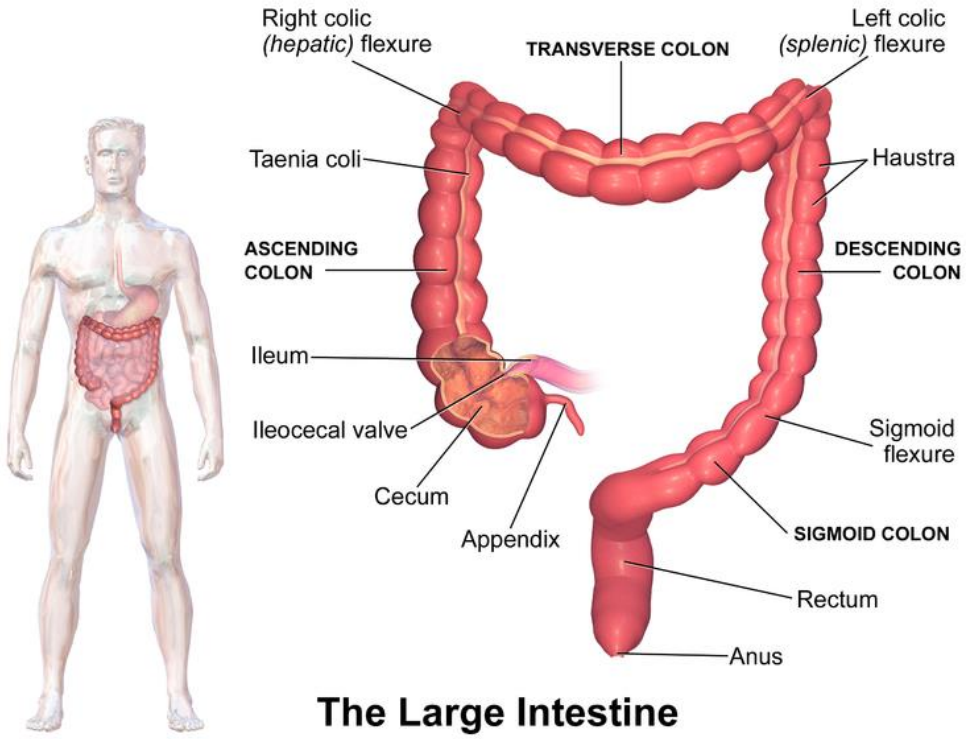


Figure 1. Anatomical diagram of the large intestine. The large intestine comprises several components which can be summarised as the left colon consisting of half of the transverse colon, descending colon, and sigmoid colon whilst the ride side has half of the transverse colon, ascending colon

and cecum. The rectum attaches to the sigmoid colon on the left side. [2]

The rectum and colon are histologically similar but differ in many ways. The rectum has a differing blood supply, levels of hormone receptors, pH, nerve supply, exposure to faecal matter, embryonic origin and lacks protection from the serosa[3]. The rectum is also in a relatively fixed location in comparison to the rest of the large intestine. The combination of these factors results in major differences for treatment, progression, metastatic spread, local invasive potential, and outcome for rectal cancer compared to colorectal cancer[3]. Moreover, the location of the rectum allows easier acquisition of biopsies and access for screening methods.

ii. Colorectal cancer epidemiology in the UK

Colorectal cancer (CRC) is the fourth most common cancer in the UK[4]. In 2014, CRC accounted for 41,265 new cancer cases; 22,844 men (55%) and 18,421 women (45%); making it the third most common cancer for each gender[5]. 15,903 bowel cancer deaths were reported in 2014 and the 10-year survival rate was 57%[4]. Approximately 11,547 cases of CRC cases were located in the rectum; 7310 men (32%) and 4237 women (23%)[5]. Moreover, a further 7% of the male cases

and 6% of the female cases were located at the rectosigmoid junction.

iii. Rectal cancer molecular pathology

Rectal cancer arises from the glandular epithelium of the mucosa[6]. Most rectal cancers are sporadic but, in some instances, can also be hereditary[7]. Rectal cancer like most cancers, progress over a long-time course from a small benign lesion to a malignant neoplasm. During this period, mutations are acquired that are subject to a Darwinian evolutionary process, typically through interactions with the surrounding environment [8]. Rectal cancer was previously thought to follow a linear molecular model whereby particular mutations are acquired at specific points of tumour progression. In recent years, the model has been challenged by non-linear molecular progression models[9]. However, the new models are largely still up for debate. For simplicity, this thesis will focus on the existing long-standing linear molecular models that are categorised into the same important sub-groups as non-linear molecular models. In all instances, key mutations contribute to the acquisition of the hallmarks of cancer as shown below in figure 2.

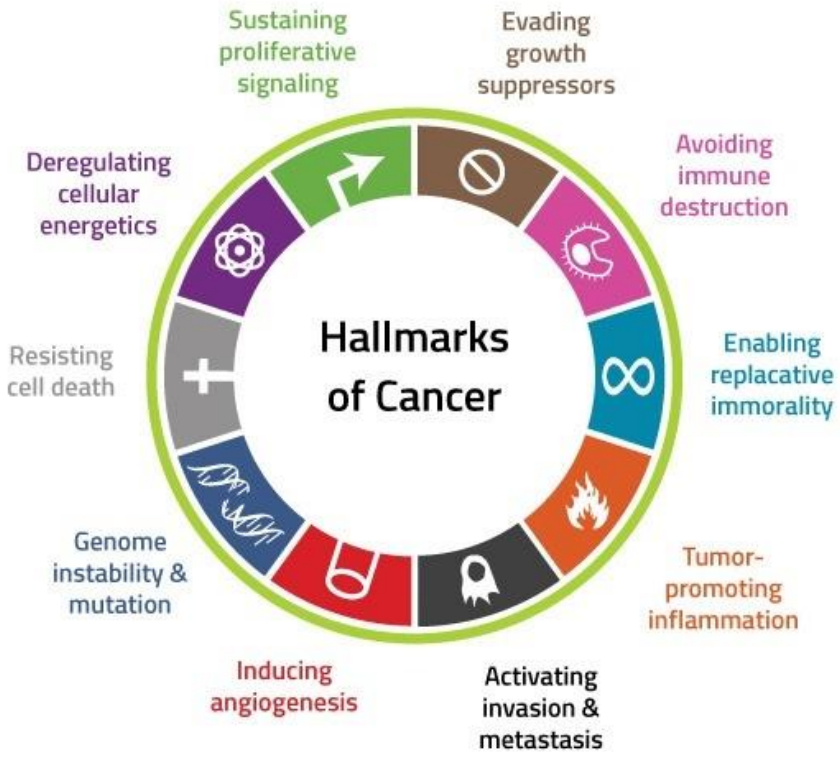


Figure 2. The hallmarks of cancer. The hallmarks of cancer are key factors which characterise cancer by contributing to the process of tumorigenesis. [10]

Acquisition of the hallmarks ultimately results in uncontrolled tumour growth and invasive spread both locally and at distant sites[10]. As shown below in figure 3, rectal cancer progresses along the usual cancer development model which begins with a long initiation phase, followed by three phases of increasing speed: promotion, progression, and metastasis. As the tumour acquires more hallmarks and mutations, it becomes genetically more unstable with the chance for the acquisition of further

new mutations becoming easier, resulting in increased aggression[10].

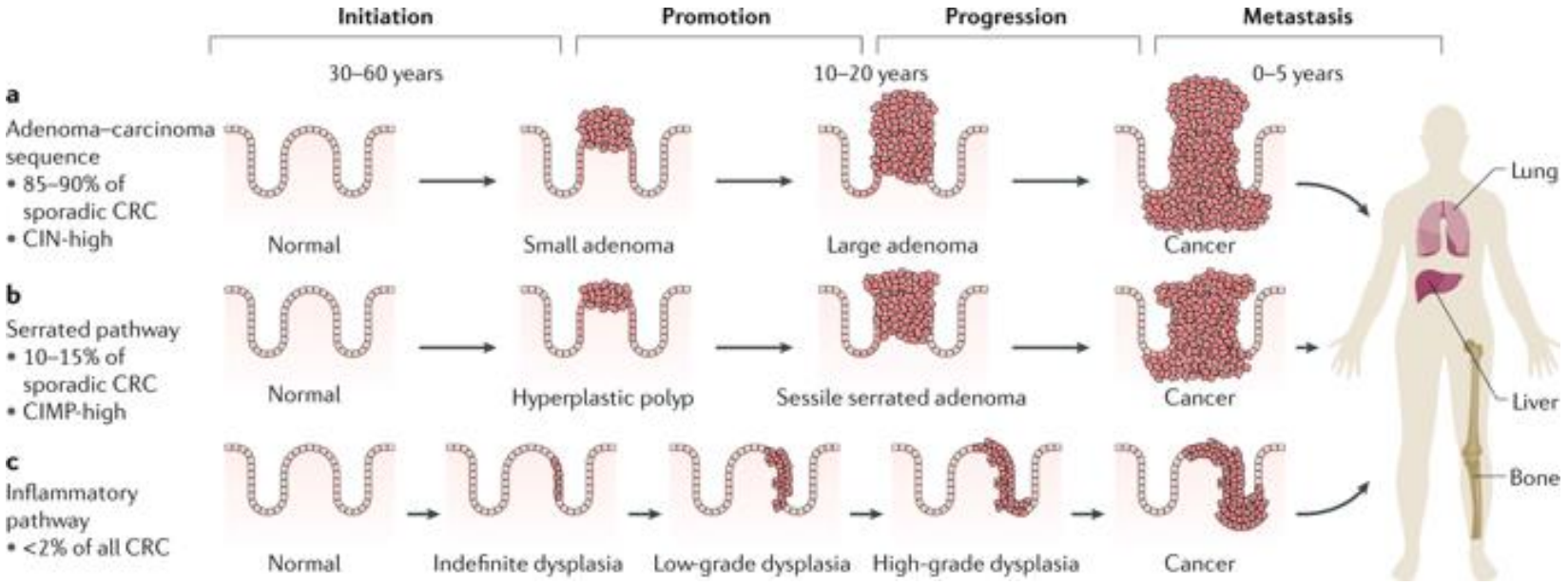


Figure 3. Colorectal cancer development pathways. This is one of the most common molecular classifications for colorectal cancer which splits tumours into 3 pathways: The adenoma-carcinoma sequence (85-90% of CRC), the serrated pathway (10-15% of CRC) and the inflammatory pathway (<2% of CRC). Each process is described as having a different cancer development model which is shown in the figure from left to right in order of tumour development beginning with initiation then leading onto promotion, progression and finally metastasis. [11]

Over the years many different molecular classifications of colorectal cancer have been described although there are two which appear to be the most used. I) CRC tumours can be characterised into three major separate pathways that are responsible for cancer development: the conventional pathway, the serrated pathway and the microsatellite instability pathway as shown below in figure 4.

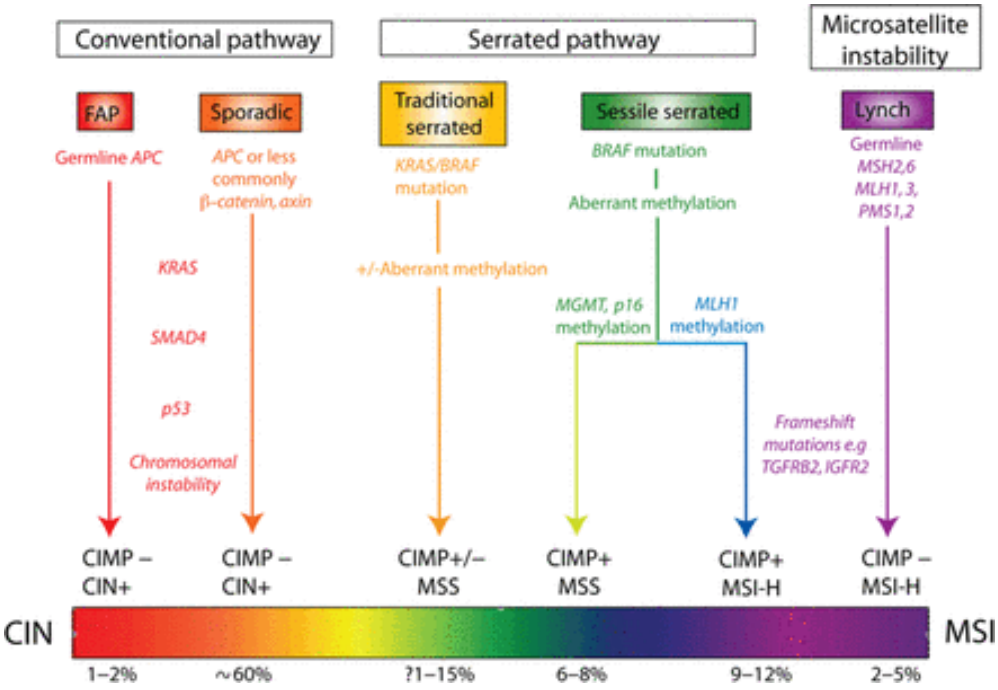


Figure 4. Colorectal cancer development pathways including molecular features. There are three molecular pathways highlighted in the white box at the top of the diagram which is the conventional pathway (including hereditary pathway, Familial Adenomatous Polyposis – FAP in red; and sporadic pathway in orange), serrated pathway (including traditional serrated in yellow and sessile serrated in green) and

microsatellite instability pathway (including lynch syndrome in purple). Each coloured arrow represents the pathway with key molecular features highlighted. The earlier the feature is shown on the arrow, the earlier the change is expected to occur. At the bottom of the arrow, the molecular phenotype is described which may be one of the following: chromosomal instability (CIN), CpG island methylator phenotype (CIMP), microsatellite stable (MSS) and Microsatellite instability – high (MSI-H). The (+) next to one of these phenotypes refers to being positive for the phenotype whilst (-) refers to being negative for the phenotype. At the very bottom of the diagram, below the sliding scale is a representation of what proportion of patients are expected to have the specific phenotype. [12]

The three pathways are founded upon disruption to a mixture of 3 different mechanisms. The first mechanism, chromosomal instability (CIN), is found in 65-70% of colorectal tumours. CIN initially occurs with a sporadic or inherited mutation in the APC gene[12]. CIN refers to an advanced state of chromosomal change, whereby large or complete losses and gains occur to chromosomes at high frequency[12]. This may result in aneuploidy, loss of heterozygosity or high levels of amplification[12]. CIN usually progresses with a sequential accumulation of mutations to key oncogenes or tumour

suppressor genes[12]. The second mechanism, CpG island methylator phenotype (CIMP), occurs in 10-15% of colorectal tumours[12]. CIMP refers to a state of hypermethylated CpG islands, which in cancer is associated with the inactivation of tumour-suppressor genes, typically through disruption to the promoter region*[12]*. CIMP is identified predominantly by hypermethylation of either MLH1, P16 or MGMT*[12]*. Disruption of MLH1 is a leading cause of the third mechanism and is why the two mechanisms can overlap frequently. The third mechanism, microsatellite instability (MSI), occurs in 10-15% of colorectal cancers and is strongly associated with Lynch Syndrome but occurs mostly in the sessile serrated pathway when caused by CIMP-based MLH1 hypermethylation. MSI occurs when mismatch-repair (MMR) mechanisms are impaired, usually by changes to the MLH1, PMS2, MSH2 and MSH6 proteins that are clinically identified by immunohistochemistry*[12]*. MMR impairment results in insufficient repair of spontaneous errors in DNA replication, especially base mismatches and indel loops*[12]*. Inefficient repair leads to genetic hypermutatability which causes activation of key oncogenes such as BRAF through mutational selective pressure*[12]*. Loss of MMR results in failure to repair indel loops (IDL) which often occur at small repeat nucleotide sequences, called microsatellites that occur throughout the

genome[13]. The failure to repair IDLs results in multiple alleles of different sizes and this phenomenon is called microsatellite instability[13]. Certain locations such as the loci BAT25 and BAT26 consist of very long runs of repeat sequences and are therefore very good markers for MSI[13]. Like CIN, there are varying levels to which MSI occurs. MSI is typically categorised into MSI-high and MSI-low[13]. Tumours without MSI, which accounts for most rectal cancers, are classified as microsatellite stable (MSS)[13].

ii) Based on genetic and gene expression differences, a new classification called the consensus molecular subtype (CMS), has been proposed[14]. Solid adenocarcinomas like rectal cancer always contain a plethora of molecular changes that are unique to the patient. Four subtypes of the CMS exist called CMS1-4 as shown below in table 1[14].

Table 1. Consensus molecular subtypes (CMS) to categorise colorectal cancer. A new widely regarded model for CRC is the CMS. The CMS is based on omic scale data to categorise CRC into 4 distinct categories (CMS1-4) based on molecular features, pathways involved and microenvironment composition. The table also suggests therapeutic targets based on the molecular information and microenvironment composition. The proportion of CRC tumours with CMS1 is

14%, CMS2 is 37%, CMS3 is 13%, CMS4 is 23% and the remaining tumours are mixed or indeterminate. Molecular features include MSI, hypermutation, somatic copy number alterations (SCNA) – split into high and low, CIMP - split into high and low, and MSS. Microenvironment composition is based on levels of high, low, or intermediate stromal infiltration and immune infiltration. [14]

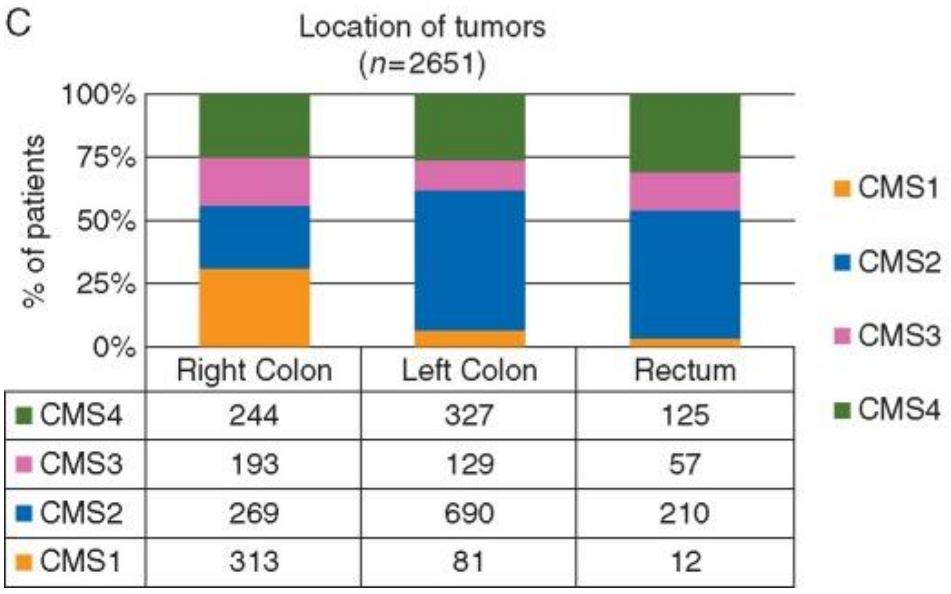
%	CRC subtypes	Clinical and prognostic associations	Molecular features	Pathways involved	Microenvironment composition	Potential target therapy
14	***CMS1*** **(MSI-Immune)**	• Good prognosis • Poor survival after relapse • Frequently diagnosed in females • Higher histopathological grade • Right-sided lesions (77%)	• MSI (76%) • Hypermutation (94%) • SCNA LOW • CIMP HIGH • Enriched *BRAF*-mut (42%)	• RTK and MAPK pathways activation	• HIGH immune infiltration • INTERMEDIATE stromal infiltration	• Anti-checkpoint immunotherapies • PD-1/PD-L1block in MSI patients
37	***CMS2*** **(Canonical)**	• Better survival rates after relapse • Left-sided lesions (59%)	• MSS (99%) • SCNA HIGH (*HNF4a*-amp) • CIMP LOW	• Wnt and MYC hyper-activation • Low Mesenchymal, EMT, TGFβ activation • High miR-17–92 cluster	• LOW immune infiltration • LOW stromal infiltration	• Enhanced tumours' immunogenicity • Cancer vaccines • Adoptive T cells therapy
13	***CMS3*** **(Metabolic)**	• Right-sided lesions (51%)	• MSI (16%) • Hypermutation (28%) • SCNA LOW • CIMP LOW • Enriched *KRAS*-mut (68%)	• Metabolic burst • RTK and MAPK-activation • Low VEGF/VEGFR • Low Mesenchymal, EMT, TGFβ activation • Low let-7 miR family	• LOW immune infiltration • LOW stromal infiltration	• Enhanced tumours' immunogenicity • Cancer vaccines • Adoptive T cells therapy
23	***CMS4*** **(Mesenchymal)**	• Worse relapse free and overall survival • Diagnosed at more advanced stages (III and IV)	• MSS (92%) • SCNA HIGH • CIMP LOW	• Angiogenic signature (High VEGF/VEGFR) • Inflammatory signature • Immunosuppressive signature • Mesenchymal, EMT, TGFβ activation • Low miR-200 family	• INTERMEDIATE immune infiltration • HIGH stromal infiltration	• Anti-angiogenic • Anti-inflammatory • Anti-checkpoint immunotherapies • Anti-fibrotic agents
6	***Mixed***					
7	***Indeterminate***					

The CMS refines the classical characterisations down to 4 groups to simplify stratification but also to expand on the breadth of information available to enhance personalisation of treatment. Since the discovery of the complete curative effects of Ipilimumab in some melanoma patients and the invention of CAR T-cell therapy, there has been a large wave on interest in immunotherapy in the hope that curative effects can be replicated in other cancers[15]. This raises the possibility of including "immune biomarkers" into the CMS classification[14].

In recent years, researchers have discovered that different tumours have different immune signatures. The type of immune signature can be identified by immunohistochemistry with markers such as CD3 (all T cells) and CD8 (cytotoxic T cells), that are used in the Immunoscore[16]. While the Immunoscore refers to tumour prognosis, it is also possible to use immune biomarkers to predict response to immunotherapy[16]. Tumours with low numbers of tumour infiltrating immune cells are characterised as immune 'cold' and classically have low PD-L1 and minimal expression of surface neoantigens[16]. Tumours with high PD-L1 expression on tumour associated immune cells, high presence of tumour-infiltrating lymphocytes and a pre-existing tumour immune response are characterised as immune 'hot'[16]. The CMS

defines group 2 (canonical) and group 3 (metabolic) as immune cold tumours, whilst group 1 (MSI-immune) and group 4 (mesenchymal) are immune hot; with the latter being weaker in immune activity than group 1[14]. Categorisation based on immunity leads to enhanced patient stratification for immunotherapy moving forward. It is likely that CMS1 and 4 will be treated by boosting the existing immune response with anti-checkpoint inhibitors[14]. CMS2 and 3 may be treated by cancer vaccines and CAR-T cell therapy to induce an immune response[14]. Radiotherapy is thought to be a good candidate for combination with immunotherapy as it breaks down barriers to entry for immune cells, releases potential target antigens and draws a large immune response to the site for damage of normal cells in the surrounding area[17].

Colon and rectal cancer are in the most part grouped together when analysed which can make it difficult to categorise data for either individually. As mentioned in section 1.A.i, the rectum differs in several important ways anatomically, but this is also the case on a molecular level. Fontana et al completed a comprehensive analysis of CMS in 2019 and as part of this, categorised the CMS across 2651 patients based on tumour location as seen in figure 5 below[18].



	Right Colon	Left Colon	Rectum
CMS4	244	327	125
CMS3	193	129	57
CMS2	269	690	210
CMS1	313	81	12

Figure 5. Anatomical distribution of consensus molecular subtypes in colorectal cancer. The bar chart at the top of the figure splits CRC into three categories: Right colon, left colon and rectum. The bar chart has 4 colours referring to the % of patients with CMS1 (orange), CMS2 (blue), CMS3 (pink) and CMS4 (green). The table below shows how many patients were in each category. The total number of tumours is 2651 of which 1019 are in the right colon, 1227 are in the left colon and 404 are in the rectum. [18]

Differences between the left and right side of the colon and rectum are widely known but in this context it is apparent that rectal cancers are similar in CMS distribution to the left colon and completely different to the right colon[18]. CMS1 has the lowest percentage of patients for the rectum but the highest for the right colon, showing a complete switch in the number

of patients with MSI-immune tumours. Generally it is known that MSI is strongly associated with right-sided colonic tumours compared to the left-side and rectum[19]. This difference reduces treatment availability for rectal cancer patients as the majority of the patients are immune-cold or mildly immune-hot. As most patients are undergoing chemoradiotherapy however, the combination with immunotherapy for the immune-cold majority could be favorable in this instance. This is also favorable for chemoradiotherapy treatment as MSI tumours are known to be resistant to 5-FU[20]. The overwhelming majority of rectal tumours are CMS2 and CMS4 that are both categorised as MSS, SCNA high and CIMP low but differ in microenvironment composition.

iv. Current clinical diagnostics and screening methods

Rectal cancer is currently diagnosed by a combination of diagnostic histopathology and extensive involvement of imaging technologies such as CT scan, MRI, PET-CT and colonoscopy/sigmoidoscopy to determine primary tumour location, presence of metastasis, nodal status or response to treatment[21]. Genetic testing of KRAS, BRAF and NRAS by pyrosequencing and immunohistochemistry of MMR proteins aids in characterising patients by molecular phenotype to determine their prognosis and course of treatment[21]. In

addition, blood-based carcinoembryonic antigen (CEA) may be used to determine cancer presence and response to treatment[21], but is known to produce false-positives and have relatively low sensitivity and thus is only useful in certain scenarios[22]. Below in figure 6 shows the current clinical pathway for rectal cancer diagnostics[21].

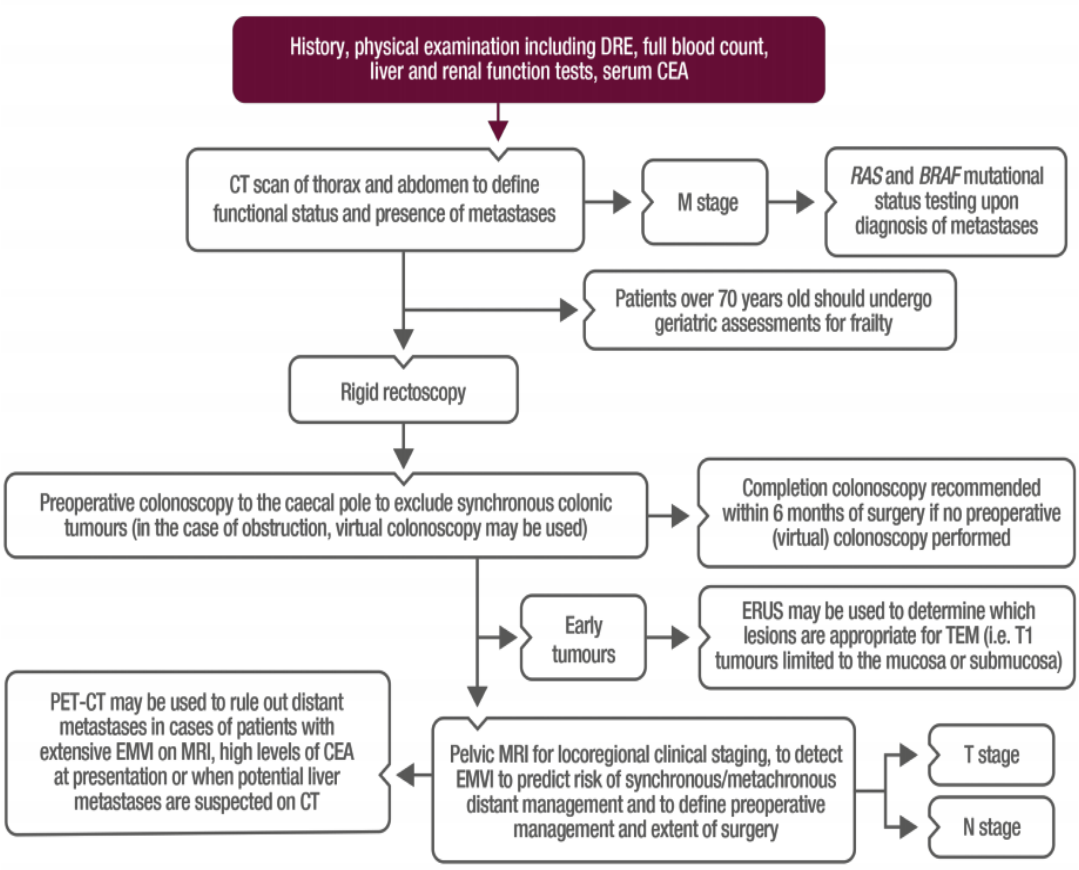


Figure 6. Colorectal cancer diagnostic pathway according to European Society for Medical Oncology (ESMO). The diagnostic pathway is represented by flow chart which begins at the top of the figure in red. The diagnostic pathway uses a number of imaging technologies and histology to define tumour depth

(T), node status (N) and metastasis status (M) which ultimately enable characterisation of tumour stage. [21]

The diagnostic pathway produces several key features both on a molecular level as discussed in section 1.A.iii and anatomically which determine tumour stage and ultimately treatment strategy. The T status refers to tumour depth penetration and is classified based on the number of layers of the rectal wall which have been invaded by the tumour[21]. The N status refers to the number of lymph nodes which the tumour has metastasised to and the M status refers to the presence or absence of metastasis at a distant site[21]. Ultimately, T and N are redundant in staging when metastasis is present as shown below in figure 7[23].

Stage 0:	Tis	N0	M0
Stage I:	T1 - T2	N0	M0
Stage IIA:	T3	N0	M0
Stage IIB:	T4a	N0	M0
Stage IIC:	T4b	N0	M0
Stage IIIA:	T1 - T2	N1 / N1c	M0
	T1	N2a	M0
Stage IIIB:	T3 - T4a	N1 / N1c	M0
	T2 - T3	N2a	M0
	T1 - T2	N2b	M0
Stage IIIC:	T4a	N2a	M0
	T3 - T4a	N2b	M0
	T4b	N1 - N2	M0
Stage IVA:	any T	any N	M1a
Stage IVB:	any T	any N	M1b
Stage IVC:	any T	any N	M1c

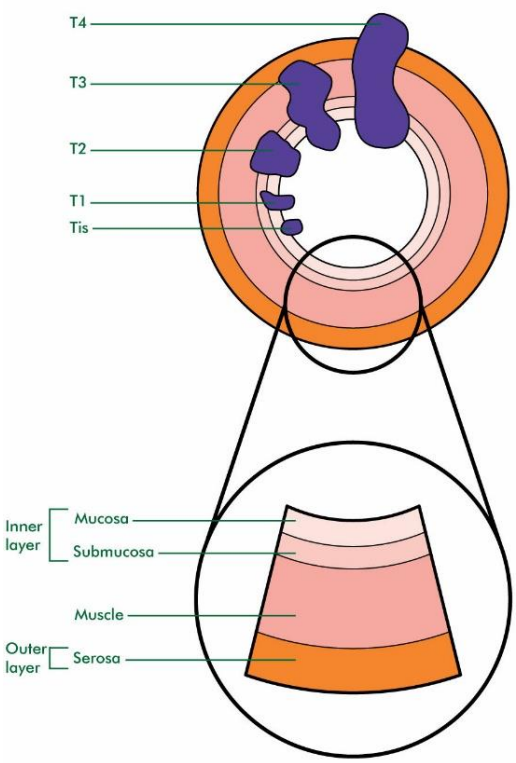


Figure 7. Colorectal cancer staging system. T = Tumour depth which indicates how far the tumour has penetrated through muscle layers from the mucosa. N = Node status which refers to how many lymph nodes have been metastasised to. M = presence or absence of metastasis in a distant organ. Keys that are not described by the diagram on the right which defines T stage based on penetration through the inner and outer layers of the bowel are as follows: T4a = tumour invades through the visceral peritoneum; T4b = tumour directly invades or adheres to other adjacent organs or structures. M0 = no distant metastasis; M1 = distant metastasis present; M1a = metastasis confined to 1 organ or site without peritoneal metastasis; M1b = metastasis to 2 or more sites or

organs without peritoneal metastasis; M1c = metastasis to the peritoneal surface alone or with other site or organ metastasis. N0 = No lymph node metastasis; N1a = 1 lymph node metastasis; N1b = 2-3 lymph node metastasis; N1c = no lymph nodes positive but tumour deposits in subserosa, mesentery or nonperitonealized pericolic or perirectal/mestorectal tissues; N2a = 4-6 lymph node metastasis; N2b = 7 or more lymph node metastasis. [23]

Staging is usually combined for both the colon and rectum but it is important to note that the majority of the rectum is not covered in serosa and thus T3 is classed as infiltration through the outer most muscle layer and T4 can only be classified if there is invasion into adjacent organs/structures (such as bladder, prostate etc)[21]. Staging is the most important factor in determining treatment strategies for a patient and is discussed in more detail in section 1.B.

Since 2006, patients at age 55 and over in the UK have been offered bowel cancer screening on the NHS[24]. The Bowelscope cancer screening program deploys flexible sigmoidoscopy to search for abnormal growths which, if found, can be interrogated by diagnostic histopathology of a biopsy[24]. The Bowel Cancer Screening Program (BCSP) will screen patients between the ages of 60 – 75 and offers full

bowel colonoscopy if there is evidence of bleeding (through positive stool tests such as Faecal Immunochemical Test or Faecal Occult Blood Test – see below) [24]. The screening programs have dramatically shifted the stage of presentation of disease. Almost half the patients with CRC identified through the screening programs are Stage 1 disease whilst around 1% are Stage 4 disease[25]. The frequency of Stage 2 and 3 has remained like that in the non-screened population[25]. Bowel screening has in turn improved survival of colorectal cancer patients dramatically and makes a strong case for screening and monitoring of disease to improve outcome for all cancers[25].

The bowel cancer screening program relies on two different fecal-based screening tests, the Fecal Immunohistochemical test (FIT) and Guaic-Based Fecal Occult Blood test (gFOBT) [24]. These are used to check for bowel bleeding on consecutive stool samples. FIT and gFOBT are highly sensitive for detecting colorectal cancer in the bowel as bleeding is a common symptom[24]. The two assays are followed up by colonoscopy to confirm tumour presence[24].

B: Radiotherapy treatment for rectal cancer

Radiotherapy in rectal cancer

Radiotherapy is a long-established therapeutic modality for rectal cancer[26, 27]. Worldwide it is generally accepted that early-stage or favourable disease (cT1-2 and some early cT3) is treated with surgery alone; whereas radiotherapy is highly recommended trimodally combined with chemotherapy and surgery for intermediate and locally advanced tumours (cT3 onwards)[21]. Many different treatment regimens which incorporate radiotherapy for rectal cancer exist and a small minority of patients may be treated with radiotherapy alone[27]. However, in the UK, the optimal treatment regimens to be used are generally decided by each local colorectal cancer multidisciplinary team (MDT)[27]. There are many differences between regimens and the main differences are as follows: the presence or absence of chemotherapy; the type of chemotherapeutic agents that are used; the application of radiotherapy or chemoradiotherapy (CRT) in either a long or short course; application of radiotherapy or CRT in an adjuvant or neoadjuvant manor and the time between radiotherapy or CRT and surgery. All factors can play a role in the patients' overall response and are usually chosen in a stage dependent fashion[21].

All patients that have completed trimodal therapy will undergo investigative histopathological analysis for their resection specimens[21]. A number of tumour regression grading

systems exist which aim to provide prognostic and predictive information based on histology[28]. Estimated quantification of therapy-induced fibrosis and the percentage of post-treatment tumour cells in comparison to the pre-treatment site are characterised by histopathologists[28]. Most systems characterise these major observations into a specific grade. The four most common systems used are Mandard, Becker, Dworak and Rödel[28]. The level of regression is considered a direct indicator for 5-year OS and DFS chances[28]. Many studies indicate that a pathological complete response (pCR) can be obtained for up to 20% of rectal cancer patients treated with neoadjuvant CRT[26, 29-32]. Moreover, up to 60% of patients can achieve tumour downstaging[30, 33, 34]. A systematic review and meta-analysis demonstrated that patients obtaining a pCR after neoadjuvant CRT (n=484) had a higher 5-year DFS (83.3%, 95% CI 78.8-87.0) compared to those without a pCR (65.6%, 95% CI 63.6-68.0) (747/2263; HR 0.44, 95% CI 0.34-0.57; p<0.0001)[31, 35]. Unfortunately, 8.7% of patients, despite achieving pCR still develop metastasis [35].

Radiotherapy across the UK NHS and Nottingham's approach

According to Morris et al, who compiled data for all patients in the UK NHS (n = 9201) diagnosed with rectal cancer between 1st April 2009 and 31st December 2010; 49.3% (n = 4585) received radiotherapy[27]. The treatment regimen used the most in this period was long-course CRT with surgery, which was defined as 25-30 radiation centre visits in total[27]. As shown below in fig 1. there is large variation between local area teams of the UK NHS with the type of radiotherapy regimens that are used. This is primarily through lack of regulatory standardisation. The MDT has flexibility in what it sees fit based on current guidelines and this leads to variation in regimens based on what the MDT deems to be the most important factors.

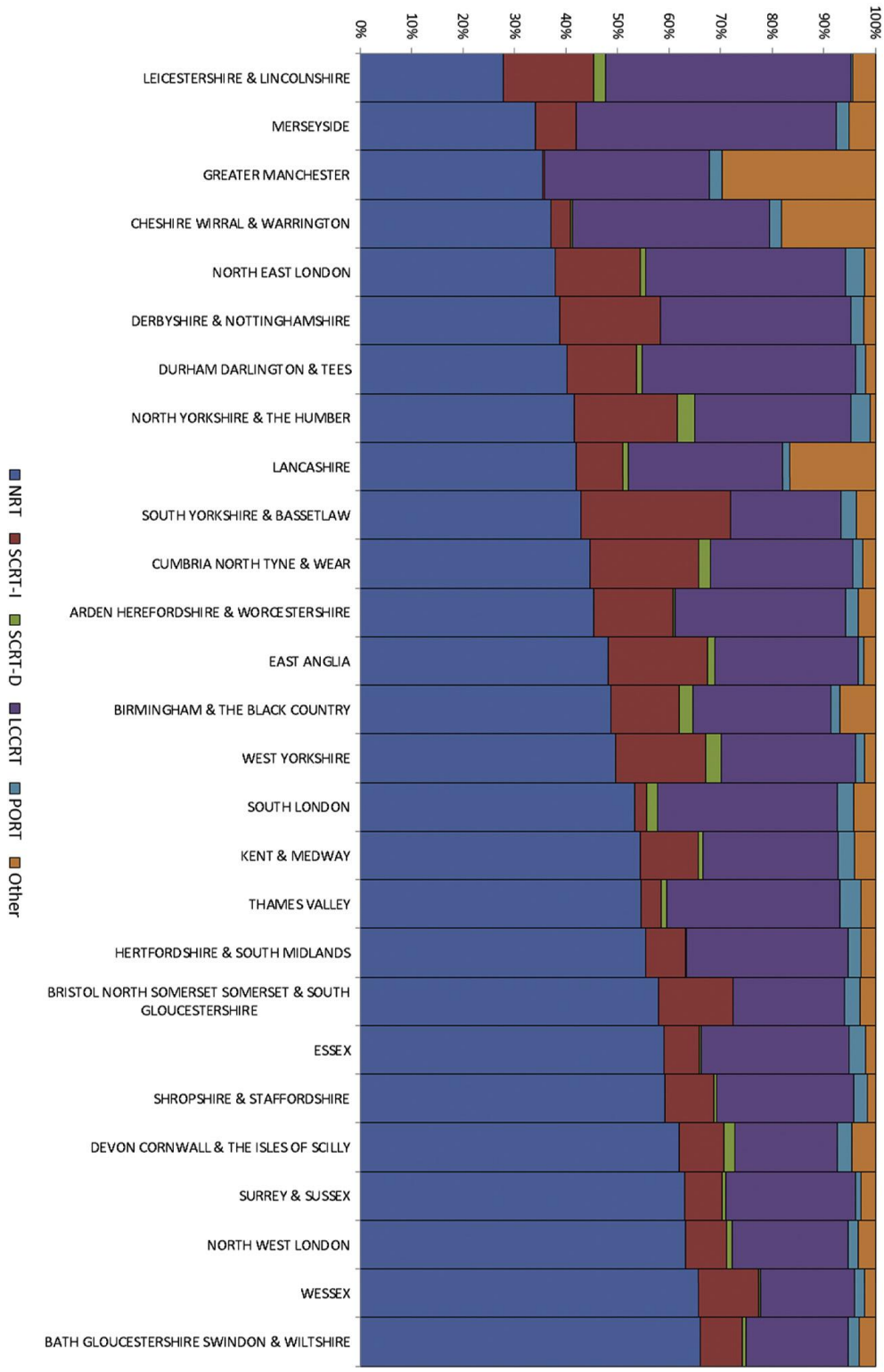


Figure 8. The proportion of each treatment strategy used for rectal cancer patients across all teams in the NHS. NRT (dark blue)= No radiotherapy; SCRT-I (red) = short-course

radiotherapy followed by immediate surgery (5 radiotherapy centre visits); SCRT-D (green) = short-course radiotherapy followed by delayed surgery (5 radiotherapy centre visits); LCCRT (purple) = long-course chemoradiotherapy (25-30 radiotherapy centre visits); PORT (light blue) = post-operative radiotherapy (received radiotherapy up to 1 year after surgery); other = other radiotherapy (non-standard attendance pattern).[27]

As the whole of this study will be conducted on patients from hospitals in Nottingham, it is appropriate to focus on the Nottingham system. Dhadda et al published a paper in 2011 describing the methodology and patient outcomes for 158 patients treated at Nottingham hospitals in relation to the Mandard score. Nottingham has treated locally advanced rectal cancer with neoadjuvant capecitabine-based chemoradiotherapy with surgery since May 2004[36]. Prior to this, chemotherapy was not included in the treatment regimen. The usual radiotherapy treatment plan includes 50Gy of radiation delivered by 10Mv of photons over the course of twenty-five 2Gy/day fractions, 5 days a week[36]. Chemotherapy begins early; one cycle of capecitabine (1250mg/m2 twice a day for 14 days out of 21 days) is given before radiotherapy begins. During radiotherapy, the dosage is reduced to 825mg/m2[36]. Surgery was carried out 6-8 weeks

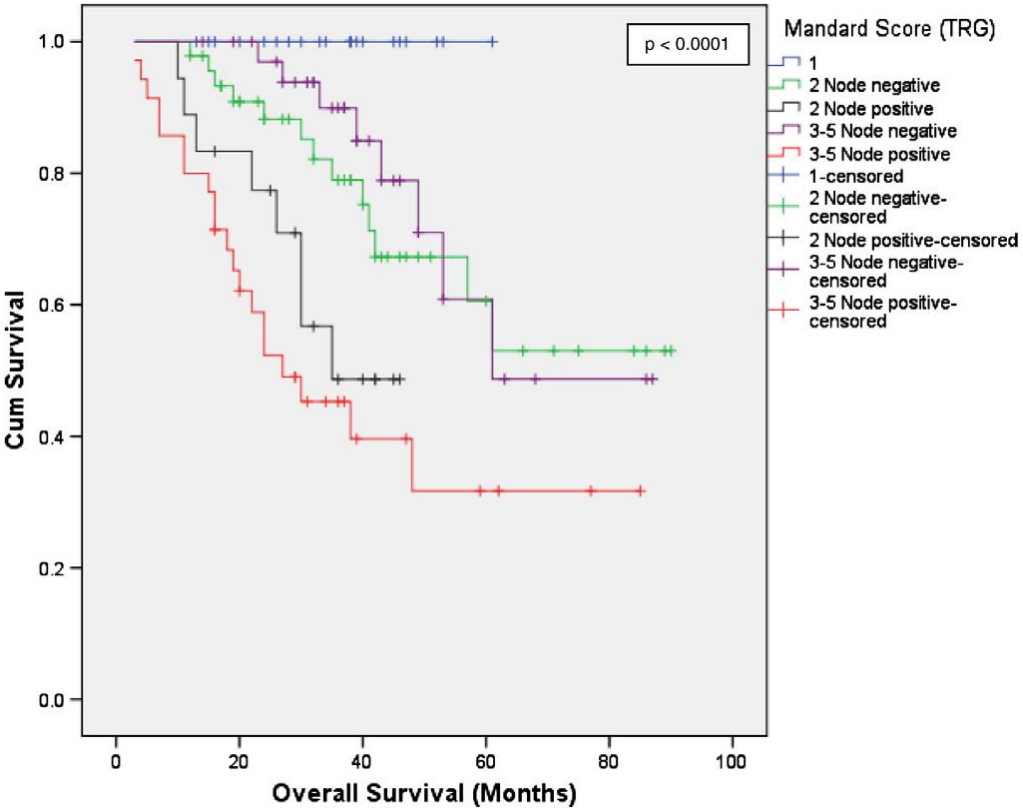
after the end of concurrent chemoradiotherapy[36]. As shown in figure 8, Nottingham & Derbyshire ranked 6 out of 27 (low to high) for no radiotherapy, showing favour towards the use of pre-operative radiotherapy for treatment. Moreover, the use of post-operative radiotherapy and alternative regimens, deemed 'other' was exceptionally low compared to other areas and comprised as little as 4% of the overall regimens with a larger focus on long-course CRT over short-course CRT by more than 2-fold.

As described above radiation is given in fractions. Fractionation uses small doses of low-linear energy transfer (LET) ionising radiation (i.e., β particles) as opposed to large doses of high-LET radiation (i.e., α particles and protons)[37-40]. High-LET radiation can cause tremendous damage to tumour cells but is less targeted and results in high levels of normal tissue damage and typically results in radiation-induced cancer[38-41]. However, since the recent developments of targeted proton beam therapy[42], high-LET radiation is currently under review in the UK and is available at two NHS centres[43]. In general, low-LET radiation is preferred as it can be targeted using intensity modulation to minimise normal tissue damage[44]. Fractionated radiotherapy leads to several benefits. Firstly, cells will either die or obtain sublethal damage or potentially lethal

damage[38-40]. Typically, tumour cells suppress repair pathways and cannot repair the damage, but normal cells can, leading to preservation[38-40]. Secondly, S-phase cells are radioresistant whereas cells in G2 and M phase are sensitive[38-40]. A delay in radiotherapy allows cells to redistribute through the cell cycle into more sensitive phases to increase efficacy[38-40]. Thirdly, tumours are typically hypoxic in their cores due to lack of blood supply via lacking angiogenesis[38-40]. Hypoxic cells are resistant to radiation[38-40]. Thus, killing cells on the outside of the tumour allows oxygenation of some hypoxic cells in the core in a stepwise manor[38-40].

The study by Dhadda presents a modification of the Mandard system and includes nodal status. The grading system is as follows: TRG 1: complete response with absence of residual cancer and fibrosis extending through the wall; TRG 2: presence of residual tumour cells scattered through the fibrosis; TRG 3: increase in the number of residual cancer cells, with fibrosis predominant; TRG 4: macroscopic tumour; absence of regressive changes; any node positive within irradiated volume[36]. TRG5 was removed to further simplify the system. As can be seen below in figure 9, patients stratified into specific response grades according to histopathological data show a trend; the higher the grade the

lower the overall survival and disease-free survival becomes over time.



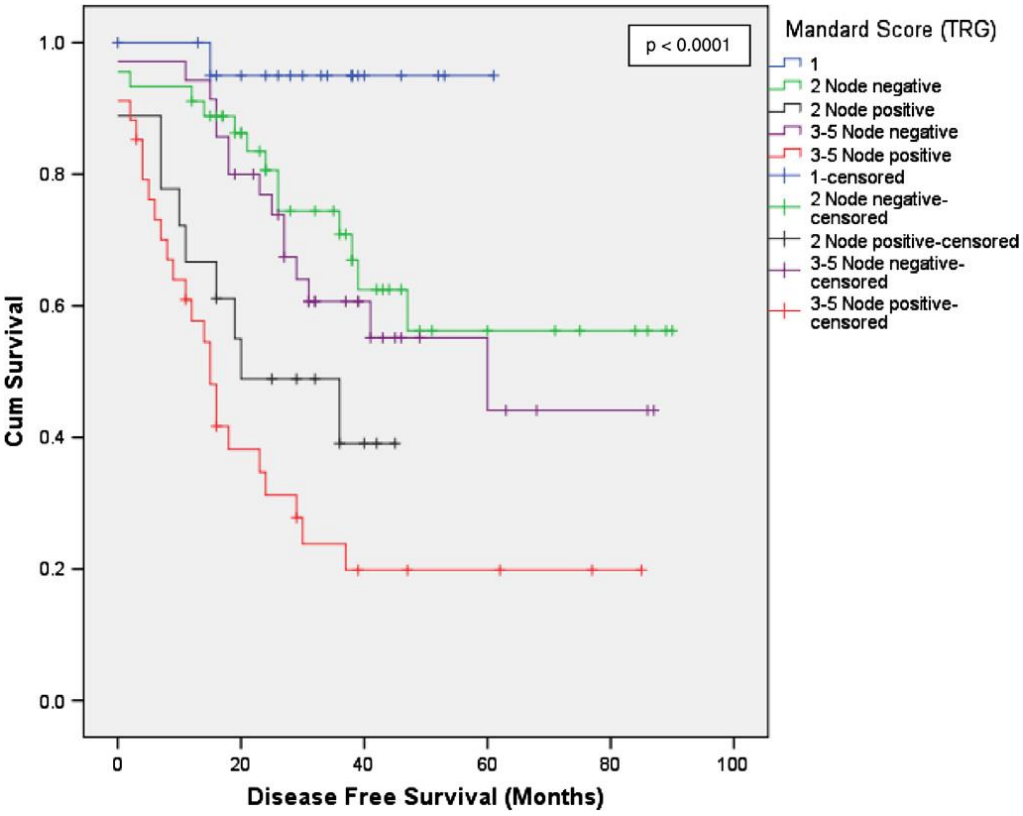


Figure 9. Survival outcomes according to the Mandard score stratified by nodal status for 158 patients treated at Nottingham hospitals.[36] The top diagram plots cumulative survival against overall survival in months. The bottom diagram plots cumulative survival against disease-free survival in months. The Kaplan-Meier curves are colour coded and the reference for each colour is found to the right-hand side of the graph that are based on TRG and node status.

It is clear from the Nottingham study that overall responses to neoadjuvant CRT with surgery ranges from total response to complete lack of response with some patients partially responding[36]. Most cancers are prone to recurrence either locally or at a distant site which is a leading cause of cancer-related death[45]. Typically, local recurrence occurs due to inadequate surgery margins or resistance to treatment whilst metastases either resist treatment or were not detected during diagnosis and treatment[45]. It is estimated that 25% of all rectal cancer patients receiving neoadjuvant CRT will develop distant metastases and this is now the leading cause of death whilst local recurrence rates have dropped to around 5-6%[46]. Currently, clinicians are unable to effectively detect minimal residual disease in real-time by radiology[47] and, until recently, the management of rectal cancer inevitably involved surgery. However, since up to 20% of patients may

achieve complete regression, patients are offered a "watch-and-wait" protocol whereby both the clinician and the patient wait to see if the cancer returns[48]. As well as causing stress to the patient, this scenario causes unnecessary lag time for institution of effective treatment[48].

There is a need for accurate prediction of tumour response to CRT to prevent unnecessary major surgery for patients in whom the tumour has disappeared[48]. Moreover, there is a need to determine patients who are not responding to therapy so that they may get surgery sooner or other forms of treatment[48]. There are two possible strategies i.e., real time surveillance to monitor tumour recurrence after treatment or identification of robust predictive markers. These are discussed in sections 1.C and 1.D.

C: Real-time monitoring of tumour response via liquid biopsy

i. Introduction to cell-free tumour DNA (cfDNA)

Currently, 20% of rectal cancer patients completely respond to chemoradiotherapy and the other 80% range from minimal response to no response[26, 29-34]. Clinicians are unable to effectively monitor response to therapy in real-time and patients who completely resist treatment are still required to complete therapy whilst the watch-and-wait scenario occurs as

response cannot be predicted[48]. Real-time treatment response monitoring may allow patients to be quickly stratified according to response so that they do not receive treatment unnecessarily and can be moved to another potentially effective treatment quicker. Thus, stratifying patients into the correct pathway may lead to increased survival rates. Patients that do not respond have time wasted in a critical window for therapy and, furthermore, the treatment on its own may be mutagenic and could cause the tumour to become increasingly aggressive. The side effects that the patient will endure are also extremely unpleasant and need to be minimised when treatment is ineffective. Real-time monitoring would also allow surveillance of cancer patients to determine tumour recurrence[49]. If there is no recurrence, then no treatment is required. However, if recurrence does occur, earlier diagnosis of tumour recurrence would allow earlier treatment. Early clinical intervention may improve outcome since treatment can take place when the recurrent tumour burden is low and thus there is greater chance of eradicating the tumour.

Liquid biopsies (blood samples taken from the arm) are a rapidly emerging solution to monitor patients in real-time[49]. Many types of liquid biopsy analyte exist such as circulating cell-free nucleic acids (cfNA), circulating endothelial cells, circulating tumour cells, circulating exosomes and tumour-

educated platelets[50]. Whilst many types of liquid biopsy analyte exist and can provide useful information, this thesis will focus specifically on cfNA analysis. CfNA includes exosomes as they are extracted concomitantly but exosomes can be specifically isolated and analysed. Exosome research will not be included in this thesis.

Tumour genetics and the tumour environment are highly volatile and these cause tumour cells to die[51]. The tumour cell death causes release of nucleic acids into the blood stream which can then be detected in blood samples as cfNA. The amount of tumour-derived cfNA depends on several factors such as tumour size, turnover rate, and vascularity. The umbrella of cfNA includes circulating free DNA (cfDNA), circulating free microRNA (cf-miRNA) and circulating free RNA (cfRNA)[52].

The total cfNA detected in a liquid biopsy has two sources – the tumour and the rest of the body[50]. The tumour derived portion is specifically referred to as circulating tumour Nucleic acid (ctNA). Thus, circulating DNA derived from a tumour is referred to as circulating tumour DNA (ctDNA). Detection of ctDNA (through the presence of mutation) can confirm tumour presence and provide information which can lead to either treatment stratification, detection of minimal residual disease,

early detection of cancer prior to diagnosis or identification of resistance mutations[49].

CtDNA is currently the most promising liquid biopsy analyte for real-time detection of minimal residual disease and treatment-response monitoring[49], mainly due to significant advances in extraction and detection. Other types of liquid biopsy will require significant advancements and comparative analysis with ctDNA to prove their validity[49]. Current research suggests that ctDNA detection is a more sensitive method of cancer surveillance compared with existing biomarkers[53] or imaging technologies[47, 53, 54] and it is indicative of poor prognosis[49, 55]. Tumour specific DNA features are broad and can be detected singularly or as a mixture. Mutation status, copy number, microsatellite status and methylation are examples of DNA variants that are typically analysed[56].

Treatment response monitoring requires quantification analysis to determine if the tumour is being eradicated or resisting treatment. The quantity of extracted cfDNA as well as the ratio of the chosen tumour specific feature to compared to normal cfDNA present can be plotted over time to create a decay curve that can determine response to therapy. For example, as shown below if figure 10, if mutation load goes down this

would indicate response to therapy and vice versa with increased mutation load.

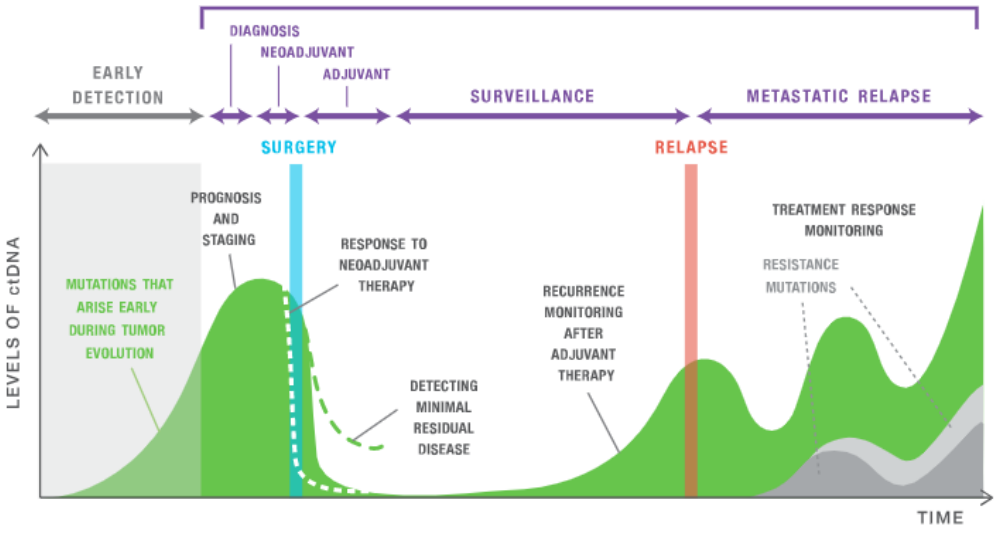


Figure 10. Diagram of predicted typical circulating tumour DNA expression throughout cancer progression and treatment. This diagram is based on known data regarding ctDNA. It represents the multi-faceted approach that liquid biopsy is now thought to be able to undertake involving treatment-response monitoring, detection of minimal residual disease, early detection, and determination of prognosis. The diagram represents a timeline of predicted cfDNA level fluctuations for undiagnosed cancer through diagnosis, neoadjuvant therapy, surgery, adjuvant therapy, post adjuvant therapy and metastatic relapse. The green wave represents the levels of ctDNA for early mutations whereas the grey wave represents acquired resistance mutations after relapse.[57]

Complete loss of mutation load over a long time course may indicate pCR and this may be even more effective after CRT ceases; as mentioned previously, a watch and wait approach before surgery is becoming increasingly common[30].

ii. Introduction to circulating tumour DNA (ctDNA) studies

The applicability of ctDNA for both research and clinical purposes has expanded immensely over the past 5 years[58]. The number of publications for liquid biopsy has been rising dramatically year on year and this has generated a plethora of data. A key issue for liquid biopsy publications pre-2016 has been the lack of appropriate extraction methodologies as they did not exist. The second major issue has been the generation of artefactual data from the method of interrogation. Next generation sequencing technology (NGS) has been frequently used to generate a large proportion of the data in the space but without error correction technology and high depth reads. In turn, this has resulted in incorrect reporting of mutations when not appropriately matched with high frequency mutations in FFPE tissue[59]. In addition, it is becoming widely accepted that studies analysing cfDNA alone are incorrectly calling sequence changes that occur via clonal haematopoiesis of white blood cells as mutations [60]. In 2020, there is now consensus that paired FFPE tissue or white

blood should be sequenced along with cfDNA, ideally both, to ensure accurate calling unless the mutation is a known driver. Known drivers are good targets since the chance of the mutation being real is statistically very high. Although caution should be taken if a particular treatment is targeting a protein with that mutation as it increases the chance for negative selection. High mutant allele frequency passenger mutations are desirable to ensure the mutation is not in error and to avoid negative selection. This has led to an increase in personalised testing to increase the number of mutations targeted per patient. In turn this reduces the number of false positive results and increases sensitivity.

There are numerous papers which describe the clinical utility of ctDNA. Reece et al in 2019 performed a systematic review of 92 clinical studies on ctDNA in CRC[61]. The study has shown that ctDNA is a reliable method of determining tumour burden and can be used to monitor tumour clearance and response to chemotherapy[61]. Moreover, the study describes that ctDNA presence post-treatment correlates with recurrence and can be used to determine emerging resistance mutations in the tumour[61]. This analysis provides overwhelming evidence for the clinical utility of ctDNA in CRC[61]. As discussed previously in section 1.A.iii, colon cancer and rectal cancer differ in many ways both anatomically and on a

molecular level. Since most studies are combining the two cancers and the overall majority of CRC is colon cancer, it cannot be assumed that rectal cancer is also behaving in the same manor when analysed alone. Huang et al in 2019 reported that colon cancer patients (8.25ng, n=26) had marginally higher ctDNA yield than rectal cancer patients (n=11, 7.51ng)[62]. Moreover, mutations in pre-treatment liquid biopsies which were concordant with the primary tumour were detected more frequently in colon cancer patients (24/26, 92.3%) compared with rectal cancer patients (5/11, 45.5%)[62]. Although the study by Huang et al is very small, there are currently no large-scale studies comparing the two types of cancer with sufficient power. A differing blood supply, molecular distribution, and the addition of radiotherapy to the course of treatment could potentially have an effect and should be taken into consideration. There are far fewer publications relating to rectal cancer specifically, especially including radiotherapy[49]. A systematic review by Massihnia et al in 2019 found 25 studies relating to ctDNA in rectal cancer[49]. Many of these studies were small scale with several large-scale studies included. Overall, most data were inconclusive, especially concerning treatment response monitoring of radiotherapy and early detection. This is possibly due to the artefactual data generation described previously.

However, most studies correlated ctDNA detection with worse prognosis and the larger-scale studies provided deeper insights into potential utility[49]. There is still room for improvement in both the method of analysis, extraction of analyte and ultimately, evidence that ctDNA analysis can improve outcome. Several papers providing strong supporting data for liquid biopsy are as follows:

In 2019, Tie et al conducted serial analysis on 462 liquid biopsies from 159 rectal cancer patients undergoing neoadjuvant chemoradiotherapy followed by surgery[63]. The study used SafeSeq[64], the most frequently used error correction method for NGS. Several important conclusions have been made by this study. Firstly, a higher percentage of patients had ctDNA present pre-surgery (77%) compared with the study by Huang et al in 2019 (45.5%)[63]. This shows that detection of pre-surgery mutations is undoubtedly different depending on either method of interrogation, cohort, or sample processing. Confirming that most studies are currently not definitive as further progress can be made. Secondly, there was a reduction in ctDNA post-surgery as the number of overall patients with positive ctDNA dropped from 77% to 12%, demonstrating surgical clearance of the primary rectal tumour[63]. Post-chemoradiotherapy the number of patients dropped further to 8.3% showing that ctDNA can

potentially determine treatment response[63]. This is further validated by the massive difference in 3-year recurrence free survival which was 33% for patients where ctDNA was detected post-surgery and 87% for patients with no ctDNA presence post-surgery[63]. Significantly worse recurrence-free survival was witnessed if ctDNA was detected after surgery (HR 6.6) or chemoradiotherapy (HR 13.0) [63]. This allowed stratification into high or low risk groups and was independent of clinicopathological risk factors, showing that ctDNA can be prognostic for rectal cancer patients[63].

Research by Shaw et al in 2019 demonstrated that 89% of patients (n=16/18) with breast cancer relapse could be detected on average 8.9 months prior to clinical relapse when followed up by personalised deep sequencing based ctDNA analysis with a maximum lead time of 2 years[47]. Moreover, all patients which were ctDNA negative (n=31) did not relapse[47]. Whilst on a small number of patients, these results are extremely significant. Patients which relapse are at high risk as the tumour typically comes back with increased resistance and aggressiveness. The generation of additional lead time could provide a critical window for therapeutic intervention which would otherwise be lost. This may improve outcome based on fundamental principles.

A study on ctDNA by Tie et al in 2016 from 230 colon cancer patients has also shown further clinical validity when compared to CEA[53]. Personalised targeted deep sequencing was used to detect ctDNA and compared with CEA for 27 patients who recurred as shown below in figure 11[53].

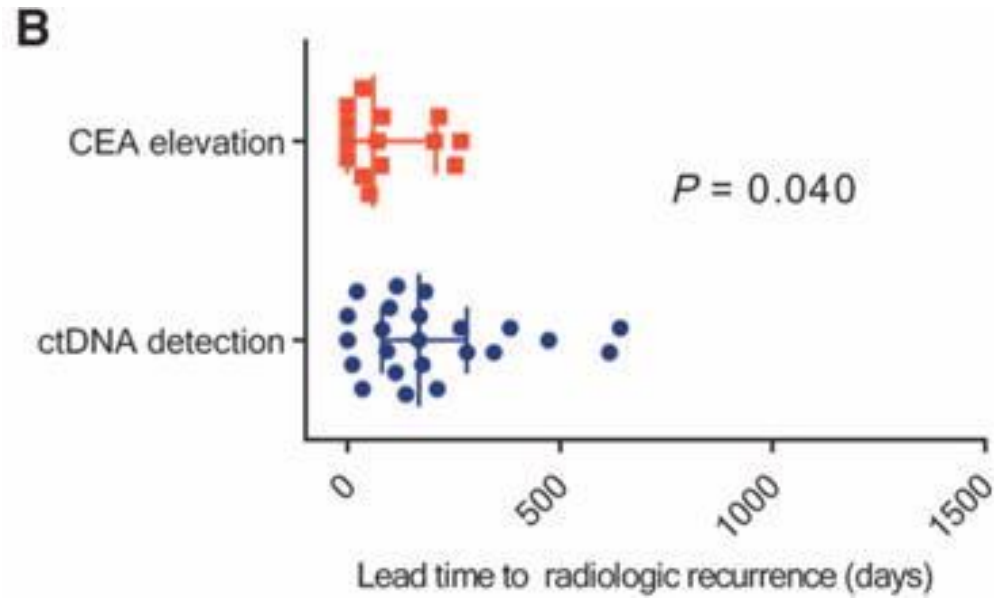


Figure 11. Scatter plot comparing the lead time to radiological recurrence in days for carcinoembryonic antigen and circulating tumour DNA testing. Line represents median with error bars. [53]

The average lead time to radiological recurrence for CEA was 61 days compared to 167 for ctDNA[53]. Moreover, at the time of radiological recurrence, 23/27 patients were ctDNA positive, whilst only 11/27 patients were CEA elevated[53]. The study demonstrates that in most patients, a longer critical window for treatment is being generated by ctDNA analysis

and it is vastly superior to CEA which is the current gold standard blood test for CRC treatment response monitoring[53]. This is extremely promising considering liquid biopsy research is still in the early phases and has plenty of room for improvement.

Research by Reinert et al in 2019 describes very high (88.5%) detection of mutations in pre-operative CRC liquid biopsies by personalised NGS[65]. This study was enriched for tumours with stage>2 and T>3. 81/130 patients had N>1, 76/130 had unknown metastasis status, 81/130 patients were stage 3 and no patients were classed as stage 4. 87.5% (14/16) of relapsed cases contained positive ctDNA and were independently associated after adjusting for clinicopathological risk factors as shown below in figure 12[65]. Whilst the detection is skewed by later stage cases, it shows that when appropriate extraction and personalised detection is carried out on a best-case scenario, very high detection rates can be obtained. This should translate to earlier stage cancer where the detection rates may be lower, but if ctDNA is present it will signify poor prognosis or an opportunity that needs to be acted upon. Similar results were obtained to the breast cancer study by Shaw et al in 2019 as serial analysis was able to predict recurrence 8.7 months earlier than radiology on average with a maximum of 16.5 months, demonstrating that

ctDNA analysis could also generate massive windows of opportunity for therapeutic intervention in CRC.

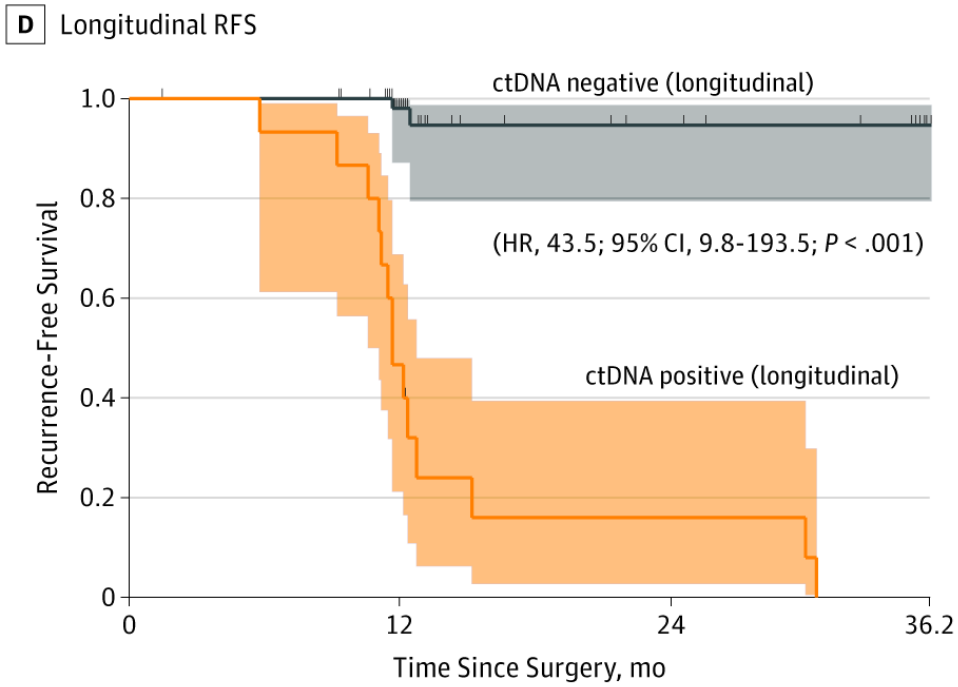


Figure 12. Comparison of 36-month recurrence-free survival between post-surgery circulating tumour DNA positive (n=14) and negative patients (n= 61). Kaplan-Meier are finished when patient proportion <10%. 95% confidence intervals noted by shaded areas. HR = hazard ratio. [65]

Overall, these studies show that ctDNA can be definitively prognostic, potentially predictive, generate large lead times over radiology by detecting the presence of recurrence and potentially play a role in treatment response monitoring. Although, the least evidence is given for the latter. A key point to consider is that different methods of interrogation may

result in a higher percentage of patients with pre-treatment ctDNA positivity. Demonstrating that different approaches, especially personalised, may improve sensitivity and is not necessarily based on absence of ctDNA production by the tumour.

iii. Introduction to mutation detection methods

CfDNA is reported to be highly fragmented with a mean fragment size of 130-166bp which correlates with the inter-nucleosome distance in DNA[66]. There may be exceptionally low concentration (1ng-100ng/ml)[67] and there may be a low of the percentage of tumour biomarker present[68, 69]. The combination of these factors pushes the mutation detection method to the very extremes.

Tumour-derived DNA can be obtained directly from tumour tissue or from bodily fluids such as liquid biopsies, urine, faeces and even saliva [56]. Unless obtained from cell lines in tissue culture, tumour-derived DNA will always contain a mixture of DNA originating from tumour cells and non-tumour cells. The latter may include DNA originating from stromal cells in a tumour or, in the case of bodily fluids, DNA originating from other regions of the body.

The ratio of tumour cell: non-tumour cell DNA is highly variable and in certain cases (such as bodily fluids or small

biopsies), the tumour cell DNA may comprise a tiny proportion of the total DNA content [70, 71]. In such cases, the detection of mutant alleles can become extremely problematic and has been likened to “finding a needle in a haystack” [72]. The probability of finding this "needle" is dependent on technical performance of the assay used, in particular, the limit-of-detection. This refers to the number of mutant alleles which must be present in the overall pool of nucleic acid for them to be detected. For example, fluorescence-based Sanger sequencing has a limit of detection of 20%, meaning that mutations will not be detected in samples which contain < 20% tumour cells; disregarding heterogeneity in which it may be even lower [73].

Over the years, several methods have been used to detect low levels of mutant sequence. These have ranged from the development of novel chemicals which have much greater sensitivity than standard fluorescent sequencing compounds [73] through to the use of water-oil emulsion droplet technology such as digital droplet PCR[74]. Recently, the development of massively parallel sequencing technologies (also known as Next Generation Sequencing, NGS) has allowed detection of low-level mutant alleles by performing huge numbers of sequencing reactions simultaneously [75]. In addition to improved sequencing technology, screening

methods have been developed, such as high resolution melting (HRM) [76] and denaturing high purity liquid chromatography (dHPLC) [77], which allow mutant sequences to be detected at a lower limit of detection than fluorescence-based Sanger sequencing. The following PCR methods are also alternatives that are based around detection of specific mutations or enrichment of mutations within the DNA pool. The amplification refractory mutation system (ARMS) [78] depends on the use of primers that are tailored for specific mutant sequences, allowing highly specific and sensitive detection of single sequence variants; wild-type blocking PCR is an enrichment system that implements a non-extendable wild-type specific probe into the conventional PCR reaction to inhibit wild-type DNA amplification[79], allowing only mutant DNA to amplify; CO-amplification at Lower Denaturation temperature PCR (COLD-PCR) is a sensitive method recently introduced to enrich the minor variant alleles in a mixture of mutant and wild-type sequences [80] and has been combined with LNA incorporated wild-type blocking probes to enhance sensitivity[81].

All the methods listed above are compared below in table 2 and have distinct advantages and disadvantages. Some methods are better for screening of unknown mutations such as PCR-HRM, COLD-PCR and sequencing, whilst other methods

are better for detection of known hot spot mutations, such as ARMS PCR, wild-type blocking PCR and digital droplet PCR (ddPCR). Each method has a different cost, expertise requirement, time to process and limit of detection.

Table 2. Comparison of methods for detecting mutations. Comparisons include the limit of detection which may range if there are modifications of the technique, the technical ease and cost/time to process that are both particularly important in a clinical scenario. Preferable methods for ctDNA analysis are mentioned. For example, SafeSeq and CAPP-seq replaces NGS as it is known to be preferable for ctDNA analysis by reducing errors and enhancing the limit of detection.

Technique	Brief description	Limit of detection	Technical ease and cost	Closed or open tube	Time to process
ARMS PCR [78]	Mutation-specific primers which amplify mutant DNA and can be visualised by qPCR amplification plot. Has been combined with a number of modifications to improve sensitivity, such as wild-type blocking probe, penultimate mismatches and fluorescent probes.	0.001-2%	Real time PCR machine required only. Basic regular PCR optimisation. Basic regular qPCR analysis. As cheap as regular PCR.	Closed	<1 day
Wild-type blocking [79]	Incorporation of wild-type specific blocking probe into the PCR for mutation enrichment. The probe inhibits amplification of wild-	0.01-1% depending on detection method	Real time PCR machine required plus mutation detection technique. Optimisation of blocking probe required. Blocking probe can be expensive	Closed or open depending on mutation detection technique.	<1 day

	type DNA by binding it directly, blocking access for primers to bind and causing mutant DNA only to amplify.		initially but very cheap in the long run.		
High resolution melting analysis [76]	Melting of DNA slowly with double-stranded binding dyes can show differences in melting temperature between mutant DNA and wild-type DNA. Mutant DNA is typically accompanied by wild-type DNA and thus heteroduplexes can form which can lower the amplicon melting temperature.	5-10%	Requires real-time PCR machine with HRM software or separate instrument which adds additional cost initially but relatively cheap in the long run. Melting plots require some expertise to analyse.	Open or closed depending on integration with PCR machine or separate instrument.	<1 day
Sanger sequencing [73]	Sanger is a form of sequencing which selectively incorporates fluorescent chain-terminating dNTPs which can be visualised by capillary	20%	Requires sequencing machine and reagents as well as PCR clean up kit on top of PCR machine. More costly than PCR methods but relatively cheap in the long run.	Open	2 days

	electrophoresis to form a sequence.				
Pyrosequencing [73]	Pyrosequencing is a form of sequencing which rather than chain terminating, adds fluorescent bases 1 at a time and detects which base was incorporated at which position in the sequence as the DNA polymerase copies the template.	5%	Requires sequencing machine and reagents as well as PCR clean up kit on top of PCR machine. More costly than PCR methods but relatively cheap in the long run.	Open	2 days
DdPCR [74, 82]	DdPCR incorporates water-oil emulsion technology with probes to reduce PCR errors, enhance specificity by reducing competition and allows direct end-point quantification by screening fluorescence created by each droplet via capillary.	0.001 – 2.99% depending on target and protocol	Requires expensive ddPCR machine. Requires probe design and optimisation. More in depth and complicated analysis.	Closed	1 day

SafeSeq [64] and CAPP-seq [83]		0.025% - 1% depending on target and protocol	Requires very expensive NGS machine which is also costly to run. Requires PCR machine. Standard sample preparation for targeted deep sequencing requires large optimisation. More in depth and complicated analysis which requires bioinformatician. Requires expensive PC and storage system for data.	Open	1-2 weeks depending on the set up.
Enhanced-Ice-COLD PCR [81]		0.05-0.1%	Requires an expensive pyrosequencing machine as well as PCR machine plus associated reagents cost. Requires optimisation of costly blocking probe and Tc. Relatively cheap long term.	Open	2 days

Methods such as Pyrosequencing (5% MAF LOD), Sanger sequencing (10-20% MAF LOD), standard NGS (3-5% MAF LOD) [59] and high-resolution melting analysis (5% MAF) are not sensitive enough for ctDNA detection where the MAF is frequently below 1%. Modified NGS protocols such as SafeSeq [64] which is the most popular method for liquid biopsy sequencing analysis use molecular barcoding to reduce errors and high read depth to increase sensitivity to levels required for ctDNA (i.e., 60000-100000x). Molecular barcoding adds unique molecular identifiers (unique tags to the PCR primer) to each of the DNA template molecules. This allows correction of PCR amplification bias and more importantly allows a PCR error to be distinguished from a true mutation since the exact amplicon can be tracked in the sequencing data. I.e., if all the reads for an amplicon containing a unique tag have the mutation then the mutation is a true positive, if less than 100% of the specific amplicons contain the mutation then it is more likely a PCR error. However, this does have disadvantages such as insufficient labelling of every DNA molecule and spurious cross-hybridization which will reduce the sensitivity and is unavoidable as an end-point PCR detection system. The main flaw with modified NGS protocols is that they are too expensive for the UK NHS in terms of cost per sample for longitudinal sample tracking.

Digital-droplet PCR[74, 82] has also been a favourite for PCR based ctDNA analysis. However, whilst boasting impressive limits of detection, many researchers are moving away from ddPCR as it is known to struggle with false positives at low copy number (≤10) [84], which is essential for sensitive ctDNA analysis[68, 69]. Digital-droplet PCR is an end-point analysis which has heavy reliance on fluorescent probes[84]. End-point PCR is subject to analysis of PCR errors and fluorescent probes may bind the incorrect target sequence causing a false positive result[84].

Considering the expensive nature of the methods listed above, the logical choice for a rapid, inexpensive, ultrasensitive and closed tube test for serial liquid biopsy monitoring are one of the following methods: ARMS qPCR variants[72, 85], Enhanced Improved and Complete Enrichment COLD-PCR (E-ICE COLD-PCR)[81] and wild-type blocking PCR[79] or a mixture thereof. These methods fit the typical criteria for a clinical test and are more likely to be adopted moving forward. This conclusion has been validated by the fact that the only two FDA approved liquid biopsy tests, Therascreen PIK3CA RCQ kit by Qiagen[86] for breast cancer and COBAS EGFR kit by Roche[87] for non-small lung cell cancer are ARMS PCR, wild-type blocking PCR and real-time PCR based[88]. Although, Signatera[65] which is based on NGS is currently

being reviewed by the FDA[89]. Below in table 3 compares extension of the methods listed in table 2 above and which have received significant attention in the liquid biopsy sphere.

Table 3. Comparison of liquid biopsy specific methods in the public domain. A summary of the most highly recognised methods in the public domain for liquid biopsy testing that have received media attention or clinical approval.

Method name	Method description
Therascreen PIK3CA RCQ PCR kit (Qiagen) [86]	FDA approved real-time PCR test that identifies 11 SNV mutations in PIK3CA
Cobas EGFR Mutation Test v2 (Roche) [87]	FDA approved real-time PCR test that identifies 42 SNV and indel mutations in EGFR exons 18, 19, 20 and 21 plus T790M
IntPlex ARMS qPCR [85]	Real-time PCR system that utilises wild type blocking probes combined with low annealing temperature primers. Currently designed for KRAS, NRAS, PIK3CA, BRAF and EGFR.

Signatera personalised next generation sequencing [57, 89]	Whole exome sequencing of FFPE tumour followed by patient specific PCR amplification of 16 somatic mutations. Mutation detection carried out by deep sequencing.
GRAIL [90]	Direct sequencing of 508 genes in liquid biopsy and matching white blood cell DNA at 60,000 depth. Utilises unique molecular identifiers for increased specificity and more sensitive variant allele calling.
CancerLocator [91]	A large methylation profiling dataset of primary tumour and healthy liquid biopsy samples was generated to compare for informative features. Based on the dataset a patient's liquid biopsy can undergo genome-wide profiling to determine cancer presence and location of the tumour of origin.
CancerSEEK [92]	Deep sequencing of a 61-amplicon panel designed specifically to detect the majority of ovarian,

	liver, stomach, pancreatic, oesophageal, colorectal, lung and breast cancers combined with 8 known protein biomarkers to detect location.
Biocartis Idylla[93]	Cartridge system which integrates cell-free DNA extraction and real-time PCR through microfluidics. Currently available to detect BRAF V600E in liquid biopsy.

The methods above, especially cancerSEEK[92], GRAIL[92] and cancerLocator [91] have received significant media attention and praise. All three methods are using sequencing based technologies with slightly different modifications and targets. These methods are fantastic for cancer screening whereby broad coverage is required. However, for detection of minimal residual disease and treatment-response monitoring, they are far too expensive and time consuming due to the complexity of the protocols.

Signatera[57, 89] has adopted an increasingly favoured approach which is to create personalised tests for each patient. The personalised approach centres around primary

tumour profiling and a custom sequencing panel is based on the mutational profile. This maximises sensitivity and specificity as multiple targets that are known to signify the patient's tumour can be screened per liquid biopsy rather than relying upon random chance-based hotspot statistics whereby some patients may only have one tumour specific marker or in a large proportion, no tumour specific markers present at all. A personalised approach reduces the need for a wide panel which is known to reduce throughput and the assays limit of detection. However, a key issue remains that sequencing still presents many challenges clinically.

IntPlex ARMS qPCR[85], COBAS[87], Therascreen[86] and Idylla[93] are all PCR-based methods for liquid biopsy analysis and focus on hotspot mutations. The methods are ultrasensitive, rapid, quick, and cheap that are perfect for clinical use. A major drawback however is the over-reliance on hotspot mutations. This reduces the overall pool of patients which can benefit from liquid biopsy and reduces the sensitivity that the assay can achieve in treatment-response monitoring and detection of minimal residual disease. These methods need to be assessed to see if they are scalable for a personalised approach as it is likely that this will have the highest impact.

D: Predictive tests for radiotherapy

i. Introduction to predictive tests for radiotherapy

Currently, 20% of rectal cancer patients treated with chemoradiotherapy can achieve a pathological complete response when the surgical resection specimen is reviewed histologically[26, 29-32]. In approximately 60% there is partial response which can achieve tumour down-staging, and, in the remainder, there is no obvious response seen[30, 33, 34]. However, it is not possible to determine how well patients will respond before assigning treatment. Thus, there is a large proportion of patients whose treatment could be improved – the complete responders could avoid having surgery after chemoradiotherapy while the non-responders could be given an alternative therapy[94]. Determining which patients would respond beforehand by biomarker analysis of biopsy specimens, either pre-treatment or at the beginning of radiotherapy would allow enhanced patient stratification which may improve outcome[94]. Researchers have been screening pre-treatment tumour tissue for biomarkers which may predict response to therapy for decades[94]. Many biomarkers have been identified at either the DNA, RNA, or protein level to predict response[94]. However, these biomarkers have failed to make it to the clinic[94]. This is mainly due to most studies being retrospective in nature and insufficiently robust. These factors make it difficult

to draw firm conclusions as different treatment plans may be used or the biomarker may not give a clear-cut difference between responders and non-responders[94]. Moreover, there may be multiple conflicting reports about the same biomarker. Many studies have focussed on a single favoured biomarker; however, the trend is now moving towards multiple biomarker analysis with the ever-increasing use of new and improved high throughput technology[95]. A key issue with all biomarker analysis, is the method used to draw the conclusion. Each method has its downsides and, the wider the net is cast, the higher the potential for inaccuracies[95]. This is the case even with new advanced techniques such as next generation sequencing or proteomic analysis methods like reverse phase protein array or mass spectrometry[95]. This has led to a catch-22 scenario with biomarker analysis. A biomarker is identified by high throughput technology and is then analysed in a specific manor. The pitfalls of the second method used may then come into play also which leads to further potential inaccuracy. For example, immunohistochemistry can be very subjective with both the method of analysis and the antibody used combined with the optimisation. Moreover, RT-qPCR and mutation analysis may depend on the primer design, how well the primer is optimised and validated as well as sensitivity of the assay. This heterogeneity within the methods used, combined with

small scale studies, contributes to the large amount of conflicting data in the public domain.

The number of studies on rectal cancer pre-treatment tissue to determine response has been low compared to other cancers, primarily because rectal cancer research is usually paired with colon cancer research. Many of the studies have been small scale, insufficiently robust and given conflicting results which is likely why these biomarkers have failed to reach the clinic.

ii. Introduction to redox-homeostasis proteome as a predictive marker for radiotherapy

The thioredoxin and glutathione family of proteins in a multitude of cancers have shown positive association with tumour grade after radiotherapy. Prof. Stewart Martin at the University of Nottingham has found thioredoxin proteins to be predictive or prognostic across multiple cancers including pancreatic, breast and brain[96-99]. This led to the development of drugs targeting the overexpression of several thioredoxin proteins which once downregulated were able to produce radiosensitivity. It is possible that these proteins may also be involved in rectal cancer. A study involving 994 tumours of different types and 353 normal tissue samples, analysed expression of 285 genes associated with anti-oxidative stress mechanisms[100]. The study used gene

microarrays and discovered that thioredoxin and glutathione pathway genes show higher levels of expression in many cancers compared with normal tissue[100]. Higher levels of expression in breast and lung cancer correlate with poor response to radiotherapy[100].

Thioredoxin and glutathione proteins are a core part of the redox-homeostasis proteome[97]. Redox is shorthand for the process of reduction and oxidation. Redox is common in molecular biology, especially in mitochondria when carrying out metabolic processes[97]. A common by-product of these reactions is reactive oxygen species (ROS) [97]. Typical redox reactions carried out are found in figure 13 below.

$$O_2 + e^- \rightarrow O_2^{\bullet -} \qquad \text{(Reaction 1)}$$

$$2O_2^{\bullet -} + 2H^+ \rightarrow H_2O_2 + O_2 \qquad \text{(Reaction 2)}$$

$$H_2O_2 + e^- \rightarrow {}^{\bullet}OH + OH^- \qquad \text{(Reaction 3)}$$

Figure 13. Redox reactions. These reactions generate the following common ROS: the superoxide anion ($O_2^{\bullet -}$) is created in reaction 1, hydrogen peroxide (H_2O_2) is created in reaction 2 and the hydroxyl radical ($^{\bullet}OH$) is created in reaction 3. [97]

The hydroxyl radical ($^{\bullet}OH$), superoxide anion ($O_2^{\bullet -}$), and hydrogen peroxide (H_2O_2) are common ROS which carry free radicals (unpaired electrons) and are produced in the reactions

listed in figure 13[97]. Another common source of ROS is NADPH which is processed as part of the pentose phosphate pathway (this runs parallel to glycolysis) [97]. ROS can be either beneficial or detrimental to the cell, with the latter being of most relevance to cancer[97]. At higher concentrations, which is termed oxidative stress, ROS can cause serious damage to DNA, proteins, and polyunsaturated fatty acids in lipids which may lead to cell death[97]. Levels of ROS are regulated by both non-enzymatic antioxidants and antioxidant enzymes such as: ROS-dependent enzymes - superoxide dismutase (SOD) [97], catalase (CAT) [97], peroxiredoxin (PRX) [97, 101] and glutathione peroxidase (GPx) [97]; glutathione system enzymes – glutathione reductase (GR) [97]; thioredoxin system enzymes – thioredoxin (TRX) [97], thioredoxin reductase (TRXr) [97] and thioredoxin-interacting protein (TXNIP) [97]; and glutaredoxin system enzymes – glutaredoxin (Grx) [97]. The balance between antioxidants and ROS is termed redox homeostasis[97]. Proteins involved in redox-homeostasis such as the antioxidant enzymes list above are referred to as members of the redox-homeostasis proteome[97]. The way in which these enzymes interact with each other to balance ROS is represented in figure 14.

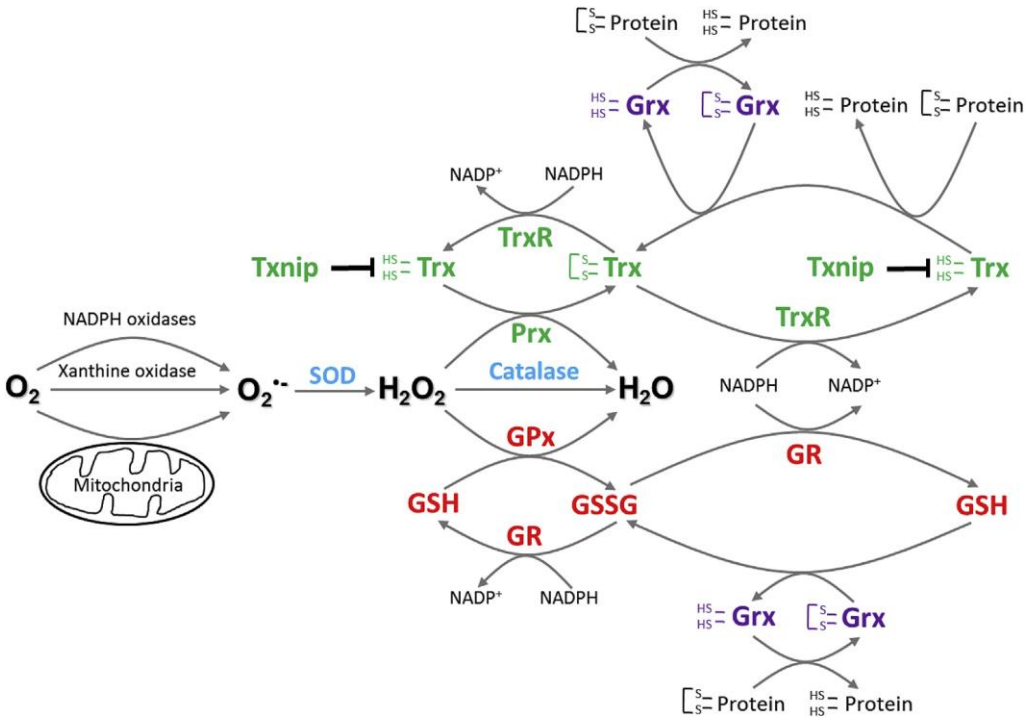


Figure 14. Pathways of the redox-homeostasis proteome to regulate ROS and their interactions with each other. The pathway begins with the generation of superoxide anion ($O_2^{\bullet-}$) in the mitochondria which is converted to hydrogen peroxide (H_2O_2) by superoxide dismutase (SOD). This may be further converted to water by catalase (CAT) and Peroxiredoxin (PRX). Members of the glutathione system: glutathione peroxidase (GPx), reduced glutathione (GSH), oxidised glutathione (GSSG), glutathione reductase (GR). Members of the thioredoxin system: thioredoxin (Trx), thioredoxin reductase (TrxR), thioredoxin-interacting protein (Txnip). Glutaredoxin (Grx). Nicotinamide adenine dinucleotide phosphate (NADPH) 'HS' represents the reduced form of the protein; 'S' denotes the oxidised form. [97]

As discussed in section 1.B, radiotherapy is administered as fractionated low-LET radiotherapy, typically intensity modulated [37-40, 44]. Although low-LET ionising radiation can cause direct DNA damage (both single and double stranded breaks), the main route of damage is by the production of ROS through the radiolysis of water and other molecules[97]. The hydroxyl radical (●OH) is responsible for most of the damage caused[97]. The main issue with treating some tumours with radiotherapy is that they are already under oxidative stress, typically through increased metabolic activity[97]. In response, cancer cells increase the expression and/or activity of the redox-homeostasis proteome to survive[97]. In turn, the tumour can minimise the damaging effects of radiotherapy by quickly dealing with increased levels of ROS[97]. It may be that in rectal cancer, patients that do not respond to radiotherapy have members of the redox-homeostasis proteome upregulated or stabilised and potentially this could predict treatment outcome.

E: Radiosensitising chemotherapy

Radiosensitising chemotherapy aims to increase tumour cell death by assisting the mechanisms of radiation-induced cell death. Oxaliplatin has been a prime candidate as a radiosensitising agent since this effect was discovered[102, 103]. However, the mechanism of action for its

radiosensitising effect is by and large unknown, other than its ability to induce double strand breaks in sub-lethally damaged cells[102, 103]. NSABP R-04 studied the addition of oxaliplatin but found no large significant improvements to treatment and considerable increases in toxicity[104].

Cancer immunotherapy has generated novel targets to increase radiosensitisation[105]. Radiotherapy induces systemic cancer cell immunity and recruit's inflammatory cells. Therefore, targeting inflammatory signalling pathways to reduce pro-survival signals could induce radiosensitivity[105].

There are several natural and synthetic agents that have been discovered or developed which have been found to induce radiosensitivity. Curcumin, Genistein and Quercetin are all natural compounds which have the ability to induce radiosensitivity by targeting important cell signalling proteins along with many other properties[106]. For example, Quercetin targets the Ataxia Telangiectasia Mutated protein which is involved in mediating the DNA-damage response[107]. If ATM is reduced, radiation-induced DNA damage is repaired at a reduced rate leading to increased cell death[40]. A reduction in DNA repair is especially effective with fractionated low-LET ionising radiation as a large proportion of cells will only attain sub-lethal damage or

potentially-lethal damage[38-40]. Synthetic agents have been developed to target many proteins that are important in preventing radiation-induced cell death, especially DDR[40].

In cancer, redox buffering systems are frequently upregulated, and this can form a radioprotective effect by reducing the amount of ROS and in turn the main form of radiation-induced cell death[97]. The glutathione and thioredoxin buffering systems are two of the main systems implicated in ROS removal[97]. Targeting either family of proteins in combination with radiotherapy may lead to increased radiosensitivity due to a rise in the level of ROS present in the cell[97]. This may be even greater in combination although side-effects are always a limiting factor.

Below in figure 15 represents members of the redox-homeostasis proteome which have been shown in breast cancer to be involved in the regulation of radiosensitivity[97]. These proteins can be downregulated to or upregulated to increase radiosensitivity.

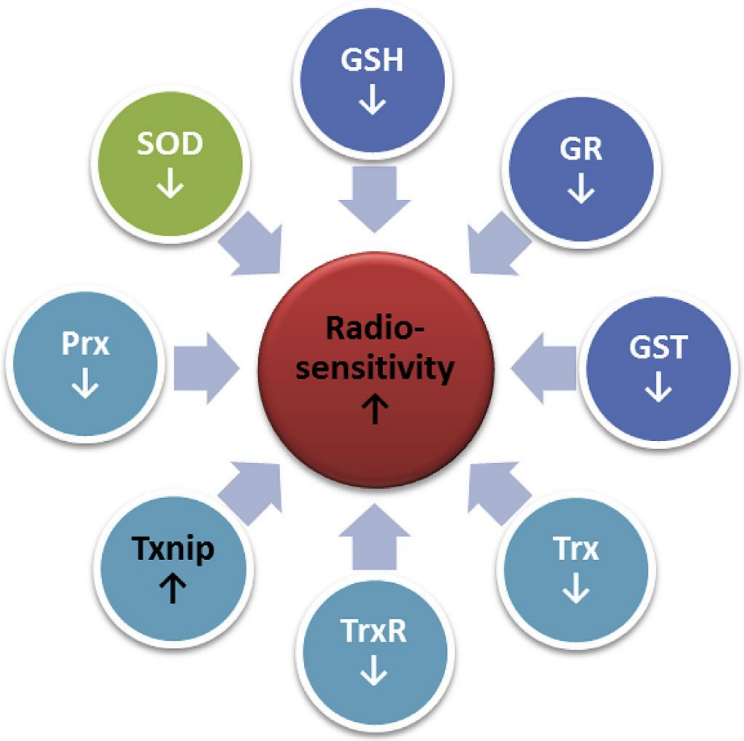


Figure 15. Redox proteins linked to radiosensitivity[97]. The diagram represents protein expression in relation to causing increased radio-sensitivity. If the arrow points downwards then decreased expression increases radiosensitivity and if the arrow points upwards, then increased expression increases radiosensitivity of the cell. Acronyms are as follows: reduced glutathione (GSH), glutathione reductase (GR), glutathione S-transferase (GST), thioredoxin (Trx), thioredoxin reductase (TrxR), thioredoxin-interacting protein (Txnip), peroxiredoxin (Prx) and superoxide dismutase (SOD). [97]

F: The aims and hypotheses of the study

The Nottingham study by Dhadda et al[36] and many other studies mentioned previously, raised several questions which

could potentially be answered by further research. The overarching question is "is there a way of improving management of rectal cancers?". This question led to a whole host of potential areas of investigation which could generate improvements to the current standard of care for rectal cancer. These areas of investigation are as follows: Firstly, it may be possible that patients who currently attain a partial or complete lack of response may be converted to responders with novel radiosensitising chemotherapy[108]. As mentioned previously, although the number of patients was low, the Nottingham study by Dhadda et al[36] found no OS and DFS benefit with chemotherapy addition, it is possible that alternative radiosensitisers, such as glutathione or thioredoxin inhibitors[97, 98], may be a good replacement for capecitabine. Secondly, before neoadjuvant CRT takes place, there is no certainty as to which patients will obtain either a pCR or tumour downstaging[94]. The development of a pre-treatment predictive test from biopsy specimens would allow patients to be stratified into the correct treatment pathway from the outset; thereby avoiding unnecessary short-term and long-term side-effects and potential time lost where another treatment regimen may generate greater success[94]. It is highly likely that a proportion of patients will either respond to radiotherapy only; respond to chemotherapy only; respond to

chemoradiotherapy or none at all and their tumours molecular biology could be the main factor responsible for this. Furthermore, where long course CRT is concerned and the 'watch and wait' approach is becoming increasingly used[30]; a predictive test may allow a stronger more evidence-based decision for the time to begin surgery for each individual patient based on their predicted response. Thirdly, as predictive tests would not always be 100% effective, a complementary real-time blood-based monitoring of response to therapy would be required[49]. As heterogeneity and the nature of the ever-evolving cancer cell cause havoc with therapy response[109]; real-time monitoring would allow faster action to be taken for a change of therapy or depending on the amount of time treated with CRT, surgery to take place sooner. Conversely, this may allow surgery to be delayed long term, allowing organ preservation, which as mentioned above, goes hand in hand with a predictive test. These three areas of need could generate significant advances in rectal cancer management if successful.

Aims of the project

1. Investigate the predictive value of glutathione, thioredoxin and related redox biomarkers for radiation response in rectal

cancer and if successful investigate the marker as a radiosensitiser in rectal cell lines and primary cell lines.

2. Evaluate the utility of cfDNA as a marker of tumour response to radiotherapy in rectal cancer.

Hypotheses

1. Thioredoxin, Glutathione or related redox biomarkers will be predictive of rectal cancer response to chemoradiotherapy based on published data.
2. Circulating tumour DNA will be highly fragmented, low in concentration and mutant allele frequency.
3. PCR based methods will be able to detect 0.1% mutant allele frequency reliably and detect mutations with low template input for cell-free DNA analysis.
4. Liquid biopsy for Rectal cancer patients undergoing chemoradiotherapy will allow response to be measured in real-time.

2. Materials and methods

A: Sample sets

i. Blood samples

26 colorectal cancer patients were recruited to this study. All 26 patients were early stage and had their tumours removed by routine surgery without neoadjuvant therapy in either 2015 or 2016 at the Queen's Medical Centre situated in Nottingham. Blood was collected and stored in EDTA tubes pre-operatively on the morning of surgery. Pre-operative samples were named day 1 and post-operative samples were collected the following day which was named day 2 onwards (weekends permitting). Plasma was separated from blood within 2 hours after collection by another PhD student (Abutaleb Asiri, AA) who had complete Hepatitis B vaccinations. For plasma separation, blood samples were centrifuged at 1900g for 10mins at room temperature and transferred into 1.5ml Eppendorf tubes. The Eppendorf tubes were then centrifuged at 16,000g for 10mins at 4°C to remove the remaining cell debris. The plasma was stored at -80°C until cfDNA extraction took place.

ii. Formalin-Fixed Paraffin-Embedded (FFPE) samples and ethics

1. Rectal cancer FFPE tissue for radiation response prediction

The diagnostic biopsies (and corresponding Haematoxylin and Eosin (H&E) stained slides) of 50 cases of rectal cancer who

had undergone Long Course Chemoradiotherapy (LCCRT) between 2018 and 2019 were collected from the Queen's Medical Centre archive. Ethics were obtained to allow molecular testing on any colorectal cancer sample (REC reference number: 05/Q1605/66). Only biopsies with a reported Mandard score in the final resection specimen were chosen. In this study we classed Mandard 1 and 2 as responders and Mandard 3-5 as non-responders[110]. Whilst there is no standard for grouping of Mandard score, the previous Nottingham study has grouped Mandard grading in this way. Moreover, Mandard 3 represents a cut-off point that can be considered no response since an increase in tumour cells is seen whilst this does not occur in Mandard 1 and 2 groups. Mandard Twenty-six cases were Mandard 1/2 and 24 were Mandard 3-5. Further clinicopathological information was not collected as it was not relevant to the study. H&E stain is routinely performed for diagnosis in histopathology labs and allowed tumour tissue to be distinguished from normal tissue. A trained pathologist (Dr. Wakkas Fadhil) analysed the slides to determine the presence of tumour cells within pre-surgery biopsies. Only blocks with at least 10% tumour cellularity were used for the study to prevent bias.

2. Colorectal FFPE tissue from patients undergoing surgery to determine mutational profile for circulating tumour DNA testing

The corresponding tumour blocks or biopsies to match the blood samples mentioned in 2.A.i were collected from the Queen's Medical Centre FFPE archive in 2016 for all the patients included in the final analysis (n=21). Haematoxylin and Eosin stained slides were also collected and analysed by a fully trained pathologist (Dr. Wakkas Fadhil) to determine the presence and percentage of tumour in each FFPE block. The block with the highest tumour percentage was chosen for downstream mutational analysis to increase the chances of detection. FFPE blocks were not processed until the plasma had undergone final analysis to stop bias in result interpretation.

B: Tissue culture

i. Cell maintenance

Cells were grown for the purposes of DNA or RNA extraction only. A list of the cell lines used as standards for specific mutations of interest are found in supplementary table 6. Cell line identity was confirmed by STR genotyping undertaken by Dr T Raposo. The cell lines HEK293T, MDA-MB-231, MCF-7, SW620 and HCT116 were also grown for RNA extraction to optimise qPCR primers. All cells were maintained at 37°C in a

5% CO2 atmosphere and cultivated in Dulbecco's Modified Eagle's Medium (DMEM) (Thermo Fisher Scientific) supplemented with 10% foetal bovine serum (FBS) (Sigma) under sterile conditions. Cells were checked each day for infection and confluency under a microscope. Media was replaced once or twice a week depending on the cell line growth rate to maintain optimum growth conditions.

ii. Passaging cells

Cells were passaged when 70-80% confluency was reached and thrown away after 30 passages was reached. Media was removed and cells washed twice with phosphate buffered saline (PBS). Cells were incubated for up to 5 minutes with 1 ml of trypsin EDTA (Sigma) until they were completely detached from the flask surface. The trypsin was neutralised by the addition of 9 ml of DMEM (containing FBS). Cells in suspension were pelleted by centrifugation at 1,500 rpm for 5 minutes. The media was then discarded, and the cells resuspended with 10ml media. One ml cell suspension was transferred to a new flask containing 9 ml of DMEM (containing FBS).

C: Nucleic acid extraction methods

i. Cell line DNA extraction

DNA from cell lines (see supplementary Table 6 for list of cell lines used) was extracted using the GenElute mammalian genomic DNA miniprep kit (Sigma-Aldrich, U.S.A) using the manufacturers protocol. In brief, cells were pelleted when 70-80% confluency was reached. Cells were stored at -80°C until extraction took place. Cells were resuspended and no RNase treatment was used. Cells were lysed and protein digested by proteinase K for 10 minutes at 70°C. Proteinase K (provided in the kit) and lysis buffer was doubled if the pellet was large to ensure low protein content in the elution. Column preparation solution was added to the column for maximum DNA binding. 100% ethanol was mixed with the lysate and loaded to the spin column. Two washes were carried out and then a 3-minute centrifugation step was carried out at 16,000g to dry the column. The columns were incubated for 10 minutes with 50µl elution solution at room temperature and then centrifuged at 16,000g for 1 minute. Two elution's were performed to maximise DNA yield. 30µl of elution solution was used in the second instance. The two elution's were kept separate and not mixed. DNA elution's were stored at -20°C and diluted with nuclease-free water to a working concentration when appropriate.

ii. Cell line RNA extraction

RNA from cell lines was extracted using the GenElute mammalian total RNA miniprep kit (Sigma-Aldrich, U.S.A) using the manufacturers protocol. In brief, cells were pelleted at 70-80% confluency and the media were removed. Cells were stored at -80°C until extraction took place. Cells were lysed with 500µl of a lysis solution/2-ME mixture and vortexed until clumps were removed. 500µl of 70% ethanol was added to the lysate and not filtered. 700µl of the lysate/ethanol mix was loaded onto the column at a time and centrifuged at 16,000g for 15 seconds. The entire lysate/ethanol mix was passed through the column, requiring multiple repeats of the step. 250µl of wash solution 1 wash was applied to the column and centrifuged at 16,000g for 15 seconds. On-column DNase (provided in the kit) digestion was carried out to ensure removal of DNA contamination for the downstream RT-qPCR analysis. The DNase 1/digest buffer mix was incubated on the column at room temperature for 15 minutes. 250µl of wash solution 1 was applied to the column and centrifuged at 16,000g for 15 seconds. Two further 500µl column washes were carried out and a 2-minute centrifugation step at 16,000g was performed to dry the column. Only one elution was performed with 100µl of elution solution and a 5-minute incubation step at room temperature. The column was

centrifuged at 16,000g for 1 minute to release the RNA from the column. RNA elution's were stored at -80°C and diluted with nuclease-free water to a working concentration when appropriate.

iii. FFPE tissue DNA extraction

Twenty µm thick FFPE tissue sections cut from resection blocks were deparaffinised with 2 rounds of xylene and rehydrated with 2 rounds of graded alcohol (100% and 96%). Ten µm thick sections cut from biopsies were processed using Qiagen's deparaffinisation solution to prevent substantial loss of tissue.

FFPE tissue DNA was extracted via the use of Qiagen's QIAamp DNA FFPE tissue kit using the manufacturer's protocol. In brief, digestion with proteinase K (provided in the Qiagen kit with recommended concentration) took place until all tissue was completely lysed to ensure the adequate presence of potential mutant target DNA. Depending on the number of sections per case processed at one time; digestion may take up to 2 days to complete. Up to 4 resection or biopsy specimens were used. The amount of initial proteinase K and buffer ATL was scaled up with 20µl additional proteinase K (provided in the kit) and 180µl buffer ATL per section added. Samples were incubated at 56°C and agitated at 300rpm using an Eppendorf thermomixer. An additional 20µl proteinase K

and 180µl buffer ATL was added twice a day to ensure digestion activity. Once complete lysis was achieved, the sample was incubated at 90°C for 1 hour. 100% ethanol and buffer AL were added to the lysate and scaled to roughly equal the volume of buffer ATL and proteinase K added. The entire lysate was passed through the column by centrifugation, 500µl at a time; this would require multiple centrifugation steps. The column was washed twice and centrifuged for 3 minutes to ensure the membrane was dry. Two elution's were carried out. The first elution used 50µl elution solution and the second used 20µl. Elution solution was incubated on the column for 10 minutes at room temperature before centrifugation at 16,000g for 1 minute. DNA was stored at -20°C and diluted with nuclease-free water to a working concentration when appropriate.

iv. FFPE tissue RNA extraction

Ten µm thick sections cut from biopsies were processed using Qiagen's deparaffinization solution to prevent substantial loss of tissue.

FFPE tissue RNA was extracted via the use of Qiagen's RNeasy FFPE kit using the manufacturer's protocol. In brief, up to 3 tissue sections were digested with 20µl proteinase K (provided in the kit) and 240µl buffer PKD for 15 minutes at 56°C.

Reversal of cross-linking was carried out by incubating the sample for 15 minutes at 80°C. The uncoloured lower phase was then removed into a new tube and incubated on ice for 3 minutes. Following this a 15-minute centrifugation step took place at 16,000g. The supernatant was transferred to a new tube and 25µl DNase booster buffer plus 10µl DNase I (provided in the kit) was added to digest DNA for 15 minutes at room temperature. Following this, 500µl of buffer RBC and 1200µl ethanol was mixed with the lysate to adjust binding conditions. The entire lysate, 700µl at a time, was drawn through a spin-column. The column was washed twice and following this a 5-minute centrifugation step took place at 16,000g with the lids open to dry the membrane. Only one elution was carried out using 14µl of RNase-free water with a 5-minute incubation time at room temperature. This maximised the RNA concentration considering the expected low concentration of RNA with a biopsy specimen.

v. Cell-free DNA extraction from plasma using Qiagen vacuum manifold

The Qiagen vacuum manifold was set-up as instructed by the manufacturer's guidelines. The Qiagen circulating nucleic acid kit was performed in accordance with the manufacturer's guidelines. In brief, 1ml of plasma was used for each patient as the supply was limited. Ideally, 10ml of plasma should be

used, especially for early-stage patients to ensure the presence of ctDNA in the eluate. With new magnetic kits it may be possible to extract from volumes greater than 10ml. In essence, the more plasma that is processed the greater the chances of detecting ctDNA. The protocol was followed for the processing of 1ml. Plasma was mixed with proteinase K (provided in the kit) and buffer ACL (containing carrier RNA) and incubated at 60°C for 30 minutes with a large floor standing incubator shaker (Sanyo, JP) set at 300rpm. Following this, the lysate was mixed with buffer ACB and placed on ice for 5 minutes. The spin column was then attached to the vacuum manifold with a tube extender to increase volume capacity. The lysate was then drawn through the spin column completely by vacuum. Two rounds of wash and one round of 100% ethanol was then used to remove contaminants. The spin column was removed from the manifold and incubated at 56°C for 10 minutes to dry the membrane. One elution was performed by incubation with 40µl buffer AVE (20µl for post-surgery samples) for 10 minutes at room temperature. In the future, 20µl would be preferable for all samples due to concentration. The column was centrifuged at 16,000g for 1 minute. Eluates were stored at -20°C and not diluted due to low concentration.

D: Nucleic acid quantification, fragmentation analysis and generation of positive mutant spike-in controls

i. Cell line and FFPE tissue DNA/RNA quantification

Cell line and FFPE tissue DNA and RNA was quantified via the use of the Nanodrop 2000c spectrophotometer (Thermofisher, UK). 2µl of sample was used per reading to ensure the droplet was central. Moreover, samples were run in duplicate to avoid errors. If the quantification did not match between duplicates, a third reading would be taken. This was especially important for RNA where the quantification was crucial for downstream RT-qPCR analysis.

Cell line DNA and RNA was considered acceptable with 230/260 and 260/280 ratios of between 1.7-1.9. A minimum of 50ng/µl was considered acceptable for cell line nucleic acid. FFPE tissue DNA and RNA were considered acceptable with 230/260 and 260/280 ratios of between 1.5-2.3. A minimum of 20ng/µl in the final eluate was considered acceptable for FFPE tissue nucleic acid. A less strict criterion was applied to FFPE tissue DNA and RNA as the quality is known to be reduced compared with cell lines due to age, fixation and additional contaminants present within tissue [111].

ii. CfDNA quantification by DeNovix fluorometry

cfDNA was quantified using the DeNovix fluorometer combined with the DeNovix dsDNA high sensitivity assay kit (5pg/µl to 250ng/µl) which uses a double-stranded binding dye for detection. In brief, 0pg/µl, 5pg/µl and 25ng/µl standards were prepared alongside samples. A working solution was prepared of assay buffer/dye mix. 190µl of the mix was added to 10µl of the 3 standards initially and vortexed. Following this, the standards were incubated away from light for 5 minutes and then vortexed again. The standards were measured on the fluorometer and then the steps above were repeated for the samples. 1µl of sample was added to 199µl of mix to preserve the sample. Samples were processed 3 at a time to ensure the dye did not fade. The samples were quantified and a minimum reading of 100pg/µl was considered acceptable for analysis.

iii. Fragmentation analysis by qPCR

CfDNA contains fragmented DNA[66] and, to define the maximum size of amplifiable fragment in the cfDNA, 2 different primer sets were designed which generated amplicon sizes of 105bp or 194bp (supplementary table 1). Cycling conditions (Stratagene Mx3005p, Agilent, U.S.A) are as follows: (95°C 5min) x1; (95°C 30sec, 55°C 15sec, 72°C 15 sec) x40 plus standard melt-curve. Reaction conditions: 2x Hotshot Diamond mastermix (ClentLifeScience, UK); 20x

EvaGreen Dye in water (Biotium, inc., U.S.A); 250nM of each primer (Eurofins, UK); 5ng cfDNA template total quantified by DeNovix), addition of Nuclease-Free water (Qiagen, DE) to give a total reaction volume of 20µl.

iv. Generation of spike-in positive controls

DNA was diluted to a final concentration with nuclease free water (Qiagen, Germany). DNA from cell lines with known mutations (supplementary table 6) was spiked into HEK293T cell line DNA (globally wild-type). Samples were prepared containing mutant alleles frequencies (MAF) ranging from 50% down to 0.06%. In general, 40ng/µl was used to prepare down to 0.06% MAF (4 mutant copies). 4 mutant copies were established as a reliable minimum mutant copy threshold for general limit of detection testing due to Poisson statistics[72]. Statistically, some reactions may have 1 copy, and this was our reasoning behind using 4[72].

When testing 1 mutant copy detection for the purposes of sensitivity demonstration, 60ng/µl of both mutant (supplementary table 6) and HEK293T wild-type DNA was prepared. Mutant DNA was serially diluted to give 1 expected copy or 0.01% based on 6pg per genome.

E: General PCR methodologies

i. DNA specific primer design

1. Primers for mutation detection by high-resolution melting analysis, quantification of cfDNA and fragmentation analysis

Primers that generate shorter amplicons (<150bp) were designed to specifically amplify the fragmented DNA template usually found in cfDNA[66] and FFPE tissue[111]. Primers (supp. Table 2) were designed according to standard rules except for amplicon length[112]. Gel images confirming amplicon sizes are found in supplementary figure 1 and 2. Primers designed for fragmentation analysis are found in supplementary table 1. To speed up the design process, the following free online software was used. Cosmic (catalogue of somatic mutations in cancer - https://cancer.sanger.ac.uk/cosmic) was used to find the most common mutation hotspots for each chosen gene in the large intestine. Genes were chosen based on previous data[113]. Ensembl (https://www.ensembl.org/index.html) was used to browse the genome for hotspot specific DNA sequence. Primer3 version 4.0.0 was used for the initial primer design [114]. The cap on amplicon length (<150bp) was inserted into the software's general primer picking conditions. The concentration of reagents used in the PCR reaction was also

inserted for primer3 design. Primers were designed to have an annealing temperature of 60°C. All other conditions remained consistent with recommended primer3 conditions. As all primers also had the dual function of being used for high-resolution melting analysis it was essential that the mutation of interest was placed centrally between the forward and reverse primer. MFEprimer (https://mfeprimer3.igenetech.com/spec) which is a thermodynamics-based program for checking PCR primer specificity was used to check the following: primer melting temperature (Tm), amplicon formation, chance of amplicon formation (PPC%). PPC% should be more than 80% and in the correct chromosome, no other amplicon formation with ΔG (the change in Gibbs-free energy) >10, primer dimer ΔG <-6 and Tm <50°C, hairpin ΔG <-6 and Tm <50°C. Experimental settings were altered to replicate the PCR reaction conditions. The most specific primer pair was always chosen and if no results were presented then primer 3 conditions were changed to become less strict. The melting temperature between the forward and reverse primer was increased to 3°C and intronic sequence would be used in addition to exonic. Exonic sequence was always preferable as it was more likely to be unchanged between patients.

2. HOT_ARMS primer design and modification

Highly optimised Ta Amplification refractory mutation system PCR (HOT_ARMS PCR) is a modification of the previously established ARMS PCR protocol. This protocol can selectively amplify mutant DNA by designing primers to have the mutant base on the 3’ site of one of the primer pairs, thereby preventing wild-type DNA amplification. An in depth description of HOT_ARMS PCR can be found in section 5.A. Primers were designed to produce short amplicons (60 – 100bp) due to cfDNA fragmentation [66] and one of the primers contained a mutation specific base at the 3’ end of the sequence[78]. The other primer contained wild-type sequence only. Primer sequences are found in supplementary table 5. Two types of mutation were tested i.e., Single Nucleotide Variants (SNV) and deletion mutations. In both cases, a novel sequence is created. Examples of how the mutation specific primers are designed for different mutation scenarios as shown below.

Hypothetical wild-type sequence of 30 nucleotides is mutated in different ways to model the different possibilities which would occur in practice. Mutations are highlighted and both the original and the new mutated sequence are shown. The ARMS specific primer which would work only on the mutant sequence are shown.

Hypothetical sequence:

GATCGGATTGGCAAATTAGCCGTAGGCCGG

a) Example SNV: Nucleotide highlighted in bold changes from T to C

Original sequence: GATCGGAT**T**GGCAAATTAGCCGTAGGCCGG

New sequence: GATCGGAT**C**GGCAAATTAGCCGTAGGCCGG

Primer sequence: GATCGGATC

b) Example 1 base deletion: Nucleotide highlighted in bold deleted

Original sequence: GATCGGAT**T**GGCAAATTAGCCGTAGGCCGG

New sequence: GATCGGATGGCAAATTAGCCGTAGGCCGG

Primer sequence: GATCGGATG

c) Example 9 base deletion: Nucleotides highlighted in bold deleted

Original sequence: GATCGGAT**TGGCAAATT**AGCCGTAGGCCGG

New sequence: GATCGGATAGCCGTAGGCCGG

Primer: GATCGGATA

To speed up the design process, the following free online software was used. Cosmic (catalogue of somatic mutations in cancer: https://cancer.sanger.ac.uk/cosmic) was used to find

the most common mutation hotspots for each chosen gene in the large intestine. Genes were chosen based on previous data[113]. Ensembl (https://www.ensembl.org/index.html) was used to browse the genome for hotspot specific DNA sequence.

HOT_ARMS PCR works on the principle that a high Ta will improve the specificity of the PCR. As a rule of thumb, the Ta is usually 5°C lower that the melting temperature (Tm). The definition of Tm is the temperature at which 50% of the DNA is melted (i.e., the strands are separated) and this is the physical property DNA used by primers design software packages. Primers were initially designed in Primer 3 [114] mostly according to the standard rules [115] and then modifications were made to raise the Tm/Ta.

The main design rules for HOT_ARMS were as follows: (i) minimum primer length 20nt, optimum primer length 25nt and maximum primer length 30nt; (ii) minimum GC content 30%, optimum GC content 45%, maximum GC content 60%; (iii) minimum amplicon length 60nt, maximum amplicon length 110nt; (iv) primer dimer ΔG <-6 and Tm <60°C; (vi) hairpin ΔG <-6 and Tm <60°C; (vii) 3’ base is specific for the mutation in either the forward or reverse primer; (viii) max Tm difference between forward and reverse primer 3°C; (ix)

minimum annealing temperature 63°C and maximum annealing temperature 68°C.

Both the primers need to be modified to increase the Tm. There are two main approaches available i.e., addition of a sequence tag to the primers or incorporation of modified bases into the primers.

(i) Addition of a tag: Extending the length of a primer by adding extra sequences in the form of a 5’ tag will raise the Tm of the primer. The greatest increase will be achieved if the tag contains a high proportion of Guanine and Cytosine bases. Previous work has incorporated tags to primers to develop multiplexed HRM protocols [116, 117]. One tag was tested: a 10-base tag with the sequence 5’-gggccggccc-3’ (predicted to raise the Tm by approximately 20^{o}C). The tags were added to both the forward and reverse primers.

(ii) Incorporation of modified nucleic acids: Synthetic nucleic acids (such as Locked Nucleic Acids (LNA) and Bridged Nucleic Acids (BNA) have a much greater binding affinity for the paired base on the opposite DNA strand than naturally occurring nucleic acids [118-120]. Incorporation of synthetic nucleic acids into primers/probes has been used to increase the specificity [118, 121, 122] although the high binding affinity stabilises double stranded DNA thereby increasing the

Tm slightly. Only LNA's were tested here (although BNA would be predicted to have the same effect). Primers were designed in accordance with the rules above and 1 LNA was included at the 3' end of the ARMS primer and up to 2 LNA's in the wild-type primer at the 5' end in accordance with published recommendations [118, 121-123]. More than 1 LNA is not recommended in the ARMS primer due to efficiency reasons. 2 LNA's can be incorporated into the 5' end of the wild-type primer to increase the annealing temperature in line with the ARMS primer.

iii) Incorporation of wild-type blocker[85, 124]: Wild-type blockers are designed as described in section 2.E.3 to overlap the ARMS primer by 5 nucleotides. 6 or more LNA's are incorporated of which LNA's must be included on mutant bases and the others should surround the mutant base.

MFEprimer (https://mfeprimer3.igenetech.com/spec) which is a thermodynamics-based program for checking PCR primer specificity was used to check the following: primer melting temperature (Tm), amplicon formation, chance of amplicon formation (PPC%) -which should be >80% and in the correct chromosome, primer dimer ΔG <-6 and Tm <50°C, hairpin ΔG <-6 and Tm <50°C. Experimental settings were altered to replicate the PCR reaction conditions. The most specific primer

pair was always chosen and if no results were presented then primer 3 conditions were changed to become less strict. The melting temperature between the forward and reverse primer was increased to 3°C and intronic sequence would be used rather than only exonic. Exonic sequence was always preferable as it was more likely to be unchanged between patients.

3. HOT_PI PCR Primer design and modification

A full description of the Highly Optimised Ta Probe-inhibited PCR (HOT_PI PCR) process can be found in section 6.A. In brief, HOT_PI PCR works on the principle of wild-type blocking PCR. Wild-type blocking PCR places the mutation of interest between the forward and reverse primer (i.e., in the centre of the amplicon). In addition, a wild-type blocking probe is added. The probe is in essence a wild-type sequence primer that overlaps the mutation of interest, preferably in the centre, and is placed in between the forward and reverse primer. Moreover, the blocking probe must overlap the forward or reverse primer by 5 nucleotides depending on the strand it is targeting (i.e., forward strand blocking probe must overlap forward strand primer) and bind at a temperature at least 5°C higher than the primer. In addition, the blocking probe must have a phosphate group on the 3’ end to prevent extension by the polymerase.

In HOT_PI PCR, primers are designed to produce short amplicons (60 – 100bp) due to cfDNA fragmentation [66]. For wild-type blocking PCR this presents a challenge since care must be taken to ensure that the blocking probe is binding to a specific region. The primer and probe sequence context in essence are locked with limited flexibility compared to when targeting larger amplicons where small adjustments can be made to ensure appropriate binding. If the probe does not bind a specific region then the wild-type blocker will bind other sequence resulting in a reduction in blocking efficiency. However, this is somewhat reduced in HOT_PCR since it increases the specificity of the probe. The probe specificity must still be checked by MFEprimer (https://mfeprimer3.igenetech.com/spec) to ensure it is binding to the correct region in order to maximise blocking efficiency. Moreover, it is important using MFEprimer to check for bases that are prone to dimerisation or hairpins - LNA's should not be placed on these bases.

The probe was designed to cover the whole region of interest generated between the two primers and to also extend 5bp into each primer binding site. 6 LNA bases were added to the probe to improve binding affinity further and clamp the region. The probe contained LNA bases on the hotspot regions of codon 12 and 13. LNAs were added solely to improve clamping

of the region rather than to increase mismatch temperature. A phosphate modification was added to the 3’ end of the primer to prevent polymerase extension.

HOT_PI PCR works on the principle that a high Ta will improve the specificity of the PCR. As a rule of thumb, the Ta is usually 5°C lower that the melting temperature (Tm). The definition of Tm is the temperature at which 50% of the DNA is melted (i.e., the strands are separated) and this is the physical property DNA used by primers design software packages. Primers were initially designed in Primer 3 [114] mostly according to the standard rules [115] and then modifications were made to raise the Tm/Ta.

The main design rules for HOT_ARMS were as follows: (i) minimum primer length 20nt, optimum primer length 25nt and maximum primer length 30nt; (ii) minimum GC content 30%, optimum GC content 45%, maximum GC content 60%; (iii) minimum amplicon length 60nt, maximum amplicon length 110nt; (iv) primer dimer ΔG <-6 and Tm <60°C; (vi) hairpin ΔG <-6 and Tm <60°C; (vii) 3’ base is specific for the mutation in either the forward or reverse primer; (viii) max Tm difference between forward and reverse primer 3°C; (ix) minimum annealing temperature 63°C and maximum annealing temperature 68°C.

Both the primers need to be modified to increase the Tm. This has been achieved as with HOT_ARMS PCR by addition of a 5' tag: Extending the length of a primer by adding extra sequences in the form of a 5' tag will raise the Tm of the primer. The greatest increase will be achieved if the tag contains a high proportion of Guanine and Cytosine bases. Previous work has incorporated tags to primers in order to develop multiplexed HRM protocols [116, 117]. One tag was tested: a 10-base tag with the sequence 5'-gggccggccc-3' (predicted to raise the Tm by approximately 20^{o}C). The tags were added to both the forward and reverse primers.

MFEprimer (https://mfeprimer3.igenetech.com/spec) which is a thermodynamics-based program for checking PCR primer specificity was used to check the following: primer melting temperature (Tm), amplicon formation, chance of amplicon formation (PPC%) -which should be >80% and in the correct chromosome, primer dimer ΔG <-6 and Tm <50°C, hairpin ΔG <-6 and Tm <50°C. Experimental settings were altered to replicate the PCR reaction conditions. The most specific primer pair and blocking probe was always chosen and if no results were presented then primer 3 conditions were changed to become less strict.

ii. RNA specific primer design

qPCR primers were designed almost in the same way as primers in section 2.E.i.1. However, a few minor changes were made. A list of transcripts associated with the gene of interest was found using Ensembl (https://www.ensembl.org/index.html). The transcripts were compared using NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) which is an online program to search for sequence similarity. Only regions which were found in all transcripts were targeted to ensure that quantification was not transcript dependent and was associated with the whole gene. An example is shown below for TP53 in figure 16:

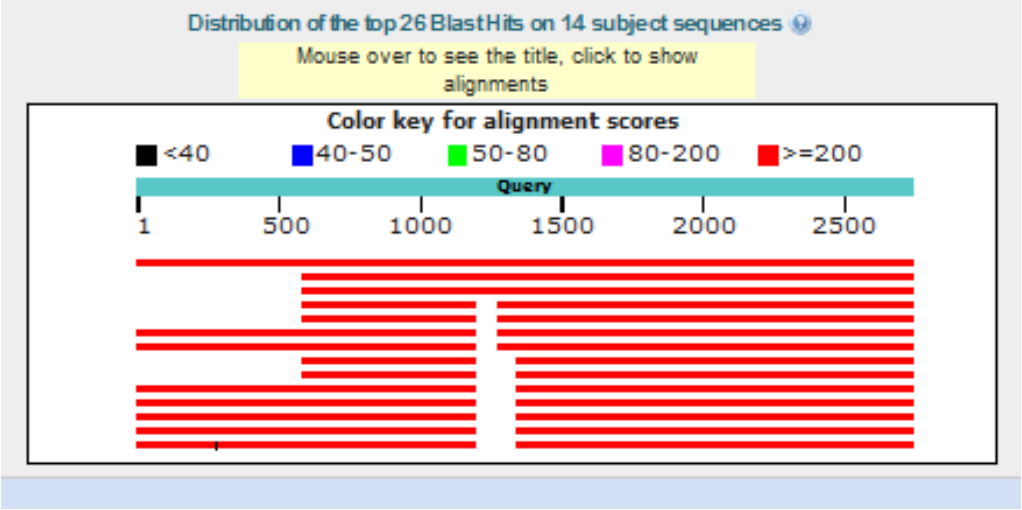


Figure 16. TP53 example of transcript alignment for maximum gene expression analysis specificity. All available transcripts were compared by BLAST against one other transcript to find

regions which were conserved. If the transcript is the same the red line will be continuous, if the transcript is missing some sequence it will show which area has been deleted by leaving a gap. The area which is deleted can be recognised by using the scale above the red lines which denotes that sequence position in base pairs.

As can be seen in figure 16 above, some transcripts do not align fully. The first transcript on Ensembl was used as the comparator. The first transcript would be the sequence used for designing. With TP53, the primer was designed to bind from base 1500 onwards to ensure that there was perfect sequence similarity between all transcripts. This process was carried out for all gene's and the region targeted is shown in supplementary table 8 along with the primer sequences. Exon-exon junction spanning primers were preferable but if MFEprimer (https://mfeprimer3.igenetech.com/spec) which is a program to determine primer specificity did not show significant specificity then primers were placed in separate neighbouring exons which had an intron between them greater than 500 bases.

iii. DNA specific primer optimisation

A 2-step system (details below) was used for primer optimisation. Initially gradient PCR followed by high resolution

melting analysis (HRMA) was performed to determine the optimal primer annealing temperature (Ta). The optimal Ta was classed as the temperature which gave the highest fluorescent melting peak which was also in the correct location by HRMA. The online software uMelt (https://dna-utah.org/umelt/umelt.html) was used to determine the approximate location of the correct melting temperature location. Gel electrophoresis (supp. Fig 1 and 2) was used for the second step of optimisation for absolute proof of amplicon size and absence of primer dimer. Only PCR products generated by the optimal Ta were analysed by gel electrophoresis.

Step 1: Gradient PCR cycling conditions (Peqlab Primus96 Advanced Gradient Thermal Cycler) are as follows: (95°C 5min) x1; (95°C 10sec, 60+/-12°C 30sec, 72°C 20 sec) x40. Reaction conditions: 5µl Hotshot Diamond mastermix (Clent Life Science); 1µl 20 x EvaGreen Dye in water (Biotium, inc.); 250nM of each primer (Eurofins); 40ng cell line DNA template; addition of Nuclease-Free water (Qiagen) to give a total reaction volume of 10µl. HRM and analysis was performed on the Idaho Technology Inc. LightScanner-96 Hi-Res melting system as described in section 2.F.i.

Step 2: Electrophoresis was performed by mini gel (50ml of 3% agarose with SYBR safe (Thermofisher)) at 100v until the dye front reached the bottom of the whole gel for maximum separation of small fragments. PCR product or DNA ladder (Promega 100bp ladder or Thermofisher 25bp ladder) was loaded into the gel with 6x DNA gel loading dye (Thermofisher). The gel was visualised using a Bio-Rad gel documentation system.

iv. RT-qPCR

1. Reverse transcription

Reverse transcription was carried out with reagents from multiple companies. Firstly, 2.5µl of 100µM random hexamer (ThermoFisher, U.S.A) was mixed with 1µg of RNA and made up to a final reaction volume of 15µl with nuclease-free water (Qiagen, DE). This mixture was then placed on ice for 5 minutes. At the same time a bulk mix containing 5µl 10x reverse transcription buffer (Promega, U.S.A), 1.25µl 10mM dNTP mix (ThermoFisher, U.S.A), 1µl (200 units) M-MLV reverse transcriptase (Promega, U.S.A), 0.625µl RNasin plus ribonuclease inhibitor and made up to a final volume of 10µl per reaction. The mastermix was distributed to the reaction wells and incubated for 1 hour at 37°C and afterwards a further 10 minutes at 95°C.

2. Primer optimisation

RNA-specific primers were optimised using a 3-phase protocol. The first phase was to check for expression in a range of cell lines. Melting patterns were predicted using the online software uMelt (https://dna-utah.org/umelt/umelt.html) and validated the melting patterns produced experimentally. If the melting patterns matched, then the primer would move to the second phase of optimisation. Phase 1 cycling conditions (Stratagene Mx3005p, Agilent, U.S.A) are as follows: (95°C 5min) x1; (95°C 30sec, 55°C 15sec, 72°C 15 sec) x40 plus standard melt-curve. Reaction conditions: 5µl Hotshot Diamond mastermix (ClentLifeScience, UK); 0.5µl 20x EvaGreen Dye in water (Biotium, inc., U.S.A); 250nM of each primer (Eurofins, UK); 1µl cDNA template @ 10ng/µl; addition of Nuclease-Free water (Qiagen, DE) to give a total reaction volume of 10µl.

The second phase used gel electrophoresis to confirm the specificity of the product. Electrophoresis was performed by mini gel (50ml of 3% agarose with SYBR safe (Thermofisher)) at 100v until the dye front reached the bottom of the whole gel for maximum separation of small fragments. PCR product or DNA ladder (Promega 100bp ladder or Thermofisher 25bp ladder) was loaded into the gel with 6x DNA gel loading dye (Thermofisher). The gel was visualised using a Bio-Rad gel

documentation system. Acceptable primers would produce one band in the reaction containing cDNA and no band for the NTC reaction. Primers would be designed again if they failed this checkpoint. This only occurred for 2 targets.

The third phase tested the efficiency of the primer for complete validation. The highest expressing cell-line from phase one was used for each target individually. The respective cell-line was serially diluted 2-fold with nuclease-free water from 40ng to 312.5pg (total 8 standards). Standards were amplified using the same cycling and reaction conditions as phase one in triplicate. PCR efficiency was automatically calculated by the Stratagene MX3005P software. The acceptable efficiency range was 85%-115%. The primer concentration was altered if the efficiency fell out of this range. If the efficiency was below 85% then the primer concentration was increased and vice versa with more than 115%. Melt-curves for all standards were checked for primer specificity. Primers were classed as unacceptable if they failed to produce specific melt-curves for 312.5pg.

3. Method of analysis

Data from RT-qPCR of the Rectal FFPE biopsy samples were analysed using the Livak method[125]. In brief samples were analysed using the $2^{-\Delta Ct}$ method via Microsoft Excel. ΔCt

values were calculated by taking the average Ct value for 3 technical replicates of the target and subtracting the geometric mean of the average Ct value from 3 technical replicates of 3 housekeeping genes: B2M, RPS23 and PPIA. Once the ΔCt was calculated, $2^{-\Delta Ct}$ was applied to calculate fold gene expression. The first 20 samples contained 10 responders to radiotherapy and 10 non-responders; the $2^{-\Delta Ct}$ values were separated into these 2 categories and transferred to GraphPad Prism (https://www.graphpad.com/scientific-software/prism/) which is a statistical software package to perform a non-parametric Mann-Whitney test as data could not be transformed to a bell-shaped distribution. A significant difference was determined by a P-value of ≤0.05 with 95% confidence.

F: Mutation detection methodologies

i. PCR combined with high resolution melting analysis

HRM and analysis was performed on the Idaho Technology Inc. LightScanner-96 Hi-Res melting system. The PCR products were first transferred into a LightScanner 96-well hard-shell plate followed by the addition of a 20µl mineral oil overlay. Before HRM, plates were spun down in a Megafuge centrifuge (2000 rpm, 5min). HRM was performed between 65 and 95°C with sample equilibration at 62°C. Exposure was set to 'Auto' and data was captured at a ramp rate of 0.10°C/sec. The acquired melting data was analysed with the LightScanner

Call-IT software version 2.0.0.1.331 using expert scanning and high sensitivity.

ii. Sanger sequencing and squirrel PCR

For KRAS exon 2, inner primers were tagged with a squirrel tail of which Useq primers bound the 5’ end of, both primers are found in supp. table 1. The squirrel primer assay[117] for HOT_PI PCR begins with a PCR on the generated amplicon. Cycling conditions (Peqlab Primus96 Advanced Gradient Thermal Cycler): (95°C 5min) x1 (95°C 10sec, 60°C 30sec, 72°C 20 sec) x30. Reaction conditions: 5µl Hotshot Diamond mastermix (Clent Life Science); 1µl 20 x EvaGreen Dye in water (Biotium, inc.); 250nM of each squirrel primer (Eurofins); 1µl undiluted PCR product as template; addition of Nuclease-Free water (Qiagen) to give a total reaction volume of 10µl. All reactions were carried out in duplicate, including negative controls.

For sequencing, PCR products were purified using the GenElute PCR Clean-up kit. The purified products were diluted to 3ng/µl following initial quantification in a NanoDrop 2000 UV Spectrophotometer. Sequencing was performed with the dye terminator chemistry (BigDye version 3.1) on the 3130xl ABI PRISM Genetic Analyzer at the University of Nottingham DNA sequencing facility. All reactions were carried out in duplicate,

including negative controls. Sequence was analysed manually via FinchTV version 1.4 (https://finchtv.software.informer.com/1.4/).

iii. HOT_ARMS PCR

1. HOT_ARMS PCR protocol and optimisation

Unless otherwise described, all sections for HOT_ARMS PCR use the following reaction conditions and cycling program. All reactions were undertaken in 0.2ml tubes with caps (Agilent, U.S.A). Each reaction was performed in a final volume of 20µl which contained the following components: 2X HotShot Diamond mastermix (Clent Life Science, U.K); 20x EvaGreen dye in water (Biotium, U.S.A); 40ng DNA template or 5µl plasma eluate; 375nM each primer and Nuclease-Free water (Qiagen, Germany). Primers (supplementary table 5) were purchased from Eurofins (Luxembourg). A HOT_ARMS primer design guide is found in section 2.E.i.2. Two no template control reactions were also added to determine the presence of potential contamination.

Gradient PCR for annealing temperature optimisation was performed (using a Primus 96 advanced gradient PCR machine, Peqlab, Germany) with a Ta 62+/-10°C to ascertain the range at which the PCR would be specific. HOT_ARMS cycling conditions (Stratagene Mx3005p, Agilent, U.S.A) are as

follows: (95°C 5min) x1; (95°C 30sec, 71°C 15sec, 72°C 15 sec) x50 plus standard melt-curve.

2. Reaction conditions for additional modifications

Mismatches were introduced into the 3' penultimate base of the ARMS primer to enhance sensitivity[126, 127]. Primers with mismatches were ordered as regular primers from Eurofins. Further information on mismatch incorporation is found in section 7.C.i as the most preferable mismatch is still being determined for each SNV. Reaction conditions are described in section 2.F.i.1. The final cycling conditions for mismatch incorporated HOT_ARMS primers are as follows (Stratagene Mx3005p, Agilent, U.S.A) are as follows: (95°C 5min) x1; (95°C 30sec, 65°C 15sec, 72°C 15 sec) x50 plus standard melt-curve.

LNA's were incorporated into the ARMS primer on the 3' base to enhance sensitivity[121]. Further information on LNA incorporation and design is found in section 2.E.i.2. Primers with LNA's were ordered from Eurogentec. Reaction conditions are described in section 2.F.i.1. The final cycling conditions for LNA incorporated HOT_ARMS primers are as follows (Stratagene Mx3005p, Agilent, U.S.A) are as follows: (95°C 5min) x1; (95°C 30sec, 72°C 30 sec) x50 plus standard melt-curve.

Wild-type blocking probes with up to 10 LNA's were used to enhance sensitivity (supplementary table 5) by inhibiting ARMS primer binding to wild-type template[85, 124]. A design guide for wild-type blocking probes is found in section 2.E.i.3. Reaction conditions (plus addition of wild-type blocking probe) and cycling conditions are described in section 2.F.iv.

3. Rapid cycling

Rapid PCR cycling was achieved using the Qiagen fast cycling mastermix and magnetic induction cycler (MIC). The MIC uses magnetic induction and a centrifuge style set up for tubes to speed up ramping and capture fluorescent data from multiple reactions faster. HOT_ARMS assays with ΔCt >15 (40ng total DNA) between 50% MAF and wild-type can undergo rapid 30-minute testing without additional modifications. HOT_ARMS assays with lower specificity; ΔCt <15 between 50% MAF and wild-type will obtain lower sensitivities and require 3' LNA.

All reactions were undertaken in 0.2ml tubes with caps (Agilent, U.S.A). Each reaction was performed in a final volume of 20µl which contained the following components: 2X fast cycling mastermix (Qiagen, DE); 20x EvaGreen dye in water (Biotium, U.S.A); 40ng DNA template; 375nM each primer and Nuclease-Free water (Qiagen, DE).

The final rapid cycling conditions are as follows (MIC, Bio Molecular Systems, AU) are as follows: (95°C 5min) x1; (98°C 1sec, 68°C 1sec,) x40 plus standard melt-curve.

iv. HOT_PI PCR

HOT_PI PCR protocol

HOT_PI PCR is described as in section 6.A. Primers and blocking probes were designed as described in section 2.E.i.3. Details of primers and blocking probes for HOT_PI are found in supplementary table 5. All reactions were undertaken in 0.2ml tubes with caps (Agilent, U.S.A). Each reaction was performed in a final volume of 10µl which contained the following components: 2X HotShot Diamond mastermix (Clent Life Science, U.K); EvaGreen dye 20x in water (Biotium, U.S.A); LNA wild-type blocking probe @100nM (Eurogentec); 40ng DNA template; 375nM forward and reverse primer and Nuclease-Free water (Qiagen, Germany). Standard primers were purchased from Eurofins (Luxembourg) and LNA probes were purchased from Eurogentec (Belgium).

PCR was performed on the Stratagene MX3005P real-time machine. The threshold for detection was set using the machine's default parameters (10 standard deviations away from the mean of the baseline fluorescence). The following cycling parameters are used for HOT_PI: (95°C/5 min) X1 /

(95°C/30sec; 71°C/30sec) X10/ (95°C/30sec; 75°C/30sec) X40 / (72°C/5min) X1. PCR products were analysed either by high resolution melting analysis (section 2.F.i) or Sanger sequencing with squirrel primers (2.F.ii).

v. IntPlex ARMS qPCR

IntPlex[85] primer sequences and Ta are found in supp. table 5. Cycling conditions (Stratagene Mx3005p): (95°C 5min) x1 (95°C 10sec, 60°C 30sec, 72°C 20 sec) x30. Reaction conditions: 5µl Hotshot Diamond mastermix (Clent Life Science); 1µl 20 x EvaGreen Dye in water (Biotium, inc.); 900nM of each primer (Eurofins); 3600nM of blocker (Eurofins); 40ng cell line DNA; addition of Nuclease-Free water to give a total reaction volume of 10µl. HRM was performed via the standard system settings on the Stratagene Mx3005p to determine the presence or absence of any non-specific product generation.

3. Predicting rectal cancer radiation responsiveness by biomarker analysis

3. Predicting rectal cancer radiation responsiveness by biomarker analysis

A: Introduction

Currently, 20% of rectal cancer patients treated with chemoradiotherapy can achieve a pathological complete response when the surgical resection specimen is reviewed histologically[26, 29-32]. In approximately 60% there is partial response which can achieve tumour down-staging, and in the remainder, there is no obvious response seen[30, 33, 34]. However, it is not possible to determine how patients will respond before assigning treatment. Thus, there is a large proportion of patients whose treatment could be improved – the complete responders could avoid having surgery after chemoradiotherapy while the non-responders could be given an alternative therapy. Determining which patients would respond beforehand by biomarker analysis of biopsy specimens, either pre-treatment or at the beginning of radiotherapy would allow enhanced patient stratification which may improve outcome[94].

Published data suggest that varying expression of antioxidant enzymes, particularly members of the glutathione and thioredoxin pathways can predict response to radiotherapy in cancer[97, 98, 100]. These two protein families have shown important roles both In-Vitro and In-Vivo, for regulation of radiosensitivity[97, 98, 100]. Prof. Stewart Martin at the

3. Predicting rectal cancer radiation responsiveness by biomarker analysis

University of Nottingham has found thioredoxin proteins to be predictive or prognostic across multiple cancers including pancreatic, breast and brain[96-99]. This led to the development of drugs targeting the overexpression of several thioredoxin proteins, in particular Thioredoxin-interacting protein (TXNIP) which once downregulated was able to produce radiosensitivity in cells which were previously resistant. It is possible that these proteins may also be predictive or prognostic in rectal cancer. If they are upregulated, it will present the opportunity to perform clonogenic assays on rectal cancer cell lines which are known to be radioresistant such as SW837 and primary cell lines which could have radioresistance induced. The clonogenic assays would allow assessment of the developed drugs in rectal cancer to improve radiation response if these markers are overexpressed.

This chapter investigates the mRNA expression of antioxidant enzymes, particularly glutathione and thioredoxin pathway members in pre-treatment biopsies with matched tumour regression grading scores by histology to determine if they are predictive in rectal cancer. Whilst previous work by Prof. Martin has focused on the expression of these markers at the protein level. The funds required to perform immunohistochemistry on more than 20 targets is too great for a PhD Project. Moreover,

immunohistochemistry requires a single FFPE biopsy section per biomarker and the Queen's Medical Centre Pathology department will not be able to spare more than two sections per patient. The remainder of the biopsy block will be required if further diagnostic work needs to be performed. Due to these restrictions, RNA extraction followed by RT-qPCR was a logical choice to obtain data on many genes from the thioredoxin and glutathione families as well as related redox genes. Whilst this does not preclude protein stabilisation from causing overexpression of a protein, many proteins are up or down regulated at the RNA level. This allows the bulk of the work to be completed at low cost with minimal precious material utilised and if specific markers are found to be predictive, they can be validated at the protein level by immunohistochemistry.

Aims

1. Generate a cohort of rectal cancer patients that have undergone chemoradiotherapy. These patients must have pre-treatment biopsies available with matched tumour regression grading and resection specimens from the Queen's Medical Centre Pathology archive. Equal numbers of responders (Mandard 1 and 2) and non-responders (Mandard 3-5) are required.

3. Predicting rectal cancer radiation responsiveness by biomarker analysis

2. Assess the cohort in aim 1 by RT-qPCR for differing expression of thioredoxin, glutathione and related markers between responders and non-responders of chemoradiotherapy.
3. Perform Immunohistochemistry to validate any genes shown to be predictive by RT-qPCR at the protein level.
4. Perform clonogenic assays on SW837 and primary cell lines with radioresistance induced to see if specific markers shown to be predictive and overexpressed can cause radiosensitivity when down regulated by gene knockdown.

Hypotheses

1. Members of the thioredoxin and glutathione pathway will be predictive of response to chemoradiotherapy in rectal cancer patients.
2. Thioredoxin and glutathione proteins which are shown to be predictive and overexpressed will induce radiosensitivity in radioresistant rectal cancer cell lines when gene knockdown is used.

B: Molecular analysis of rectal cancer specimens to predict radiation response

i. Quality control analysis of rectal biopsy RNA by qPCR

It is widely known that nucleic acids extracted from FFPE samples are fragmented to approximately 150-400bp in most

cases and the bases themselves can also be damaged (for example formalin fixation can cause cytosine deamination which can then result in sequencing artifacts)[111]. DNA, RNA, and protein can also be cross-linked which reduces the amount of available template for PCR[111]. Typically, cross-linking can be partially reversed by incubation at high temperatures in the extraction process (i.e., 50-60°C with proteinase K digestion), but some cross-linking will remain and influence downstream PCR. In addition, there are PCR inhibitors in the fixative and paraffin wax that are often co-eluted with the nucleic acid from FFPE tissue[111].

All these factors are particularly important with RNA expression analysis as it adds bias to the quantification result if not dealt with strictly[128]. This is a key reason that many published papers offer conflicting results. There is no consensus of how this should be undertaken. However, research does suggest that the use of multiple housekeeping genes[129, 130] is a necessity and amplicons should be <150bp in length to avoid issues with fragmentation[128]. To avoid this, very strict quality control criteria were set to minimise the impact of formalin fixation and in turn increase the chances of a result which is both true and meaningful.

3. Predicting rectal cancer radiation responsiveness by biomarker analysis

FFPE samples were analysed in 4 phases from extract to result to ensure adequate QC was undertaken. In the first phase, 50 individual patient FFPE biopsy samples which had undergone radiotherapy and had a reported Mandard score[110] were extracted by Qiagen RNeasy FFPE kit and converted to cDNA by MLV reverse transcriptase. In this study responders were classed as Mandard 1 and 2 and non-responders were classed as Mandard 3-5. 26 patients were Mandard 1/2 and 24 patients were Mandard 3-5.

CDNA from all 50 samples underwent qPCR with the housekeeping genes RPS23, B2M, PPIA, RPLP0 and HPRT1 as an initial quality control check. 10ng cDNA was loaded into 10µl reactions. As the samples were biopsy specimens, some contained less than 1µg of RNA in total and further sections could not be obtained as they were required for clinical purposes. This meant that a preserving strategy was put in place so that screening of all intended targets could be undertaken. All Ct-values for the housekeeping genes roughly correlated between markers and so the data provided by B2M was used to remove poor quality samples. Any sample with a Ct-value higher than 26 (n=11) was excluded from analysis as shown in figure 17 (before) and figure 18 (after) below. This experiment was repeated twice.

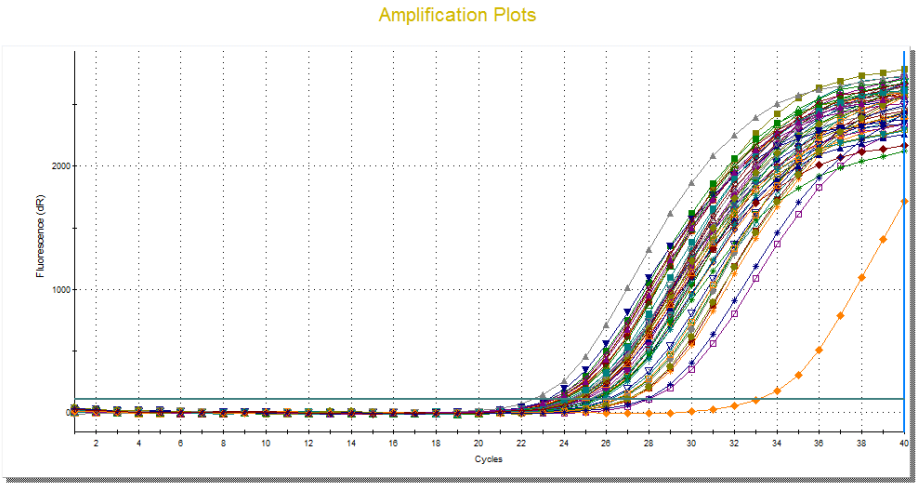


Figure 17. Rectal biopsy RT-qPCR amplification plot of 50 rectal biopsy samples for housekeeping gene B2M before quality control exclusion (excluding NTC – samples shown as single Ct-value).

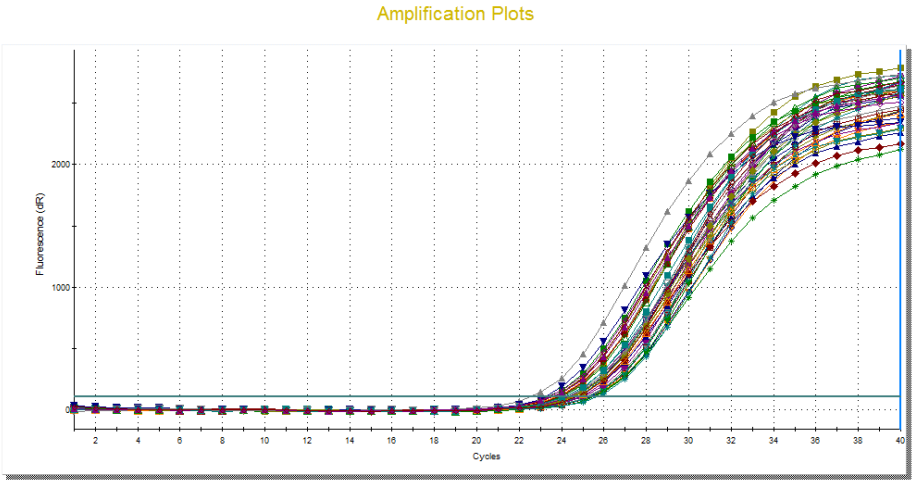


Figure 18. Rectal biopsy RT-qPCR amplification plot of 39 rectal biopsy samples after quality control exclusion to remove samples with low amounts of amplifiable cDNA (excluding NTC – samples shown as single Ct-value). 11 samples were

removed from the total 50 by using a Ct-value cut-off exclusion criterion of 26 cycles.

After the first phase of analysis, 39 out of 50 samples remained. The range of Ct-values were much tighter than before increasing confidence that the variation in expression analysis would be more robust. The maximum difference in Ct-value between the samples with highest and lowest Ct was 2.85. This indicated that similar quantities of template were being loaded into each reaction.

In the second phase, the samples were further quality control checked with the epithelial related gene EPCAM[131]. The presence of these markers indicates the level of epithelial cell content. Although the role of stroma in influencing the response to neo-adjuvant therapy cannot be ignored, it was felt that superficial biopsies of rectal tumours may have stromal features caused by tumour ulceration and may therefore not cancer specific. The focus here was on features of cells of epithelial origin and this QC step was aimed to ensure that the epithelial content should be as high as possible. This ensured samples with low tumour epithelium content in the extract were excluded. Any sample with a Ct-value higher than 28 (n=9) was excluded from analysis as shown in figure 19 (before) and figure 20 (after) below.

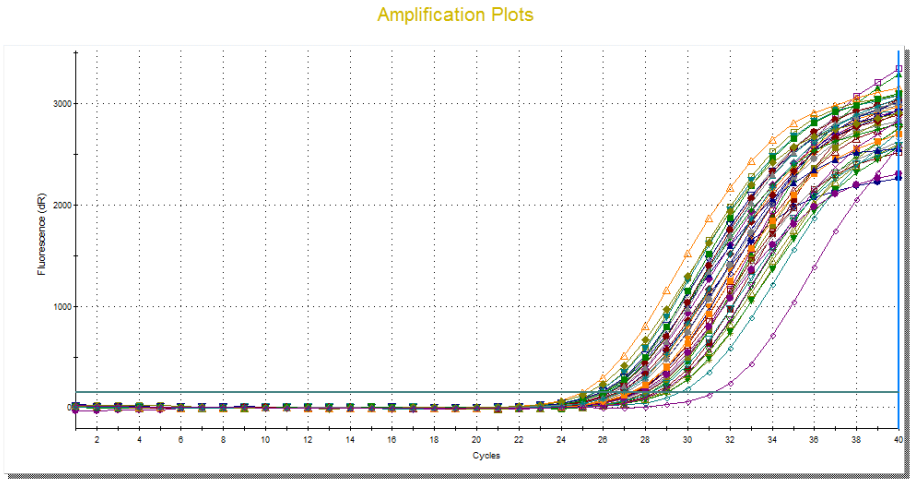


Figure 19. Rectal biopsy RT-qPCR amplification plot of 39 rectal biopsy samples for epithelial specific marker EPCAM before quality control exclusion to remove samples with low epithelial content (excluding NTC – samples shown as single Ct-value).

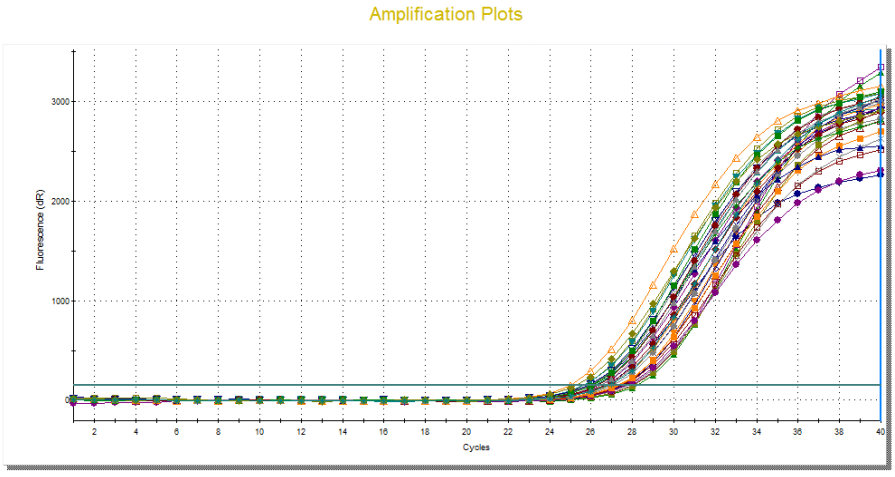


Figure 20. Rectal biopsy RT-qPCR amplification plot of 30 rectal biopsy samples after EPCAM quality control exclusion (excluding NTC – samples shown as single Ct-value). Ct-value

cut-off exclusion criteria set at 28 cycles which removed 9 samples in total with low epithelial content out of the total 39.

After the second phase, 30 out of 39 samples remained. The Ct-values were much tighter than before and the maximum difference in Ct-value between the samples with highest and lowest Ct was 3.12. This increased confidence that the variation in expression analysis would be more robust.

ii. RT-qPCR based gene expression profiling

In the third phase, expression profiling was carried out using a smaller sample set. As biopsy tissue is small, the RNA yield can be low. To preserve the RNA obtained for more in-depth analysis later, samples with total RNA yields >2µg were used first for general profiling of genes in triplicate reactions (10ng total cDNA). This left 20 samples for initial analysis: 10 responders (Mandard 1-2) and 10 non-responders (Mandard 3-5). To further preserve RNA, five of these samples along with cell line cDNA from HEK293T (Kidney), MDA-MB-231 (Breast), HCT116 (Colon), Huvec (Endothelial cell from umbilical cord) and HT29 (Colon) were used to screen the primers for expression. If the average Ct value of the samples was below 35, the gene moved to the next phase including the 20 samples listed above.

3. Predicting rectal cancer radiation responsiveness by biomarker analysis

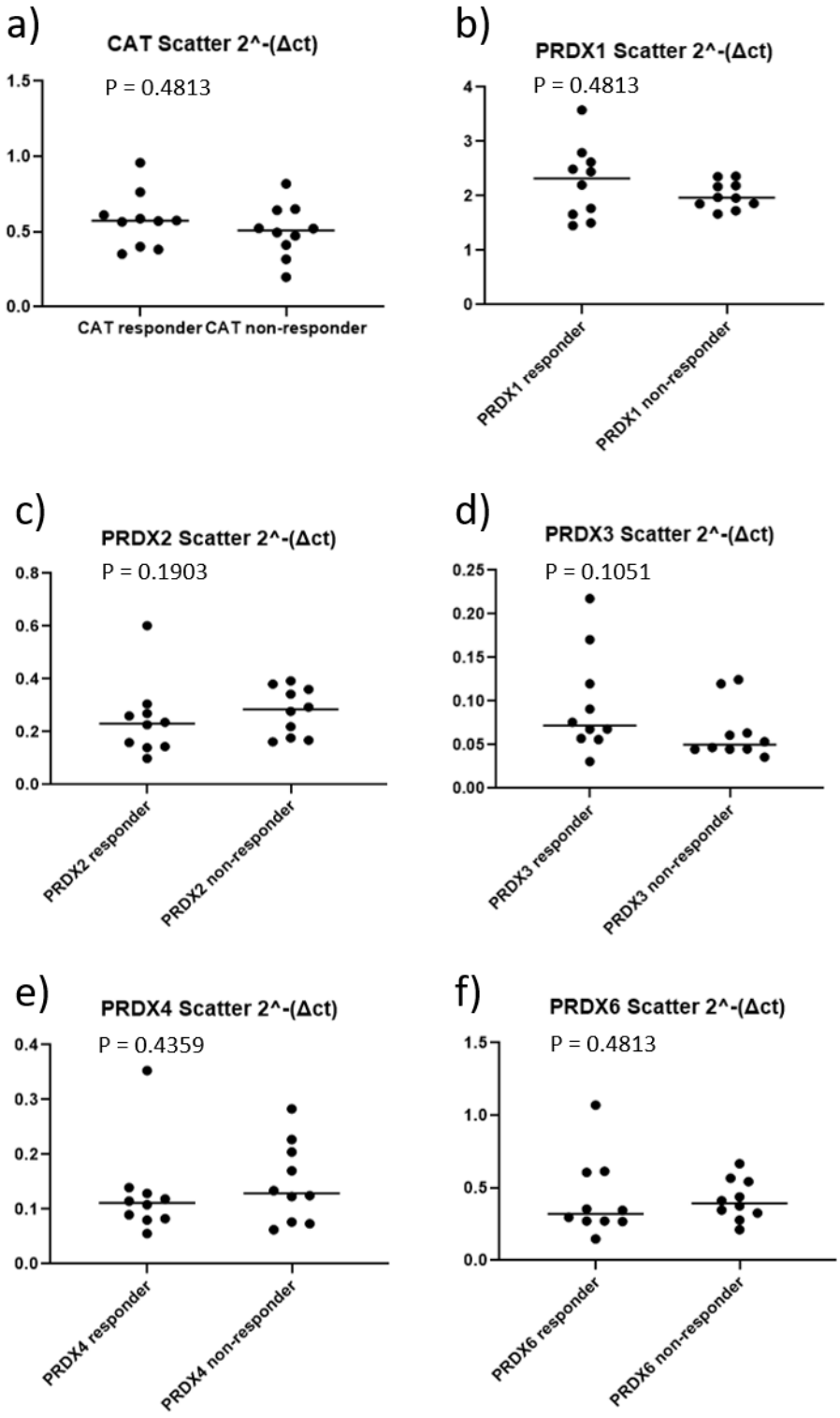
In total, 26 genes which were crucial to the thioredoxin and glutathione pathways or linked to them were targeted (supplementary table 8) and are as follows: ROS-dependent enzymes- SOD1 (superoxide dismutase), SOD3 (Superoxide dismutase 3), CAT (Catalase), PRDX1 (peroxiredoxin 1), PRDX2 (peroxiredoxin 2), PRDX3 (peroxiredoxin 3), PRDX4 (peroxiredoxin 4), PRDX6 (peroxiredoxin 6), GPX2 (glutathione peroxidase), GPX3 (glutathione peroxidase 3), GPX4 (glutathione peroxidase 4), GPX5 (glutathione peroxidase 5), GPX6 (glutathione peroxidase 6) and GPX7 (glutathione peroxidase 7); Glutathione system enzymes - GSR (Glutathione-Disulphide Reductase), GSS (Glutathione synthetase), GSTP1 (Glutathione S-transferase pi), GSTZ1 (Glutathione S-Transferase Zeta 1), GCLC (Glutamate-Cysteine Ligase Catalytic Subunit) and GCLM (Glutamate-Cysteine Ligase Modifier Subunit); thioredoxin system enzymes - TXNDR1 (Thioredoxin reductase 1), TXNIP (thioredoxin-interacting protein) and TXN2 (Thioredoxin); and glutaredoxin system enzymes - GLRX (Glutaredoxin-1), GLRX3 (Glutaredoxin 3), GLRX5 (Glutaredoxin-5). Every assay was designed to bind most transcript variants as shown in supplementary table 8. Whilst there is no consensus on this in the public domain, this was another way to limit the number of potential confounders in the final analysis. Every assay had

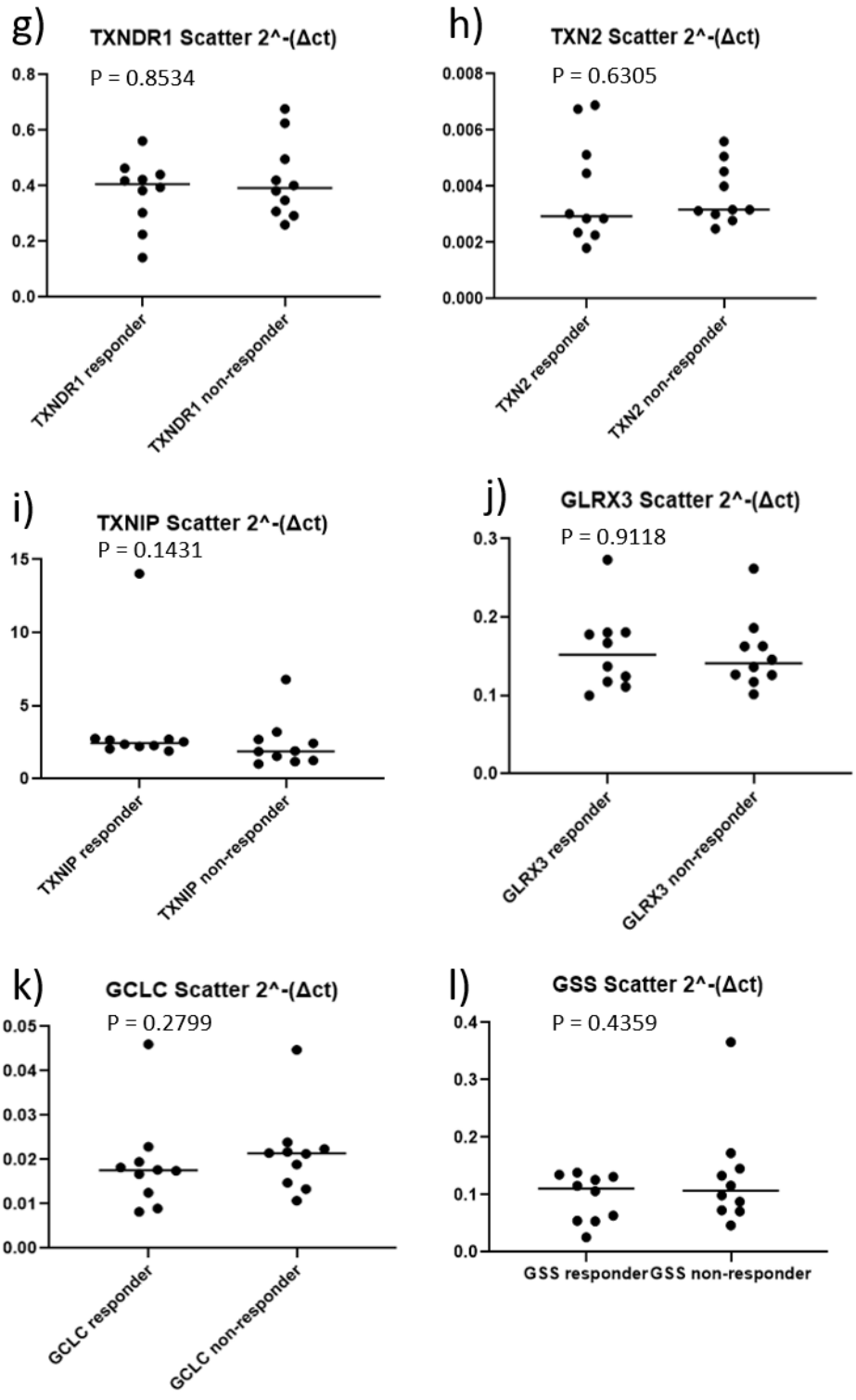
an efficiency between 85% and 115% as shown in supplementary figure 3-8.

5 rectal biopsy samples were used to screen the genes for general expression in rectal tissue before moving to analyse the full cohort of patients to preserve material. Ten genes which amplified specifically on colon cancer cell line DNA failed to generate Ct values below 35 in rectal cancer samples. The genes with low expression by RT-qPCR analysis are as follows: GPX3, GPX4, GPX5, GPX6, GPX7, GCLM, GLRX, GLRX5, GSTZ1 and SOD3. Assays which gave Ct-values in the 5 rectal biopsy samples below 35 (n=14) and were considered appropriate for analysis were as follows: PRDX1, PRDX2, PRDX3, PRDX4, PRDX6, GPX2, TXNDR1, CAT, GLRX3, GSR, GSS, GCLC, TXNIP and TXN2. These genes underwent qPCR with the 20 samples containing >2µg total RNA (10 responders, 10 non-responders). Data was analysed using the Livak $2^{-\Delta Ct}$ method as described in section 2.E.vi.2. The Ct values on all samples for 3 housekeepers - B2M, RPS23 and PPIA was used as a reference. The geometric mean of the housekeepers Ct value for each sample was used for the Livak method calculation to ensure that quantification and analysis was robust. In all instances, data was skewed and could not be transformed to a bell-shaped curve. Thus, the non-parametric Mann-Whitney

test was used to compare the Ct-values of the responders with the non-responders to determine if there were significant differences which may indicate a predictive marker. None of the genes gave a P-value <0.05. This was surprising considering previous reports, although some were based on immunohistochemistry data and qPCR does not account for protein stabilisation. TXNIP presented the lowest P-value of 0.1431 and so it was decided to take TXNIP to the next stage of analysis based on previous microarray reports to see if increased power would have an effect. The scatter plots of the $2^{-\Delta Ct}$ values for responders and non-responders are shown below in figure 21 for all genes run on the 20-sample set (10 responders, 10 non-responders to chemoradiotherapy).

3. Predicting rectal cancer radiation responsiveness by biomarker analysis




g)
TXNDR1 Scatter 2^-(Δct)
P = 0.8534
0.8
0.6
0.4
0.2
0.0
TXNDR1 responder
TXNDR1 non-responder
h)
TXN2 Scatter 2^-(Δct)
P = 0.6305
0.008
0.006
0.004
0.002
0.000
TXN2 responder
TXN2 non-responder
i)
TXNIP Scatter 2^-(Δct)
P = 0.1431
15
10
5
0
TXNIP responder
TXNIP non-responder
j)
GLRX3 Scatter 2^-(Δct)
P = 0.9118
0.3
0.2
0.1
0.0
GLRX3 responder
GLRX3 non-responder
k)
GCLC Scatter 2^-(Δct)
P = 0.2799
0.05
0.04
0.03
0.02
0.01
0.00
GCLC responder
GCLC non-responder
l)
GSS Scatter 2^-(Δct)
P = 0.4359
0.4
0.3
0.2
0.1
0.0
GSS responder
GSS non-responder

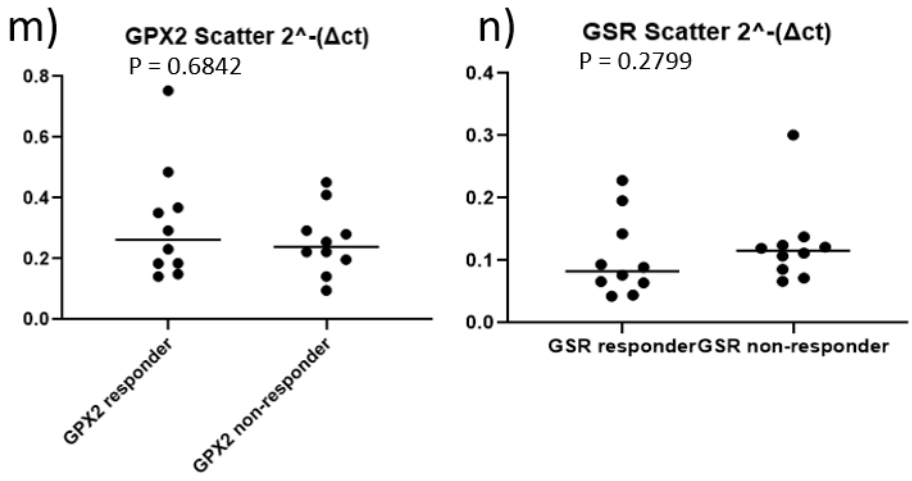


Figure 21. Scatter plots of $2^{-\Delta Ct}$ values for glutathione, thioredoxin and related pathway genes to compare rectal cancer responders and non-responders of chemoradiotherapy.

In the fourth phase TXNIP was investigated in further depth as the P-value was the lowest out of all samples at 0.01431 and it was hoped since a drug was developed that this marker would be overexpressed in non-responders so that it could provide radiosensitivity. Samples with RNA <2µg were included in the analysis as well as samples >2µg and all ran in triplicate. The total number of samples in this phase was 30, as shown in section 3.B.i.1 after determining epithelial content as a QC check. Experiments were also completed three times. Moreover, the amount of cDNA in the PCR reaction was increased to 40ng. However, the P value for TXNIP was still >0.05.

C: Discussion

Published studies had previously shown that members of the glutathione and thioredoxin system pathways could predict response to radiotherapy in a multitude of cancers, particularly breast cancer and lung cancer[97, 98, 100]. These pathways had previously not been studied to a large extent in rectal cancer regarding radiation response. This study aimed to identify if any key players of the redox-homeostasis proteome could predict response to radiotherapy. Twenty-six RT-qPCR assays specific to fragmented RNA were designed to evaluate gene expression of 4 different sets of redox-homeostasis enyzmes: ROS-dependent enzymes, glutathione system enzymes, thioredoxin system enzymes and glutaredoxin system enzymes. RT-qPCR is a high-throughput low-cost method to investigate potential biomarkers compared with antibody-based analysis and was well suited to the budget and time restrictions of the PhD. We had planned to use immunohistochemistry as a follow up to RT-qPCR for validation.

This study has determined that there was no difference in RNA expression for 16 of the 26 total genes investigated between rectal cancer responders and non-responders to radiotherapy. The 16 genes are as follows: PRDX1, PRDX2, PRDX3, PRDX4,

PRDX6, GPX2, TXNDR1, CAT, SOD1, GSTP1, GLRX3, GSR, GSS, GCLC, TXNIP and TXN2. The remaining 10 genes, GPX3, GPX4, GPX5, GPX6, GPX7, GCLM, GLRX, GLRX5, GSTZ1 and SOD3 were not highly expressed and could not be evaluated. Predictive markers must show clear differences between responders and non-responders. Thus, a larger sample set would not be of additional value. Whilst there was no difference in expression for these key genes, this does not preclude other genes from the redox-homeostasis proteome from the 4 pathways being predictive but does rule these specific key gene's out. Moreover, it does not preclude these enzymes from being predictive at a protein level since expression analysis cannot determine protein expression that is controlled by stabilisation mechanisms[132]. Antibody-based analysis such as reverse phase protein array or immunohistochemistry could be utilised in the future to determine if these key redox-homeostasis proteins are predictive. This would however require a larger funding pool as antibodies are expensive and do not always work on FFPE biopsy sections, requiring multiple rounds of antibody screening for a specific protein.

As our funding pool was low and we had success with liquid biopsy method development (section 4, 5 and 6), the work

was abandoned. We had also planned to test inhibitors of glutathione and thioredoxin proteins as a radiosensitising agent but as these results were negative this route was also abandoned.

4. Investigation of circulating cell-free DNA characteristics

4: Investigation of circulating cell-free DNA characteristics

A. Introduction

In 2016, when the project first began, the availability of circulating tumour DNA products and academic research on the area was limited. Initial research in 2016 indicated that cell-free DNA was highly fragmented[66], low in concentration[67] and extremely low in mutant allele frequency[68, 69].

In the first instance, it was necessary to optimise the extraction, quantification, and method of analysis for detecting mutations in cfDNA. Since a new liquid biopsy study on pre-and post-surgery colorectal cancer patients was beginning, this generated a sample set for initial investigation into cfDNA characteristics. This has allowed many initial failures to be rectified to create a useable protocol for analysis of cell-free DNA from rectal cancer patients treated with neoadjuvant radiotherapy.

This chapter discusses the key issues with analysing cfDNA by PCR. Key issues include fragmentation, limited template concentration and low mutant allele frequencies. The latter includes investigation of published ultrasensitive methods which could potentially overcome these factors such as

enhanced-ICE-COLD PCR[81, 133] and IntPlex ARMS qPCR[85].

Aims

1. Investigate circulating tumour DNA characteristics such as low concentration and fragmentation.
2. Investigate Enhanced-ICE COLD PCR and IntPlex ARMS qPCR as ultrasensitive methods to detect mutations as low as 0.1% mutant allele frequency.

Hypotheses

1. Cell-free DNA will be highly fragmented with a mean fragment size of 130-166bp and as such will struggle to amplify 200bp amplicons.
2. Cell-free DNA will be low in concentration (1-100ng/ml).
3. Enhanced-ICE COLD PCR and IntPlex ARMS qPCR will be able to detect 0.1% mutant allele frequency samples.

B: Results - Initial optimisation of cfDNA

i. CfDNA extraction optimisation

The first place to begin optimisation was the method of extraction and the way in which to quantify the product. CfDNA from 1ml patient pre- and post-surgery plasma samples (n=49) was extracted with the Qiagen circulating nucleic acid kit. Only 1ml could be spared but in an ideal

scenario up to 5ml of plasma can be extracted using this kit which would enhance the chances for detecting ctDNA by increasing the cfDNA concentration. The volume of plasma used, elution volume and DNA concentration of each sample can be found in supp. Table 3. These factors have been taken into consideration to calculate the total DNA (ng) per ml of plasma which can be found in table 4 below. The calculation is as follows: (elution volume (µl) x DNA concentration (ng/ml)) ÷ volume of plasma (ml). Pre-operative samples were named day 1 and eluted in 40µl. Post-operative samples were collected the following day which was named day 2 onwards (weekends permitting) as described in section 2.A.i and eluted in 20µl as it was expected that detection could be enhanced since total ctDNA would be lower after surgery.

Table 4. Quantification of cfDNA from CRC pre- and post-surgery patient liquid biopsies extracted with the Qiagen circulating nucleic acid kit. Volume of plasma used, elution volume and DNA concentration of each sample can be found in supp. Table 3. Pre-operative samples (n=25) were named day 1 and post-operative samples (n=24) were collected the following day which was named day 2 onwards (weekends permitting). Dukes stage is reported alongside the total DNA per ml of plasma that was obtained.

Sample	Total DNA per ml of plasma (ng/ml)	Dukes' stage	Sample	Total DNA per ml of plasma (ng/ml)	Dukes' stage
2 Day 1	161.06	B	15 Day 1	59.60	A
2 Day 2	28.6		15 Day 2	46.20	
3 Day 1	153.45	C1	16 Day 1	37.60	B
3 Day 4	53.4		16 Day 2	61.00	
4 Day 1	244	B	17 Day 1	95.33	B
4 Day 2	72.2		17 Day 2	63.00	
5 Day 1	51.6	C1	18 Day 1	56.00	B
5 Day 2	83		18 Day 2	39.60	
6 Day 1	294	B	19 Day 1	42.80	A
6 Day 2	98		19 Day 2	40.20	
7 Day 1	79.6	B	20 Day 1	54.40	A
7 Day 2	40.4		20 Day 2	72.20	

8 Day 1	73.6	A	21 Day 1	59.20	B
8 Day 2	37		21 Day 2	37.80	
9 Day 1	70.8	Unknown	22 Day 1	74.4	A
9 Day 2	52.6		22 Day 2	58.2	
10 Day 1	66	C1	23 Day 1	9.32	Unknown
10 Day 2	52.2		23 Day 2	5.96	
11 Day 1	50.40	B	24 Day 1	4.72	Unknown
11 Day 2	37.40		24 Day 2	3.78	
12 Day 1	57.20	A	25 Day 1	12.08	Unknown
12 Day 2	36.40		25 Day 2	51.8	
13 Day 1	42.40	B	26 Day 1	15.96	Unknown
13 Day 2	18.48				
14 Day 1	45.20	A			
14 Day 2	19.34				

As shown in table 4, most patients are Dukes' stage A (n = 7) and B (n = 10). A few patients were Dukes' stage C cases (n = 3), and no patients were Dukes' stage D. Furthermore, five cases had unknown staging characteristics. Later stage cases are expected to produce more ctDNA since they are generally larger tumours with an increased shedding rate. Whilst this sample set is small, stage A cases (mean = 52.05 ng/ml, n=7, lowest = 42.8ng/ml - highest = 74.4ng/ml) generally have lower cfDNA yield compared with stage B (mean = 111.96, n=10, lowest = 37.6ng/ml – highest = 294ng/ml) and C (mean = 90.35, n=3, lowest = 51.6ng/ml – highest = 153.45 ng/ml). However, it is widely accepted that this is not a reliable way of indicating presence of cancer as cell-free DNA can be produced naturally at different rates in individuals and other conditions may elevate cfDNA levels. Qiagen states a minimum of 1ng/ml and a maximum of 100ng/ml for extracted cfDNA is to be expected in their handbook[67]. Most samples (n = 45) fall within this normal range. However, the following four samples fall out of this normal range by a large proportion which is promising for downstream analysis: sample 2 day 1 (161.06ng/ml), sample 3 day 1 (153.45 ng/ml), sample 4 day 1 (244 ng/ml) and 6 day 1 (294ng/ml). The minimum value obtained was 3.78ng/ml by sample 24 day 2 whilst the maximum was 294ng/ml by sample 6 day 1.

Whilst these values are expected, this gave total concentrations of 0.118ng/µl (3.78ng/ml) and 7.35ng/µl (294ng/ml). This data set shows that cfDNA can be present at extremely low concentrations which presents a large challenge for both the detection of ctDNA and the method of analysis overall since many methods have specific input requirements which are greater than many of the samples in table 4. This further reinforces that PCR is likely a method which is well suited to cfDNA analysis since it will not result in sample exclusion due to the amount of cfDNA obtained. This will broaden access to testing via this route in comparison to others.

ii. Testing the effect of low DNA concentration on PCR

The next logical step was to test the effect of exceptionally low DNA concentration on the PCR reaction. A gradient of 40ng, 20ng, 10ng, 7.5ng, 5ng, 2.5ng, 1ng, 0.5ng and 0.1ng total DNA per reaction (20µl total volume) was tested with the KRAS exon 2, 92bp amplicon which is described in supplementary table 2. All reactions gave Ct values below 35 and generated specific amplification as shown by the melting peaks in figure 22 below.

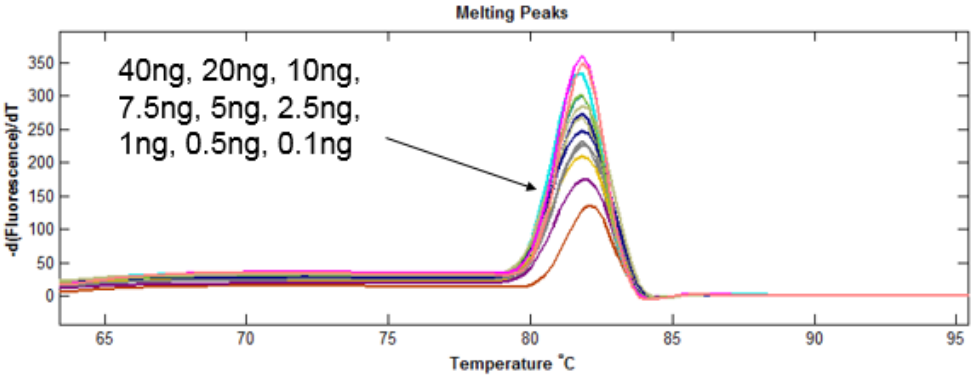


Figure 22. Melting analysis of gradient total DNA per reaction to test the reliable limit of amplification. A gradient of 40ng, 20ng, 10ng, 7.5ng, 5ng, 2.5ng, 1ng, 0.5ng and 0.1ng total DNA per reaction (20µl total volume) was tested in duplicate. The melting plot shows the fluorescence generated by double-stranded binding dye (Eva Green) plotted against the temperature. All DNA inputs melted at the same overall temperature; some melting peaks are lower as they did not obtain plateau phase since the PCR was 40 cycles in total.

This data shows that 0.1ng total DNA per reaction (20µl total volume) is sufficient for real-time PCR, end-point HRMA or other types of PCR analysis and that this should not be a limiting factor so long as the DNA is of sufficient quality.

iii. Fragment size analysis of extracted cfDNA

In section 4.B.i, it was determined that extraction of cfDNA from 1ml of plasma was possible as fluorometric quantification validated the presence of double-stranded DNA. In the next step it was key to determine the amount of fragmentation to

4. Investigation of circulating cell-free DNA characteristics

determine if cfDNA could be amplified reliably by PCR with Ct <35. CfDNA is reported to have a mean fragment size of 130-166bp[66] and it is possible using the Qiagen circulating nucleic acid kit to extract fragment sizes of up to 1000bp[67]. Thus, this experiment set out to determine how many samples contained longer fragments which could be amplified. Two sets of primers were used to validate the level of fragmentation by qPCR. One primer set generated a 105bp amplicon (KRAS exon 4) and another generated a 194bp amplicon (TP53 exon 5). Both sets of primers are found in supplementary table 1. Since the downstream method of analysis was to be PCR, methods for analysing fragmentation such as Tapestation or Bioanalyser are less representative as they do not account for amplifiable fragmented DNA. 5ng of cfDNA template from 13 different samples (sample 2-14 day 1) was loaded into separate 20µl reactions. No-amplification was described as a Ct> 35 cycles. Melting peaks were analysed to ensure the products were specific.

All 13 samples amplified 105bp amplicons with Ct-value ranging from 27-31 and 10 out of 13 samples (Ct 32-35) amplified 194bp amplicons. This indicates that larger fragments are extracted in most samples but amplification over 146bp would be unreliable in all samples (Ct >35). This data shows that the PCR assay should be designed to

generate amplicons <130-166bp[66] and ideally as short as possible to enhance detection of ctDNA fragments which E-ICE COLD PCR and IntPlex ARMS qPCR are designed for.

iv. Analysis of ultra-sensitive mutation detection methods for liquid biopsy

As discussed in 1.C.iii, a rapid, inexpensive, ultrasensitive, and closed tube test was required for serial treatment response monitoring and detection of minimal residual disease. Several methods were highlighted that have these characteristics: IntPlex ARMS qPCR[85] and Enhanced ICE-COLD PCR[81, 133]. Other methods described in section 1.C.iii which are now FDA approved were previously not available in 2017 and as such could not be evaluated. The threshold limit of detection for a method was set at 0.1% as many publications backed up the requirement of this cut-off as more research was completed by 2017[68, 69].

Enhanced ICE-COLD PCR[81, 133] (E-ICE COLD PCR) was a logical choice as prior work had been undertaken which achieved a limit of detection of 1.5% mutant allele frequency for KRAS exon 2[80]. E-ICE COLD PCR[81, 133] is a method combining two enrichment methods, co-amplification at lower denaturation temperature (COLD) [80] and wild-type blocking PCR[79] with LNA inclusion (Enhanced ICE). E-ICE COLD PCR

is method of enrichment which allows amplification of mutant DNA by inhibiting wild-type amplification in two ways. Firstly, COLD-PCR[80] uses a critical denaturation temperature over regular 95°C denaturation temperature. The critical denaturation temperature is lower than that of wild-type homoduplexes[80]. This means that there is preferential amplification of heteroduplexes, reducing the pool of wild-type DNA[80]. Secondly, wild-type blocking probes (ICE) bind to wild-type DNA at a higher temperature than the primers which in turn block primer access to the wild-type template, causing amplification of mutant template preferentially[79]. In the enhanced version (E-ICE) [81, 133] locked nucleic acids (LNA's), that are modified nucleotides known to enhance binding strength and specificity over regular nucleotides are included in the wild-type blocking probe to enhance binding capacity to the wild-type template[122]. The published data shows a limit of detection of 0.05% MAF for KRAS exon 2 codons 12/13 which is suitable for liquid biopsy[81]. In practice, the KRAS E-ICE COLD PCR assay showed that amplification of 40ng HEK293T wild-type DNA was delayed by 5 cycles compared to 40ng 50% MAF HCT116 DNA (KRAS exon 2 c.38G>A), representing mutation enrichment. However, when the products were sequenced it appeared that many PCR errors were introduced as shown below in figure

23. Thus, the method was not taken forward as it was still unreliable after multiple repeats.

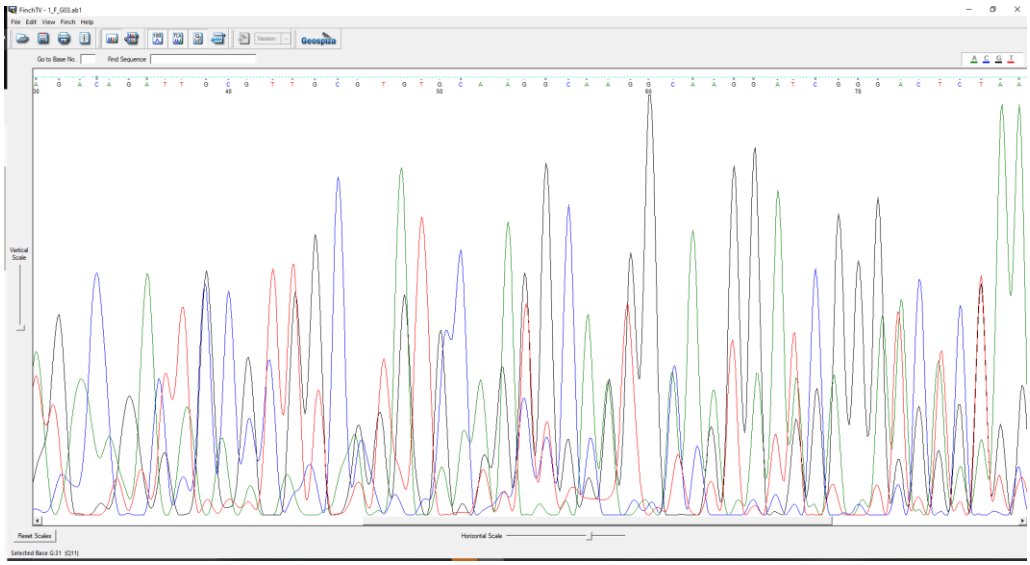


Figure 23. Multiple sequencing errors after amplification of KRAS mutant c.38G>A HCT116 DNA by E-ICE COLD PCR. This experiment was repeated multiple times with errors introduced each time.

IntPlex is an ARMS qPCR with low Tm mutation-specific primers and high Tm nucleotide blocker for the ultra-sensitive detection of mutant alleles[85, 134]. The low Tm ARMS primers allow the blocking probe to bind on effectively as there is a large difference between probe and primer annealing temperature[85]. IntPlex uses a double-stranded binding dye such as SYBR green to determine amplification[85] and relies on the delayed amplification of wild-type DNA to show the presence of mutant DNA by Ct-

value. IntPlex is now commercially available for liquid biopsy mutation detection[135] and published data suggest that it is able to achieve a limit of detection as low as 0.01%[85, 134]. The IntPlex team were of few to publish papers claiming over 90% concordance between FFPE tumour DNA and cfDNA in 2017[134]. This made ARMS PCR a serious contender as a cfDNA analysis method since it provided significantly higher concordance than other methods at the time.

To test the limit of detection of the system, three of the published[85] IntPlex assays targeting KRAS c.38G>A (G13D), c.35G>A (G12D) and c.34G>T (G12C) were tested with positive spike-in controls. The controls were generated by use of colorectal cancer cell lines HCT 116 (KRAS exon 2, c.38G>A, G13D+), GP2D (KRAS exon 2, c.35G>A G12D+) and SW1116 (KRAS exon 2, c34G>A G12A+). HEK293T was used a global wild-type background for dilution. 1% mutant allele frequency spike-in controls were initially prepared by mixing 1-part mutant DNA with 99 parts wild type DNA. A 2-fold serial dilution of the 1% spike was prepared to give two low MAF positive controls of 0.5% and 0.125% which would enable limit of detection testing that is comparable to ctDNA. Mutant allele frequencies which are higher are not necessary to investigate since yes/no answers are a priority over quantification. Two no template control reactions were also

added to determine the presence of contamination. Primer sequences are described in supplementary table 5. The experimental conditions are described in section 2.F.v. All reactions were running in triplicate with 40ng of template and the Ct values for the respective positive and negative controls of all 3 assays are found in table 5 below.

Table 5. LOD results for three common KRAS exon 2 sequence changes in colorectal cancer. Successful detection of 0.1% mutant allele frequency for 3 major sequence changes in KRAS exon 2; c.38G>A (G13D), c.35G>A (G12D) and c.34G>T (G12A) using IntPlex.

Target	50% MAF	0.5% MAF	0.125% MAF	0% MAF
KRAS c.38G>A (G13D)	26.87	35.06	35.41	37.51
KRAS c.35G>A (G12D)	35.50	39.92	40.49	48.42
KRAS c.34G>T (G12A)	34.43	38.50	38.62	46.51

The limit of detection was classed as a ΔCt between the mutant and wild type of at least 2 in 3 replicates. In practice, if the ΔCt of more than 2 is always maintained then the result

is reliable when standards are run alongside. The Ct-values described in table 5 are an average of the 3 replicates. The ΔCt between 0.1% MAF and wild-type DNA are sufficiently far apart to characterise these assays as having a 0.125% limit of detection which would be sufficient for analysis of ctDNA. The c.38G>A (G13D) assay had a ΔCt of 2.1, the c.35G>A (G12D) assay had a ΔCt of 7.93 and the c.34G>T (G12A) assay had a ΔCt of 7.89.

The high levels of blocking probe pushed the Ct value for the c.35G>A (G12D) assay to 35.50, the c.38G>A (G13D) to 26.87 and the c.34G>T (G12A) to 34.43. The Ct-values are far beyond 30 cycles for the 50% MAF which was extremely inefficient and will be detrimental for cfDNA analysis where concentrations are low. The IntPlex[85] system relied heavily on wild-type blockers for mutation detection rather than the primer itself and this factor gives the system a flaw. We saw the potential in ARMS PCR[72, 78, 85, 126] and wild-type blocking PCR[79, 133] and this led us to develop our own methods as discussed in chapter 5 (ARMS qPCR) and chapter 6 (wild-type blocking PCR).

C: Discussion

Several important conclusions have been generated by this study about the characteristics of cfDNA, methods for

extraction and methods for mutation analysis. The concentration of cfDNA is exceptionally low in some samples and in this study, the minimum extracted quantity from the circulating nucleic acid kit was 3.78ng/ml (0.118ng/µl). Qiagen reports in its handbook that the minimum can be as low as 1ng/ml[67] (0.025ng/µl with 40µl elution). This research has concluded that 0.1ng total DNA per PCR reaction (20µl reaction volume) can be amplified by PCR for analysis which is sufficient if 5µl of eluate is loaded (equal to 0.125ng).

The fragmented nature of cfDNA is something that cannot be prevented as this occurs when cell-death-mediated DNases are activated during apoptosis or necrosis[136]. This study has confirmed the fragmented nature of cfDNA in section 4.B.iii. Extracts from the Qiagen circulating nucleic acid kit can generate fragments long enough to amplify 194bp amplicons in 10 out of 13 samples and amplification of 105bp amplicons is reliable as all 13 samples were amplified with a Ct value below 35. This data indicates that primers should be designed to be less than the average ctDNA fragment size of 130-166bp[66] and ideally as short as possible to enhance detection of ctDNA.

cfDNA has been shown to carry a wide-range of variant allele frequency which may be very low in cancer patients[68]. It is

difficult to obtain an average expected value before conducting analysis as published studies have used a wide range of method's and patient samples for mutation detection. It is generally accepted that later stage, larger tumours, or metastatic patients will shed more circulating tumour DNA than early-stage patient's due to increase in tumour cell death rate[137, 138]. This will increase the percentage of mutant alleles to a large degree as well as the quantity of cfDNA that can be extracted. However, it is now widely known in published literature that cfDNA variant allele percentages can be as low as 0.01% in some patients and this is the largest challenge that cfDNA presents for analysis[68, 69].

During this research, IntPlex ARMS qPCR[85], and Enhanced ICE-COLD PCR[81] was investigated to see if the limit of detection was sufficient to detect mutations in liquid biopsy. Enhanced-ICE COLD PCR presented mutation enrichment by amplification plot but when the products were sequencing, artefactual sequence not resembling KRAS exon 2 was present on multiple repeats and so the method was abandoned.

IntPlex ARMS qPCR was able to reliably detect 0.125% MAF in 3 KRAS assays, c.38G>A (G13D), c.35G>A (G12D) and c.34G>T (G12A). The limit of detection was sufficient but the reliance on high concentrations of primer (900nM) and high

concentrations of wild-type blocking probe (3600nM) was extreme. The potential for multiplexing was limited due to the high concentrations as it would likely generate dimers. Moreover, the over reliance on wild-type blocking probes which rely on specific sequence context to work effectively would not always be available when developing new assays as the ARMS primer locks in a small amount of sequence which is further hindered by the fragmented nature of cfDNA. If the target sequence is not favourable, then the blocker may not block efficiently, or it may bind to other areas of the genome and reduce the effect. It is likely that this approach will not fare well across multiple genes and this heavily reduces the potential scalability for the method. This is further compounded by the increased optimisation that wild-type blocking probes bring. Moreover, the other 4 KRAS hotspot mutation IntPlex assays were tested but failed to produce any difference between mutant and wild-type Ct values. There was also a significant efficiency drop as 2 assays, c.35G>A (G12D) and c.34G>T (G12A) were amplifying 50% MAF positive controls at 35.50 and 34.43, respectively. This is a large issue when trying to amplify samples with low template concentrations such as in liquid biopsy where efficiency needs to be high to generate specific signals.

4. Investigation of circulating cell-free DNA characteristics

The above methods had issues but the potential of ARMS qPCR[72, 78, 85, 126] and wild-type blocking PCR[79, 133] for liquid biopsy analysis is massive. The only solution left was to develop a new method as is discussed in greater depth in the next section of the thesis, chapter 5 (ARMS PCR) and chapter 6 (wild-type blocking PCR). As discussed in section 1.C.iii a personalised approach is also favoured for liquid biopsy sensitivity which is further discussed in the future plans section (section 7.C.iii).

As mentioned previously, improvement to extraction may be critical to improving ctDNA detection. Recent advancements have occurred since 2016/2017 which look promising to improve ctDNA detection. Combined DNA/RNA extraction kits have been shown to improve detection by increasing the number of target copies by including RNA extraction. A study by Krug et al presented large increases in sensitivity with detection of EGFR activating mutations increasing from 26% to 74% and EGFR T790M increasing from 19% to 31% when comparing DNA only to RNA/DNA combined[139]. In this study, 1ml of plasma was used for cfDNA extraction, although it is possible using the Qiagen circulating nucleic acid kit[67] to extract from 5ml and with the new Qiagen magnetic kit[140] to extract from 10ml. The magnetic kit also has the capability to extract from more than 10ml, although it is not

recommended in the manual, the reagents and process are scalable. Hence, generating higher total DNA yield in the eluate for PCR. The magnetic kit has lower yield per ml compared with the circulating nucleic acid kit but the option to increase plasma input can give higher yield per extraction overall shown by Dieffenbach et al in 2018[141]. The other problem with the magnetic kit is that it reduces the fragment size to approximately 170bp[67, 141] whereas the circulating nucleic acid kit gives the capability of extracting larger fragments up to 1000bp[67, 141]. Whilst these kits allow the option for increased plasma volume; it remains difficult to obtain larger volumes of plasma since patients are bled daily for standard of care analysis. This leaves a small minority for research but translates well for the clinic once priority is given.

Research into the presence of ctDNA within urine cfDNA for solid tumours may be a viable alternative to plasma to increase the concentration. Using urine would allow an increased volume since it is easy to obtain. It may be possible using the Qiagen magnetic extraction kit to extract cfDNA from large volumes of urine. Whilst the kit has been specifically marketed for plasma (up to 10ml), the extraction kit is flexible and realistically larger volumes than 10ml could be extracted from using 50ml tubes. The magnetic beads bind

170bp DNA fragments (ctDNA size) and the initial processing steps are compatible with urine. Moreover, although cfDNA is excreted through the urine and has been shown as a good marker in bladder cancer[142]; there is no evidence yet to suggest that solid tumour DNA will be as abundant in urine as plasma. The ease of acquiring the fluid may be a factor which counter's this though; especially when using an ultrasensitive method of mutation detection. A large study to confirm the presence of ctDNA in urine whilst comparing against matched plasma samples would allow proof of principle.

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

A: Introduction to HOT_ARMS PCR and the Highly Optimised annealing Temperature_PCR (HOT_PCR) umbrella

i. Why is a new method of cell-free DNA analysis required?

Detection of ctDNA holds a great deal of promise for the management of Rectal cancer. There are some methodological constraints including low levels of cfDNA[67], low mutant allele frequency[68, 69] and small DNA fragment size[66, 141]. If ctDNA is to be used for monitoring treatment response and cancer surveillance, then repeated analysis will need to be performed, adding further constraints such as cost. We aimed to develop a test which could be used in the outpatient department and would therefore need a rapid sample-to-test result time of approximately 2-3 hours.

All the commercially available tests and tests in the research domain were either too expensive for longitudinal testing, not sensitive enough, not reliable enough or not scalable enough (section 1.C.iii). Thus, we sought to develop a test which would overcome most of the constraints of ctDNA testing for both this research and liquid biopsy testing clinically. We could see the potential of ARMS-PCR and wild-type blocking PCR to

fulfil these criteria from work carried out in section 4 which followed through into further method development in section 5 and 6.

Aims

1. Develop an ultrasensitive test which is reliable, scalable, and low cost for longitudinal cell-free DNA analysis.

2. Use the method generated in aim 1 to investigate liquid biopsies from rectal cancer patients undergoing chemoradiotherapy. This will enable response to be measured in real-time.

Hypotheses

1. A key issue with ARMS-PCR is wild-type DNA amplification which reduces sensitivity, this may be overcome by raising the annealing temperature of the reaction to reduce non-specific priming. This may be done through primer length or the addition of modified nucleotides such as LNA. We have recently published work with 5’ tags [76, 117] to change amplicon melting temperature for multiplex high-resolution melting analysis and we see the potential for these to be applied to ARMS primers to raise the primer annealing temperature.

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

2. ARMS-PCR once modified will be a robust and inexpensive solution for cell-free DNA testing which will be able to detect ctDNA mutations in rectal cancer patients' liquid biopsies.
3. CtDNA will aid in tracking patient response to chemoradiotherapy in real-time.

ii. What is HOT_ARMS PCR and the HOT_PCR umbrella?

We have developed novel liquid biopsy specific methods under an umbrella term we call HOT_PCR (High Optimised Ta_ PCR). This currently comprises 2 different methods based on the same principle and with different advantages: HOT_PI (probe-inhibited) PCR and HOT_ARMS (amplification refractory mutation system) PCR. HOT_PI PCR is discussed in chapter 6. HOT_ARMS PCR is an ultra-sensitive real-time PCR based mutation detection system specifically designed for cell-free DNA which can scale across the whole genome with detection as low as 1 mutant copy.

1. What is ARMS PCR?

The Amplification Refractory Mutation System (ARMS) [143] depends on the use of primers that are tailored to specific mutant sequences. ARMS PCR was initially designed to test for Single Nucleotide Variants (SNVs). It works on the principle

that the 3’ base of a primer needs to be bound to the template for the polymerase enzyme to begin extension. It follows from this that if the mutant and wild type sequence differ by just one base, a mutation-specific primer can be designed with the 3’ base being complementary to the mutant sequence. Thus, if a mutation specific primer anneals to wild type DNA, there should be no primer extension and amplification would be mutation specific.

2. What is the problem with ARMS PCR?

Theoretically, ARMS PCR should result in amplification of only the mutant allele with no amplification from the wild-type allele. However, in practice, aberrant priming of the wild-type template by ARMS primers does occur and can be referred to as ‘leakiness’ in the system or wild-type ‘bleed-through’. The bleed through is a problem with low mutant allele frequency (MAF) as the amplification signal from non-specific priming of wild-type DNA cannot be distinguished from the signal produced by low levels of mutant alleles. Since the inception of ARMS PCR[78], several modifications have been introduced to improve the limit of detection and reduce wild-type bleed through. These include mismatches on the penultimate 3’ base[126, 127], wild-type blocking probes[85, 120, 124],

modified polymerases[72], 3’ modified nucleic acids[121] and modified mastermix components. Whilst these modifications do work, they can generate additional issues. For example, penultimate 3’ base mismatches, wild-type blocking probes and 3’ modified nucleic acids can drastically reduce the efficiency of the PCR reaction. The efficiency reduction can impact the limit of detection, limit of quantification, quantification capability and dynamic range of template input. Moreover, all modifications tend to either increase the need for optimisation or increase cost.

3. What is different about HOT_ARMS PCR?

There are several factors which affect the stability of base pairs in DNA and which can therefore influence the chances of non-specific annealing [144, 145]. As the modifications listed in section 5.A.ii.2 are not satisfactory on their own, we reasoned that the inherent instability of non-specific base-pairing could be enhanced by performing PCR at a high annealing temperature (Ta). The increase in free energy would prevent inappropriate base-pairing and non-specific primer extension. In contrast, perfect base matches would be stable and therefore the specificity of the PCR would be improved. This study shows that by a simple modification of

the primers to increase the Ta >65^{o}C with a 5' tag of complementary nucleotides on both the forward and reverse primer, the specificity of ARMS PCR can be increased. We call this modification Highly Optimised Ta ARMS (HOT_ARMS) PCR.

B: HOT_ARMS PCR - Optimisation and validation of the methodology

i. HOT_PCR: raising the annealing temperature (Ta) through primer modification

We reasoned that issues with ARMS PCR could be counteracted by increasing the primer annealing temperature (Ta). Generally, the higher the Ta the greater the specificity. With HOT_PCR this specificity is achieved with a Ta ≥ 65°C and could be achieved by altering the primer design and adding a 5' tag of nucleotides to the forward and reverse primer.

The primers can be modified in several ways to achieve a high annealing temperature and, in our hands, the best results were obtained by adding a 10 base GC tag onto each primer as shown below in figure 24.

Figure 24. Diagram of HOT_ PCR modification. The primer includes two sections, the tag section which is a universal GC-rich sequence applied to both the forward and reverse primer; and the primer segment which is typically 18-30nt in length and designed to have an annealing temperature of 65°C.

This tag increases the annealing temperature, on average, 6°C. Additionally, the primers can be made longer, and this achieves an annealing temperature of 65°C without the aid of the tag. With the tag incorporated, however, the annealing temperature can be at least 71°C. Incorporation of a tag and modified nucleic acid such as locked nucleic acid (LNA)[121-123] to increase the annealing temperature further is also possible.

Primers tagged with a 17 base GC tag were also tested but the efficiency was reduced compared to the 10 base GC tag primers (data not shown). The primers tagged with a 17 base GC tag were not tested further. Similarly, primers containing 3-6 LNAs to increase the annealing temperature as an alternative to a tag were found to be less efficient than the 10

base GC tag primers (data not shown) and these were therefore not tested further. Each LNA alone increases the annealing temperature by approximately 1-2 °C but there was reduced PCR efficiency. More than 3 LNA's inhibited amplification which agrees with the current literature[122, 123]. Longer primers which generated a similar annealing temperature to HOT_ARMS were found to be very inefficient and generated increased wild-type bleed through.

We reasoned that the implementation of the HOT_ PCR modification in all PCR can also increase selectivity for PCR amplicons over DNA template. We reasoned that the tagged sequence would not be present in the DNA sequence and so the tag would only be partially binding to the DNA. Thus, when an amplicon is formed and the primer is incorporated, amplicon sequence would contain the tag, allowing the primer to fully bind amplicons in cycle 2 onwards. In theory if the primer could then bind the amplicon at a higher temperature than the DNA template, this would cause selective amplification of the amplicons generated in early cycles and reduce the amount of non-specific amplification in PCR. At lower cycles, errors are low as the polymerase has not had chance to make many copies. Moreover, in ARMS PCR the

specificity is higher in the initial cycles as the ARMS primers have had less chance to non-specifically prime wild-type DNA. Similarly, with wild-type blocking, some wild-type amplification occurs throughout the PCR and so in the early cycles there are fewer wild-type copies for the wild-type blocking probe to occupy leading to greater chance of blocking. In addition, increases in specificity and selectivity may result in less frequent failure of PCR; reduced formation of non-specific products; increased multiplexing capability and increased amplification of areas of the genome containing difficult template which have high similarity with other sequences.

The theory was tested by using a 2-phase touch up protocol which includes 10 cycles at a lower annealing temperature (71°C) followed by 40 cycles at a higher annealing temperature (72-80°C). In this case, the second annealing temperature was higher than the maximum annealing temperature that the primers could generate amplification with in a one phase protocol. An example is shown in figure 25 below.

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

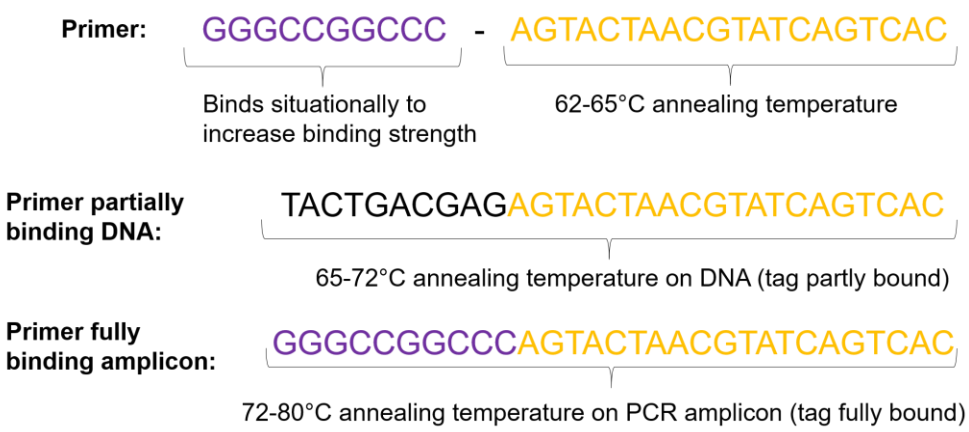


Figure 25. HOT_ PCR method of increasing selectivity of PCR amplicons. The primer which is described in figure 25 shows the GC tag section (purple) on the left and specific primer segment on the right (orange). When the tagged primer binds DNA at 65-72°C depending on the mastermix used, the tag sequence is not present which is represented by the sequence to the left (black). Therefore, there will only be minor binding to G's and C's which will increase the stability of the primer and in turn, increase the annealing temperature by up to 6°C over a primer without a tag. When an amplicon is formed, the whole primer including tag is incorporated and therefore the primer can fully bind at full strength approximately 72-80°C.

The experiment has shown that tagged primers can generate a 2-phase touch-up cycling PCR with phase 1 having a lower maximum potential annealing temperature than phase 2. During phase 2 the raised annealing temperature which is

beyond the annealing temperature that can be achieved in phase 1 will provide preferential amplification of amplicons rather than DNA. A more detailed explanation is as follows: In cycle 1 as amplicons are forming, the maximum annealing temperature may be 71°C as the specific section of the primer is binding fully to the template and the tag is partially binding the template. The tag does not form a natural part of the DNA sequence in the targeted location. In cycles 2 onwards, the annealing temperature may be increased to 72-80°C because the tagged primer is incorporated into the amplicon. A new sequence is generated (the amplicon) which includes the tag. This amplicon sequence can now be bound fully by a tagged primer. Thus, generating a higher annealing temperature than in cycle 1. Since polymerase activity is between 68-80°C, extension can still occur. Cycling can alternate between 95°C and 72-80°C with the annealing temperature varying based on the primer sequence. This also greatly reduces the amount of time spent ramping up and down which in turn speeds up the PCR protocol. Other than speed, no benefit is achieved by step-up HOT_ARMS PCR. One-stage HOT_ARMS PCR at 71°C is equivalent. This shows that selectivity is also occurring in a one-stage protocol. This will occur due to amplicons being primed first at a higher annealing temperature than the DNA

in cycle 2 onwards. The step-up PCR experiment does however show that this occurs.

ii. HOT_ARMS PCR is superior to ARMS PCR

It was important to establish that using the tagged primers with ARMS PCR (hence HOT_ARMS PCR) gave the anticipated benefits over the existing ARMS protocol and to show that HOT_ARMS PCR could achieve our target limit of detection of at least 0.1% MAF.

Primers were designed for 6 different mutations in codons 12/13 of KRAS exon 2 (supplementary table 5 – HOT_ARMS1 (KRAS c.38G>A, G13D), HOT_ARMS2b (KRAS c.35G>T, G12C), HOT_ARMS 4 (KRAS c.35G>A, G12V), HOT_ARMS5 (KRAS c.34G>C, G12R), HOT_ARMS6 (KRAS c.34G>A, G12S) and HOT_ARMS7 (KRAS c.35C>G, G12A). KRAS exon 2 codons 12/13 are mutated in approximately 40-50% of colorectal cancers. Gradient PCR (55-75°C) of both HOT_ARMS (tagged) assays and ARMS (untagged) assays was initially undertaken to determine the annealing temperatures which would form a single melting product. The sequence was the same for each primer set except the addition of a 10-base GC tag. Cell line DNA containing the specific mutations for each assay (supplementary table 6) showed that the

maximum annealing temperature of all 6 HOT_ARMS assays was 71°C and the maximum for ARMS assays was 65°C. Both sets of assays worked at all annealing temperatures down to 55°C.

Firstly, the ability to detect mutations from cell line DNA was tested and compared against the non-specific background amplification of HEK293T cell line DNA. HEK293T is globally wild-type and cell lines containing mutations specific to each assay were used (supplementary table 6). Using the normal rules and nomenclature of real-time PCR, detection was defined by the Ct value. Mutant DNA is expected to produce a lower Ct value through specific priming compared to the non-specific priming of wild-type DNA. For each test, 3 replicates were performed, and the mean Ct value of the replicates was used for analysis. Initially primers tagged with a 10 base GC tag were tested at 65°C and 71°C and compared with untagged primers at 65°C and 55°C for HOT_ARMS1 (KRAS c.38G>A, G13D), HOT_ARMS2b (KRAS c.34G>T, G12C), HOT_ARMS 4 (KRAS c.35G>A, G12D), HOT_ARMS5 (KRAS c.34G>C, G12R), HOT_ARMS6 (KRAS c.34G>A, G12S) and HOT_ARMS7 (KRAS c.35G>C, G12A). The results of this experiment are shown in table 6 below. Data is represented as

a ΔCt as in every instance the mutant DNA amplified at a lower Ct value than the wild-type DNA. ΔCt represents the Ct difference between the mutant DNA (40ng 50% MAF in 10µl reaction volume) and wild-type DNA (40ng in 20µl reaction volume). The larger the ΔCt the less non-specific wild-type amplification occurs.

Table 6. Comparison of tagged versus untagged primers. Tagged primers (HOT_ARMS1, 2b, 4, 5, 6, 7) are amplified at 71°C (maximum one-stage) and 65°C whereas untagged primers are amplified at 65°C (maximum annealing temperature) and 55°C to show the difference in Ct value (ΔCt) between the mutant DNA (40ng 50% MAF in 20µl reaction volume) and wild-type DNA (40ng in 20µl reaction volume).

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

HOT_ARMS assay (KRAS)	Mutation (KRAS)	Tagged ΔCt @71˚C	Tagged ΔCt @65˚C	Untagged ΔCt @65˚C	Untagged ΔCt @55˚C
HOT_ARMS1	C.38G>A	12.395	7.27	8.425	4.05
HOT_ARMS2b	C.35G>T	14.160	12.015	10.175	4.595
HOT_ARMS4	C.35G>A	10.665	6.015	7.035	3.125
HOT_ARMS5	C.34G>C	>24.655	16.58	15.535	7.63
HOT_ARMS6	C.34G>A	13.395	7.695	5.385	2.725
HOT_ARMS7	C.35C>G	16.18	13.34	12.05	4.82

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

The HOT_ARMS primers were compared with the same primer sequences without the 10 base GC tags. The non-tagged primer element was designed to produce a Ta of 65°C which is out of the ordinary. Typically, maximum annealing temperatures are 60°C.

When ARMS PCR (untagged) was performed at a Ta of 55°C, the assays performed very poorly and presented a ∆Ct value between the mutant and wild-type DNA of 2.725 - 7.63. When ARMS PCR (untagged) was performed at a Ta of 65°C, the assays performed better than at 55°C with a ΔCt of 5.385 – 15.535. In each instance, the ΔCt was higher (2.66 - 7.905) compared to 55°C showing that a higher annealing temperature reduces the non-specific priming of wild-type DNA. HOT_ARMS assays performed similarly to ARMS assays at 65°C with ΔCt ranging between -1.155 and 2.31. However, when Ta was increased to 71°C the HOT_ARMS assays always outperformed ARMS assays at 65°C (3.63 - >9.12). HOT_ARMS5 did not amplify over 50 cycles and so it is assumed the ΔCt is >9.12. When comparing HOT_ARMS assays at 71°C to ARMS assays at 55°C the benefit of high annealing temperature is the easiest to recognise. The smallest ΔCt increase was 7.53 whilst the largest was

>17.025. The ΔCt range of ARMS assays at 55°C (2.66 - 7.905) is dwarfed by the ΔCt range of HOT_ARMS assays at 71°C (10.665 – 24.665). As it was determined that HOT_ARMS could provide enhanced ΔCt, the next step was to design more assays to see what the ΔCt was like for multiple genes and multiple sequence changes. An extension to the experiment above is shown below in table 7 which compares amplification of 40ng mutant DNA 50% MAF to 40ng HEK293T wild-type DNA as a ΔCt. This was completed for 7 new assays HOT_ARMS3a (KRAS c.35C>A in reverse, G12V), HOT_ARMS8 (PIK3CA c.1624G>A, E542K), HOT_ARMS9 (PIK3CA c.1633G>A, E545K), HOT_ARMS10 (PIK3CA c.3140A>G, H1047R), HOT_ARMS 11 (APC c.4287_4296delAACCATGCCA, p.Q1429fs*41), HOT_ARMS 12 (BRAF c.1799 A>T in reverse, V600E) and HOT_ARMS13 (EGFR c.2369C>T in reverse, T790M). The mutations listed above are common mutations in cancer, PIK3CA c.1624G>A, (E542K), PIK3CA c.1633G>A (E545K) and PIK3CA c.3140A>G, (H1047R) combined are found in approximately 10%-15% of CRC[146]. BRAF c.1799 A>T (V600E) is found in approximately 10-15% of CRC[146]. EGFR c.2369C>T (T790M) was tested as the sequence change differed to other HOT_ARMS assays and is found in approximately 60% of non-small cell lung cancers[146]. An

APC deletion found in the common CRC cell line SW1116 was also tested to see if ARMS PCR could be extended to detect deletions as well as single nucleotide variants. In turn, this would broaden the pool of available sequence variants for liquid biopsy testing. All thirteen HOT_ARMS assays were included to summarise the data.

Table 7. Comparison of ΔCt values to demonstrate HOT_ARMS performance. HOT_ARMS primers were test at 71°C annealing temperature to determine the ΔCt between the mutant DNA (40ng 50% MAF in 20µl reaction volume) and wild-type DNA (40ng in 20µl reaction volume) for HOT_ARMS1-13. The direction which gave the best ΔCt is shown for the sequence change.

Ref	**Direction**	**Protein sequence/common name**	**Sequence change**	**40ng total DNA – ΔCt 50% MAF vs wild-type**
HOT_ARMS1	Forward	KRAS G13D	c.38G>A	12.395
HOT_ARMS2a	Reverse	KRAS G12C	c.34C>A (in reverse)	>24.955
HOT_ARMS3a	Reverse	KRAS G12V	c.35C>A (in reverse)	>24.385
HOT_ARMS4	Forward	KRAS G12D	c.35G>A	10.665
HOT_ARMS5	Reverse	KRAS G12R	c.34C>G (in reverse)	>24.655
HOT_ARMS6	Forward	KRAS G12S	c.34G>A	13.395
HOT_ARMS7	Reverse	KRAS G12A	c.35C>G (in reverse)	16.18
HOT_ARMS8	Forward	PIK3CA E542K	c.1624G>A	16.15
HOT_ARMS9	Forward	PIK3CA E545K	c.1633G>A	11.825
HOT_ARMS10	Forward	PIK3CA H1047R	c.3140A>G	12.315

HOT_ARMS11	Forward	APC p.Q1429fs*41	c.4287_4296delAACCATGCCA	14.67
HOT_ARMS12	Reverse	BRAF V600E	c.1799A>T (in reverse)	>22.87
HOT_ARMS13	Reverse	EGFR T790M	c.2369C>T (in reverse)	12.37

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

Tests were performed at least in triplicate and all 13 primers showed specific amplification of the mutant sequence (containing either 50% or 100% mutant alleles). The mean Ct value was around 26 cycles and so efficiency was not altered unlike with the IntPlex data shown in section 4.B.iv. It is expected that wild-type bleed through will occur even with mutation-specific primers [72, 78] and non-specific amplification of HEK293T DNA was varied. Every HOT_ARMS primer pair showed a difference in Ct (∆Ct value) between the mutant DNA and wild type of at least 10.665 cycles (HOT_ARMS4, KRAS c.35G>A, G12D) and a maximum of >24.955. Four assays, HOT_ARMS2a (KRAS c.34G>T, G12C), HOT_ARMS3a (KRAS c.35G>T, G12V), HOT_ARMS5 (KRAS c.34G>C, G12R) and HOT_ARMS12 (BRAF c.1799T>A, V600E) gave no wild-type amplification over 50 cycles. Thus, the ΔCt is given a more than sign (>). It was noted that assays producing lower ΔCt were typically G>A or A>G variants whilst assays producing higher ΔCt were G>T, G>C or T>A. HOT_ARMS primers were tested in both forward and reverse format. The direction which produced the least bleed-through was chosen as the direction of choice for the final assay.

iii. HOT_ARMS PCR has a very low limit of detection

It was important to establish the limit of detection as this was critical for detecting mutations in liquid biopsies. DNA from cell lines containing known mutations was spiked into HEK293T DNA (Supplementary Table 6) to produce mutant allele frequencies (MAF) ranging from 100% in the case of homozygous cell lines or 50% in the case of heterozygous cell lines to 0.06% MAF. This was prepared using 40ng/µl DNA giving a total mutant copy number of 6666 (homozygous) or 3333 (heterozygous) down to 4 copies per µl. One microlitre was added per 20µl reaction. At the same time, we decided that circulating tumour DNA produces low yield; hence, performance may be enhanced by performing PCR in replicates with lower total DNA per 20µl reaction or by screening for multiple targets with lower total DNA per 20µl reaction. A set of spike-in controls was generated with 5ng/µl DNA ranging from 100% MAF in the case of homozygous cell lines or 50% in the case of heterozygous cell lines. The total mutant copy number was 833 mutant copies (100% MAF homozygous), 417 mutant copies (50% MAF heterozygous) down to 4 mutant copies (0.5% MAF) per µl. Four mutant copy detection was chosen as it should, despite the Poisson distribution, result in 1 or 2 mutant alleles present in every

reaction. This gave an indication of how easy it was for the assay to detect extremely low mutant copy number between 40ng total input (20µl reaction) and 5ng total DNA input (20µl reaction).

Limit of detection tests were performed in duplicate using the 10 base GC tag only and a test was called positive if there was a ΔCt value between test sample and HEK293T DNA of ≥ 2 over 3 replicates. Optimisation of the primer concentration showed that a concentration of 375nM per reaction gave the best limit of detection. 375nM was therefore chosen as the final primer concentration. The limit of detection for all 13 HOT_ARMS assays are shown in table 8 below along with ΔCt values between 4 mutant copies (0.06% MAF) in a background of 40ng total DNA (20µl reaction) and 40ng HEK293T wild-type DNA (20µl reaction). This is compared with ΔCt of 4 mutant copies (0.5% MAF) in a background of 5ng total DNA (20µl reaction) and 5ng HEK293T wild-type DNA (10µl reaction).

Table 8. Comparison of HOT_ARMS assay sensitivity. The sensitivity of each assay was tested and ΔCt values for 4 mutant copies in a background of 40ng total DNA (0.06%

MAF) or 5ng total DNA 0.5% MAF) compared to wild-type DNA

(40ng or 5ng respectively).

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

Ref	**Direction**	**Protein sequence/common name**	**Sequence change**	**40ng total DNA - ΔCt 4 mutant copies (0.06%) vs wild-type**	**5ng total DNA - ΔCt 4 mutant copies (0.5%) vs wild-type**	**Sensitivity achieved**
HOT_ARMS1	Forward	KRAS G13D	c.38G>A	2.64	5.56	0.06%
HOT_ARMS2a	Reverse	KRAS G12C	c.34C>A	7.6075	10.065	0.06%
HOT_ARMS3a	Reverse	KRAS G12V	c.35C>A	11.97	>13.12	0.06%
HOT_ARMS4	Forward	KRAS G12D	c.35G>A	1.66	5.51	0.1%
HOT_ARMS5	Reverse	KRAS G12R	c.34C>G	7.45	8.70	0.06%
HOT_ARMS6	Forward	KRAS G12S	c.34G>A	1.54	2.35	0.1%
HOT_ARMS7	Reverse	KRAS G12A	c.35C>G	4.10	9.46	0.06%
HOT_ARMS8	Forward	PIK3CA E542K	c.1624G>A	5.45	9.28	0.06%
HOT_ARMS9	Forward	PIK3CA E545K	c.1633G>A	1.6	1.97	0.1%

HOT_ARMS10	Forward	PIK3CA H1047R	c.3140A>G	1.23	3.97	0.1%
HOT_ARMS11	Forward	APC p.Q1429fs*41	c.4287_4296delAACCATGCCA	3.16	4.85	0.06%
HOT_ARMS12	Reverse	BRAF V600E	c.1799A>T	6.58	11.41	0.06%
HOT_ARMS13	Reverse	EGFR T790M	c.2369C>T	1.26	2.11	0.1%

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

The limit of detection for HOT_ARMS was variable ranging from 0.1% MAF for HOT_ARMS4 (KRAS c.35G>A, G12D), HOT_ARMS6 (KRAS c.34G>A, G12S), HOT_ARMS9 (PIK3CA c.1633G>A, E545K), HOT_ARMS10 (PIK3CA c.3140A>G, H1047R) and HOT_ARMS13 (EGFR c.2369 C>T, T790M) down to 0.06% MAF for HOT_ARMS1 (KRAS c.38G>A, G13D), HOT_ARMS2a (KRAS c.34G>A, G12C), HOT_ARMS3a (KRAS c.35G>T, G12V), HOT_ARMS5 (KRAS c.34G>C, G12R), HOT_ARMS7 (KRASc.35G>C, G12A), HOT_ARMS8 (PIK3CA c.1624G>A, E542K), HOT_ARMS11 (APC c.4287_4296delAACCATGCCA, APC p.Q1429fs*41) and HOT_ARMS12 (BRAF c.1799T>A, V600E). HOT_ARMS 4, 6, 8, 9, 10 and 13 which detect G>A, A>G, C>T or T>C SNVs performed more poorly in comparison to the other assays which detected 0.06% MAF. All assays could detect 0.1% MAF which means HOT_ARMS can be used for liquid biopsy testing. Moreover, detection of 4 mutant copies was easier with 5ng total DNA input compared with 40ng and is more realistic for the majority of ccfDNA extracted from liquid biopsies. Thus, it may be beneficial to have several reactions with lower input and would likely occur in practice with replicates. The increase in ΔCt between 4 copies and wild-type ranged from 0.37 to 5.36. This was particularly important for some assays which

present more bleed through such as HOT_ARMS1 (ΔCt 2.64) and HOT_ARMS4 (ΔCt 1.66) which had increases in ΔCt by 2.92 (ΔCt 5.56) and 3.85 (ΔCt 5.51) respectively, easing the detection of low mutant copy number.

iv. HOT_ARMS3a can detect 1 mutant copy

It was important to ascertain how sensitive HOT_ARMS could be. Liquid biopsy samples are now known to contain as little as 1 mutant copy. The following experiment aimed to establish if this was possible by HOT_ARMS. Section 2.F.iii.3 fully describes the experiment for 1-copy detection which was originally performed by the SNPase-ARMS qPCR team[72]. A slightly modified version of this experiment was performed on HOT_ARMS3a (KRAS c.35 G>A, G12V) using in a continuous total of 10,000 copies in a 10µl reaction by log scale (10^4(n=4), 10^3(n=4), 10^2(n=4), 10(n=4), 3(n=24), 1(n=24) and 0 copies (n=24)). The MAF was 100%, 10%, 1%, 0.1%, 0.03%, 0.01% and 0% respectively.

Three copies amplified in 100% of reactions (n=24) and 1 expected copy amplified in 70% of reactions (n=24). This is expected as DNA template would follow Poisson distribution statistics when dispensed at such a low concentration. It would be expected, by chance, that 1 or more mutant copies

would be dispensed into 70% of the reactions whilst 30% of the reactions would have no mutant copies. HOT_ARMS was able to replicate these statistics - a feat obtained by few mutation detection systems. HOT_ARMS3a (KRAS c.35 G>A, G12V) was chosen as it had a large Δct. However, it can be expected that many of the other assays will be able to detect 1 copy as the variation in Ct-value was minimal and like detection of 4 mutant copies as shown below in figure 26. To the far right of the figure describes amplification of ≥1 copy, this includes amplification of both 1 copy and 3 copies, proving that the Ct values are similar.

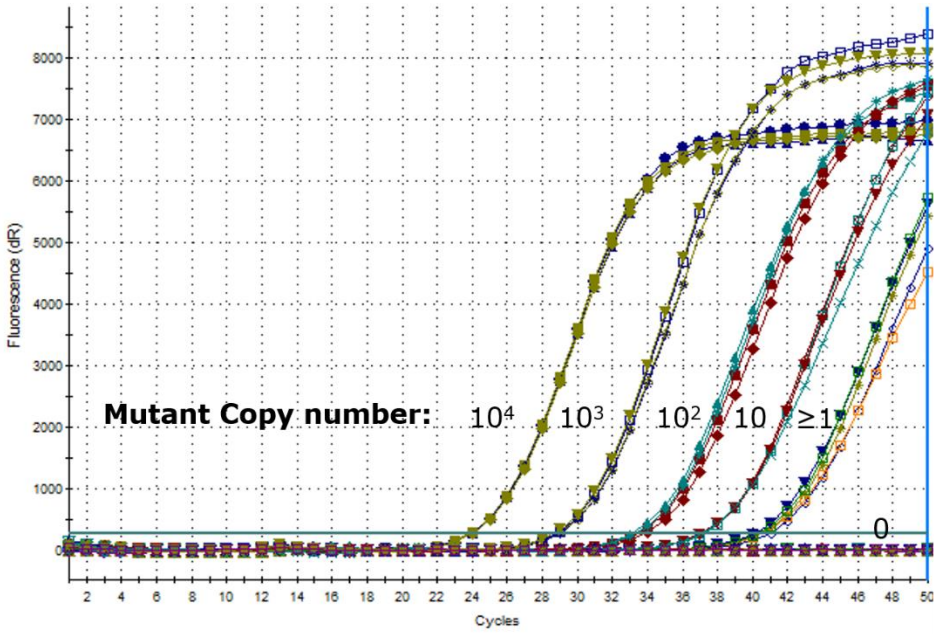


Figure 26. Dynamic range and single copy detection by HOT_ARMS3a. HOT_ARMS3a assay (c.35 G>A, G12V) was

tested to see if it could detect 1 mutant copy by using the SNPase ARMS qPCR experiment set up for validation[72]. A continuous total of 10,000 copies in a 10µl reaction was added and mutant copies were altered by log scale (10^4(n=4), 10^3(n=4), 10^2(n=4), 10(n=4), 3(n=24), 1(n=24) and 0 copies (n=24)) which is shown on the amplification plot to the left-hand side of the amplification replicate cluster. The MAF was 100%, 10%, 1%, 0.1%, 0.03%, 0.01% and 0% respectively.

Whilst 1 copy detection was successful, this experiment also showed the ability of HOT_ARMS to be potentially quantitative. Amplification plots were very tight for each set copy number and so the next experiment set to prove that HOT_ARMS had this capability.

v. HOT_ARMS PCR has a wide dynamic range and excellent precision

To test the robustness of the assay, the short-term precision (also known as the intra-assay variability) was evaluated, which is best evaluated through measuring the coefficient of variation.

Eight replicates were performed for the HOT_ARMS1 (KRAS c.38G>A, G13D) assay of spiked-in template (40ng total

HCT116 DNA in a 20µl reaction) containing MAFs ranging from 50% down to 0.09% (table 9). For 50%MAF to 1.6%MAF, there was excellent short-term precision with a coefficient of variation <1% across all templates and a maximum Ct range (i.e., highest value Ct – lowest value Ct) of 0.72. At the more dilute templates with 0.78%MAF and 0.09%MAF there was slight loss of precision. The respective %CVs were 1.02 and 1.36 and Ct ranges were 1.10 and 1.38. The loss of precision is due to Poisson distribution and occurs in all PCR based tests. Poisson distribution is a random distribution that occurs when pipetting template into PCR wells which is more noticeable at low copy number. In one reaction you may pipette 1 copy and in another you may pipette 2 or 3 due to the random distribution of the copies in the liquid you are pipetting from. This results in subtle Ct value changes when amplifying low mutant allele frequency samples which is unavoidable and presents as a loss in precision.

Table 9. HOT_ARMS PCR short term precision. Eight replicates were performed for the HOT_ARMS1 (KRAS c.38G>A, G13D) assay of spiked-in template (40ng total DNA in a 20µl reaction) containing MAFs ranging from 50% down to 0.09%. The minimum Ct value, maximum Ct value, mean Ct value

and range of Ct values for the 8 replicates for the respective

mutant allele frequency is shown.

MAF (%)	**50.00**	**25.00**	**12.50**	**6.25**	**3.13**	**1.56**	**0.78**	**0.09**
Total mutant copies per reaction	**3333**	**1666**	**833**	**416**	**208**	**104**	**52**	**6**
Mean Ct	26.26	27.45	28.72	29.84	31.03	31.85	33.37	35.79
Min Ct	26.18	27.04	28.47	29.5	30.77	31.52	32.79	35.02
Max Ct	26.42	27.68	28.99	30.16	31.24	32.24	33.89	36.4
Range of Ct	0.24	0.64	0.52	0.66	0.47	0.72	1.1	1.38
CV%	0.30	0.74	0.58	0.76	0.51	0.92	1.02	1.36

The templates had been diluted twofold and so the mean Ct value for each MAF was plotted against the log2 (1/MAF) to enable the dynamic range to be assessed. Over this range, the slope of the curve was 1.07 (Figure 27) demonstrating a PCR efficiency of 105% with an intercept at Ct = 25.43. Thus, the PCR remained efficient over a large range of MAF and, with this level of efficiency, HOT_ARMS PCR could probably be used for mutant allele quantification.

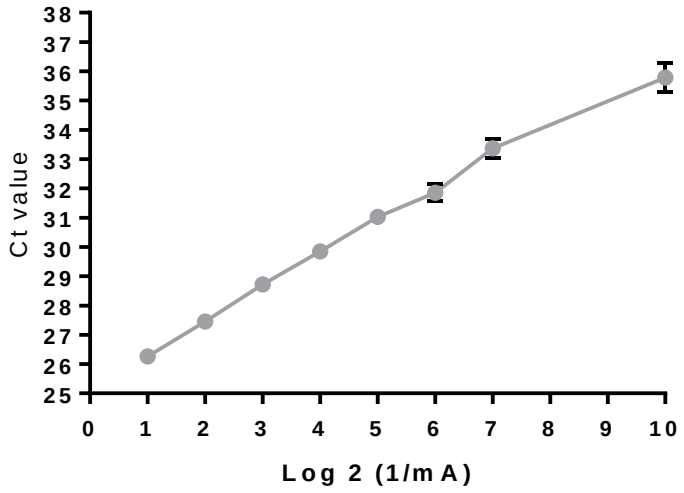


Figure 27. HOT_ARMS1 (KRAS c.38G>A, G13D) assay linear dynamic range and efficiency plotted from data in table 9. The templates in table 9 had been diluted twofold and so the mean Ct value for each MAF was plotted against the log2 (1/MAF).

vi. Further modifications to increase HOT_ARMS ΔCt between 4 mutant copies and wild-type

The data in this study had shown that, of the different strategies tested, the lone optimal primer modification was a 10 base GC tag at the 5' end of both primers. However, it was important to continue development of the assay to increase ΔCt for weaker assays such as HOT_ARMS 1, 4, 6, 8, 9, 10 and 13 which detect G>A, A>G, C>T or T>C SNVs. This would allow the assays to translate more readily onto other machines, increasing the overall robustness of the assays. A list of 36 modifications tested during the study is found in supplementary table 7. Only three modifications which provided benefit are discussed here.

From first principles, the transition mutations i.e., G>A, A>G, C>T and T>C are more likely to bleed through as the mismatch will have less free energy and therefore is more likely to anneal and allow the polymerase to work. This was reflected in our data as the assays for the SNVs which were not included in this group, did not require improvement. Thus, there are 12 possible types of SNV that can occur meaning that HOT_ARMS alone was sufficient for 8 out of 12 and that

we felt further modifications may be beneficial for the remaining 4 out of 12 to increase robustness.

To overcome the problems with the transition mutations, HOT_ARMS was combined with additional modifications. Unlike other modifications, HOT_ARMS is novel in that it did not cause disruption to the PCR efficiency, allowing its combination with other modifications to reduce wild-type bleed through further for enhanced low mutant copy detection.

HOT_ARMS was combined with 3’ locked nucleic acids (LNA)[121] to improve the primer specificity. Inclusion of LNA’s, that are modified nucleic analogues that replace DNA bases in the primer, are known to improve binding specificity[122, 123], especially on the 3’ base of an ARMS primer[121]. The 3’ LNA can be effective at reducing bleed through but in this study, it was found to be extremely target dependent. With some targets, the amplification of mutant DNA would be completely abolished. In other instances, inclusion of 3’ LNA was able to increase ΔCt between 0.06% and wild-type by up to 4 cycles for HOT_ARMS1 (KRAS c38G>A, G13D) and HOT_ARMS15 (ESR1 c.1613 A>G, D538G). It was concluded that 3’ LNA’s should not be

integrated universally as they would increase optimisation and cost. However, 3' LNA could still be used as a reserve modification in instances where a specific single mutation needs to be detected if the sequence permits. Examples include liquid biopsy companion diagnostics with small molecular drug inhibitors where resistance mutations may develop, causing the pool of nucleotides to contain more than one mutation on the same base.

Adding a mismatch at the 3' penultimate base is widely known to be able to reduce wild-type bleed-through with ARMS[126, 127]. During this study, 3' penultimate mismatches were investigated briefly to see if integration would reduce bleed-through for HOT_ARMS. However, the mismatches were very target-dependent in their effect. Mismatches which can also be referred to as 'wobbles' caused the primer annealing temperature to drop approximately 5-6°C. In some instances, amplification efficiency was reduced which did not allow the amplification of low mutant allele frequencies, reducing the limit of detection. In other instances, the limit of detection was reduced or unchanged compared with regular HOT_ARMS. Occasionally though, integration of a mismatch would improve the ΔCt between 0.06% and wild-type by up to 6 cycles for

some assays. As we had empirically verified the fact that transition mutation (G>A, A>G, C>T and T>C) mismatches are comparatively stable, we began to apply these principles to penultimate base mismatches for a more subtle HOT_ARMS primer design. This reduced the need for optimisation to find the best mismatch and increased the likelihood of the mismatch reducing bleed-through compared with HOT_ARMS. Future plans are described in section 7.D which contains more details regarding wobbles and our plan to integrate them with HOT_ARMS for weaker assays to improve ΔCt.

Wild-type blockers have been used to prevent mis-priming of wild-type DNA by the ARMS primer as shown by the IntPlex[85] method in section 4.B.iv. Wild-type blockers are probes which bind specifically to wild-type DNA on the same strand as the ARMS primer at a higher temperature than the primers, thereby blocking access of the ARMS primer to wild-type template. The integration of 3’ phosphate prevents polymerase extension so that the blocker does not interfere with amplification signals.

Wild-type blockers were successfully implemented for KRAS exon 2 c12/13 HOT_ARMS assays 1-7 and HOT_ARMS15 (ESR1 c.1613 A>G, D538G). The blocker would only provide a

reduction in wild-type amplification with at least 6 LNA's included and the PCR efficiency of mutant DNA amplification was always affected. The blocker is described in supplementary table 5. Below in table 10 shows the results for HOT_ARMS4 (KRAS c.35G>A, G12D) and HOT_ARMS6 (KRAS c.34G>A, G12S). 10 LNA's are included in the blocker as it gave the greatest reduction in wild-type bleed through. As the previous limit of detection was 0.1% MAF for both assays as shown in section 5.B.iii, 40ng of 50% MAF and 40ng of 0.1% MAF spike-in control was compared against 40ng of HEK293T wild-type DNA with different concentrations of 10 LNA blocker in a 20µl reaction. Cell line DNA for the specific mutation required for spike in controls is found in supplementary table 6.

Table 10. Demonstration of wild type blocking probe addition to HOT_ARMS PCR to further enhance ΔCt values.

HOT_ARMS4 (KRAS c.35G>A, G12D) and HOT_ARMS6 (KRAS c.34G>A, G12S) are tested with different concentrations of wild-type blocking probe. Results are reported as ΔCt between 0.1% MAF and wild-type and 50% MAF and wild-type. Ct values are also given for 50% MAF, 0.1% MAF and wild-type. Other MAF are not included as the key difference was to see if

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

ΔCt could be improved at low MAF (0.1%) compared with the regular HOT_ARMS assay without blocker (0nM). 50% MAF is included to see what the overall effect on PCR efficiency is.

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

Target	**Wild-type Blocker concentration**	**ΔCt 50 % MAF & wild-type**	**ΔCt 0.1% MAF & wildtype**	**Ct:50% MAF**	**Ct: 0.1% MAF**	**Ct: wild-type**
HOT_ARMS 4 c.35G>A (G12D)	0nM	10.665	1.405	26.435	35.695	37.100
	30nM	14.436	5.311	31.10	40.225	45.536
HOT_ARMS 6 c.34G>A (G12S)	0nM	13.395	2.855	25.980	36.520	39.375
	20nM	19.62	8.45	30.38	41.55	48.874

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

Both HOT_ARMS4 (KRAS c.35G>A, G12D) and HOT_ARMS6 (KRAS c.34G>A, G12S) contain G>A SNV's which typically present more wild-type amplification. The 10 LNA wild-type blocker increased ΔCt values between 0.1% and wild-type from 1.405 to 5.311 for HOT_ARMS4 and 2.855 to 8.45 for HOT_ARMS6. Although it must be noted that there was an increase in Ct-value for the mutant DNA of approximately 5 cycles, but this did not affect the amplification of low mutant allele frequency. HOT_ARMS combined with a wild-type blocker containing 10 LNA's only required a concentration of 20-30nM which was drastically less compared to the 3600nM required by the IntPlex system[85]. The specificity of HOT_ARMS and LNA's in a wild-type blocker combined means that a lower blocker concentration is required to have the desired effect, reducing the overall complexity of the reaction. Moreover, the primer concentration (375nM) was unchanged compared to 900nM used in IntPlex. However, whilst this modification is useful, it does increase cost and optimisation. Therefore, this modification will be used as a reserve modification in instances where there are issues with bleed-through and a particular mutation needs to be detected at a single copy level for companion diagnostics.

All three modifications, 3' LNA, wild-type blockers and mismatches typically reduced PCR efficiency but still allowed the interpretation of the assay to be a dichotomous "yes/no" depending on the presence/absence of amplification but the downside was that HOT_ARMS became semi-quantitative. However, it may be necessary to reduce the bleed-through as much as possible to maximise the reliability of the liquid biopsy test as it is translated to other machines. It is important to note that modifications will only be required when particular mutations are required to be detected such as in companion diagnostics or generic hotspot testing. If a personalised approach is undertaken, then these assays may not need to be developed since mutations which are easier to detect can be selected from sequencing the primary tumour as discussed in greater detail in section 7.C.iii.

vii. HOT_ARMS PCR can use a rapid cycling program effectively

A rapid liquid biopsy test for out-patient departments or GP clinics could be transformative for cancer patient management as it would reduce the number of required visits to the oncologist. This experiment set out to see if the time taken for HOT_ARMS to cycle could be reduced whilst maintaining

enough sensitivity for ctDNA detection. HOT_ARMS12 (BRAF c.1799 A>T, V600E) was combined with Qiagen rapid cycling mastermix and a thermocycler called the magnetic induction cycler (MIC) which was capable of rapid data collection, melting analysis and ramping. The MIC uses magnetic induction and a centrifuge style set up for tubes to speed up ramping and capture fluorescent data from multiple reactions faster. Positive spike-in controls were generated by mixing 40ng/µl HT29 cell line DNA (BRAF c.1799 T>A, V600E) with 40ng/µl HEK293T wild-type DNA to generate reactions containing MAF of 50% (n=1), 25% (n=1), 12.5% (n=1), 6.25% (n=10, 1% (n=2), 0.06% (n=5, 4 mutant copy) and 0% (n=5). 1µl of control was added to a 20µl reaction. Figure 28 below represents the amplification plot for HOT_ARMS12 (BRAF c.1799T>A, V600E) using a 30-minute PCR program

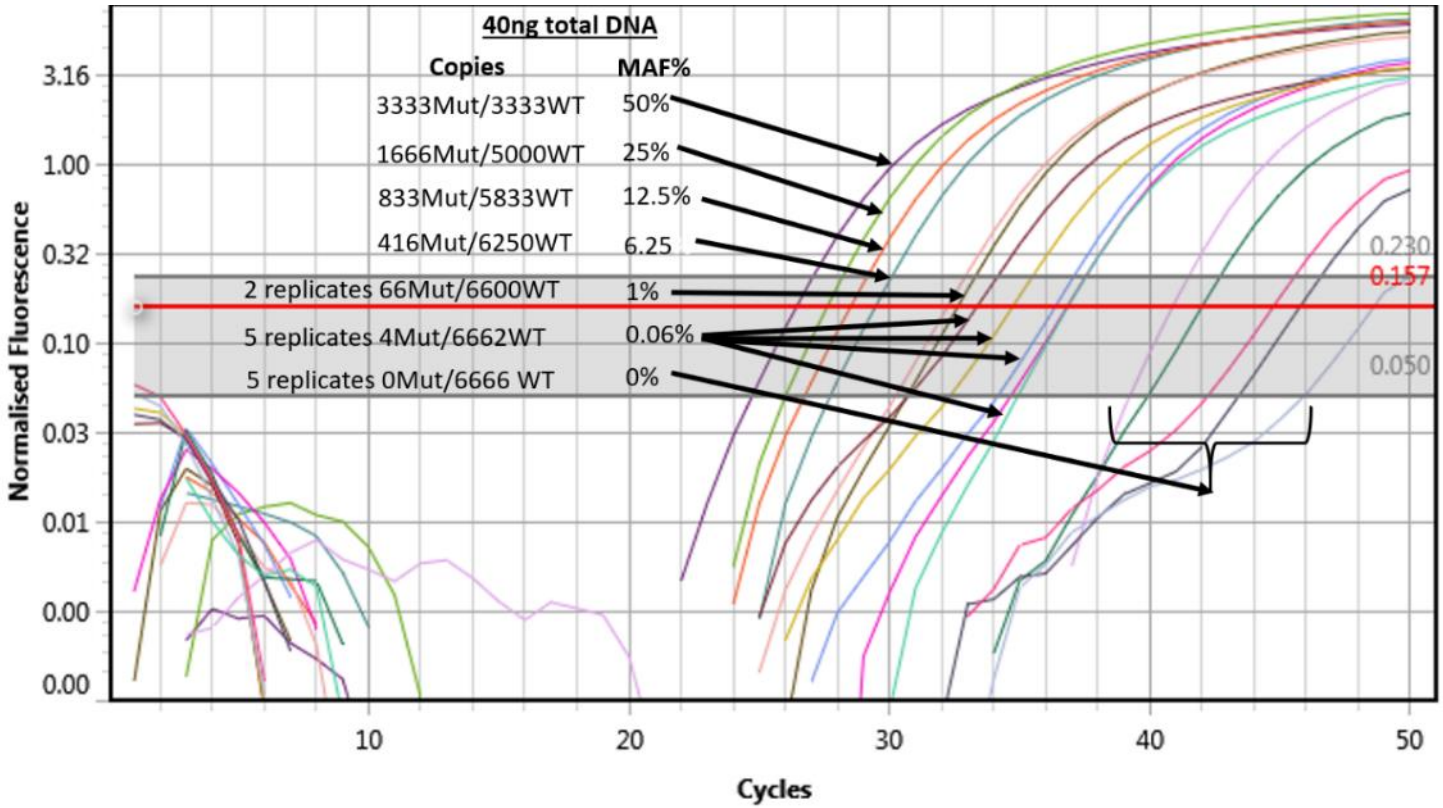


Figure 28. Amplification plot for HOT_ARMS12 (BRAF c.1799T>A, V600E) rapid testing. Positive spike-in controls with MAF of 50% (n=1), 25% (n=1), 12.5% (n=1), 6.25% (n=10, 1% (n=2), 0.06% (n=5, 4 mutant copy) and 0% (n=5) are shown by arrow on the amplification plot.

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

HOT_ARMS12 (c.1799T>A, V600E) was able to detect 0.06%MAF with a minimum ΔCt of 4 between 0.06% and wild-type. Moreover, the wild-type DNA reaction replicates were very random and spanned between a Ct-value of 40.45 and 48.27 whilst 0.06% MAF replicates were far tighter, ranging between 33.53 and 36.45. Rapid testing is made possible by the extremely high specificity HOT_ARMS primers achieve. By inference, HOT_ARMS assays with ΔCt >15 (40ng total DNA) between 50% MAF and wild-type can undergo rapid 30-minute testing without additional modifications. Since there is an increase in wild-type bleed through though, HOT_ARMS assays with lower specificity; ΔCt <15 between 50% MAF and wild-type will obtain lower sensitivities. HOT_ARMS1 (ΔCt 12.395) achieved 0.5% MAF detection (33 copies, 40ng DNA per reaction) with a rapid cycling program. However, with the addition of 3’ LNA the specificity was increased to 0.2% MAF. The incorporation of wobbles with a fast-cycling program still needs to be investigated as this may give a more universal solution to make HOT_ARMS assays with ΔCt <15 utilise a rapid cycling program.

viii. HOT_ARMS can work effectively on a variety of template with low DNA input

After establishing and validating the HOT_ARMS protocol using cell line DNA; it was important to confirm the assays dynamic range for cfDNA and FFPE DNA before moving on to mutation detection with these sample types which is presented in the next chapter. Both types of template are extremely fragmented and have additional PCR inhibitors present when compared with cell line DNA. Moreover, FFPE DNA is of poor sequence quality and cfDNA generally produces low yield. To stress test the assays accordingly, cfDNA, FFPE DNA and cell line DNA were diluted to 2ng/µl, 1ng/µl, 250pg/µl, 125pg/µl and 62.5pg/µl. 1µl of DNA was added to a 15µl total reaction volume. The DNA was amplified using the 10nt GC-tagged KRAS exon 2 wild-type primers used in the HOT_PI protocol (supplementary table 5). These primers span codons 12/13 and form a 92bp amplicon. As shown below in table 11, all dilutions, including 62.5pg (approx. 10 genome copies) gave a Ct value below 35. Furthermore, the melting peaks in figure 29 below confirm the presence of the amplicon with a single melting peak. This confirms that the modification has no effect on the ability to amplify low DNA input from cfDNA and FFPE DNA.

Table 11. Amplification of cfDNA, FFPE and cell line DNA with wild-type modified primers. FFPE DNA is of poor sequence quality and cfDNA generally produces low yield. To stress test the assays accordingly, cfDNA, FFPE DNA and cell line DNA were diluted to 2ng/µl, 1ng/µl, 250pg/µl, 125pg/µl and 62.5pg/µl. Replicate Ct values are shown in the table below. The FFPE samples and cfDNA samples are from the pre- and post-surgery sample set reported in 4.B.i.

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

Sample type	2ng total DNA Ct	1ng total DNA Ct	250pg total DNA Ct	250pg total DNA Ct	125pg total DNA Ct	125pg total DNA Ct	62.5pg total DNA Ct	62.5pg total DNA Ct
FFPE DNA sample 18	31.06	30.66	31.62	31.62	32.47	32.35	32.00	32.50
FFPE DNA sample 19	32.10	31.73	32.88	34.09	32.09	32.48	32.86	33.83
FFPE DNA sample 20	31.51	31.30	32.34	32.29	32.69	33.22	33.44	33.37
cfDNA sample 18	31.94	32.40	32.39	33.02	33.12	33.66	33.36	32.53
cfDNA sample 19	32.26	31.67	31.59	31.10	31.56	31.76	31.77	31.78
cfDNA sample 20	30.33	30.58	30.96	30.77	31.55	30.42	32.81	32.10
HEK293T cell line DNA	29.67	30.22	30.44	30.62	31.79	30.54	31.70	34.22

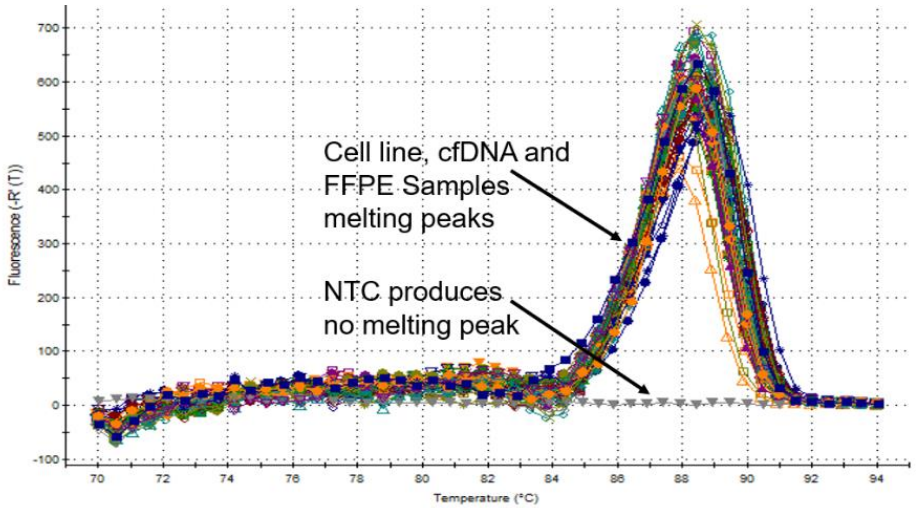


Figure 29. Clean and specific melting peaks for low DNA input amplification of cfDNA, FFPE DNA and cell line DNA with flat no template control (NTC). All replicates from table 11 are shown.

ix. Validation of HOT_ARMS with Qiagen and LifeArc

Over the course of 2018 and 2019, several companies became interested in HOT_ARMS PCR. In particular, Qiagen and LifeArc opened negotiations with the University of Nottingham Tech transfer office which instigated a series of experiments to prove the validity of HOT_ARMS as an ultrasensitive method for liquid biopsy.

Qiagen wanted us to stress test HOT_ARMS2a (c.34G>T, G12C) on liquid biopsy samples which were particularly low in copy number in their hands by Therascreen[147] testing.

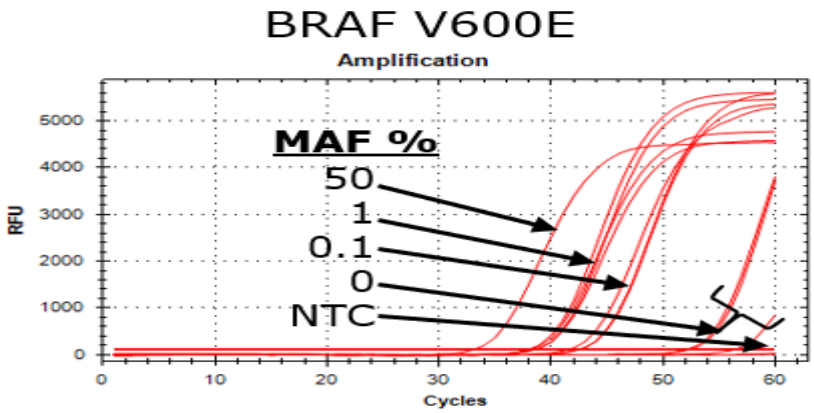
5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

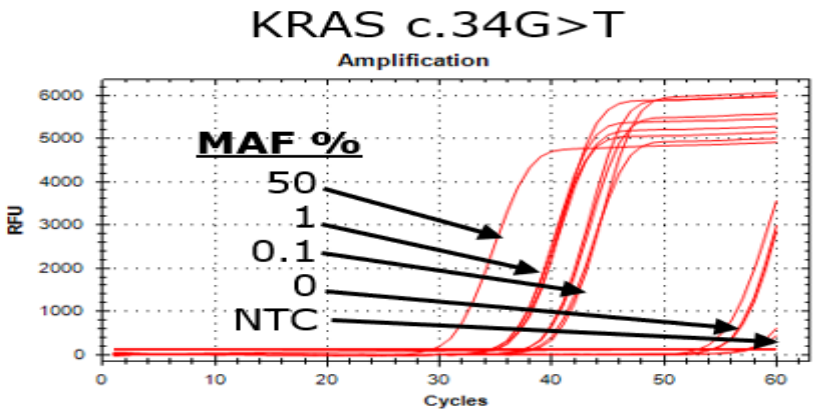
LifeArc requested a demonstration of HOT_ARMS2a (KRAS c.34G>T, G12C), HOT_ARMS3a (KRAS c.35G>T, G12V) and HOT_ARMS12 (BRAF c.1799T>A, V600E). LifeArc also requested the development of two new assays targeted against the ESR1 mutations E380Q (c.1138 G>C) and D538G (c.1613 A>G) to show that HOT_ARMS could work on different targets with quick turnaround in development.

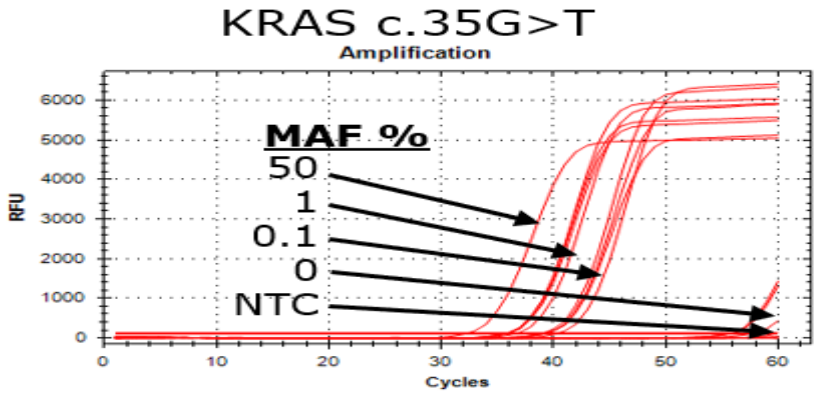
The LifeArc demonstration took place at their lab in Edinburgh on a Bio-Rad CFX96 PCR machine which HOT_ARMS assays had not previously been tested on. Therefore HOT_ARMS was stress-tested in a new lab with a different PCR machine, although the operator (James Hassall) remained the same. LifeArc provided ESR1 standards which mimicked cfDNA characteristics. The standards were gBlocks which are double-stranded DNA gene fragments manufactured by IDT (https://eu.idtdna.com/pages/products/genes-and-gene-fragments/double-stranded-dna-fragments/gblocks-gene-fragments) to have a fragment size of 170bp and contain the mutation of choice. LifeArc previously quantified the gBlocks by ddPCR to determine copy number. The ESR1 gBlocks, both c.1138 G>C (E380Q) and c.1613 A>G (D538G) were mixed with wild-type gBlock to form spike in controls of 50%, 1%,

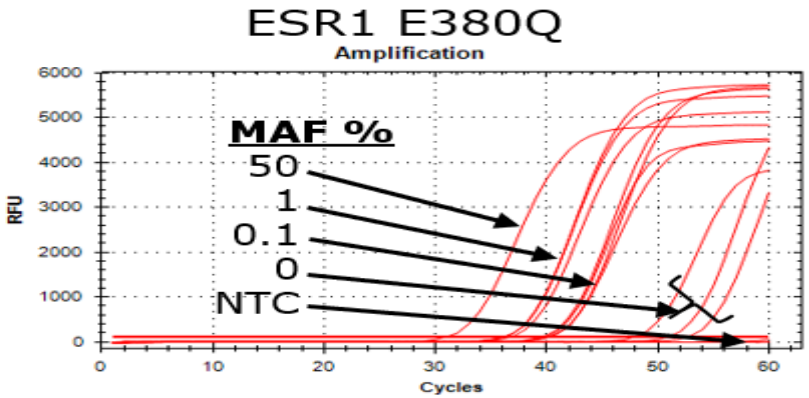
0.1% and 0%. 10,000 copies of each standard were added to separate reactions in triplicate. In addition, for ESR1 c.1613A>G (D538G), D538Y standards were tested in triplicate to determine the assays cross-reactivity with other mutations on the same nucleotide position. HOT_ARMS14 (ESR1, c.1138 G>C, E380Q) was tested as regular HOT_ARMS but HOT_ARMS15 (ESR1, c.1613 A>G, D538G) had the addition of a wild-type blocker (with 6 LNA included targeted to same strand as HOT_ARMS primer) and 3’ LNA on the HOT_ARMS primer to reduce bleed-through further as the single nucleotide variant was A>G. At the same time, HOT_ARMS2a (KRAS c.34G>T, G12C), HOT_ARMS3a (KRAS c.35G>T, G12V) and HOT_ARMS12 (BRAF c.1799T>A, V600E) were tested in a similar fashion by triplicate analysis of 50% MAF, 1% MAF, 0.1% MAF and 0% MAF. These standards were generated by mixing 60ng mutation specific cell line DNA (supplementary table 6) with HEK293T wild-type DNA. Below, figure 30 shows the amplification plots for the two new ESR1 HOT_ARMS assays and the existing HOT_ARMS assays.

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy









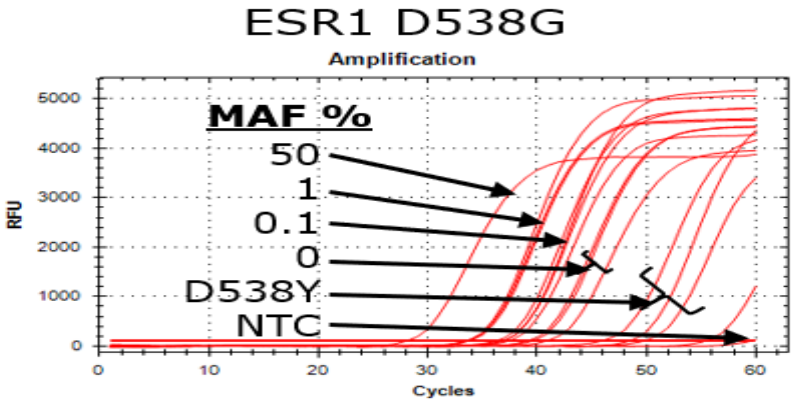


Figure 30. Amplification plots for HOT_ARMS assays tested at LifeArc. HOT_ARMS14 (ESR1, c.1138 G>C, E380Q) was tested as regular HOT_ARMS but HOT_ARMS15 (ESR1, c.1613 A>G, D538G) had the addition of a wild-type blocker (with 6 LNA included targeted to same strand as HOT_ARMS primer) and 3' LNA on the HOT_ARMS primer to reduce bleed-through further as the single nucleotide variant was A>G. HOT_ARMS2a (KRAS c.34G>T, G12C), HOT_ARMS3a (KRAS c.35G>T, G12V) and HOT_ARMS12 (BRAF c.1799T>A, V600E) were also tested. All assays were tested in triplicate with 50% MAF, 1% MAF, 0.1% MAF and 0% MAF. In addition, for ESR1 D538G, D538Y standards were tested in triplicate to determine the assays cross-reactivity with other mutations on the same nucleotide position. Positive-controls amplification is shown by arrows on the amplification plots. NTC = No template control (water only loaded, no template).

All assays performed exceptionally well and were able to detect 1% mutant allele frequency and 0.1% mutant allele frequency in a mixture of 10,000 copies total (60ng) with ease. Moreover, the D538G assay did not heavily cross-react with D538Y sequence and amplification was further delayed compared to wild-type DNA. Considering the D538G assay is

one of the weaker assays, this shows that HOT_ARMS does not cross-react and can be used in instances such as companion diagnostics where it is essential to know the mutation for determining treatment intervention. Considering all assays performed well, this provides validation that HOT_ARMS is robust as the assays can be transferred to other labs and PCR machines.

Similarly, Qiagen wanted to test HOT_ARMS2a (KRAS c.34G>T, G12C). The experiments focused on the compatibility of the HOT_ARMS assays with patient-derived cfDNA. Qiagen chose several particularly difficult samples from their archive with low mutant copy numbers to thoroughly test the assay. In all instances, samples were blinded and compared against the results of the current Qiagen Therascreen[147] assay. To fall in line with Therascreen, the DNA concentration was unknown before testing and 5µl of sample was added to each 15µl reaction. Positive control reactions containing 50% MAF, 0.06% MAF and 0% MAF were added in triplicate to ensure the assay was performing correctly. Below in table 12 shows the results of the HOT_ARMS KRAS c.34G>T (G12C) assay. The cut-off for

mutant samples was 40 which was determined by the Ct-values obtained for 0.06%.

Table 12. Comparison of HOT_ARMS2a (KRAS c.34G>T, G12C) screening results on cfDNA samples with Qiagen's current Therascreen assay for KRAS c.34G>T (G12C). Cut off for mutant – Ct value <40. Fifty PCR cycles were completed as so if no amplification was seen the Ct value is described as >50.

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

Sample name	**HOT_ARMS2a Ct value**	**Mutation status HOT_ARMS2a**	**Mutation status Qiagen**
1	38.46	Mutant	Mutant
2	>50	Wild	Wild
3	>50	Wild	Wild
4	>50	Wild	Wild
5	>50	Wild	Wild
6	>50	Wild	Wild
7	>50	Wild	Wild
8	38.87	Mutant	Mutant
9	>50	Wild	Wild
10	>50	Wild	Wild
11	37.01	Mutant	Mutant

HOT_ARMS was 100% concordant with Therascreen for all 11 samples. 3 samples contained a mutation, and the remaining 8 samples were wild-type. Ct-values for samples were not provided by Qiagen which makes it difficult to draw complete direct comparisons. HOT_ARMS presented complete loss of wild-type amplification on wild-type samples as no amplification presented over 50 cycles. The results of this experiment show that HOT_ARMS2a (KRAS c.34G>T, G12C) can be used to determine the mutation status of liquid biopsy samples that are known to contain low copy number. This is further demonstrated by the fact these samples presented Ct-values between 37.01 and 38.46 which registered at a similar level to some 0.06% MAF positive control replicates.

x. Further validation of HOT_ARMS on cfDNA extracted from CRC pre-surgery liquid biopsies

To further validate HOT_ARMS, the pre-surgery liquid biopsies discussed in section 4.B.i (sample 2 day 1 to sample 22 day 1) were used to determine if HOT_ARMS could detect mutations in cfDNA samples. Whilst some evidence was already presented in 5.B.ix, it was important to determine if mutations could be detected in liquid biopsies processed in the Nottingham lab.

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

FFPE samples were retrieved from the Queen's Medical Centre pathology archive which corresponded to the pre-surgery liquid biopsies. FFPE samples were characterised to determine which mutations to test in the liquid biopsy. FFPE samples were extracted using the Qiagen DNA FFPE kit and diluted to 20ng/µl for PCR. 20ng of FFPE DNA in a 20µl reaction was amplified by PCR for 40 cycles with primers that generated short amplicons <150bp targeting the following frequently mutated CRC exons: KRAS exon 2, KRAS exon 3, KRAS exon 4, PIK3CA exon 1, PIK3CA exon 9, PIK3CA exon 20 and BRAF exon 15. Many of these regions have HOT_ARMS assays targeted to hotspot mutations. Primers (supplementary table 2) were validated by gel electrophoresis in supplementary figure 1 and 2. Mutation screening (table 13 below) was carried out by high-resolution melting analysis to speed up the process. If an exon was considered mutant by high-resolution melting analysis, the corresponding HOT_ARMS primer to that exon was used on that sample to determine if the HOT_ARMS assay hotspot mutation was present. High-resolution melting analysis is a method which determines the presence of mutation by shift in melting patterns due to the form of heteroduplexes[76]. Heteroduplexes depend on (i) the presence of both wild type and mutant alleles of gene in a

DNA sample and (ii) little difference in sequence between the wild type and mutant DNA (such as SNV) [76]. They are characterised by wild-type ssDNA binding to mutant ssDNA; if the DNA is melted and then allowed to anneal, the high homology between the wild and mutant DNA will enable annealing[76]. However, there will be a mismatch at the site of the mutation[76]. The mismatch causes physical disruption of the DNA duplex which causes instability compared to a homoduplex (i.e., two wild-type alleles binding together or two mutant alleles binding together) [76]. This instability causes a significant reduction in amplicon melting temperature which can be determined by a high-resolution melting analysis machine such as the Lightscanner from Biofire defence. Assays are designed to place the potential mutation in the centre of the amplicon to increase the destabilising effect of the mismatch.

Table 13. Final mutation status of FFPE blocks for surgery case samples. Primers that generated short amplicons <150bp targeting the following frequently mutated CRC exons: KRAS exon 2, KRAS exon 3, KRAS exon 4, PIK3CA exon 1, PIK3CA exon 9, PIK3CA exon 20 and BRAF exon 15 underwent high-resolution melting analysis for FFPE sample 2-22 which

corresponded to pre-surgery liquid biopsy sample 2-22. If a mutation was found, a HOT_ARMS assay corresponding to that exon was used to determine the sequence change. This was not possible with KRAS exon 4 since a HOT_ARMS assay was not developed but is left for future work.

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

Sample	FFPE Mutation status HRM (KRAS exon 2, KRAS exon 3, KRAS exon 4, PIK3CA exon 1, PIK3CA exon 9, PIK3CA exon 20, BRAF exon 15)	Sequence change HOT_ARMS
2	Wild all	N/A
3	KRAS exon 2	KRAS c.35G>T
4	KRAS exon 2	KRAS c.34G>T
5	KRAS exon 2	KRAS c.34G>T
6	Wild all	N/A
7	Wild all	N/A
8	Wild all	N/A
9	Wild all	N/A
10	Wild all	N/A
11	KRAS exon 2	KRAS c.38G>A
12	KRAS exon 2	KRAS c.35G>A
13	Wild all	N/A

14	KRAS exon 2	KRAS c.35G>A
15	KRAS exon 4	N/A
16	Wild all	N/A
17	Wild all	N/A
18	PIK3CA exon 9	PIK3CA c.1633G>A
19	Wild all	N/A
20	Wild all	N/A
21	KRAS exon 2	KRAS c.35G>T
22	KRAS exon 2	KRAS c.35G>A

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

Overall, 8 samples harboured a mutation in KRAS exon 2, sample 15 harboured a mutation in KRAS exon 4 and sample 18 harboured a mutation in PIK3CA exon 9. This gave a further 2 samples to monitor by liquid biopsy.

In total, 9 out of 10 cases containing a mutation in the primary tumour FFPE block were screened in the corresponding liquid biopsy by HOT_ARMS. An assay for KRAS exon 4 had not been developed at this stage and the liquid biopsy was not tested further. This will be completed as a further validation study in the future. As part of the final analysis, 5 HOT_ARMS assays were utilised; HOT_ARMS3a (KRAS c.35G>T, G12V), HOT_ARMS2a (KRAS c.34G>T, G12C), HOT_ARMS4 (KRAS c.35G>A, G12D), HOT_ARMS1 (KRAS c.38G>A, G13D) and HOT_ARMS9 (PIK3CA c.1633G>A, E545K). 5ng of cfDNA quantified by DeNovix was added to each reaction (20µl reaction volume) and carried out in triplicate. 3 randomised wild-type pre-surgery liquid biopsy samples which were negative in the FFPE sample for the respective mutation were included per assay as negative controls to check for false-positive amplification. Spike-in controls for 50% MAF, 1% MAF, 0.06% MAF and 0% MAF were run in duplicate to determine if the assay was running

normally and to determine Ct cut-offs. Table 14 below shows the final analysis of the liquid biopsies for the 9 primary tumours with KRAS exon 2 or PIK3CA exon 9 mutations.

Table 14. Final analysis of pre-surgery liquid biopsies. Mutations in corresponding FFPE tissue DNA to the liquid biopsies were determined initially by HRMA and followed by determination of SNV by HOT_ARMS PCR. Mutations were then followed into pre-surgery liquid biopsy by HOT_ARMS.

Sample	**FFPE sample mutation**	**Pre-surgery mutation status HOT_ARMS**
3	KRAS c.35G>T	Wild
4	KRAS c.34G>T	Mutant
5	KRAS c.34G>T	Mutant
11	KRAS c.38 G>A	Wild
12	KRAS c.35G>A	Mutant
14	KRAS c.35G>A	Mutant
18	PIK3CA c.1633G>A	Mutant

21	KRAS c.35G>T	Mutant
22	KRAS c.35G>A	Wild

In total 6 out of 9 samples (66%) contained a mutation in the pre-surgery liquid biopsy and the remaining 3 samples were confirmed as wild-type. This presents further validation that HOT_ARMS can detect mutations in liquid biopsies.

C: Discussion

The Amplification Refractory Mutation System (ARMS) was originally described as a means of genotyping single nucleotide variants (SNVs) without the need for formal sequencing [78, 143]. It has subsequently been used for mutation detection in cancer and, since the PCR is mutation-specific, it can be very sensitive. However, with standard ARMS PCR, there is still low-level base-pairing at the 3' end between the mutation-specific primer and the wild-type sequence despite being mismatched. This allows mis-priming of wild-type DNA by the mutant primer with consequent non-specific amplification. Discriminating non-specific from specific amplification when MAF is low can be problematic.

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

In High Optimised Ta ARMS (HOT_ARMS) PCR, the specificity of the PCR is increased by modifying the primers to raise the annealing temperature (Ta). The increased kinetic energy of the primers reduces non-specific 3' base-pairing and the tag increases selectivity for the mutant amplicon. The primers can be modified in several ways, but the best results were obtained by adding a 10 base GC tag onto each primer. This simple adaptation of ARMS PCR with a "working" Ta of 71°C generates an unrivalled all-round PCR-based method of detecting low frequency mutant alleles which is scalable, ultrasensitive, and robust.

Whilst others have increased the specificity of ARMS primers with tags previously [148, 149]. HOT_ARMS differs by modifying both sets of primers rather than the ARMS primer alone to increase the overall annealing temperature of the reaction. This allows selection of mutant PCR amplicons generated in early cycles for amplification and the high temperature will increase primer annealing specificity. Previous interactions between ARMS primers and tags have aimed to increase specificity of the ARMS primer alone without modifying the annealing temperature, this has resulted in variable sequence dependent systems that require heavy

optimisation and have limited scalability. These systems have either incorporated a 5’ tag which generated hairpin loops [149] to prevent amplification of wild-type DNA or have introduced tags into the centre of the primer to destabilise the 3’ end [148]. Whilst these systems use tags, the underlying principle of using a high annealing temperature to cause increased specificity and selectivity for mutant amplicons is novel. Moreover, the tag used in HOT_ARMS is always the same and does not need optimisation. Therefore, this is a solution which is scalable since it is not sequence context dependent.

Fifteen different mutations, located in 6 different genes and 8 exons; including both SNVs and deletion mutations, were tested. We believe we are the first group to use ARMS PCR for detection of deletions and the results were comparable with detection of SNVs. The primers required no more optimisation than regular primers and in fact, after the first two sets of primers, optimisation was found to be minimal as the primers worked “off the shelf”. HOT_ARMS primers have been shown to detect indels which we believe to be novel but not unexpected. Many cancers contain indels, gene fusion events, rearrangements, copy number changes or mixture thereof as

opposed to primarily SNV's[7]. As long as a novel sequence is created, it should be detectable by HOT_ARMS PCR. Development of HOT_ARMS for other types of mutation is discussed in section 7.C.ii.

The robustness of the methodology is reflected in the fact that all 15 assays with little or no optimisation had a limit of detection of 0.1% MAF, making it suitable for mutation detection of liquid biopsies. Moreover, approximately half of the assays had a limit of detection of 0.06% and HOT_ARMS3a was shown to detect 1 mutant copy at 0.01%, demonstrating that the technology is truly ultrasensitive. Additional wild-type bleed-through reduction if necessary, could be achieved by combination with 3' LNA, penultimate mismatch or wild-type blocker but it is likely due to the low concentration of cfDNA that this will not be required often. This was possible due to the high specificity of HOT_ARMS and the fact that PCR efficiency was unaffected. The results of the demonstration at LifeArc further demonstrated the robustness of the technology. Five HOT_ARMS assays, including the generation of two new ESR1 assays were able to reliably detect 0.1%MAF at LifeArc labs in Edinburgh. This experiment verified that HOT_ARMS can be performed in a different lab on

a different machine with a variety of templates. In addition, successful cross-reactivity experiments have shown the potential of HOT_ARMS to be used in scenarios such as companion diagnostics where resistance mutations may be acquired, leading to the pool of nucleic acids containing multiple variants. This level of robustness with such a simple modification which does not affect PCR efficiency for all types of sequence change makes HOT_ARMS extremely scalable across the genome as it is not tied to specific sequence compositions. This is a novel factor that has not been achieved previously with ARMS technology as other types of modification to the system which have resulted in significant sensitivity increases have had significant downsides. Other modification types make the method optimisation heavy and tied to specific sequence changes or sequence compositions. HOT_ARMS working at a high temperature with a lack of change in PCR efficiency means that it can theoretically be deployed anywhere that a regular PCR primer can be designed with any sequence change on the 3’ end. Regardless of ultrasensitivity, this alone makes HOT_ARMS unique and capable of different ctDNA detection strategies which can be personalised (further discussed in section 7.C.iii).

5. Development of Highly Optimised annealing Temperature_Amplification Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy

HOT_ARMS was successfully further modified by combination with a rapid cycling mastermix. HOT_ARMS 12 (BRAF c.1799A>T, V600E) was used as an example and could detect 0.06% (4 mutant copies) with a 30-minute cycling program including melt-curve analysis on the MIC PCR machine. To our knowledge this is the fastest ultrasensitive assay developed for BRAF c.1799A>T (V600E). The specificity of HOT_ARMS is what allows HOT_ARMS rapid cycling to occur as the impact of lower annealing times and lower specificity mastermix compositions is counteracted by the high annealing temperature. Rapid cycling programs such as this would theoretically allow HOT_ARMS to be deployed within the setting of an out-patient department or GP clinic. This would not only allow patients greater access to testing but would also release a proportion of the burden experienced by radiology departments and if placed in GP clinics, releasing the burden on the out-patient department, and saving the NHS considerable amounts of money.

Unfortunately, from 2018 onwards when we had completed HOT_ARMS validation on cell line template, the biobank had entered a crisis which resulted in complete restriction of sample access for researchers. We had planned to collect

liquid biopsy samples from rectal cancer patients undergoing radiotherapy with Dr. Eliot Chadwick, who is a consultant clinical oncologist for rectal cancer patients at Nottingham City Hospital for a full evaluation of treatment response monitoring. As this could not take place, we decided to test the previous liquid biopsy samples collected for the pilot study. We decided to use these samples for further validation on liquid biopsy samples since they were the best surrogate, we could find for rectal cancer samples. Unfortunately, this does not give an idea of treatment-response monitoring, but the development of a method which can undertake such as a study is equally important for future work. HOT_ARMS was shown to be able to detect mutations in pre-surgery liquid biopsies which corresponded to the patients FFPE tissue. Screening revealed 10 out of 21 samples (48%) contained a mutation from the FFPE screening panel. The percentage is expected, although it was unusual that no BRAF exon 15 or PIK3CA exon 20 mutations were present. These mutations are usually present in 10-15% of a cohort[146]. However, the sample set is small, and most cases are early-stage Dukes A and B. Thus, the frequencies are likely to be lower even though these mutations are not stage-specific as the overall tumour mutation burden will be lower compared to late-stage

tumours that are included in the majority of mutation frequency analyses. To increase the sensitivity of screening CRC for mutations, other frequently mutated genes such as TP53, APC, SMAD4 and PTEN should be included[146]. Moving forwards, as described in section 7.C.iii, FFPE tumours may undergo mutational profiling by next generation sequencing to expand the breadth of analysis. Mutations were detected in 6 out of the 9 total pre-surgery liquid biopsies, representing a detection rate of 66% and validating further that HOT_ARMS can detect mutations in liquid biopsy. It is generally accepted that early-stage tumours will shed less ctDNA[137, 138]. Thus, a detection rate of 66% is expected and less than 100% is found commonly in studies describing ctDNA detection in pre-surgery liquid biopsies[62, 63]. HOT_ARMS was further validated on liquid biopsies by achieving 100% concordance with Therascreen[147] for the detection of KRAS c.34G>T (G12C). The samples chosen by Qiagen were known to have low total cfDNA quantity and low mutant copy number. For this reason, it was an extremely important stress test for liquid biopsy testing by HOT_ARMS. The simplicity of HOT_ARMS in comparison with Therascreen's scorpion probe-based KRAS assays makes it far superior as a scalable solution for detecting all types of sequence variant across the

genome and reducing optimisation time considerably, proving once again that HOT_ARMS is a significant advancement for liquid biopsy testing.

Overall, HOT_ARMS is an exceptionally robust, ultrasensitive, and scalable mutation detection technology which could be deployed to maximise cancer patient management. The simplicity and flexibility of HOT_ARMS due to its high specificity means it could be utilised for cancer surveillance, tumour-response monitoring, and cancer screening. HOT_ARMS presents an all-round solution which is both inexpensive and rapid. As mentioned in section 1.C.iii, NGS also presents a solution for the detection of ctDNA. However, NGS is very expensive and time-consuming which limit's the use-case for clinical implementation; especially when tracking several samples from the same patient over a long time period. HOT_ARMS has been extensively validated and has allowed us to be confident in the method moving forward into further investigative work on ctDNA, such as treatment-response monitoring of rectal cancer patients.

6. HOT_PCR: Highly Optimised annealing Temperature_Probe inhibited PCR (HOT_PI) PCR as a method to detect mutations in hotspot clusters for liquid biopsy

6. HOT_PCR: Highly Optimised annealing Temperature_Probe inhibited PCR (HOT_PI) PCR as a method to detect mutations in hotspot clusters for liquid biopsy

A: Introduction

i. Wild-type blocking PCR as a complement to HOT_ARMS PCR

HOT_ARMS PCR (section 5) is an exquisitely sensitive method of mutation detection. However, it is most efficient when there is prior knowledge of the mutation since each sequence change requires a specific primer. Whilst this is acceptable for cancer surveillance and treatment response monitoring when the tumour has been profiled, it is not efficient when the tumour has not been profiled. In this case, liquid biopsies are being screened for the presence of unknown mutations. The latter may be favourable with cancers such as colorectal cancer where mutational burden is high and associated with a high number of hotspot mutations. HOT_ARMS PCR could theoretically be used as a multiplex panel of hotspot mutations when combined with hydrolysis probe technology. This approach would require mutational profiling of the pre-treatment liquid biopsy. Patients identified with mutations could then undergo singleplex testing of the identified mutation for treatment-response monitoring and surveillance. This would save testing of the FFPE tissue, which is more expensive and consumes precious tumour material. Although the overall number of patients benefiting from this technology

would be lower compared to a personalised approach which would include FFPE profiling as discussed in 7.C.iii. HOT_ARMS could also be used as single tests for cancers that present the same SNV frequently, i.e., BRAF c.1799T>A (V600E) in melanoma, KRAS codon 12/13 mutations in pancreatic cancer and EGFR c.2369C>T (T790M) in non-small cell lung cancer[146]. However, when screening for unknown mutations, the more hotspot mutations included, the higher the sensitivity of the test for picking up a single mutation in the panel. We saw the potential of wild-type blocking as a separate method to complement to HOT_ARMS PCR (chapter 5) for the detection of mutations in hotspots clusters. This would reduce the number of HOT_ARMS assays required and increase the number of mutations that could be tested efficiently in practice. This is critical with liquid biopsy, as the amount of available DNA is limited.

Aims

1. Improve fundamental flaws with wild-type blocking PCR so that it becomes scalable, ultrasensitive and has the ability to integrate with HOT_ARMS PCR protocols for multiplexing.

6. HOT_PCR: Highly Optimised annealing Temperature_Probe inhibited PCR (HOT_PI) PCR as a method to detect mutations in hotspot clusters for liquid biopsy

Hypothesis

1. The integration of the HOT_PCR modification for wild-type blocking will improve the specificity of the blocking probe by raising the annealing temperature. This will in turn produce enhanced mutation enrichment which will allow detection of low frequency variant alleles for ctDNA detection. Moreover, it will produce wider regions of access for mutation hotspot cluster screening as the blocking probe will be increased in length. Most importantly the reaction will be able to take place at similar annealing temperatures to HOT_ARMS PCR for multiplexing integration.

ii. Highly Optimised Annealing Temperature_Probe Inhibited PCR (HOT_PI PCR) introduction

We developed our own wild-type blocking approach which used the same reagents, primer modifications and cycling protocols as HOT_ARMS PCR so that it could be integrated easily with HOT_ARMS PCR testing. This method is called Highly Optimised annealing Temperature_Probe Inhibited PCR (HOT_PI PCR). The modification used (10nt GC-tag) and benefits for HOT_PI PCR are described in section 5.A.ii.

6. HOT_PCR: Highly Optimised annealing Temperature_Probe inhibited PCR (HOT_PI) PCR as a method to detect mutations in hotspot clusters for liquid biopsy

Currently, mutation detection systems that are designed to screen for mutations across multiple nucleotides in a single reaction use conventional PCR to amplify DNA prior to analysis. Conventional PCR maintains equal efficiency for both mutant and wild-type DNA throughout a PCR program. Hence, the mutant allele frequency will remain unchanged from DNA sample to PCR amplicon. To improve the sensitivity and limits of detection further for mutation detection systems, one solution is to modify conventional PCR to enrich specifically for the mutant fraction. Mutation enrichment is a means of converting low mutant allele frequency samples into high mutant allele frequency samples[80, 150, 151]. This allows the sample to become detectable as it breaches the barrier which is the limit of detection. Moreover, borderline samples which were close to the limit of detection will become more visible. I.e., if Sanger sequencing were used as the mutation detection method for two samples which have mutant allele frequencies of 1% and 20% and conventional PCR was used to amplify them; only the 20% sample would be detected due to the 20% limit of detection[73]. However, if an enrichment protocol was used to enrich the 1% sample to 20% and the 20% sample to over 80%; then the previous 20% borderline

sample is now easy to detect and the 1% sample becomes detectable.

Wild-type blocking PCR is an enrichment system that implements a non-extendable wild-type specific probe into the conventional PCR (regular primers that are not mutation-specific)[79]. The probe typically overlaps the 3’ end of one of the primers by 5 nucleotides. The probe is designed to have a higher annealing temperature than the primers and usually incorporates locked nucleic acids (LNAs)[123] to increase binding affinity. Moreover, a phosphate group is integrated on the 3’ end to prevent polymerase extension. Thus, when the reaction is ramping down from the denaturation temperature to the annealing temperature, the probe will bind before the primer to the wild-type template thereby blocking access to the template. When a mutation is present, a probe-template mismatch will occur that will cause thermodynamic instability. The mismatch instability generated will cause the probe to lose binding affinity for the mutant template and bind approximately 8°C lower for DNA-DNA mismatches and 20°C lower for DNA-LNA mismatches[79, 122, 123, 151]. In essence, the Ta of the probe for the mutant template will be lower than the Ta of the primer. This will allow the primer to bind to the mutant template (as the probe is no longer

occupying the region of interest) resulting in specific amplification of the mutant allele.

HOT_PI PCR is a method of mutation enrichment and thus can be combined with any type of mutation detection method. Here specifically, a HOT_PI PCR assay targeting KRAS exon 2 codon 12/13 mutations is generated and combined with high-resolution melting analysis[76] to keep the test closed-tube. An alternative approach which has not been tested is the implementation of hydrolysis probes which is shown by Huang et al. Sanger sequencing with squirrel primer modifications were used to validate the assay[117].

B: HOT_PI PCR - Optimisation of the method

i. HOT_PI PCR strongly blocks wild-type DNA amplification and is concentration dependent

In the first instance, we wanted to prove that HOT_PI PCR could work effectively with the same mastermix and cycling protocols used in HOT_ARMS PCR whilst being able enrich to a high enough level to detect 0.1% MAF for liquid biopsy testing. This was critical as we were unsure if the wild-type blocking probe would still be effective when the primer annealing temperature was so high since the wild-type blocking probe is required to bind on at a higher temperature than the primers. Moreover, we investigated if a 2-phase

step-up protocol could be used as described in section 5.B.i to reduce overall cycling time.

The first place to begin was to design an assay to prove the principle. Studies had already been completed on wild-type blockers with the IntPlex system[85], E-ICE COLD PCR[81] system and combination of LNA blockers with HOT_ARMS PCR that are all evidenced in this thesis. The evidence so far suggests that in our lab wild-type blockers can prevent wild-type amplification to a degree but had been shown in a different context, i.e., with ARMS PCR and COLD-PCR. In all methods, KRAS exon 2 hotspot mutations at codons 12/13 were targeted and this presented a good starting point. Moreover, this region is perfect to target with a wild-type blocking assay as we could validate its use over multiple nucleotides. This would also reduce the number of tests required from 7 HOT_ARMS assays to 1 wild-type blocking assay. The primers used for HOT_PI PCR were the same as those used in section 5.B.viii.

Since it was previously determined that the blocking probe can influence PCR efficiency (section 5.B.vi – table 10), it was important to define the optimal probe concentration. The experiment was undertaken as a two-phase cycling protocol with the first 10 cycles having an annealing temperature of

71°C and the remaining 40 cycles were carried out at 75°C as described in section 5.B.i to be compatible with HOT_ARMS PCR. The amplicon generated was 95 bases in length excluding tag sequence to be compatible with liquid biopsy testing. As with HOT_ARMS PCR in section 5.B.vi, we began with a wild-type blocking probe containing 6 LNA's, three of which were present on the mutation hotspots c34, c35 and c38 and three which were randomly distributed as shown in supplementary table 5. The blocker was 53 bases long with 6 LNA's to raise the annealing temperature to >85°C as the maximum annealing temperature of the HOT_PI primers were 75°C in a two phase step-up reaction giving a difference in binding temperature of at least 10°C. The wild-type blocking probe was added to a 20µl reaction at concentrations of 0nM, 100nM, 200nM and 300nM. Spike-in positive controls of KRAS c.38G>A (G13D) using the cell line HCT116 in a background of HEK293T wild-type DNA was generated to produce 1% MAF and 0.1% MAF for limit of detection testing by Ct value. All samples were run in triplicate including the addition of positive control KRAS mutant cell lines, one for each nucleotide position: position 38 - HCT116 (KRAS c.38G>A, 50% MAF), position 35 - SW480 (KRAS c35G>T, 100% MAF) and position 34 - A549 (KRAS c.34G>A, 100%MAF). The PCR products of

all reactions were used for downstream mutation analysis by high-resolution melting and Sanger sequencing for validation. The results of the first experiment are shown in table 15 below.

Table 15. HOT_PI Ct values for different wild type blocking probe concentrations. Reactions took place with 3 different concentrations of blocking probe, 100nM, 200nM and 300nM. A reaction without wild-type blocking probe was also added for reference to see the effect on the wild-type blocking probe (0nM). All samples were run in triplicate including the addition of positive control KRAS mutant cell lines, one for each nucleotide position: position 38 - HCT116 (KRAS c.38G>A, 50% MAF, 1% MAF and 0.1% MAF), position 35 - SW480 (KRAS c35G>T, 100% MAF) and position 34 - A549 (KRAS c.34G>A, 100%MAF). HEK293T was used as the wild-type cell line.

Wild-type blocking probe concentration	c.35G>T 100% MAF	c.34G>A 100% MAF	c.38G>A 50% MAF	c.38 G>A 1% MAF	c.38 G>A 0.1% MAF	Wild-type	ΔCt 1% and wild-type
0nM	25.92	28.52	26.97	28.02	28.32	28.21	0.19
100nM	27.24	30.18	30.10	35.06	36.28	36.48	1.42
200nM	28.33	31.53	31.59	37.84	44.77	43.61	6.13
300nM	29.28	32.73	33.15	40.32	48.89	49.32	9.00

6. HOT_PCR: Highly Optimised annealing Temperature_Probe inhibited PCR (HOT_PI) PCR as a method to detect mutations in hotspot clusters for liquid biopsy

With no wild-type blocking probe, there is minimal difference between mutant and wild-type Ct values. The max ΔCt was 2.29. As the wild-type blocking probe concentration is increased, the max ΔCt between mutant template at 50% or 100% MAF and wild-type is increased incrementally. At 100nM, the ΔCt is between 6.3 and 9.24; at 200nM, the ΔCt is between 12.02 and 15.28; at 300nM, the ΔCt is between 16.17 and 20.04. Individually, mutant samples are consistent in that efficiency is reduced incrementally at different wild-type blocking probe concentrations. For every 100nM of wild-type blocking probe added, there is an increase of between 1 to 2 cycles for the 50% and 100% MAF samples but for the wild-type sample there is an increase of between 5 to 6 cycles. With regards lower MAF spike-in controls, 1% MAF becomes easily detectable by Ct value using concentrations of 200nM and 300nM. A ΔCt of 9 was seen between 1% MAF and wild-type with 300nM wild-type blocking probe and a ΔCt of 6.13 was seen for 200nM. 100nM blocking probe was unable to detect 1% MAF by Ct-value. Whilst the ΔCt looked promising for 200nM and 300nM, 0.1% MAF which is our threshold for ctDNA detection was not detectable. This does show the possibility for the use of fluorescent hydrolysis probes to replace the double-stranded binding dyes used here

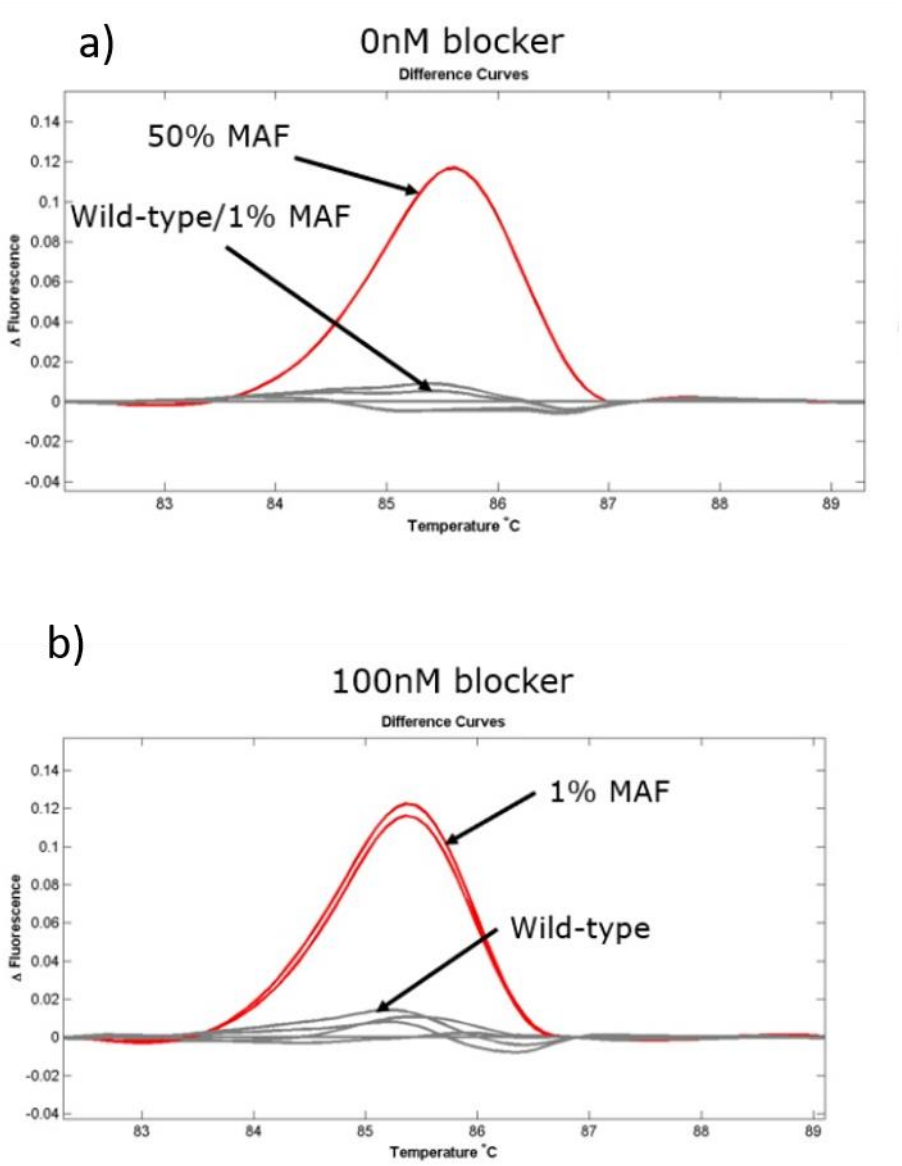
as this may lead to detection of 0.1% MAF as shown by Huang et al.

ii. HOT_PI PCR can be combined with high resolution melting analysis (HRMA) to provide a rapid, ultrasensitive, closed-tube test

As we had insufficient funding to test fluorescent hydrolysis probes for detection like with the original LNA wild-type blocking protocol by How-Kit[79], we tested the PCR products by high resolution melting analysis instead to see if 0.1% MAF could be detected by the addition of a mutation detection system. Since wild-type blocking PCR alone is a mutation enrichment method, it can be combined with other mutation detection systems such as NGS, ddPCR, sequencing and high resolution melting analysis. We deployed high resolution melting analysis on HOT_PI PCR products that had been enriched with 100nM wild-type blocking probe. All seven KRAS mutations which are the most common targeted by HOT_ARMS PCR were enriched at 50% MAF, 1% MAF and 0.1% MAF. In all instances, 50% MAF, 1% MAF and 0.1% MAF could be detected for all seven mutation types. The results are shown in figure 31 using difference plots. High resolution melting analysis has previously been described in section 5.C.x. Difference plots demonstrate the difference in

fluorescence produced by the double-stranded binding dye (Eva Green) at different temperatures. Since wild-type samples are known they can be used to set as the baseline for reference. A difference in fluorescence of at least 0.02 can be sufficient to call a sample mutant. The figures below do not show 200nM and 300nM as these samples produced aberrant melting patterns and so this method of analysis was not suitable and is shown in figure 31. High resolution melting analysis by LightScanner cannot determine homozygous mutations. When enrichment of high MAF samples occurs, samples contain MAF after enrichment of >50%. Thus, Ct-value can be used to determine mutations of high MAF samples and high-resolution melting is useful to analyse samples with low Ct-value after enrichment to see if they contain a mutation. A way to fix this issue is to use a Lightcycler-480 system by Roche which can detect homozygous mutations. However, it is mostly likely unnecessary as ctDNA presents a low proportion of overall cfDNA and MAF is typically very low.

6. HOT_PCR: Highly Optimised annealing Temperature_Probe inhibited PCR (HOT_PI) PCR as a method to detect mutations in hotspot clusters for liquid biopsy



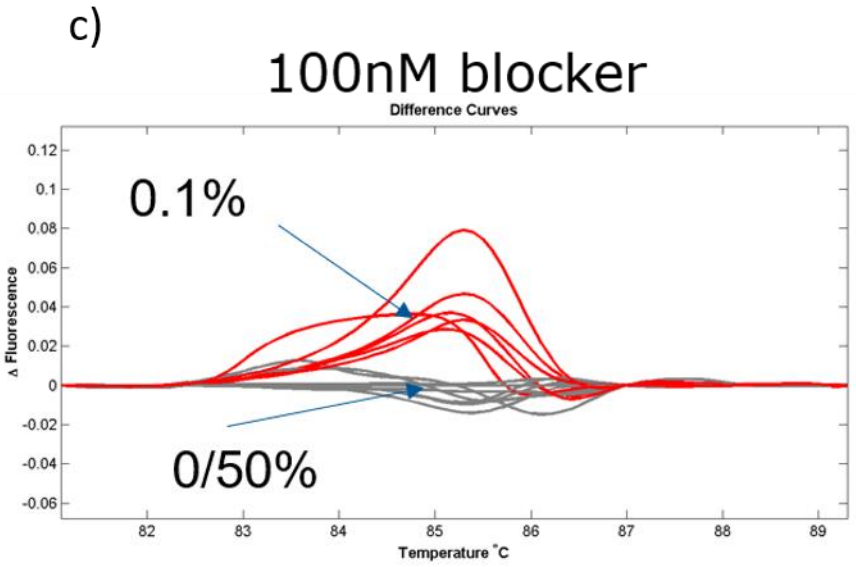


Figure 31. HOT_PI PCR mutation enrichment demonstrated by high-resolution melting analysis difference curves. High resolution melting analysis has previously been described in section 5.C.x. Difference plots demonstrate the difference in fluorescence produced by the double-stranded binding dye (Eva Green) at different temperatures. Since wild-type samples are known they can be used to set as the baseline for reference. A difference in fluorescence of at least 0.02 can be sufficient to call a sample mutant. Figure a) shows that 0.1% MAF, 1% MAF and wild-type cannot be detected as mutant without wild-type blocking mutation enrichment using standard PCR plus high-resolution melting analysis. Figure b) shows that with the addition of 100nM blocking probe to a HOT_PI PCR protocol, significant enrichment which allows 1% MAF to now be detected by high resolution melting analysis

where the standard limit of detection of high-resolution melting analysis is 6% MAF. Figure c) shows that with the addition of 100nM blocking probe to a HOT_PI PCR protocol, significant enrichment allows 0.1% MAF to now be detected by high-resolution melting analysis for all 7 common KRAS variants which HOT_ARMS PCR detects. 50% MAF can only be detected by Ct value since high-resolution melting analysis can only detect heterozygous mutation and the enrichment turns the sample near homozygous after enrichment.

Without wild-type blocking probe; 1% MAF does not differ from wild-type DNA as the limit of detection of high-resolution melting analysis with these primers is 6% which is typical for high-resolution melting. However, with 100nM wild-type blocking probe, 1% MAF shows a large reduction in melting temperature with ΔFluorescence at 0.12; representing a mutation. Moreover, 0.1% MAF (8 mutant copies) could be detected, although less easily at ΔFluorescence 0.03-0.08. This represents a >60x fold enrichment since as mentioned above the LOD is 6% MAF. This experiment determines that high-resolution melting analysis combined with HOT_PI PCR can detect 0.1% MAF which is reliable for liquid biopsy testing. High-resolution melting is likely not a true reflection of absolute LOD as this depends on the mutation detection

method of choice. The LOD may be enhanced by using other methods such as ddPCR, NGS or hydrolysis probes.

iii. HOT_PI PCR produces significant mutation enrichment and is validated with Sanger sequencing

Wild-type blocking probe concentrations of 200nM and 300nM gave superior ΔCt by qPCR between 1% MAF and wild-type compared with 100nM. However, concentrations this high gave aberrant high-resolution melting analysis data for wild-type DNA. Aberrant high-resolution melting data is shown in figure 32 below.

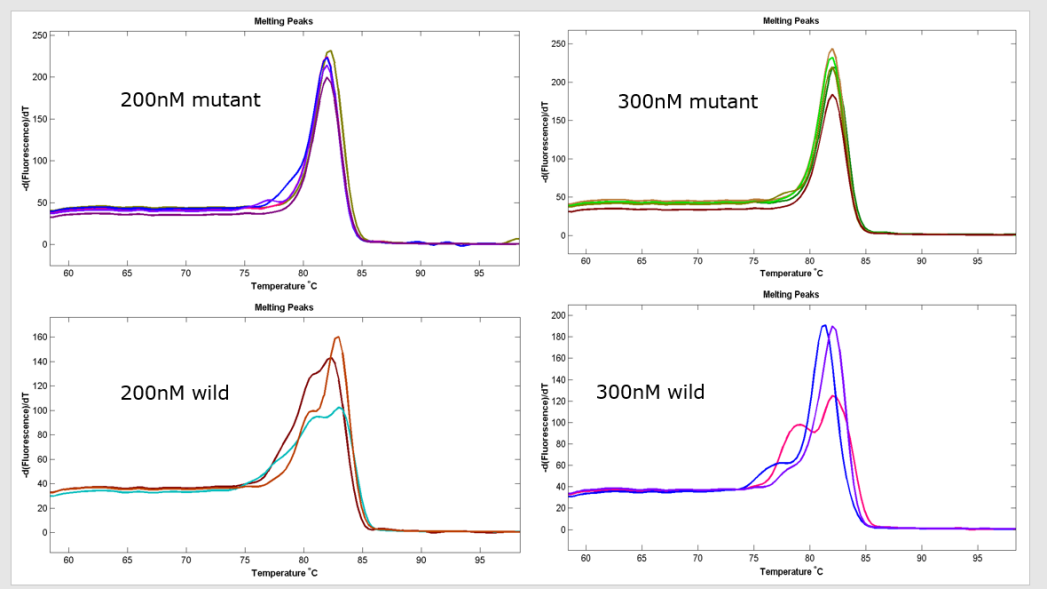


Figure 32. Aberrant melting curves for 200nM and 300nM wild-type blocking probe in HOT_PI PCR. Each concentration, 200nM wild-type blocking probe and 300nM wild-type blocking probe has two melting peak charts which show peaks for

mutant samples (50% MAF – top) and wild-type samples (HEK293T – bottom).

Due to the aberrant melting patterns, the PCR products were investigated by Sanger sequencing to determine if 100nM was producing any artefacts. Since the amplicon size was below 100bp it could not be investigated directly. We had previously described a method for performing Sanger sequencing on short amplicons which was termed Squirrel PCR[117] (section 2.F.ii).

Figure 33 below demonstrates that clean sequence can be generated with 100nM wild-type blocking probe, containing no artefacts and shows the enrichment potential of HOT_PI PCR by Sanger sequencing. Image A represents wild-type sequence from the HEK293T cell line; Image B shows a homozygous G>T (100% MAF) mutation found in the colon cancer cell line SW480 (c.35G>T) on the second DNA base in the image; Image C shows SW480 spiked into HEK293T (wild-type) at 1% mutant allele frequency (MAF); Image D shows SW480 spiked into HEK293T (wild-type) at 5% MAF; Image E shows SW480 spiked into HEK293T (wild-type) at 10% MAF; Image F shows SW480 spiked into HEK293T (wild-type) at 20% MAF. SW480 is known to be homozygous but this was also tested without enrichment and was similarly 100% MAF

6. HOT_PCR: Highly Optimised annealing Temperature_Probe inhibited PCR (HOT_PI) PCR as a method to detect mutations in hotspot clusters for liquid biopsy

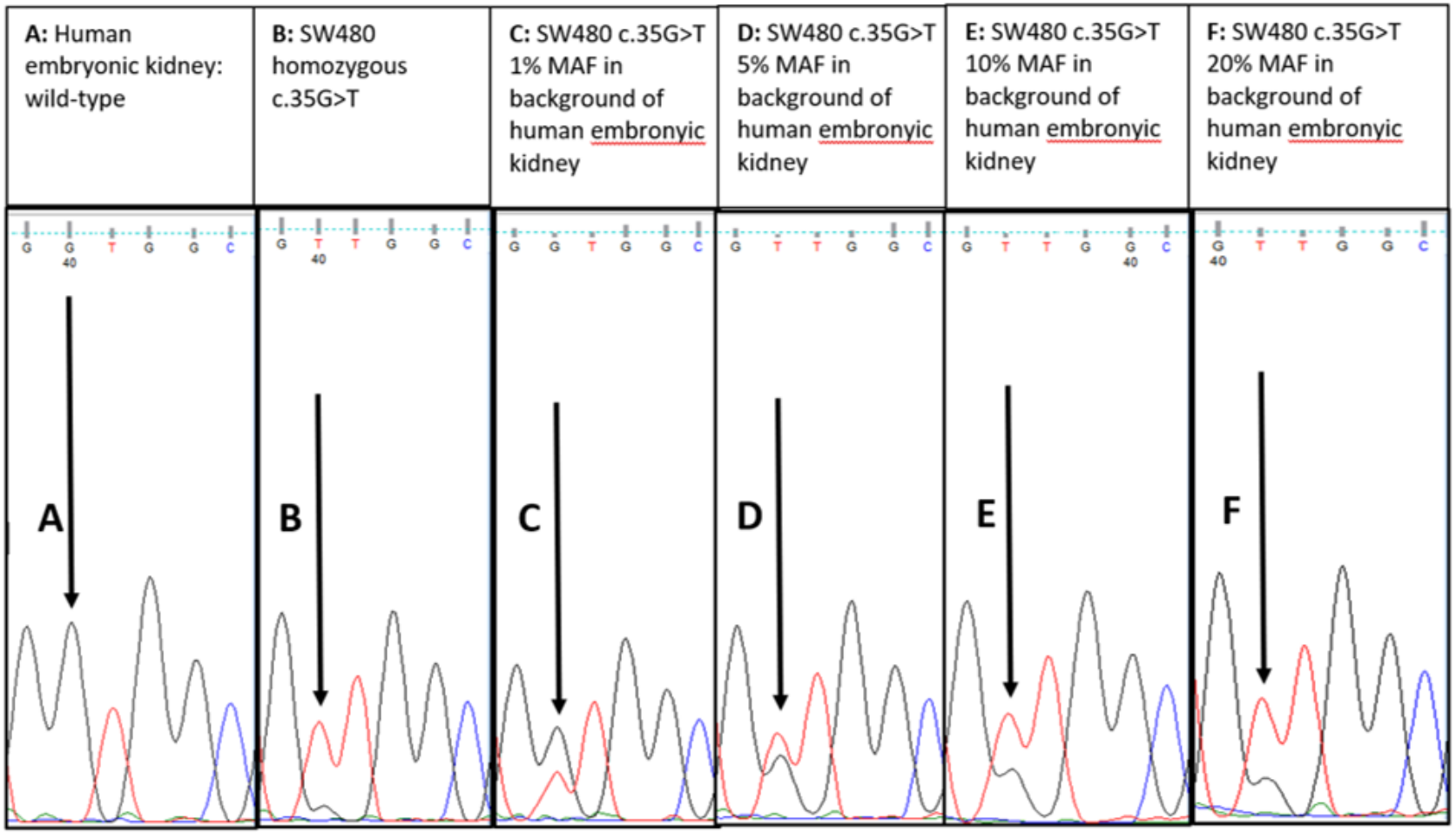


Figure 33. Sanger sequencing of a 95 base pair amplicon for KRAS exon 2 (codon 12 and 13 shown) which has undergone mutation enrichment by the highly optimised annealing temperature probe inhibited PCR system. Image A represents wild-type sequence from the HEK293T cell line; Image B shows a homozygous G>T (100% MAF) mutation found in the colon cancer cell line SW480 (c.35G>T) on the second DNA base in the image; Image C shows SW480 spiked into HEK293T (wild-type) at 1% mutant allele frequency (MAF); Image D shows SW480 spiked into HEK293T (wild-type) at 5% MAF; Image E shows SW480 spiked into HEK293T (wild-type) at 10% MAF; Image F shows SW480 spiked into HEK293T (wild-type) at 20% MAF. Arrows point to the base on interest which contains the mutation (c.35 G>T).

As the limit of detection for Sanger sequencing is around 20%, image C (1% MAF), D (5% MAF) and E (10% MAF) should show wild-type sequence and image F (20% MAF) should show a small peak for Thymine together with the peak for Guanine which would represent the mutation. HOT_PI gave substantial mutation enrichment and the 1% MAF shown in image C shows that the mutation is easily detectable. Unexpectedly, Thymine becomes the dominant peak in image

D (5% MAF), E (10% MAF) and F (20% MAF) and the wild-type DNA is now the minor allele.

C. Discussion: HOT_PI PCR and its advantages

This chapter introduced HOT_PI PCR, an additional method to complement HOT_ARMS PCR for screening of mutations in pre-treatment liquid biopsies. Whilst HOT_PI PCR requires more optimisation than HOT_ARMS PCR; in scenarios where hotspot mutation clusters exist like KRAS exon 2 codon 12 and 13 in colorectal cancer, it can reduce the number of required tests from 7 HOT_ARMS tests to 1 HOT_PI PCR test, ultimately increasing throughput whilst maintaining sensitivity. By the addition of the simple 10nt GC tag to increase the primer Ta ≥65°C combined with a 6LNA wild-type blocking probe, HOT_PI PCR mutation enrichment can cause dramatic improvements to the limit of detection of mutation detection methods such as Sanger sequencing (figure 33) and high-resolution melting analysis (figure 31). Improving Sanger sequencing from a poor limit of detection of 20% mutant allele frequency[73] to easy detection of 1% MAF and high-resolution melting analysis from 6% limit of detection[76, 113] to a reliable 0.1% MAF limit of detection. This represents >60x fold mutation enrichment. A scanning region of 48bp can be achieved since the blocker is 53bp in length and up to

6 LNA bases can be added to detect any SNV on the base of choice. It may be possible to increase this to 10 LNA bases as shown with HOT_ARMS PCR and in turn reduce the required blocker concentration. The test can be performed within 90 minutes due to reduced PCR ramping from a 2-phase step up protocol. Moreover, when combined with high-resolution melting analysis, the result can be obtained within a total time of 2 hours.

HOT_PI PCR was demonstrated using mutations in KRAS exon 2 codons 12 and 13. We found that wild-type blocking probes were only effective with LNA incorporation on the bases which were mutated. Equal mutation enrichment was achieved across all 3 commonly mutated bases with 7 sequence changes tested in total across codons 12/13. By inference it can be assumed that other sequencing technologies such as pyrosequencing[73] (5% LOD) and deep sequencing[64] (0.01-3% LOD) could undergo HOT_PI mutation enrichment before analysis and improve sensitivity even further whilst also allowing the sequence change to be known. It may also be possible to combine HOT_PI PCR with ddPCR, to our knowledge this has not been shown to date.

Overall, an extremely effective mutation enrichment system has been generated, with enrichment >60x. High resolution

melting analysis can be performed directly after PCR, meaning that the reaction can remain a closed single-tube test. This system will be able to effectively reduce the number of HOT_ARMS tests in scenario's where mutation hotspot clusters are required for screening of pre-treatment liquid biopsies to determine mutations to target for treatment response monitoring and surveillance without mutational profiling of the FFPE sample. In colorectal cancer this may be particularly effective as hotspots mutations are frequent and many are found in clusters.

7. Final discussion

A: Introduction

HOT_PCR is a combination of ultrasensitive, inexpensive, and rapid mutation analysis methods for liquid biopsy analysis. HOT_PCR is currently patent pending and has been published (WO2019145734). Search reports indicate that the method is novel and does not incur unity issues between methods. Five papers or patents were raised in total against novelty. The first method is one of our own publications surrounding the use of tags to change amplicon melting temperature for high-resolution melting analysis multiplexing [76] which led to the development of HOT_ARMS PCR; the second method is a general nested ARMS-PCR method which does not include a high annealing temperature or tags[126]; the third method is a standard ARMS-PCR for KRAS codon 12/13 variants which does not include a high annealing temperature or tags[152]; the fourth method is an ARMS-PCR method which does not include a high annealing temperature but does contain a 5’ tag on the ARMS primer which forms a hairpin loop when binding wild-type DNA, blocking polymerase access [149]; and the fifth method is an ARMS-PCR with wild-type blocker [153] which does not modify the primer annealing temperature but adds an additional annealing step for the wild-type blocker to bind which they claim enhances specificity – this was attempted with HOT_ARMS PCR before this paper

was brought forward but it failed to provide any benefit, as with the paper published for this method, the benefit shown was marginal. All five methods show that with wide search reports, HOT_PCR is novel which allows this technology the opportunity to move forward in this context towards the clinic. The HOT_PCR methods originate due to limitations with existing technologies. To complete the treatment response monitoring of rectal cancer patients by liquid biopsy outlined at the beginning of the PhD, a new method had to be generated. We now have a working method which can potentially be beneficial to patients. Without a clinical trial using intervention, it is not possible yet to claim that benefit will be achieved. However, the process is logical and initial reports show that detecting circulating tumour DNA can be indicative of response to therapy in rectal cancer.

B: Cardinal Findings and conclusions

i. HOT_ARMS PCR is a reliable method for liquid biopsy testing

The need to identify low frequency tumour-derived mutant alleles in a pool of wild-type DNA is well described in molecular diagnostics. Several methods of varying complexity have been established for the detection of very low MAF each with its own utility and advantages/disadvantages. Apart from setting a specific cycling program and purchasing GC tagged

primers, HOT_ARMS PCR does not differ from a standard real-time PCR i.e., this is a single closed-tube reaction which does not require extra mismatches, special probes, special enzymes or a nested protocol for liquid biopsy testing. It is much more sensitive than ARMS PCR and there is no other probe-free single stage closed tube method that comes close to the most common limit of detection achieved by HOT_ARMS. HOT_ARMS has a similar limit of detection as other more complicated systems published in the literature and since the PCR is efficient, it can be used in combination with standard curves to quantify MAF for treatment-response monitoring of radiotherapy for rectal cancer. Addition of wild-type blocking probes has been used to improve the limit of detection and some closed-tube systems claim to have a limit of detection as low as 0.005% MAF [72]. ARMS methods relying on wild-type blockers will however be more complex than HOT_ARMS PCR and will not be universally applicable since each mutation will require specific probes manufacturing which may not be favourable depending on sequence context. Moreover, the addition of wild-type blockers severely impacts PCR efficiency which means that it would not be suitable for treatment-response monitoring. However, in instances where wild-type blocking probes may be required such as companion diagnostics and early detection, HOT_ARMS can integrate this

adaptation along with other adaptations such as penultimate mismatch and 3’ LNA due to the HOT_ARMS modification presenting no reduction in PCR efficiency. Methods such as COLD-PCR and its derivatives [81] can enrich low frequency mutant alleles allowing detection down to a MAF of 0.1%. These do however require a second step (i.e., sequencing or mutation screening) to confirm the presence of mutations and as we have previously shown, the method is unreliable at present. Digital-droplet PCR[74, 82] has boasted impressive limits of detection but false positives at low copy number (≤10) [84] leads to a lack of low end sensitivity. Furthermore, over dependence on probes leads to a lack of scalability across multiple genes due to the probe sequence being locked by the primer location and fragmented nature of ctDNA. Digital droplet PCR is also expensive, and the machines are not widely distributed among clinical laboratories which currently are heavily invested in real-time PCR and pyrosequencing machines. Modified NGS protocols such as SafeSeq have been reported to detect MAF as low as 0.001% [64, 83] and have dominated liquid biopsy research but there are several issues moving into the clinic for serial ctDNA analysis. The main flaw with modified NGS protocols is that they are too expensive for the UK NHS in terms of cost per sample for longitudinal sample tracking. A PCR alternative will

always be cheaper, and it would be far more cost effective if NGS was only deployed in scenario's where broad-scale analysis is required such as on FFPE tissue, white blood cells or a one-time snapshot on ctDNA. The use of NGS for detection of several previously identified mutations is unnecessary and a waste of resources that could be spent in other areas of cancer patient management. Furthermore, genetics labs across the NHS already have access to PCR machines and are fully trained, allowing a much faster deployment. This has been further validated by the FDA approving PCR tests for liquid biopsy and at present there is a minimal role for NGS outside of the research arena. NGS is extremely slow in comparison to PCR and could not easily be implemented in scenario's where quick turnaround may be required, such as outpatient departments and GP surgeries which would add a significant economic advantage by not requiring further appointments to be made.

HOT_ARMS PCR can be used to detect any mutation which causes a sequence change including SNVs and indels. If the specific sequence changes can be identified in structural variants, it could be used to test for these too. As we have shown, HOT_ARMS PCR is readily applicable to all mutations and therefore patient specific primer sets can be established as soon as the mutation profile is known. Tumour surveillance

may become a major part of the cancer care pathway especially in the wake of data showing that tumour specific mutations can be detected in the cfDNA of patients up to a year before recurrence becomes clinically overt [154, 155]. Since HOT_ARMS PCR can be undertaken within two hours (even faster with rapid cycling programs) and does not require complex data interpretation, it could provide a result within the time scale of a hospital outpatient appointment. Slightly more ambitious, but equally feasible, would be the establishment of a patient-specific HOT_ARMS PCR assays for treatment response monitoring of rectal cancer patients and tumour surveillance in the primary care setting which is discussed further in section 7.C.iii.

ii. HOT_PI PCR is an effective mutation enrichment method for liquid biopsy and can complement HOT_ARMS PCR

HOT_ARMS PCR is most efficient when there is prior knowledge of the sequence changes induced by mutations. Thus, it is an inefficient method for mutation screening of pre-treatment liquid biopsies to determine mutation status unless there is a very limited spectrum of sequence change that can be targeted. In CRC, since tumours are highly driven by mutations, it may be possible to screen pre-treatment liquid biopsies rather than FFPE tissue. This may increase processing

speed and preserve precious tissue for histology. The majority of patients (roughly 80-90%) contain at least one mutation in one of the following genes: TP53, APC, KRAS, BRAF, PIK3CA[146]. It is possible that HOT_ARMS could be combined with fluorescent hydrolysis probes to form a multiplexed approach. However, for mutation hotspot clusters such as KRAS this is extremely inefficient. Fluorescent hydrolysis probe integration with ARMS primers on mutation hotspot clusters is not possible since the primers overlap. This is possible with scorpion probes[147, 156] although these have not been assessed and so we cannot determine if this is compatible with HOT_ARMS modifications. However, by using a wild-type blocking approach such as HOT_PI, the number of tests for screening can be effectively reduced. This is particularly important with liquid biopsy where the total amount of DNA extracted is very low. HOT_PI is capable of >60x mutation enrichment and when combined with high-resolution melting analysis has a limit of detection of 0.1% MAF which is suitable for liquid biopsy testing. Moreover, HOT_PI can be processed with the same mastermix and cycling protocols as HOT_ARMS for seamless integration of mutation hotspot cluster testing. This will aid in determining mutation status of patients for treatment-response monitoring of rectal cancer and it may be possible to use for early

detection of cancer, particularly in people with hereditary conditions or family history of colorectal cancer to complement colonoscopy.

iii. Common redox-homeostasis gene expression is not predictive of response to radiotherapy in rectal cancer patients

This study aimed to identify if any key players of the redox-homeostasis proteome[97] could predict response to radiotherapy. RT-qPCR expression analysis of 20 rectal biopsy samples including 10 known responders and 10 known non-responders by Mandard score has determined that there was no difference in RNA expression for 16 of the 26 total genes investigated between rectal cancer responders and non-responders to radiotherapy. The 16 genes are as follows: PRDX1, PRDX2, PRDX3, PRDX4, PRDX6, GPX2, TXNDR1, CAT, SOD1, GSTP1, GLRX3, GSR, GSS, GCLC, TXNIP and TXN2. Whilst the sample set is small, power calculations are unnecessary as a good predictive marker should give a clear-cut difference between responders and non-responders even on a small sample set. The remaining 10 genes, GPX3, GPX4, GPX5, GPX6, GPX7, GCLM, GLRX, GLRX5, GSTZ1 and SOD3 were not highly expressed and could not be evaluated. Whilst there was no difference in expression for these key genes, this does not preclude other genes from the redox-homeostasis

proteome being predictive but does rule these specific key gene's out. Moreover, it does not preclude these enzymes from being predictive at a protein level since expression analysis cannot determine protein expression that is controlled by stabilisation mechanisms[132].

C: Future perspectives

i. HOT_ARMS mismatch ranking system for reduced bleed-through

Preliminary data in this study suggest that the introduction of 3' penultimate mismatches into HOT_ARMS primers may reduce wild-type bleed through further and in turn enhance the robustness of low MAF detection below 0.1%. The results in table 7 (section 5.B.ii) describe the ΔCt values between 50% MAF and wild-type for all the HOT_ARMS assays. Table 7 has shown that certain sequence changes would generate more bleed-through than others. In particular, the transition mutations, G>A, A>G, C>T and T>C were among the worst performers and the transversion mutations such as G>T, G>C, C>A and C>G were among the best performers. This difference arises because of the comparative stability of mismatches.

There are only a small number of sources which can influence base-pairing stability and they can be distilled into mismatch-

intrinsic (i.e., the properties of the template and the primer 3' base) and mismatch-extrinsic (such as nearest neighbour, base stacking, serum Mg etc.)[157]. Regarding the MM-Intrinsic, there are two factors: the physical constraints imposed by the bases and the hydrogen bonds. Purine (A, G) are characterised by two carbon rings and pyrimidines (T, C) have a single ring as shown in figure 34 below.

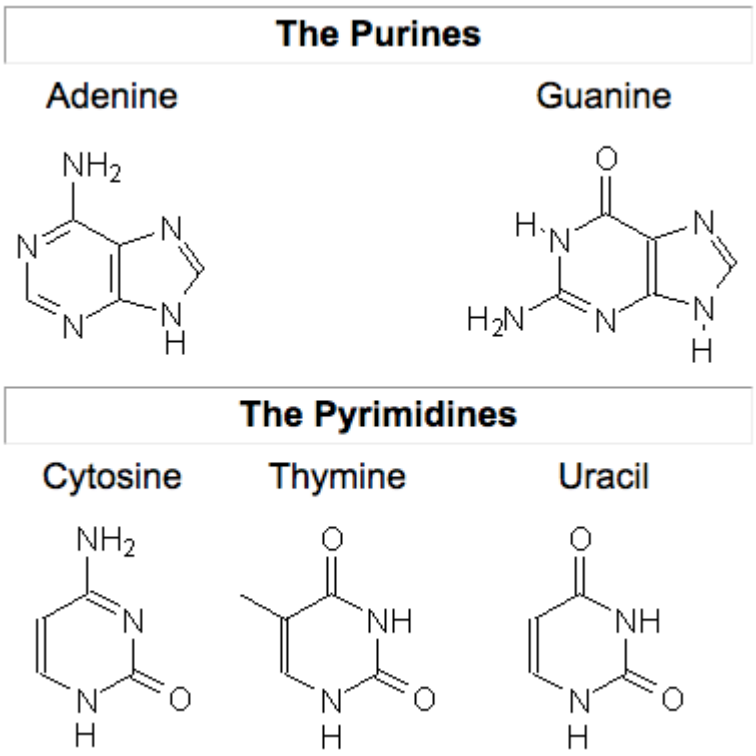


Figure 34. Purine and pyrimidine categories and structure.

In the natural state, a purine pairs up with a pyrimidine (R:Y) and maximum stability is obtained[157]. It follows from this, that transversion mutations which result in a purine:purine (R:R) or pyrimidine:pyrimidine (Y:Y) base pairing will be less

stable than transition mutations which preserve the purine:pyrimidine base pairing (R:Y base pairing such as such as A:C / T:G). Between the R:R and the Y:Y base pairings, our data show that R:R is less stable than Y:Y. This may be due to something as simple as physical space occupancy i.e., the purine pairs may occupy too much physical space and may deform the double helix and affect the adjacent bases whilst the pyrimidine pairs do not occupy the available physical space and deform the double helix in a different way. While the nature of the bases is the main factor, the formation of the hydrogen bonds can have a subtle effect i.e., the A:A base pair has less stability than the G:G base pair and is therefore less likely to show bleed through.

From this we can create a ranking system for both HOT_ARMS assays and mismatch introduction (Table 16). HOT_ARMS PCR assays can be broken down into 2 types: Type 1 mismatch – transversion mutations. These mutations will allow robust testing without needing any further optimising. Type 2 mismatch - transition mutations. These may be at risk of bleed-through and may therefore benefit from a second mismatch. Moreover, a sub-classification of categories can be used to determine the destabilizing capability of wild-type DNA primer binding at the 3’ end which is as follows: Category 1 - severely destabilizing, Category 2 - highly destabilizing,

Category 3 - moderately destabilizing, Category 4 - mildly destabilizing.

Table 16. Mismatch instability ranking system for ARMS PCR.

Type	Category	Sequence change	Actual mismatch	Ring ratio	H+ bonds
1	1	C>A	A:G - Purine:Purine	2:2	2:3
1	1	C>G	G:G - Purine:Purine	2:2	3:3
1	1	T>A	A:A - Purine:Purine	2:2	2:2
1	1	T>G	G:A - Purine:Purine	2:2	3:2
1	2	A>T	T:T - Pyrimidine:Pyrimidine	1:1	2:2
1	2	G>T	T:C - Pyrimidine:Pyrimidine	1:1	2:3
1	2	G>C	C:C - Pyrimidine:Pyrimidine	1:1	3:3
1	2	A>C	C:T - Pyrimidine:Pyrimidine	1:1	3:2
2	3	G>A	A:C - Purine:Pyrimidine	2:1	2:3
2	3	A>G	G:T - Purine:Pyrimidine	2:1	3:2
2	4	C>T	T:G - Pyrimidine:Purine	1:2	2:3
2	4	T>C	C:A - Pyrimidine:Purine	1:2	3:2

There is no known algorithm or ranking system for ARMS sequence changes currently. Our data supports the ranking system for HOT_ARMS assays but further research needs to be undertaken to confirm the effect of mismatches using this system. For the poorly performing transition mutations, mismatches can be introduced to reduce bleed-through in any of the 3 bases upstream of the 3’ end. This would leave us with 9 different mismatch combinations per primer to optimise as each base can be replaced with 3 other types of nucleotide. The ranking system could also be applied to these changes. When a G>A (A:C) change was introduced at the penultimate base, that is one of the worst performers, it was not a strong choice and resulted in no reduction in bleed-through. Recently, we introduced a C>A (A:G) mutation into one of the primers and the bleed-through was reduced by approximately 6 cycles.

The ranking system may be able to increase the performance of our weaker HOT_ARMS assays and if the modification permits, this may increase the specificity of all HOT_ARMS assays sufficiently for rapid PCR testing (30-minute cycling time). Currently only HOT_ARMS assays with ΔCt>15 can be combined with rapid cycling whilst maintaining a 0.1% MAF limit of detection. This is due to the mastermix and ramping speed reducing PCR specificity. The ranking system and

algorithm would significantly reduce the amount of optimisation required for mismatch integration and allow for an overall more streamlined approach.

ii. HOT_ARMS for other types of sequence variant

HOT_ARMS11 (APC c.4287_4296delAACCATGCCA, p.Q1429fs*41) detects a deletion by ARMS PCR which we believe to be novel but not surprising. We hypothesised that HOT_ARMS PCR could be used for the detection of all sequence variants. Future work will focus on determining experimentally if this is possible. Primers will be designed to detect these sequence variants as shown below:

a) Example insertion: Nucleotide highlighted in bold is inserted

Original sequence: GATCGGATTGGCAAATTAGCCGTAGGCCGG

New sequence: GATCGGATT**T**GGCAAATTAGCCGTAGGCCGG

Primer: GATCGGATTT

b) Example amplification: Nucleotides highlighted in bold amplified x4

Original sequence: GATCGGAT**TGGC**AAATTAGCCGTAGGCCGG

New sequence:
GATCGGATTGGC**TGGCTGGCTGGCTGGC**AAATTAGCCGTAGGCCGG

Primer: GATCGGATTGGCT

C) Example rearrangement: Nucleotides highlighted bold are moved to another site in the sequence in bold

Original sequence: GATCGGATTGGCAAATTAGCCGT**AGGCCG**G

New sequence: GATCGGATTG**AGGCCG**GCAAATTAGCCGTG

Primer sequence: GATCGGATTGA

It is highly likely that HOT_ARMS will be able to achieve detection of rearrangements, amplifications and insertions based on the detection of a small deletion by HOT_ARMS. This will generate a larger pool of mutations to screen in liquid biopsy by PCR effectively, increasing the sensitivity of liquid biopsy testing.

iii. HOT_PCR pipeline (N_CaRT)

HOT_ARMS PCR is extremely scalable due to high primer binding specificity and the ultra-sensitivity achieved across all sequence variations without complications. The combination of scalability with low cost opens the opportunity for personalised testing. We have assumed that due to overall cost, a cancer screening program by liquid biopsy is less likely to be adopted in comparison to a liquid biopsy test specifically for cancer patients. With this in mind, we decided to prioritise

tumour surveillance and tumour-response monitoring workflows.

Our approach is to build a personalised liquid biopsy mutation screening panel for each patient based on tumour mutation profiling beforehand. The pipeline we have designed, is referred to as the Nottingham Cancer Recurrence Test (N_CaRT). We have recently been awarded funds to validate the N_CaRT proof of concept via the Medical Research Council Confidence in Concept programme (May 2020-May 2021) and Midlands Innovation Commercialisation of Research Accelerator (MICRA) programme (Nov 2019-April 2020). N_CaRT utilises HOT_ARMS PCR as shown in figure 35 below.

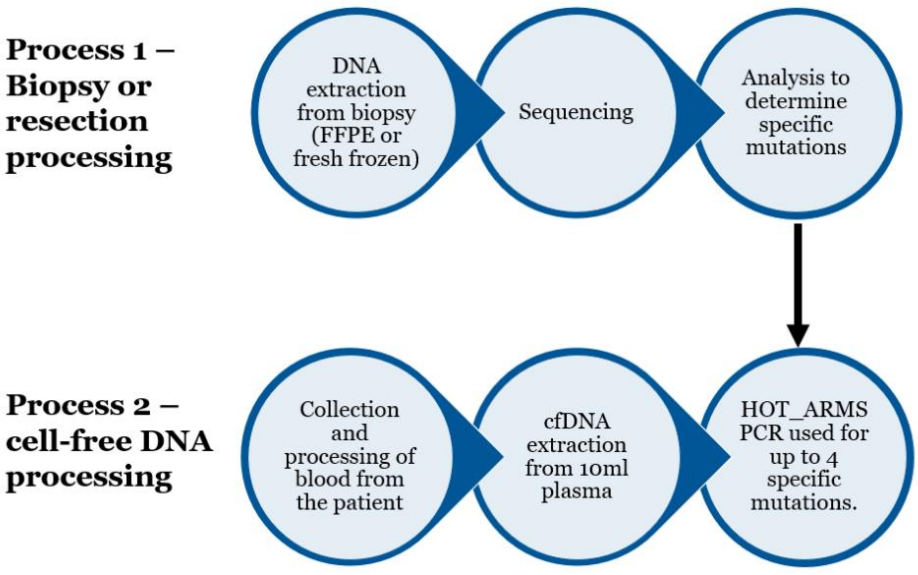


Figure 35. The process for tumour surveillance or treatment-response monitoring by HOT_ARMS PCR.

Process 1 outlines the pathway for tumour mutation profiling. Whole exome or genome sequencing will allow the maximum number of mutations to be identified for the liquid biopsy HOT_ARMS panel. However, other mutation detection methodologies can be used for smaller scale panel systems such as multiplexed HOT_ARMS PCR combined with HOT_PI PCR. Process 2 represents the pathway for cell-free DNA mutation detection. The cell-free DNA will be divided up to detect up to 4 or more specific mutations using singleplex HOT_ARMS. Multiple blood samples can also be taken and pooled to increase the amount of available DNA if necessary. A panel of 4 or more mutations will increase the chance of discovering a true-positive cancer presence[65, 90, 154]. This is especially important with most cancer treatments which can drive tumour homogeneity and where minor clone mutation signals can be lost. The further benefit of a personalised approach is that targeting several passenger mutations with high mutant allele frequency ensures that tumour heterogeneity does not impact tracking. It is unlikely that a passenger mutation will be lost and if it is present at high frequency it is likely to be in most clones. Furthermore, if several passenger mutations are targeted this increases the statistical likelihood that heterogeneity will not be an issue. Both process 1 and 2 can be carried out individually or

simultaneously apart from the last step in process 2 which requires tumour specimen mutation identification beforehand either by liquid biopsy or FFPE tissue. We hypothesise that this personalised approach will give every cancer patient the opportunity to have benefit from cfDNA analysis rather than a minority based on hotspot mutations.

Each patient will require custom panels which will individually require assay validation as new sequences are generated. Whilst cell lines cannot always provide the mutation of interest; companies such as IDT (https://eu.idtdna.com/pages/products/genes-and-gene-fragments/double-stranded-dna-fragments/gblocks-gene-fragments) are able to rapidly create contrived sequence at 170bp lengths to mimic cfDNA. These contrived sequences can be spiked into sheared cell line gDNA such as the globally wild-type HEK293T for assay validation. We envisage that this process can become streamlined over time as hotspot mutations are validated and a bank of assays are generated.

We visualise that treatment response monitoring of rectal cancer patients using HOT_ARMS will be most optimal using the personalised approach as described in figure 35. As shown below in figure 36 (top), the current pathway for radiotherapy has several critical time points for monitoring of response. We

plan to add liquid biopsy testing to each step (figure 36 bottom); encompassing pre-surgery, post-surgery, during radiotherapy and follow up.

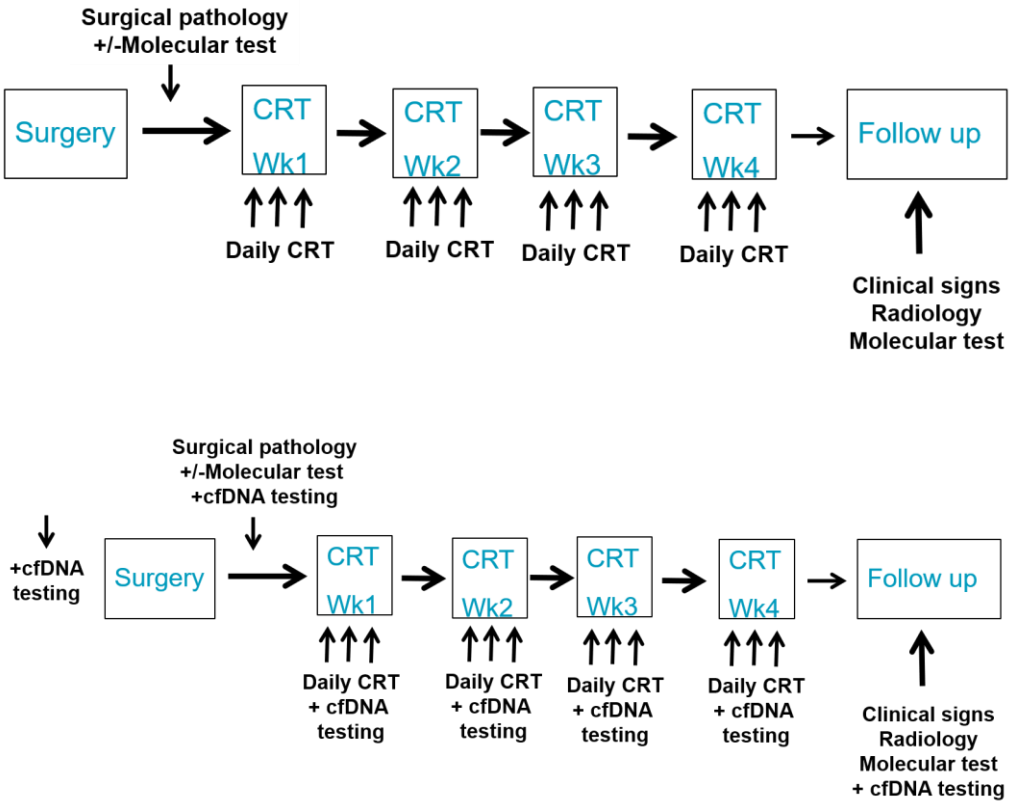


Figure 36. Current (top) and future (bottom) pathway for adjuvant rectal cancer treatment.

A massive advantage of PCR-based technology over sequencing is speed of turn around. Once patients have their tumours profiled, a personalised panel can be allocated to them, this panel can then be used for follow-up analysis at GP clinics and out-patient departments as shown below in figure 37. This process will become even more real as companies are

creating automated extraction and PCR systems for liquid biopsy which require minimal user input.

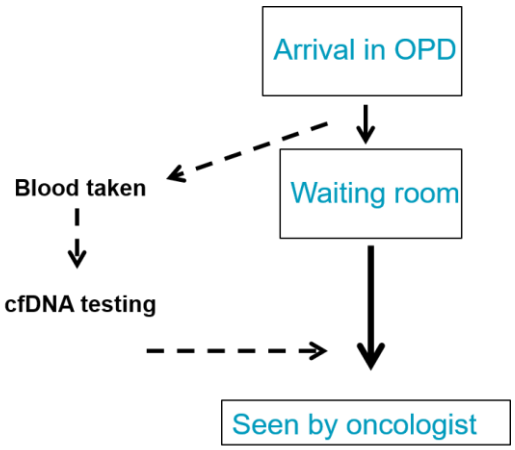


Figure 37. Future pathway for out-patient department.

iv. Further suggestions

- Investigate glutathione and thioredoxin proteins by immunohistochemistry or reverse-phase protein array to determine if they are predictive of response to radiotherapy. This would be via protein stabilisation effects.
- Investigate glutathione and thioredoxin proteins with known radioresistant rectal cancer cell lines and perform knock-out. Knock-out cell lines could be treated with radiotherapy to see if there is an increase in radiosensitivity when glutathione or thioredoxin pathway proteins are downregulated.
- Investigate KRAS mutations as a predictive marker of radiotherapy using HOT_ARMS. This could be done using the left-over cDNA from the study.

7. Final discussion

- Investigate ctDNA as a predictive marker of radiotherapy using HOT_PCR methods.

D: Limitations of the thesis research

Whilst the planned research has been carried out as well as it could have been given budget restraints and lack of sample availability, these are inevitably a limitation of the research. In relation to this the limitations and ideal scenarios are highlighted below.

Section 3 – Predicting rectal cancer responsiveness by biomarker analysis: The limitations of this chapter are the number of samples and that the analysis was undertaken by RT-qPCR only. In an ideal and realistic scenario, it would have been preferable to complete a power calculation which would identify the number of samples required. It is likely that this number would be over 100. In turn, this would increase the power of the statistical analysis carried out to identify differences in RNA expression between responders and non-responders of chemoradiotherapy. However, it is still likely that since rigorous quality control metrics were put in place, a small sample set will suffice in determining a good predictive biomarker.

Whilst RT-qPCR provides information on gene expression which can lead to changes in protein expression it does not inform what is happening to the protein downstream. The protein may have a

feedback loop which causes it to be targeted for ubiquitylation or it may be stabilised causing it to be overexpressed. The use of immunohistochemistry and/or reverse phase protein array would provide deeper insights into the glutathione, thioredoxin and related pathways expression between responders and non-responders of chemoradiotherapy for rectal cancer patients. This would allow a clear conclusion to be generated. As discussed previously, antibodies are extremely expensive and the availability of patient biopsy blocks with Mandard score is limited at Queen's Medical Centre. Furthermore, the amount of biopsy tissue per block which can be spared for research would not be sufficient for protein-based analysis.

Section 4 – Investigation of circulating cell-free DNA characteristics: The limitation of this chapter is primarily the amount of plasma used for extraction. In an ideal scenario up to 10ml of plasma should be used for cfDNA extraction to ensure that ctDNA can be detected. Unfortunately, this was not available as the surgeon's had a cap of 30ml blood per patient on surgery day of which the majority was given to other research projects. Another draw-back is the lack of comparison with other extraction kits for cfDNA. However, there was data in the public domain to suggest that the Qiagen kits were generating the highest yield and in 2017 the number of available kits was limited. In 2021, improved kits may generate increased yield

such as the DNA/RNA combined extraction kit from Qiagen or their magnetic bead-based kit. The last limitation for this chapter is that fragmentation analysis was carried out by PCR only. Whilst this is a true representation for the downstream PCR analysis, it would have been preferable to perform a Tapestation assay to get a full view of the fragment profile.

Section 5 – Development of Highly Optimised annealing Temperature_Amplifcation Refractory Mutation System PCR (HOT_ARMS PCR) as a method to detect mutations in liquid biopsy: HOT_ARMS was thoroughly validated on cell line DNA and a small number of cell-free DNA samples. A limitation of this research is that the number of cell-free DNA samples tested is small and were not from patients with Rectal cancer. In an ideal scenario, 10ml plasma from rectal cancer patients would be collected pre- and post-chemoradiotherapy to investigate the utility of ctDNA in treatment-response monitoring. Unfortunately, due to a crisis in the Biobank, this was not possible, but the development of HOT_ARMS will allow this to be investigated thoroughly in the future. It would also have been ideal to have plasma samples from normal patients to determine specificity. Limitations of the HOT_ARMS method include issues with the detection of G>A, A>G, C>T and T>C mutations. The latter two are unnecessary to detect as the ARMS primer can be designed in the opposite direction. HOT_ARMS has not completely

abolished wild-type amplification with these mutations without the aid of further modifications. Whilst penultimate mismatches are a quick solution to this, this adds significant optimisation time. However, this is only a limitation of the HOT_ARMS method against other types of mutation detection methods. It is still superior to all other ARMS methods which would struggle with these types of sequence change. The detection of these sequence changes would only be necessary with companion diagnostics where a single oncogene mutation is key. In this scenario, it is likely that time to optimise additional modifications for the protocol would be available. The development of a personalised approach for treatment-response monitoring, recurrence detection and minimal residual disease also counters this weakness as it ensures that assays can be designed for favourable mutations which can be targeted immediately with limited optimisation by ordering a single primer set. In this scenario it is key to be able to have limited optimisation so that patient primer panels can be developed quickly. At present a further limitation is that primer design needs to be completed manually. To develop a personalised system, we need to build software which is capable of automatically designing HOT_ARMS primers based on the sequencing data generated from a primary tumour. This will significantly reduce the time for pipeline completion, but it would be simple to design a piece of software

such as this. A further limitation is inherent with ARMS PCR, it can only detect a small number of mutations at a time. Generally, ARMS PCR can detect one mutation and it is possible to multiplex to detect up to 5 in a single reaction. This means that ARMS is not always useful as a screening tool and fits into specific niche's where either the mutation is already known, or it is a common hotspot. Fortunately, in rectal cancer, many hotspot mutations exist which would allow HOT_ARMS to be used for screening, although this would only allow for a limited number of mutations to be detected per patient statistically which would reduce overall sensitivity and specificity of the approach. A personalised HOT_ARMS PCR will still rely on NGS for profiling of tumour DNA which incurs an expensive up-front cost. This is a limitation of the personalised approach as either the sample will need to be sent away to an external company for sequencing or the set up will require NGS. However, for longitudinal tracking it is logical to pair the two methods to ensure that costs remain low as this is unfeasible by NGS.

Section 6 – HOT_PCR: Highly Optimised annealing Temperature_Probe Inhibited PCR (HOT_PI PCR) as a method to detect mutations in hotspot clusters for liquid biopsy: Key limitations of HOT_PI PCR development and validation is firstly the number of genes in which the method has been developed for. Whilst HOT_PI PCR has been designed for KRAS mutation

hotspot clusters which is an extremely useful target as it is mutated in 40% of rectal cancer patients; it has not yet been shown if it is as effective in other genes which means that the potential for scalability has not been determined. A further limitation of the HOT_PI PCR research is that it has not be tested out on cfDNA samples. Whilst the limitation of detection of HOT_PI PCR is sufficient and the protocol is designed for fragmented DNA, this does not necessarily mean that the approach will translate well to cfDNA. It is likely though that as high-resolution melting analysis on cell line DNA was extremely reliable, the method will likely translate. PCR-HRMA is known to be reliable on samples containing low concentration fragmented DNA as the analysis is end-point and so the important factor is the mutant allele frequency which HOT_PI can dramatically enrich.

8. References

1. Kadric, N., D. Mesic, and M. Bazardzanovic, *[Anatomy of the colon, rectum and anus].* Med Arh, 2003. **57**(3 Suppl 1): p. 29-31.
2. Wikipedia. *Large intestone*. 2020; Available from: https://en.wikipedia.org/wiki/Large_intestine#/media/File:Blausen_0604_LargeIntestine2.png.
3. Fazeli, M.S. and M.R. Keramati, *Rectal cancer: a review.* Med J Islam Repub Iran, 2015. **29**: p. 171.
4. UK, C.R. *Bowel Cancer Statistics*. 2017 [cited 2017 31st May]; Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer.
5. UK, C.R. *Bowel Cancer Incidence Statistics*. 2017 [cited 2017 31st May]; Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/incidence.
6. Fleming, M., et al., *Colorectal carcinoma: Pathologic aspects.* J Gastrointest Oncol, 2012. **3**(3): p. 153-73.
7. Shiller, M. and S. Boostrom, *The molecular basis of rectal cancer.* Clin Colon Rectal Surg, 2015. **28**(1): p. 53-60.
8. Cho, K.R. and B. Vogelstein, *Genetic alterations in the adenoma--carcinoma sequence.* Cancer, 1992. **70**(6 Suppl): p. 1727-31.
9. Papagiorgis, P., *Colorectal cancer: dichotomous or continuum model? Perhaps, a combination of both.* Gut, 2013. **62**(10): p. 1519-20.
10. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
11. Keum, N. and E. Giovannucci, *Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies.* Nat Rev Gastroenterol Hepatol, 2019. **16**(12): p. 713-732.
12. East, J.E., et al., *British Society of Gastroenterology position statement on serrated polyps in the colon and rectum.* Gut, 2017. **66**(7): p. 1181-1196.
13. Kim, J.E., et al., *Mutation Burden and I Index for Detection of Microsatellite Instability in Colorectal Cancer by Targeted Next-Generation Sequencing.* J Mol Diagn, 2019. **21**(2): p. 241-250.
14. Komor, M.A., et al., *Consensus molecular subtype classification of colorectal adenomas.* J Pathol, 2018. **246**(3): p. 266-276.
15. Jain, S. and J.I. Clark, *Ipilimumab for the treatment of melanoma.* Melanoma Manag, 2015. **2**(1): p. 33-39.
16. Angell, H.K., et al., *The Immunoscore: Colon Cancer and Beyond.* Clin Cancer Res, 2020. **26**(2): p. 332-339.
17. Pitroda, S.P., S.J. Chmura, and R.R. Weichselbaum, *Integration of radiotherapy and immunotherapy for treatment of oligometastases.* Lancet Oncol, 2019. **20**(8): p. e434-e442.

18. Fontana, E., et al., *Context matters-consensus molecular subtypes of colorectal cancer as biomarkers for clinical trials.* Ann Oncol, 2019. **30**(4): p. 520-527.
19. Sugai, T., et al., *Analysis of molecular alterations in left- and right-sided colorectal carcinomas reveals distinct pathways of carcinogenesis: proposal for new molecular profile of colorectal carcinomas.* J Mol Diagn, 2006. **8**(2): p. 193-201.
20. Jo, W.S. and J.M. Carethers, *Chemotherapeutic implications in microsatellite unstable colorectal cancer.* Cancer Biomark, 2006. **2**(1-2): p. 51-60.
21. Glimelius, B., et al., *Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Ann Oncol, 2013. **24 Suppl 6**: p. vi81-8.
22. Ardekani, A.E., et al., *False-positive elevated CEA during colon cancer surveillance: a cholecystitis case report diagnosed by PET-CT scan.* J Surg Case Rep, 2019. **2019**(6): p. rjz138.
23. support, M.c. *Staging and grading of bowel cancer*. 2020; Available from: https://www.macmillan.org.uk/cancer-information-and-support/bowel-cancer/staging-and-grading-of-bowel-cancer.
24. UK, B.c., *Bowel cancer screening.* 2020.
25. Logan, R.F., et al., *Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests.* Gut, 2012. **61**(10): p. 1439-46.
26. Gamaleldin, M., et al., *Is routine use of adjuvant chemotherapy for rectal cancer with complete pathological response justified?* Am J Surg, 2017. **213**(3): p. 478-483.
27. Morris, E.J., et al., *Wide Variation in the Use of Radiotherapy in the Management of Surgically Treated Rectal Cancer Across the English National Health Service.* Clin Oncol (R Coll Radiol), 2016. **28**(8): p. 522-31.
28. Thies, S. and R. Langer, *Tumor regression grading of gastrointestinal carcinomas after neoadjuvant treatment.* Front Oncol, 2013. **3**: p. 262.
29. Chua, Y.J., *Pathological complete response: still a relevant endpoint in rectal cancer?* The Lancet Oncology, 2010. **11**(9): p. 807-808.
30. Glynne-Jones, R. and R. Hughes, *Complete Response after Chemoradiotherapy in Rectal Cancer (Watch-and-Wait): Have we Cracked the Code?* Clin Oncol (R Coll Radiol), 2016. **28**(2): p. 152-60.
31. Martin, S.T., H.M. Heneghan, and D.C. Winter, *Systematic review and meta-analysis of outcomes following pathological complete response to neoadjuvant chemoradiotherapy for rectal cancer.* Br J Surg, 2012. **99**(7): p. 918-28.
32. Mohiuddin, M., et al., *Randomized phase II study of neoadjuvant combined-modality chemoradiation for distal rectal cancer: Radiation Therapy Oncology Group Trial 0012.* J Clin Oncol, 2006. **24**(4): p. 650-5.

33. Francois, Y., et al., *Influence of the interval between preoperative radiation therapy and surgery on downstaging and on the rate of sphincter-sparing surgery for rectal cancer: the Lyon R90-01 randomized trial.* J Clin Oncol, 1999. **17**(8): p. 2396.
34. Gerard, J.P., et al., *Clinical complete response (cCR) after neoadjuvant chemoradiotherapy and conservative treatment in rectal cancer. Findings from the ACCORD 12/PRODIGE 2 randomized trial.* Radiother Oncol, 2015. **115**(2): p. 246-52.
35. Maas, M., et al., *Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data.* The Lancet Oncology, 2010. **11**(9): p. 835-844.
36. Dhadda, A.S., et al., *Prognostic importance of Mandard tumour regression grade following pre-operative chemo/radiotherapy for locally advanced rectal cancer.* Eur J Cancer, 2011. **47**(8): p. 1138-45.
37. Behr, T.M., et al., *High-linear energy transfer (LET) alpha versus low-LET beta emitters in radioimmunotherapy of solid tumors: therapeutic efficacy and dose-limiting toxicity of 213Bi- versus 90Y-labeled CO17-1A Fab' fragments in a human colonic cancer model.* Cancer Res, 1999. **59**(11): p. 2635-43.
38. Steel, G.G., T.J. McMillan, and J.H. Peacock, *The 5Rs of radiobiology.* Int J Radiat Biol, 1989. **56**(6): p. 1045-8.
39. Harrington, K., P. Jankowska, and M. Hingorani, *Molecular biology for the radiation oncologist: the 5Rs of radiobiology meet the hallmarks of cancer.* Clin Oncol (R Coll Radiol), 2007. **19**(8): p. 561-71.
40. Beyzadeoglu, M., G. Ozyigit, and C. Ebruli, *Radiobiology.* 2010: p. 71-144.
41. Niemantsverdriet, M., et al., *High and low LET radiation differentially induce normal tissue damage signals.* Int J Radiat Oncol Biol Phys, 2012. **83**(4): p. 1291-7.
42. Hu, M., et al., *Proton beam therapy for cancer in the era of precision medicine.* J Hematol Oncol, 2018. **11**(1): p. 136.
43. NHS. *Proton beam therapy*. 2020; Available from: https://www.england.nhs.uk/commissioning/spec-services/highly-spec-services/pbt/.
44. Taylor, A. and M.E. Powell, *Intensity-modulated radiotherapy--what is it?* Cancer Imaging, 2004. **4**(2): p. 68-73.
45. Guan, X., *Cancer metastases: challenges and opportunities.* Acta Pharm Sin B, 2015. **5**(5): p. 402-18.
46. Kalyan, A., S. Rozelle, and A. Benson, 3rd, *Neoadjuvant treatment of rectal cancer: where are we now?* Gastroenterol Rep (Oxf), 2016. **4**(3): p. 206-9.
47. Coombes, R.C., et al., *Personalized Detection of Circulating Tumor DNA Antedates Breast Cancer Metastatic Recurrence.* Clin Cancer Res, 2019. **25**(14): p. 4255-4263.

48. Smith, J.J., et al., *Assessment of a Watch-and-Wait Strategy for Rectal Cancer in Patients With a Complete Response After Neoadjuvant Therapy.* JAMA Oncol, 2019. **5**(4): p. e185896.
49. Massihnia, D., et al., *Liquid biopsy for rectal cancer: A systematic review.* Cancer Treat Rev, 2019. **79**: p. 101893.
50. Mattox, A.K., et al., *Applications of liquid biopsies for cancer.* Sci Transl Med, 2019. **11**(507).
51. Gallaher, J.A., J.S. Brown, and A.R.A. Anderson, *The impact of proliferation-migration tradeoffs on phenotypic evolution in cancer.* Sci Rep, 2019. **9**(1): p. 2425.
52. Schwarzenbach, H., *Circulating nucleic acids as biomarkers in breast cancer.* Breast Cancer Res, 2013. **15**(5): p. 211.
53. Tie, J., et al., *Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer.* Sci Transl Med, 2016. **8**(346): p. 346ra92.
54. Lennon, A.M., et al., *Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention.* Science, 2020.
55. Bi, F., et al., *Circulating tumor DNA in colorectal cancer: opportunities and challenges.* Am J Transl Res, 2020. **12**(3): p. 1044-1055.
56. Wan, J.C.M., et al., *Liquid biopsies come of age: towards implementation of circulating tumour DNA.* Nat Rev Cancer, 2017. **17**(4): p. 223-238.
57. Natera, *How can Signatera be applied in clinical research and drug development?* 2020.
58. Marketsandmarkets, *Liquid Biopsy Market by Cancer Type.* 2020.
59. Pfeiffer, F., et al., *Systematic evaluation of error rates and causes in short samples in next-generation sequencing.* Sci Rep, 2018. **8**(1): p. 10950.
60. Sun, K., *Clonal hematopoiesis: background player in plasma cell-free DNA variants.* Ann Transl Med, 2019. **7**(Suppl 8): p. S384.
61. Reece, M., et al., *The Use of Circulating Tumor DNA to Monitor and Predict Response to Treatment in Colorectal Cancer.* Front Genet, 2019. **10**: p. 1118.
62. Huang, K., et al., *Circulating tumor DNA sequencing for colorectal cancers: A comparative analysis of colon cancer and rectal cancer data.* Cancer Biomark, 2019. **26**(3): p. 313-322.
63. Tie, J., et al., *Serial circulating tumour DNA analysis during multimodality treatment of locally advanced rectal cancer: a prospective biomarker study.* Gut, 2019. **68**(4): p. 663-671.
64. Kinde, I., et al., *Detection and quantification of rare mutations with massively parallel sequencing.* Proc Natl Acad Sci U S A, 2011. **108**(23): p. 9530-5.
65. Reinert, T., et al., *Analysis of Plasma Cell-Free DNA by Ultradeep Sequencing in Patients With Stages I to III Colorectal Cancer.* JAMA Oncol, 2019.

66. Underhill, H.R., et al., *Fragment Length of Circulating Tumor DNA.* PLoS Genet, 2016. **12**(7): p. e1006162.
67. Qiagen, *QIAamp circulating nucleic acid kit.* 2020.
68. Mouliere, F., et al., *Circulating Cell-Free DNA from Colorectal Cancer Patients May Reveal High KRAS or BRAF Mutation Load.* Transl Oncol, 2013. **6**(3): p. 319-28.
69. Elazezy, M. and S.A. Joosse, *Techniques of using circulating tumor DNA as a liquid biopsy component in cancer management.* Comput Struct Biotechnol J, 2018. **16**: p. 370-378.
70. Han, X., J.Y. Wang, and Y.L. Sun, *Circulating Tumor DNA as Biomarkers for Cancer Detection.* Genomics Proteomics & Bioinformatics, 2017. **15**(2): p. 59-72.
71. Mouliere, F., et al., *Circulating Cell-Free DNA from Colorectal Cancer Patients May Reveal High KRAS or BRAF Mutation Load.* Translational Oncology, 2013. **6**(3): p. 319-U281.
72. Stadler, J., et al., *SNPase-ARMS qPCR: Ultrasensitive Mutation-Based Detection of Cell-Free Tumor DNA in Melanoma Patients.* PLoS One, 2015. **10**(11): p. e0142273.
73. Tsiatis, A.C., et al., *Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications.* J Mol Diagn, 2010. **12**(4): p. 425-32.
74. Taylor, S.C., G. Laperriere, and H. Germain, *Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data.* Sci Rep, 2017. **7**(1): p. 2409.
75. Lin, M.T., et al., *Clinical validation of KRAS, BRAF, and EGFR mutation detection using next-generation sequencing.* Am J Clin Pathol, 2014. **141**(6): p. 856-66.
76. Ebili, H.O., et al., *QMC-PCRx: a novel method for rapid mutation detection.* J Clin Pathol, 2017.
77. Cohen, V., et al., *Evaluation of denaturing high-performance liquid chromatography as a rapid detection method for identification of epidermal growth factor receptor mutations in nonsmall-cell lung cancer.* Cancer, 2006. **107**(12): p. 2858-65.
78. Little, S., *Amplification-refractory mutation system (ARMS) analysis of point mutations.* Curr Protoc Hum Genet, 2001. **Chapter 9**: p. Unit 9 8.
79. Huang, J.F., et al., *Single-Tubed Wild-Type Blocking Quantitative PCR Detection Assay for the Sensitive Detection of Codon 12 and 13 KRAS Mutations.* PLoS One, 2015. **10**(12): p. e0145698.
80. Milbury, C.A., et al., *COLD-PCR: improving the sensitivity of molecular diagnostics assays.* Expert Rev Mol Diagn, 2011. **11**(2): p. 159-69.
81. How Kit, A., et al., *Sensitive detection of KRAS mutations using enhanced-ice-COLD-PCR mutation enrichment and direct sequence identification.* Hum Mutat, 2013. **34**(11): p. 1568-80.

82. O'Leary, B., et al., *Comparison of BEAMing and Droplet Digital PCR for Circulating Tumor DNA Analysis.* Clin Chem, 2019. **65**(11): p. 1405-1413.
83. Newman, A.M., et al., *An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage.* Nat Med, 2014. **20**(5): p. 548-54.
84. Rowlands, V., et al., *Optimisation of robust singleplex and multiplex droplet digital PCR assays for high confidence mutation detection in circulating tumour DNA.* Sci Rep, 2019. **9**(1): p. 12620.
85. Thierry, A.R., *A Targeted Q-PCR-Based Method for Point Mutation Testing by Analyzing Circulating DNA for Cancer Management Care.* Methods Mol Biol, 2016. **1392**: p. 1-16.
86. Martinez-Saez, O., et al., *Frequency and spectrum of PIK3CA somatic mutations in breast cancer.* Breast Cancer Res, 2020. **22**(1): p. 45.
87. Han, A.L., et al., *Comparison of cobas EGFR Mutation Test v2 and PANAMutyper-R-EGFR for Detection and Semi-Quantification of Epidermal Growth Factor Receptor Mutations in Plasma and Pleural Effusion Supernatant.* Ann Lab Med, 2019. **39**(5): p. 478-487.
88. (FDA), F.a.D.A., *List of Cleared or Approved Companion Diagnostic Devices (In Vitro and Imaging Tools).* 2020.
89. Natera, *FDA Grant Breakthrough Debice Designation to Natera's Signatera Test.* 2020.
90. Li, B.T., et al., *Ultra-deep next-generation sequencing of plasma cell-free DNA in patients with advanced lung cancers: results from the Actionable Genome Consortium.* Ann Oncol, 2019. **30**(4): p. 597-603.
91. Kang, S., et al., *CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA.* Genome Biol, 2017. **18**(1): p. 53.
92. Cohen, J.D., et al., *Detection and localization of surgically resectable cancers with a multi-analyte blood test.* Science, 2018. **359**(6378): p. 926-930.
93. Vivancos, A., et al., *Comparison of the Clinical Sensitivity of the Idylla Platform and the OncoBEAM RAS CRC Assay for KRAS Mutation Detection in Liquid Biopsy Samples.* Sci Rep, 2019. **9**(1): p. 8976.
94. Lim, S.H., et al., *Predictive and prognostic biomarkers for neoadjuvant chemoradiotherapy in locally advanced rectal cancer.* Crit Rev Oncol Hematol, 2015. **96**(1): p. 67-80.
95. Sallam, R.M., *Proteomics in cancer biomarkers discovery: challenges and applications.* Dis Markers, 2015. **2015**: p. 321370.
96. Yao, A., et al., *Thioredoxin System Protein Expression Is Associated with Poor Clinical Outcome in Adult and Paediatric Gliomas and Medulloblastomas.* Mol Neurobiol, 2020. **57**(7): p. 2889-2901.
97. Zhang, Y. and S.G. Martin, *Redox proteins and radiotherapy.* Clin Oncol (R Coll Radiol), 2014. **26**(5): p. 289-300.

98. Woolston, C.M., et al., *The prognostic and predictive power of redox protein expression for anthracycline-based chemotherapy response in locally advanced breast cancer.* Mod Pathol, 2012. **25**(8): p. 1106-16.
99. Khaled Al-Hadyan, S.S., Abed M Zaiutoun, Dileep N Lobo, Stewart Martin, *Thioredoxin system protein expression in carcinomas of the pancreas, bile duct and ampulla.*
100. Rotblat, B., et al., *Anti-oxidative stress response genes: bioinformatic analysis of their expression and relevance in multiple cancers.* Oncotarget, 2013. **4**(12): p. 2577-90.
101. Fischer, J., et al., *Peroxiredoxins in Colorectal Cancer: Predictive Biomarkers of Radiation Response and Therapeutic Targets to Increase Radiation Sensitivity?* Antioxidants (Basel), 2018. **7**(10).
102. Hill, E.J., et al., *Oxaliplatin as a radiosensitiser for upper and lower gastrointestinal tract malignancies: what have we learned from a decade of translational research?* Crit Rev Oncol Hematol, 2012. **83**(3): p. 353-87.
103. Lawrence, T.S., A.W. Blackstock, and C. McGinn, *The mechanism of action of radiosensitization of conventional chemotherapeutic agents.* Semin Radiat Oncol, 2003. **13**(1): p. 13-21.
104. Allegra, C.J., et al., *Neoadjuvant 5-FU or Capecitabine Plus Radiation With or Without Oxaliplatin in Rectal Cancer Patients: A Phase III Randomized Clinical Trial.* J Natl Cancer Inst, 2015. **107**(11).
105. Deorukhkar, A. and S. Krishnan, *Targeting inflammatory pathways for tumor radiosensitization.* Biochem Pharmacol, 2010. **80**(12): p. 1904-14.
106. Painuli, S. and N. Kumar, *Prospects in the development of natural radioprotective therapeutics with anti-cancer properties from the plants of Uttarakhand region of India.* J Ayurveda Integr Med, 2016. **7**(1): p. 62-8.
107. Lin, C., et al., *Combination of quercetin with radiotherapy enhances tumor radiosensitivity in vitro and in vivo.* Radiother Oncol, 2012. **104**(3): p. 395-400.
108. Dumont, F., A. Altmeyer, and P. Bischoff, *Radiosensitising agents for the radiotherapy of cancer: novel molecularly targeted approaches.* Expert Opin Ther Pat, 2009. **19**(6): p. 775-99.
109. Sun, X.X. and Q. Yu, *Intra-tumor heterogeneity of cancer cells and its implications for cancer treatment.* Acta Pharmacol Sin, 2015. **36**(10): p. 1219-27.
110. Dhadda, A.S., et al., *Mandard tumour regression grade, perineural invasion, circumferential resection margin and post-chemoradiation nodal status strongly predict outcome in locally advanced rectal cancer treated with preoperative chemoradiotherapy.* Clin Oncol (R Coll Radiol), 2014. **26**(4): p. 197-202.

111. Kresse, S.H., et al., *Evaluation of commercial DNA and RNA extraction methods for high-throughput sequencing of FFPE samples.* PLoS One, 2018. **13**(5): p. e0197456.
112. Dieffenbach, C.W., T.M. Lowe, and G.S. Dveksler, *General concepts for PCR primer design.* PCR Methods Appl, 1993. **3**(3): p. S30-7.
113. Fadhil, W., et al., *Quick-multiplex-consensus (QMC)-PCR followed by high-resolution melting: a simple and robust method for mutation detection in formalin-fixed paraffin-embedded tissue.* J Clin Pathol, 2010. **63**(2): p. 134-40.
114. Untergasser, A., et al., *Primer3-new capabilities and interfaces.* Nucleic Acids Research, 2012. **40**(15).
115. Dieffenbach, C.W., T.M.J. Lowe, and G.S. Dveksler, *General Concepts for Pcr Primer Design.* Pcr-Methods and Applications, 1993. **3**(3): p. S30-S37.
116. Ebili, H.O., et al., *QMC-PCRx: a novel method for rapid mutation detection.* Journal of Clinical Pathology, 2017. **70**(8): p. 702-711.
117. Ebili, H.O., et al., *"Squirrel" Primer-Based PCR Assay for Direct and Targeted Sanger Sequencing of Short Genomic Segments.* J Biomol Tech, 2017. **28**(3): p. 97-110.
118. Ballantyne, K.N., R.A. van Oorschot, and R.J. Mitchell, *Locked nucleic acids in PCR primers increase sensitivity and performance.* Genomics, 2008. **91**(3): p. 301-5.
119. Mitsuoka, Y., et al., *A bridged nucleic acid, 2',4'-BNA COC: synthesis of fully modified oligonucleotides bearing thymine, 5-methylcytosine, adenine and guanine 2',4'-BNA COC monomers and RNA-selective nucleic-acid recognition.* Nucleic Acids Res, 2009. **37**(4): p. 1225-38.
120. Paulasova, P. and F. Pellestor, *The peptide nucleic acids (PNAs): a new generation of probes for genetic and cytogenetic analyses.* Ann Genet, 2004. **47**(4): p. 349-58.
121. Latorra, D., et al., *Enhanced allele-specliflic PCR discriminationin SNP genotypling using 3 ' locked nucleic acid (LNA) primers.* Human Mutation, 2003. **22**(1): p. 79-85.
122. Levin, J.D., et al., *Position-dependent effects of locked nucleic acid (LNA) on DNA sequencing and PCR primers.* Nucleic Acids Research, 2006. **34**(20).
123. Latorra, D., K. Arar, and J.M. Hurley, *Design considerations and effects of LNA in PCR primers.* Molecular and Cellular Probes, 2003. **17**(5): p. 253-259.
124. Han, J.Y., et al., *PNA clamping-assisted fluorescence melting curve analysis for detecting EGFR and KRAS mutations in the circulating tumor DNA of patients with advanced non-small cell lung cancer.* BMC Cancer, 2016. **16**: p. 627.
125. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. **25**(4): p. 402-8.

126. Medrano, R.F. and C.A. de Oliveira, *Guidelines for the tetra-primer ARMS-PCR technique development.* Mol Biotechnol, 2014. **56**(7): p. 599-608.
127. Ayyadevara, S., J.J. Thaden, and R.J. Shmookler Reis, *Discrimination of primer 3'-nucleotide mismatch by taq DNA polymerase during polymerase chain reaction.* Anal Biochem, 2000. **284**(1): p. 11-8.
128. Zeka, F., et al., *Straightforward and sensitive RT-qPCR based gene expression analysis of FFPE samples.* Sci Rep, 2016. **6**: p. 21418.
129. Dowling, C.M., et al., *The importance of selecting the appropriate reference genes for quantitative real time PCR as illustrated using colon cancer cells and tissue.* F1000Res, 2016. **5**: p. 99.
130. Vandesompele, J., et al., *Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes.* Genome Biol, 2002. **3**(7): p. RESEARCH0034.
131. Trzpis, M., et al., *Epithelial cell adhesion molecule: more than a carcinoma marker and adhesion molecule.* Am J Pathol, 2007. **171**(2): p. 386-95.
132. Tian, Q., et al., *Integrated genomic and proteomic analyses of gene expression in Mammalian cells.* Mol Cell Proteomics, 2004. **3**(10): p. 960-9.
133. Tost, J., *The clinical potential of Enhanced-ice-COLD-PCR.* Expert Rev Mol Diagn, 2016. **16**(3): p. 265-8.
134. Thierry, A.R., et al., *Clinical validation of the detection of KRAS and BRAF mutations from circulating tumor DNA.* Nat Med, 2014. **20**(4): p. 430-5.
135. Diadx, *A Blood Test for Real-Time Diagnostics.* 2020.
136. Kitazumi, I. and M. Tsukahara, *Regulation of DNA fragmentation: the role of caspases and phosphorylation.* FEBS J, 2011. **278**(3): p. 427-41.
137. Bettegowda, C., et al., *Detection of circulating tumor DNA in early- and late-stage human malignancies.* Sci Transl Med, 2014. **6**(224): p. 224ra24.
138. Wan, J.C., et al., *Liquid biopsies come of age: towards implementation of circulating tumour DNA.* Nat Rev Cancer, 2017. **17**(4): p. 223-238.
139. Krug, A.K., et al., *Improved EGFR mutation detection using combined exosomal RNA and circulating tumor DNA in NSCLC patient plasma.* Ann Oncol, 2018. **29**(10): p. 2143.
140. Qiagen, *QIAamp MinElute ccfDNA Kits.* 2020.
141. Diefenbach, R.J., et al., *Evaluation of commercial kits for purification of circulating free DNA.* Cancer Genet, 2018. **228-229**: p. 21-27.
142. Satyal, U., A. Srivastava, and P.H. Abbosh, *Urine Biopsy-Liquid Gold for Molecular Detection and Surveillance of Bladder Cancer.* Front Oncol, 2019. **9**: p. 1266.

143. Newton, C.R., et al., *Analysis of Any Point Mutation in DNA - the Amplification Refractory Mutation System (Arms).* Nucleic Acids Research, 1989. **17**(7): p. 2503-2516.
144. Yakovchuk, P., E. Protozanova, and M.D. Frank-Kamenetskii, *Base-stacking and base-pairing contributions into thermal stability of the DNA double helix.* Nucleic Acids Research, 2006. **34**(2): p. 564-574.
145. Sundaralingam, M. and P.K. Ponnuswamy, *Stability of DNA duplexes with Watson-Crick base pairs: A predicted model.* Biochemistry, 2004. **43**(51): p. 16467-16476.
146. COSMIC, 2020.
147. Qiagen, *therascreen KRAS RCQ PCR kit.* 2020.
148. Salvatore Marras, D.V.-G., Sanjay Tyagi, Fred Russell Kramer, *Highly selective nucleic acid amplification primers*
149. Zhou, L., et al., *Enrichment and detection of rare alleles by means of snapback primers and rapid-cycle PCR.* Clin Chem, 2010. **56**(5): p. 814-22.
150. Hersh Ham-Karim, H.O.E., Wakkas Fadhil, Abutaleb Asiri, James C Hassall, Mohammad Ilyas, *COLD-HRM: A Combination of Methods to Infer the Nature of Somatic Mutations.* Journal of submicroscopic cytology and pathology, 2017. **2**(2): p. 1-9.
151. How-Kit, A. and J. Tost, *Pyrosequencing(R)-Based Identification of Low-Frequency Mutations Enriched Through Enhanced-ice-COLD-PCR.* Methods Mol Biol, 2015. **1315**: p. 83-101.
152. VILLAHERMOSA JAEN, M.L.M.D.P., Juan, *Method for detection of KRAS mutations.*
153. Qu, S., et al., *Detection of low-level DNA mutation by ARMS-blocker-Tm PCR.* Clin Biochem, 2016. **49**(3): p. 287-91.
154. Shaw, J.A., et al., *Mutation Analysis of Cell-Free DNA and Single Circulating Tumor Cells in Metastatic Breast Cancer Patients with High Circulating Tumor Cell Counts.* Clin Cancer Res, 2017. **23**(1): p. 88-96.
155. Shaw, J.A., et al., *Genomic analysis of circulating cell-free DNA infers breast cancer dormancy.* Genome Res, 2012. **22**(2): p. 220-31.
156. Thelwell, N., et al., *Mode of action and application of Scorpion primers to mutation detection.* Nucleic Acids Res, 2000. **28**(19): p. 3752-61.
157. Yakovchuk, P., E. Protozanova, and M.D. Frank-Kamenetskii, *Base-stacking and base-pairing contributions into thermal stability of the DNA double helix.* Nucleic Acids Res, 2006. **34**(2): p. 564-74.

9. Appendix

A. Primer sequences for fragmentation analysis and fragmented high-resolution melting analysis primers

Supplementary table 1. Details of primers used for circulating nucleic acid kit fragmentation assessment by qPCR and primers for the squirrel primer assay (Sequencing).

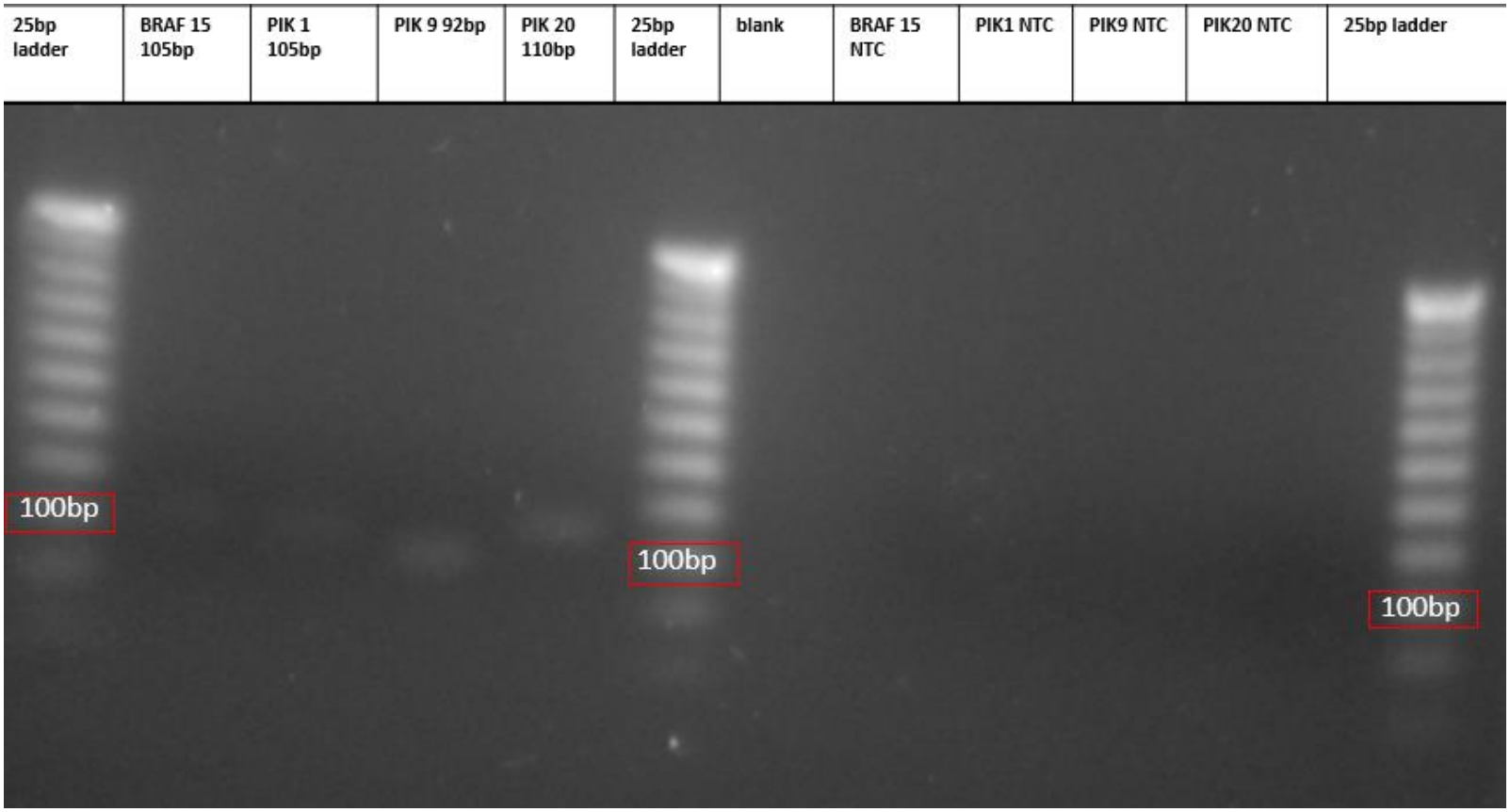
Target	Primer sequence	Amplicon length	Estimated primer Tm (F/R)	Primer Length (F/R)	GC% (F/R)
KRAS exon 4	F: 5′ GACACAAAACAGGCTCAGGACT 3′ R: 5′ CAGATCTGTATTTATTTCAGTGTTA 3′	105bp	64.4/56.8	22/25	50.0/28.0
TP53 exon 5	F: 5′ AGTACTCCCCTGCCCTCAAC 3′ R: 5′ CTGCTCACCATCGCTATCTG 3′	194bp	64.6/61.7	20/20	60.0/55.0
KRAS exon 2 seq	F: 5′ AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAATGAATATAAACTTGTGGTAGTTGG 3′ R: 5′ TAGACTCCTGATCCCTTCATTGCCCTGCATCTGACACGCAGCTGTATCGTCAAGGCACTCT 3′	142bp	79.2/84.6	64/61	40.6/52.5
Useq	F: 5′ CTGACTAAACTAGGTGCCACGT 3′ R: 5′ TAGACTCCTGATCCCTTCAT 3′	N/A	63.6/58.0	22/20	50.0/45.0

9. Appendix

Supplementary table 2. Details of primers designed to target highly fragmented DNA for purposes of high-resolution melting on FFPE DNA.

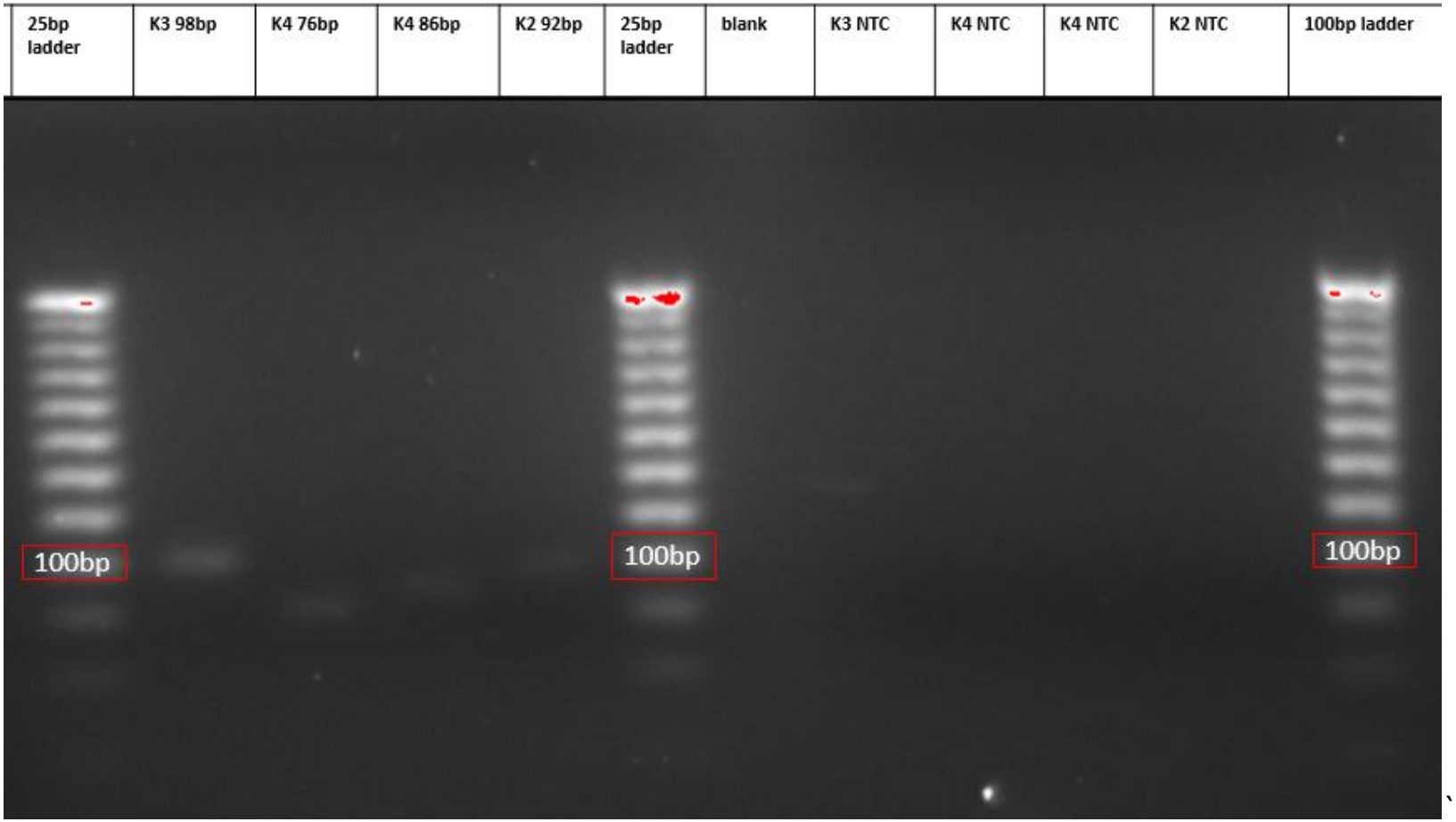
Target	Primer sequence	Mutation hotspot	Amplicon length (bp)
KRAS exon 2 (Mack et al, 2016)	F: 5’ TGAATTAGCTGTATCGTCAAGGCACT 3’ R: 5’ TTATAAGGCCTGCTGAAAATGACTGAA 3’	Codon 12/13	92
KRAS exon 3	F: 5’ TGTCTCTTGGATATTCTCGACA 3’ R: 5’ GCAAATACACAAAGAAAGCCC 3’	Codon 61	98
KRAS exon 4	F: 5’ CAAAACAGGCTCAGGACTTA 3’ R: 5’ TTCAGTGTTACTTACCTGTCT 3’	Codon 146	86
PIK3CA exon 1	F: 5’ GAAGCAGAAAGGGAAGAATTT 3’ R: 5’ TTCTTCACGGTTGCCTACT 3’	Codon 88	105
PIK3CA exon 9	F: 5’ GCTCAAAGCAATTTCTACACGAGA 3’ R: 5’ TCCATTTTAGCACTTACCTGTGAC 3’	Codon 542/545	92
PIK3CA exon 20	F: 5’ AGATAAAACTGAGCAAGAGG 3’ R: 5’ GCTGTTTAATTGTGTGGAAG 3’	Codon 1047	110
BRAF exon 15	F: 5’ TTCATGAAGACCTCACAGTAAA 3’ R: 5’ ACAAAATGGATCCAGACAACT 3’	Codon 600	105

9. Appendix



Supplementary figure 1. Gel electrophoresis for primers listed in supp. table 2.

9. Appendix



Supplementary figure 2. Gel electrophoresis for primers listed in supp. table 2.

B. CfDNA concentration data

Supplementary table 3. Concentration of cfDNA from patient blood samples extracted with the Qiagen circulating nucleic acid kit. Lists volume of plasma used and elution volume used. Concentration is determined by DeNovix high sensitivity assay kit and is representative of double-stranded DNA. To calculate the total DNA (ng) per ml of plasma: (elution volume (µl) x DNA concentration (ng/ml)) ÷ volume of plasma (ml). Pre-operative samples were named day 1 and post-operative samples were collected the following day which was named day 2 onwards (weekends permitting)

Sample (elution 1)	Vol of plasma (ml)	Vol of elution (µl)	Concentration (ng/µl)	Total DNA per ml of plasma (ng/ml)
2 Day 1	0.75	40	3.02	161.06
2 Day 2	1	20	1.43	28.6
3 Day 1	0.55	40	2.11	153.45
3 Day 4	1	20	2.67	53.4
4 Day 1	0.4	40	2.44	244
4 Day 2	1	20	3.61	72.2
5 Day 1	1	40	1.29	51.6
5 Day 2	1	20	4.15	83
6 Day 1	1	40	7.35	294
6 Day 2	1	20	4.9	98

7 Day 1	1	40	1.99	79.6
7 Day 2	1	20	2.02	40.4
8 Day 1	1	40	1.84	73.6
8 Day 2	1	20	1.85	37
9 Day 1	1	40	1.77	70.8
9 Day 2	1	20	2.63	52.6
10 Day 1	1	40	1.65	66
10 Day 2	1	20	2.61	52.2
11 Day 1	1	40	1.26	50.40
11 Day 2	1	20	1.87	37.40
12 Day 1	1	40	1.43	57.20
12 Day 2	1	20	1.82	36.40
13 Day 1	1	40	1.06	42.40
13 Day 2	1	20	0.92	18.48
14 Day 1	1	40	1.13	45.20
14 Day 2	1	20	0.96	19.34
15 Day 1	1	40	1.49	59.60
15 Day 2	1	20	2.31	46.20
16 Day 1	0.6	20	1.13	37.60
16 Day 2	0.6	20	1.83	61.00
17 Day 1	0.6	20	2.86	95.33

17 Day 2	0.6	20	1.89	63.00
18 Day 1	1	40	1.40	56.00
18 Day 2	1	20	1.98	39.60
19 Day 1	1	40	1.07	42.80
19 Day 2	1	20	2.01	40.20
20 Day 1	1	40	1.36	54.40
20 Day 2	1	20	3.61	72.20
21 Day 1	1	40	1.48	59.20
21 Day 2	1	20	1.89	37.80
22 Day 1	1	40	1.86	74.4
22 Day 2	1	20	2.91	58.2
23 Day 1	1	40	0.233	9.32
23 Day 2	1	20	0.298	5.96
24 Day 1	1	40	0.118	4.72
24 Day 2	1	20	0.189	3.78
25 Day 1	1	40	0.302	12.08
25 Day 2	1	20	2.59	51.8
26 Day 1	1	40	0.399	15.96

C. IntPlex ARMS PCR

Supplementary table 4. Details of IntPlex primers.

Common name	Mutation hotspot	Primer sequence	Mutation hotspot	Estimated primer Tm	Primer Length	GC%
G13D	c.38G>A	5′ GCACTCTTGCCTACGT 3′	c.38G>A	56.6	16	56.2%
G12D	c.35G>A	5′ CTCTTGCCTACGCCAT 3′	c.35G>A	56.3	16	56.2%
G12A	c.35G>C	5′ CTCTTGCCTACGCCAG 3′	c.35G>C	57.2	16	62.5%
Universal forward	N/A	5′ GCCTGCTGAAAATGACTGA 3′	N/A	59.3	19	47.4%
Blocker	N/A	5′ GCCTACGCCACCAGCTC-PHO 3′	N/A	63.6	17	70.6%

D. HOT_ARMS primers, modifications and cell-line sources

Supplementary table 5. Sequence of primers for HOT_ARMS primer pairs, HOT_ARMS blockers and HOT_PI primer pair and blockers. Moreover, they both pair with the same common wild-type primer. If a base is in brackets i.e., (X) it denotes the use of locked nucleic acid on that base (LNA).

Ref	**ID number**	**Direction**	**ARMS/WT**	**Target gene**	**Sequence change**	**Primer sequence (5′ to 3′)**	**Amplicon length (bp)**
HOT_ARMS1	1	Forward	ARMS	KRAS	c.38G>A	CTTGTGGTAGTTGGAGCTGGTGA	60
HOT_ARMS2a	2	Reverse	ARMS	KRAS	c.34G>T	GCACTCTTGCCTACGCCACA	60
HOT_ARMS2b	3	Forward	ARMS	KRAS	c.34G>T	TATAAACTTGTGGTAGTTGGAGCTT	66
HOT_ARMS3a	4	Reverse	ARMS	KRAS	c.35G>T	GCACTCTTGCCTACGCCAA	60
HOT_ARMS3b	5	Forward	ARMS	KRAS	c.35G>T	ATAAACTTGTGGTAGTTGGAGCTGT	65
HOT_ARMS4	6	Forward	ARMS	KRAS	c.35G>A	ATAAACTTGTGGTAGTTGGAGCTGA	65

HOT_ARMS5	7	Reverse	ARMS	KRAS	c.34G>C	GCACTCTTGCCTACGCCACG	60
HOT_ARMS6	8	Forward	ARMS	KRAS	c.34G>A	TATAAACTTGTGGTAGTTGGAGCTA	66
HOT_ARMS7	9	Reverse	ARMS	KRAS	c.35G>C	GCACTCTTGCCTACGCCAG	60
HOT_ARMS8	10	Forward	ARMS	PIK3CA	c.1624G>A	GCAATTTCTACACGAGATCCTCTCTCTA	65
HOT_ARMS9	11	Forward	ARMS	PIK3CA	c.1633G>A	GAGATCCTCTCTCTGAAATCACTA	69
HOT_ARMS10	12	Forward	ARMS	PIK3CA	c.3140A>G	CATGAAACAAATGAATGATGCACG	80
HOT_ARMS11	13	Forward	ARMS	APC	c.4287_4296 delAACCATG CCA	TGATCTTCCAGATAGCCCTGGACAC	83
HOT_ARMS12	14	Reverse	ARMS	BRAF	c.1799A>T	GGACCCACTCCATCGAGATTTCT	70
HOT_ARMS13	15	Reverse	ARMS	EGFR	c.2369C>T	CGAAGGGCATGAGCTGCA	50
HOT_ARMS14	16	Reverse	ARMS	ESR1	c.1138 G>C	GGATCTCTAGCCAGGCACATTG	69
HOT_ASMS15	17	Forward	ARMS	ESR1	c.1613 A>G	GAACGTGGTGCCCCTCTATG(A) *bracket denotes LNA	71
HOT_ARMS1,2 b,3b,4,5,6	18	Reverse	Wild type	KRAS	N/A	CTGAATTAGCTGTATCGTCAAGGCA	N/A
HOT_ARMS2a, 3a,7	19	Forward	Wild type	KRAS	N/A	GCTGAAAATGACTGAATATAAACTTGTGG T	N/A
HOT_ARMS8	20	Reverse	Wild type	PIK3CA	N/A	TGACTCCATAGAAAATCTTTCTCCTGCT	N/A
HOT_ARMS9	21	Reverse	Wild type	PIK3CA	N/A	ATTTTAGCACTTACCTGTGACTC	N/A
HOT_ARMS10	22	Reverse	Wild type	PIK3CA	N/A	CATGCTGTTTAATTGTGTGGAAGA	N/A
HOT_ARMS11	23	Reverse	Wild type	APC	N/A	ACTTCTCGCTTGGTTTGAGCTGTTT	N/A
HOT_ARMS12	24	Forward	Wild type	BRAF	N/A	TCATGAAGACCTCACAGTAAAAATAGGTG	N/A
HOT_ARMS13	25	Forward	Wild-type	EGFR	N/A	CATCTGCCTCACCTCCACCG	N/A

HOT_ARMS14	26	Forward	Wild-type	ESR1	N/A	TTTCAGGCTTTGTGGATTTGACCCT	N/A
HOT_ARMS15	27	Reverse	Wild-type	ESR1	N/A	GGCTAGTGGGCGCATGTAGG	N/A
HOT_PI1	28	Forward	Wild-type	KRAS	Exon 2	TTATAAGGCCTGCTGAAAATGACTGA	95
HOT_PI1	29	Reverse	Wild-type	KRAS	Exon 2	TTCTGAATTAGCTGTATCGTCAAGGC	95
HOT_PI blocker 1	30	Reverse	Wild-type	KRAS	Exon 2	AAGGCAC(T)CTTGCCTACG(C)(C)A(C)(C)AGCTCCAACTACCACAAGTTTA(T)ATTCAGT-PHO *brackets denotes LNA	N/A
HOT_ARMS blocker 10 LNA	31	Forward	Wild-type	KRAS	Exon 2	AGTT(G)GAGC(T)(G)(G)(T)(G)(G)(C)(G)TAGG(C)AAG-PHO *brackets denotes LNA	N/A
HOT_ARMS ESR1 (D538G) blocker	32	Forward	Wild-type	ESR1	c.1613A>G	TGGTGCCCC(T)C(T)AT(G)(A)(C)CTGC(T)GC(T)GGAGAT-PHO *brackets denotes LNA	N/A

Supplementary table 6. Cell lines used for testing HOT_ARMS PCR. This table lists the cell lines used and the mutations contained in them. Most of the cell lines were available in-house (obtained from the NCI 60 panel) whilst some had to be obtained from external commercial sources.

Gene	Protein sequence/common name	Gene sequence	Cell line harbouring mutation	Cell line tumour source	Zygosity	Cell line DNA sourced from
KRAS	p.G13D	c.38G>A	HCT 116	Colon	Heterozygous	In house

KRAS	p.G12C	c.34G>T	SW837	Colon	Heterozygous	In house
KRAS	p.G12V	c.35G>T	SW480	Colon	Homozygous	In house
KRAS	p.G12D	c.35G>A	GP2D	Colon	Heterozygous	In house
KRAS	p.G12R	c.34G>C	SW48 +G12R	Colon+knock-in	Heterozygous	Horizon discovery
KRAS	p.G12A	c.35G>C	SW111 6	Colon	Heterozygous	Public Health England
KRAS	p.G12S	c.34G>A	A549	Lung	Homozygous	In house
PIK3 CA	p.E542K	c.1624G>A	SW948	Colon	Heterozygous	Public Health England
PIK3 CA	p.E545K	c.1633G>A	MCF7	Breast	Heterozygous	In house
PIK3 CA	p.H1047R	c.3140A>G	HCT 116	Colon	Heterozygous	In house
BRAF	p.V600E	c.1799 T>A	HT-29	Colon	Heterozygous	In house
APC	p.Q1429fs*41	c.4287_4296delAACCA TGCCA	SW1116	Colon	Heterozygous	Public Health England
EGFR	p.T790M	c.2369C>T	RKO +T790M	Colon	Heterozygous	Horizon discovery

Supplementary table 7. Modifications to ARMS or HOT_ARMS PCR and the effect. Whilst developing HOT_ARMS PCR, many experiments have been carried out to improve the method. The final improvements are included in the

data. However, it is necessary in research to also show what did not work or what had minimal impact so that it is not attempted in the future or is expanded on.

Modification	Effect	Comments
Long primers (Eurofins)	Reduced sensitivity compared with HOT_ARMS primers, efficiency similar (30-50nt used)	
Q-solution (Qiagen)	Similar sensitivity, minor efficiency reduction (recommended amount added)	
DMSO (Sigma-Aldrich)	Similar sensitivity, minor efficiency reduction (1-10% used)	
Betaine (Sigma-Aldrich)	Efficiency impacted severely	
GC enhancer (ThermoFisher)	Efficiency impacted severely (recommended amount added)	
Tetramethylammonium chloride (Sigma-Aldrich)	Target-dependent minor increases in sensitivity	Increases in sensitivity not worth the increase in optimisation.
GoTaq mastermix (Promega)	Sensitivity impacted severely compared with in-house mastermix	
SensiFAST mastermix (Bioline)	Sensitivity impacted severely compared with in-house mastermix	

SNPase mastermix (Bioron)	Minor increase in sensitivity compared with in-house mastermix	Too costly and subject to change risks. Manufacturer is small.
Taqman fast advanced mastermix (ThermoFisher)	Sensitivity impacted moderately compared with in-house mastermix	
Taqman universal mastermix II (ThermoFisher)	No amplification possible at high annealing temperature	
3’ penultimate base mismatch (Eurogentec)	Reduction in annealing temperature by 6°C, resulting in loss of sensitivity. Efficiency largely impacted.	
2 LNA’s in both primers (Eurogentec)	When at 5’ end, increase in efficiency occurs. No change in annealing temperature.	
3 LNA’s in both primers (Eurogentec)	When at 5’ end, increase in efficiency occurs. Small change in annealing temperature (1°C).	
5 LNA’s in both primers (Eurogentec)	Amplification lost	
Single 3’ LNA (Eurogentec)	Target-dependent efficiency changes. Amplification sometimes lost and sometimes minor gains in sensitivity occur.	Too target-dependent to carry out more optimisation for a minor sensitivity gain. Most primers lose amplification.

Single 3’ penultimate LNA (Eurogentec)	Abolishes sensitivity. Mutant and wild-type give same ct-value.	
3’ and 3’ penultimate LNA (Eurogentec)	Abolishes sensitivity. Mutant and wild-type give same ct-value.	
Increased polymerase (HotStarTaq – Qiagen)	Reduces sensitivity above 1 unit (between 1 and 10 units used)	
Increased dNTPs (ThermoFisher)	No effect except at high concentrations it becomes inhibitory. 100-1000µM used.	
Increased magnesium (ThermoFisher)	Reduces sensitivity even with annealing temperature gains. 1-14mM used.	
3’ LNA primer (Eurogentec) plus increased magnesium (ThermoFisher) or polymerase (HotStarTaq – Qiagen)	Poor efficiency LNA primers gain efficiency but in return no sensitivity is gained with higher magnesium or more polymerase; or a combination of both.	
3’ LNA primer (Eurogentec) with increased annealing time	No effect compared to lower time (up to 2 minutes used)	
Inhibitory primer containing LNA’s and 3’ phosphate group (Eurogentec)	A blocker with the same sequence as the primer was used to reduce design complexity. Although it worked for mutations	Instead, mutations were located centrally in the blocker.

	present 3 bases away from the 3' end, mutations present in the last bases did not obtain blocking effects.	
3' LNA primer (Eurogentec) with nested pre-amplification	Efficiency increased compared to one-stage 3'LNA primers and gains in sensitivity over regular HOT_ARMS primers in a nested scenario. Will allow multiplexing in first stage and to make better use of the small amount of cfDNA that can be extracted currently.	Although this is great for research, nested procedures appear to be a none-negotiable for diagnostic work.
3' LNA primer (Eurogentec) with 2 stage cycling - reduced annealing temperature in first segment	Target-dependent mild increase in sensitivity, large increase in efficiency compared to single stage 3' LNA primer.	Too much optimisation to apply globally to the method.
3' LNA primer (Eurogentec) with 2 stage cycling - reduced annealing temperature in second segment	The same issues as single stage 3' LNA arise with no improvement.	
Mixture of 3' LNA primer (Eurogentec) plus regular HOT_ARMS primer	No effect	
PFU polymerase (Promega)	Repairs mismatch and wild-type/mutant give the same ct-value.	

DNA only blocker (Eurofins)	No blocking effect occurs	
LNA blocker with LNA not on mutant bases (Eurogentec)	No blocking effect occurs	
DNA (Eurofins) or LNA (Eurogentec) blocker plus holding step	No effect	
LNA blocker plus fast cycling (Eurogentec)	Reduced blocking compared to regular 30 second annealing and extension times.	
Fast cycling polymerase (Fast cycling mastermix – Qiagen)	Reduced sensitivity but increased efficiency. Useful for fast assays where ultra sensitivity is not required.	
Double LNA blocker – Forward and reverse (Eurogentec)	Efficiency severely impacted	
Wild primer LNA blocker (Eurogentec)	Here the blocker prevents single-sided amplification of wild-type DNA caused by the wild-type primer. Reducing the amount of wild-type DNA was proven to have an effect when reducing total template input from 40ng to 5ng. The blocking effect can be powerful but in general, unreliable.	Effect is unreliable because the ARMS primer can still bind to wild-type template on the opposite strand and cause bleed-through.

E. Rectal biopsy primer sequences and efficiency

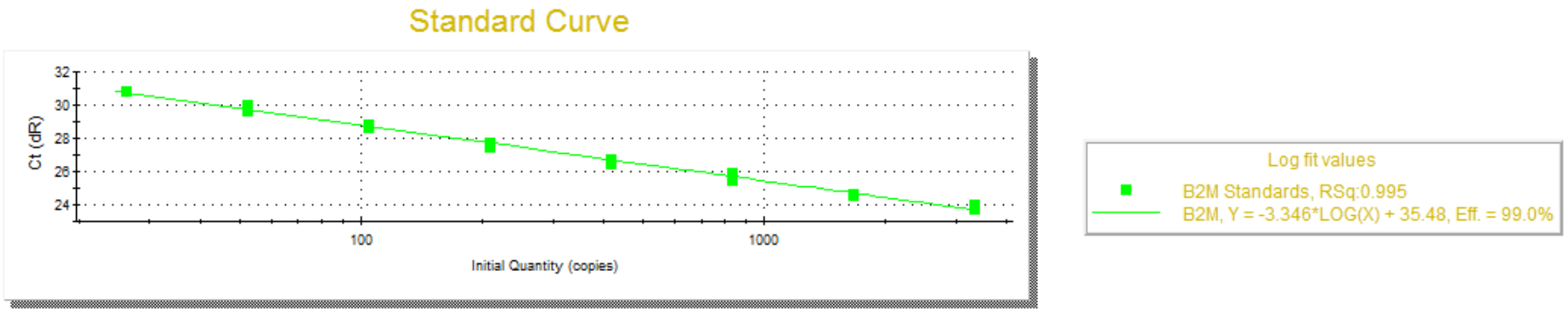
Supplementary table 8. Primers for RT-qPCR to determine rectal cancer predictive biomarkers to radiotherapy response. Transcripts used are listed (Ensembl), the first transcript in the list was used as a reference on NCBI BLAST to determine sequence similarity between the other transcripts listed. Regions which were highly conserved are listed which were targeted by the designed primers. N/A is used if only one transcript was found on Ensembl.

Ref	Protein name	Gene name	Transcripts	Transcript 1 most conserved region (bp)	Primer sequences
1	Peroxiredoxin-1	PRDX1	NM_001202431.1/NM_002574.3/NM_181696.2/NM_181697.3	114-1000	Forward Primer: ACAAACATGGGGAAGTGTGC Reverse Primer: GCTCTTTTGGACATCAGGCTT
2	Peroxiredoxin-2	PRDX2	NM_005809.6	N/A	Forward Primer: TGCCCCACCGAGATCAT Reverse Primer: GGGTGTTGATCCAAGCCAG
3	Peroxiredoxin-3	PRDX3	NM_001302272.1/NM_006793.5	622-1440 OR 54-627	Forward Primer: GGTTCCAGTCAAGCAAAATTATTC Reverse Primer: AGGTCTTTGAACTCTCCATTGA
4	Peroxiredoxin-4	PRDX4	NM_006406.2	N/A	Forward Primer: TGACCCATCAGATCTCAAAGG Reverse Primer: TCAGAGTAATTTGTCTTAGGATTCC
5	Peroxiredoxin-6	PRDX6	NM_004905.3	N/A	Forward Primer: GGAATGTTAAGTTGATTGCCC Reverse Primer: TGGGCTCTTCACAATTGTAA
6	Thioredoxin Reductase 1	TXNRD1	NM_003330.3/NM_182729.2/NM_182742.2/NM_182743.2/NM_001093771.3/NM_001261446.1/NM_001261445.1	900-4000	Forward Primer: CAGTTCGTACCAATTAAAGTTGA Reverse Primer: ATTGGTGGACTGAGCTACTA

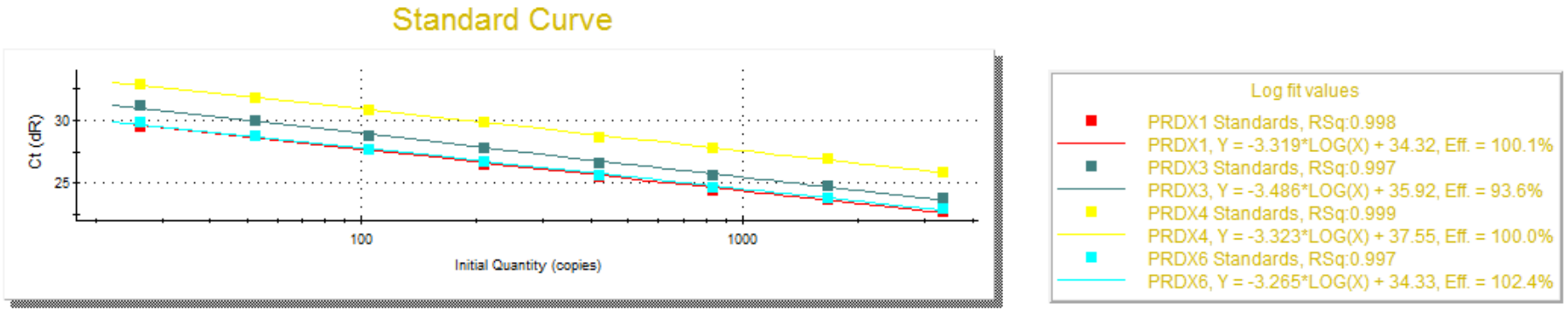
7	Thioredoxin Reductase 3	TXNRD3	NM_052883.2/NM_001173513.1/NM_001039783.1	1942-2321	Forward Primer: CATGTGGGGAGGTGTTCA Reverse Primer: AGCCTTTCTGAGTGATGTCT
8	Thioredoxin 2	TXN2	NM_012473.4	N/A	Forward Primer: AGGTTAGAGAAGATGGTGGC Reverse Primer: CCGCTGACACCTCATACT
9	Thioredoxin interacting protein	TXNIP	NM_006472.5/NM_001313972.1	614-2892	Forward Primer: AAGATCAGGCCTTCTATCCT Reverse Primer: GGAACGCTAACATAGATCAGT
10	Glutamate-Cysteine Ligase Catalytic Subunit	GCLC	NM_001498.4/NM_001197115.1	1-448 OR 481-902 OR 1014-3785	Forward Primer: CAGTTCCTGCACATCTACCAC Reverse Primer: CATGATCAAAAGATACCAACATGT
11	Glutathione Synthetase	GSS	NM_001322495.1/NM_001322494.1/NM_000178.4	1000-2700	Forward Primer: TGCCTCGTCAGTACAGTCTA Reverse Primer: CTGCACCTTCTTAGTCCCAG
12	Glutathione S-Transferase P	GSTP1	NM_000852.3	N/A	Forward Primer: GAGACCTCACCCTGTACCA Reverse Primer: TCCTTCCCATAGAGCCCAA
13	Glutathione Peroxidase 2	GPX2	NM_002083.4	N/A	Forward Primer: CAATTTGGACATCAGGAGAACT Reverse Primer: GTGAAGGTGGGCTGGTATC
14	Glutathione Peroxidase 3	GPX3	NM_002084.5/NM_001329790.1	1-150 OR 160-929 OR 990-1603	Forward Primer: TATGCTGGCAAATACGTCCT Reverse Primer: GTGCAAGCTCTTCCTGTAGT
15	Glutathione Peroxidase 4	GPX4	NM_002085.5/NM_001039848.4/NM_001039847.3	1-540 OR 579-851	Forward Primer: TTTGCCGCCTACTGAAGC Reverse Primer: CGGAAAACTCGTGCATGGA

16	Glutathione Peroxidase 5	GPX5	NM_001509.2/NM_003996.3	362-1452 OR 1-243	Forward Primer: ATCTGCTTCCCCTTCTCCTA Reverse Primer: GTGCCTTTCTCGTCTTTGTG
17	Glutathione Peroxidase 6	GPX6	NM_182701.1	N/A	Forward Primer: CTGTTTGTCAATGTGGCCG Reverse Primer: CCTCCTGTAGTGCATTCAGT
18	Glutathione Peroxidase 7	GPX7	NM_015696.5	N/A	Forward Primer: GTGTCGCTGGAGAAGTACC Reverse Primer: CGGTAGTGCTGGTCTGTG
19	Glutamate-Cysteine Ligase Modifier Subunit	GCLM	NM_002061.4/NM_001308253.1	459-4883	Forward Primer: ACTAGAAGTGCAGTTGACATG Reverse Primer: AGGAGGTGAAGCAATGATCA
20	Superoxide dismutase 1	SOD1	NM_000454.4	N/A	Forward Primer: TCCATGTTCATGAGTTTGGAGA Reverse Primer: CCCACCGTGTTTTCTGGATA
21	Catalase	CAT	NM_001752.4	N/A	Forward Primer: TGAACAGATAGCCTTCGACC Reverse Primer: GTGAGTGTCAGGATAGGCAA
22	Glutathione Reductase	GSR	NM_000637.5/NM_001195102.2/NM_001195103.2/NM_001195104.2	1-600 OR 1200-3000	Forward Primer: ACAGCTGTCCACTCTGAATT Reverse Primer: ATTCAGGCGGCTCACATAG
23	Glutaredoxin	GLRX	NM_001243658.1/NM_002064.2/NM_001118890.2/NM_001243659.1	250-500 OR 1000-1450	Forward Primer: CAAGAACGGTGCCTCGAG Reverse Primer: TTCCCCACTCTGTTGCAAAG

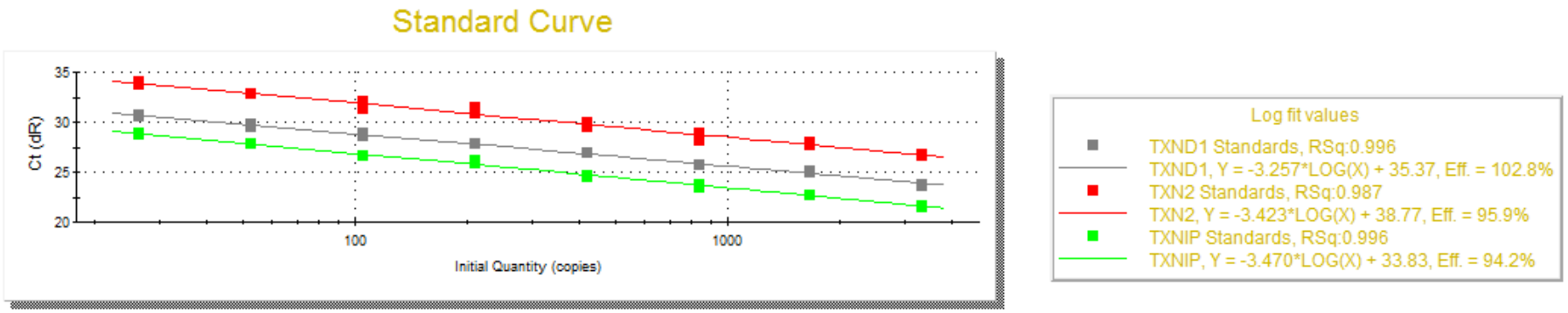
24	Glutaredoxin 3	GLRX3	NM_001199868.1/NM_001321980.1/NM_006541.5	300-1000	Forward Primer: ACAATTTGTCCCAAAGCTCC Reverse Primer: GCATCACAGAAGCTTTATTTGTC
25	Glutaredoxin 5	GLRX5	NM_016417.3	N/A	Forward Primer: CGATTACGCGGCCTACAA Reverse Primer: CCCTACAAACTCGCCATTGA
26	Glutathione S-transferase zeta 1	GSTZ1	NM_145870.3/NM_001312660.1/NM_145871.2	500-1100	Forward Primer: TGACCATGGCTGATCTGTG Reverse Primer: ATGGAGCTGATGGTAGGGTA
27	Superoxide dismutase 3	SOD3	NM_003102.3	N/A	Forward Primer: TGCAGCTCTCTTTTCAGGAG Reverse Primer: CAGGCAGGAACACAGTAGC
28	Homo sapiens caudal type homeobox 2	CDX2	NM_001265.5/NM_001354700.1	1-3517 (no difference)	Forward Primer: GTTTCACTACAGTCGCTACA Reverse Primer: TGAAACCAGATTTTAACCTGC
29	EPCAM/BEREP4	EPCAM	NM_002354.2	N/A	Forward Primer: ATCACAACGCGTTATCAACT Reverse Primer: GAGAAGAATTTTGAACCAGATCA
30	Hypoxanathine Phosphoribosyltransferase 1	HPRT1	NM_000194.3	N/A	Forward Primer: TGGTCAGGCAGTATAATCCA Reverse Primer: TGGCTTATATCCAACACTTCG
31	Ribosomal Protein Lateral Stalk Subunit P0	RPLP0	NM_053275.3/NM_001002.4	250-1250	Forward Primer: CCCTGAAGTGCTTGATATCA Reverse Primer: TGATGCAACAGTTGGGTAG
32	Beta-2-Microglobulin	B2M	NM_004048.3	N/A	Forward Primer: GCGCTACTCTCTCTTTCTG Reverse Primer: CATTCTCTGCTGGATGACG



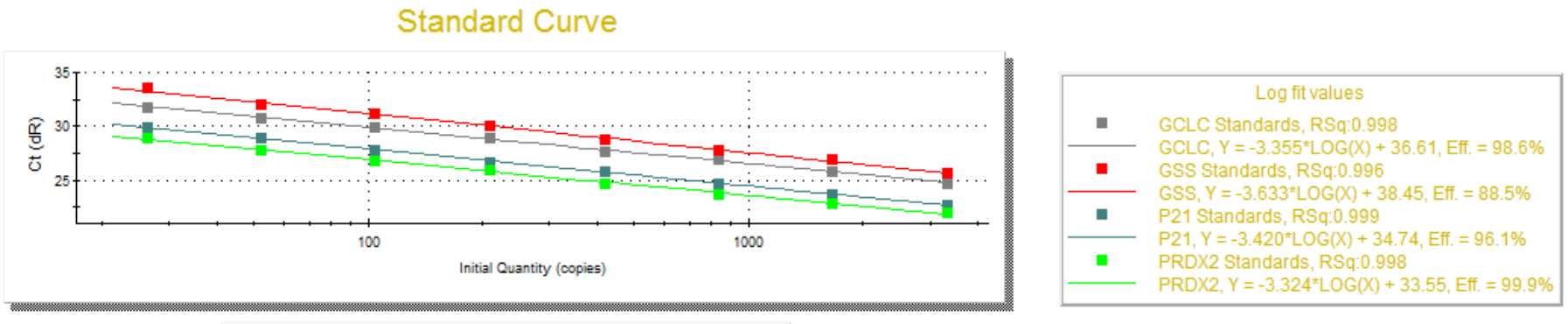
Supplementary figure 3. B2M standard curve to determine primer efficiency. Automatically calculated by Stratagene MX3005P. Efficiency values can be found in the legend on the right.



Supplementary figure 4. PRDX1, PRDX3, PRDX4 and PRDX6 standard curves to determine primer efficiency. Automatically calculated by Stratagene MX3005P. Efficiency values can be found in the legend on the right.

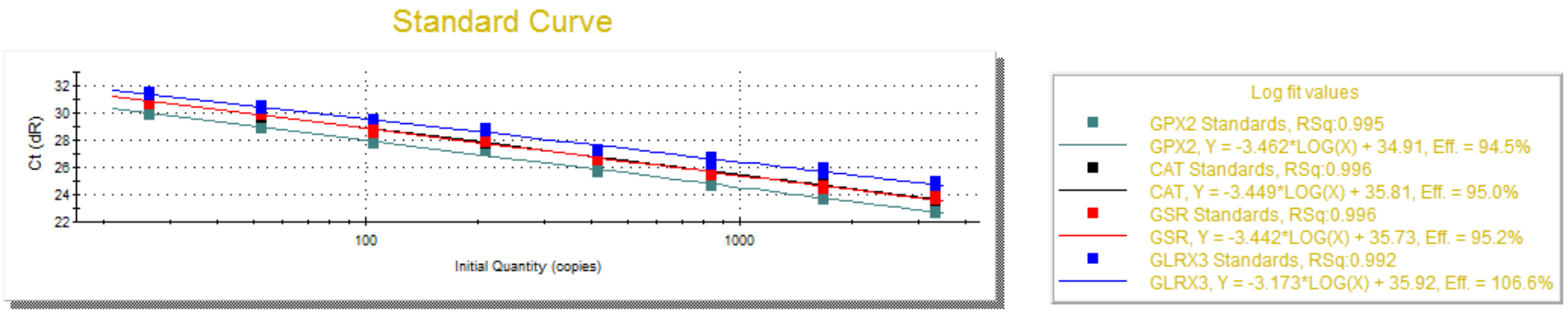


Supplementary figure 5. TXND1, TXN2 and TXNIP (final – primer concentration >250nM) standard curves to determine primer efficiency. Automatically calculated by Stratagene MX3005P. Efficiency values can be found in the legend on the right.



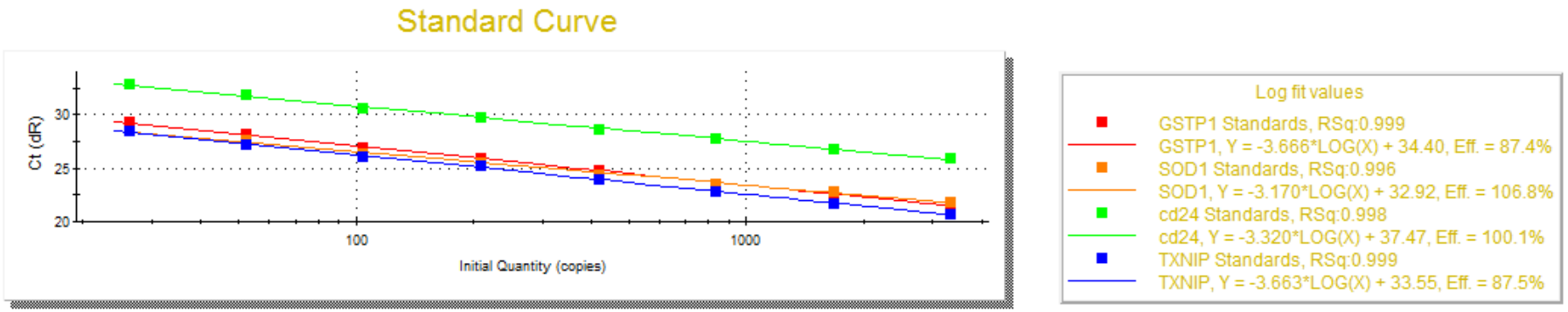
Supplementary figure 6. GCLC, GSS, P21 and PRDX2 standard curves to determine primer efficiency.

Automatically calculated by Stratagene MX3005P. Efficiency values can be found in the legend on the right.



Supplementary figure 7. GPX2, CAT, GSR and GLRX3 standard curves to determine primer efficiency.

Automatically calculated by Stratagene MX3005P. Efficiency values can be found in the legend on the right.



Supplementary figure 8. GSTP1 and SOD1 standard curves to determine primer efficiency. Automatically calculated by Stratagene MX3005P. Efficiency values can be found in the legend on the right.