Next generation circulating tumour DNA detection and subclonality analysis for gastrointestinal cancer therapy personalisation

By

Dr Sonia Mansukhani

The Institute of Cancer Research (University of London)

Doctor of Philosophy (PhD)

ABSTRACT

Background: Gastrointestinal cancers are associated with poor outcomes as many patients present with advanced disease. Intratumour heterogeneity (ITH) is a major barrier to predicting the evolution of cancers and successful treatment. ITH is represented through somatic mutations and copy number aberrations (CNAs). These can be detected through genome profiling of tumour biopsies, however this can be subject to sampling bias. Minimally invasive circulating free (cf)DNA analysis can portray genomic landscapes but highly sensitive approaches are necessary to detect mutations with low variant frequencies.

Methods: I have developed a fully customisable error-corrected targeted sequencing assay for low cfDNA input amounts to interrogate the genomic landscape of patients with metastatic colorectal cancer (mCRC).

Results: Through optimising the target enrichment technology with molecular barcode and novel duplex DNA-molecule identification tool for error correction, I have established a sensitive cfDNA-seq technology with a mutation detection limit <0.15%, which is necessary for the reliable detection of evolving drug resistant subclones. I applied this assay to mCRC patients to detect driver mutations and CNAs and investigate the subclonal genomic landscape that may drive therapy resistance.

I have also designed a translational clinical trial (ICONIC) to improve treatment approaches in localised gastro-oesophageal adenocarcinomas (GOAs). The trial utilises a novel perioperative chemo-immunotherapy combination with the aim of generating synergies by enhancing progression through the cancer-immunity cycle via the pro-immunogenic effect of docetaxel, oxaliplatin, leucovorin and

fluorouracil (FLOT) chemotherapy and simultaneous avelumab-induced release of the immune inhibitory PD-1/PD-L1 checkpoint. I have designed the translational sub-study to investigate candidate predictive biomarkers of tumour response through longitudinal blood (ctDNA) and tumour sampling. The translational study will also investigate dynamic changes ctDNA load and mutant allele frequency, changes in immune infiltrates in baseline and on-treatment biopsies and correlate neoepitope load and faecal microbiome with tumour response.

Conclusions: Overall, the error-corrected cfDNA-seq assay with a customisable target region enables broad insights into cancer genomes and evolution.

DECLARATION

I declare that the work presented in this thesis is my own. I have stated the contribution from other parties within the following acknowledgements section and at relevant points in the text.

_____

Dr Sonia Mansukhani

ACKNOWLEDGMENTS

I would like to thank Marco Gerlinger for the opportunity to join his team and for his guidance throughout. I enjoyed working with every member of the lab but would most like to thank Louise Barber and Beatrice Griffiths for teaching countless laboratory skills and for their support and patience throughout. Your immense scientific knowledge and insight in the challenges of research were invaluable. I am incredibly grateful to my mentors Naureen Starling and Professor David Cunningham.

I would like to thank Mike Hubank for his support in orthogonal cfDNA sequencing by AVENIO, Dimi Kleftogiannis for his bioinformatic support and Katharina von Loga for her work on ICONIC tumour biopsies. I would like to thank the Royal Marsden Clinical Trial's Unit for their work in the set-up of the ICONIC trial. I would also like to thank all my co-authors for their contributions.

I am grateful to Cancer Research UK for funding the PhD and to all the patients and healthy donors who contributed to this study and without whom this project would not have been possible.

To my parents and family, thank you for all your support, dedication and blessings over the years that has allowed me to be who I am.

Finally, to my husband and my son for their love and constant support and for keeping me sane through all the late nights and early mornings.

ABBREVIATIONS

5FU	5- fluorouracil
ADCC	Antibody-dependent cell mediated cytotoxicity
APCs	Antigen presenting cells
BAM	Binary alignment map
BEAMing	Beads, emulsion, simplification, magnets
BL	Baseline
Bp	Base pair
BRC	Biomedical Research Centre for Cancer
CALR	Calreticulin
CAPOX	Capecitabine and oxaliplatin
CAPP-Seq	Cancer personalised profiling deep sequencing
CEA	Carcinoembryonic antigen
CF	Cisplatin + 5FU
cfDNA	Circulating free DNA
cfDNA-seq	Circulating free DNA sequencing
CH	Chromosome
CI	Confidence interval
CIMP	CpG island methylator phenotype
CNVs	Copy number variations
CPS	Combined positive score
CRC	Colorectal cancer
ctDNA	Circulating tumour DNA
CTC	Circulating tumour cells
CTLA-4	Cytotoxic T-lymphocyte associated protein 4
DAMPs	Damage-associated molecular patterns
DC	Dendritic cells
ddPCR	Droplet digital polymerase chain reaction
DFS	Disease free survival
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
DuDe	Duplex DeDuplication software
ECF	Epirubicin, cisplatin and 5FU
ECX	Epirubicin, cisplatin and capecitabine
EDTA	Ethylenediaminetetraacetic acid

EGFR	Epidermal growth factor receptor
EMT	Epithelial to mesenchymal transition
EOX	Epirubicin, oxaliplatin and capecitabine
FFCD	Fédération Francophone de Cancérologie Digestive
FISH	Fluorescence in situ hybridisation
FLOT	Docetaxel, oxaliplatin, leucovorin and fluorouracil
FLOT-A	Docetaxel, oxaliplatin, leucovorin and fluorouracil with avelumab
FOLFI	5FU + irinotecan
FOLFOX	5FU + oxaliplatin
FOLFOX-A	5FU, leucovorin, oxaliplatin with avelumab
FOLFOXIRI	5FU + irinotecan + oxaliplatin
FOrMAT	Feasibility of a Molecular Characterisation Approach to Treatment
FP	5-fluorouracil + cisplatin
GOA	Gastro-oesophageal adenocarcinomas
GOJ	Gastro-oesophageal junction
GE	Genome equivalents
GSEA	Gene Set Enrichment Analysis
HD	Healthy donor
HGF	Hepatocyte growth factor
HMGB1	High-mobility-group box 1
HR	Hazard ratio
HMW	High molecular weight
ICD	Immunogenic cell death
ICGC	The International Cancer Genome Consortium
ICONIC	Immuno-chemotherapy in operable oesophageal and gastric cancer
IDMC	Independent data monitoring committee
IFN-□	Interferon gamma
IGV	Integrated Genome Viewer
IHC	Immunohistochemistry
Indels	Insertions and deletions
ITH	Intratumour heterogeneity
Kb	Kilobase
LAG-3	Lymphocyte activation gene 3
LcWGS	Low coverage whole genome sequencing
LOH	Loss of heterozygosity

M6P	Mannose-6-phosphate
MAD	Maximum administered dose
MAGIC	Medical Research Council Adjuvant Gastric Infusional Chemotherapy
MBC	Molecular barcode
mCRC	Metastatic colorectal cancer
MDSC	Myeloid derived suppressor cells
Ml	Millilitres
MMR	Mismatch repair
MMRd	Mismatch repair deficiency
MRD	Minimal residual disease
MSI	Microsatellite instability
MUT	Mutant
NER	Nucleotide excision repair
NGS	Next generation sequencing
ng	Nanograms
ng/ml	Nanograms per millilitre
nm	Nanomolar
ORR	Overall response rate
OS	Overall survival
PBS	Phosphate-buffered saline
pCR	Pathological complete response
PCR	Polymerase chain reaction
PD	Progressive disease
PD-1	Programmed cell death receptor-1
PD-L1	Programmed cell death-1 ligand
PDO	Patient derived organoid
PFS	Progression free survival
Pg	Picograms
QC	Quality control
QPCR	Quantitative polymerase chain reaction
RECIST	Response evaluation criteria in solid tumours
RMH	The Royal Marsden Hospital
RNA	Ribonucleic acid
RTK	Receptor tyrosine kinases
SCNAs	Somatic copy number aberrations
SRC	Safety review committee

SNP	Single nucleotide polymorphism
SNV	Single nucleotide variants
TCGA	The Cancer Genome Atlas
TILs	Tumour infiltrating lymphocytes
TNM	Tumour node metastasis
TPU	Tumour profiling unit
TSG	Tumour suppressor genes
VAF	Variant allele frequency
VCQT	Variant call quality threshold
WES	Whole exome sequencing
WGS	Whole genome sequencing
WT	Wild type
XI	Consensus family size

PUBLICATIONS ARISING FROM WORK REPORTED IN THIS THESIS

Ultra-sensitive mutation detection and genome-wide DNA copy number reconstruction by error-corrected circulating tumor DNA sequencing. **Mansukhani S**, Barber LJ, Kleftogiannis D, Moorcraft SY, Davidson M et al. *Clinical Chemistry* 2018; 64 (11): 1626-1635.

ICONIC: Peri-operative immune-chemotherapy in operable oesophageal and gastric cancer. **Mansukhani S**, Davidson M, Gillbanks A, Peckitt C et al. *Journal of Clinical Oncology.* 2018; 36 (15_suppl).

Molecular subtypes and novel genetic mechanisms of primary and acquired anti-EGFR resistance in colorectal cancer in the PROSPECT-C biomarker trial. Khan KH, Woolston A, Spain G, Barber LJ, Patil Y, Griffiths B, GonzalezExposito R, **Mansukhani S** et al. *Cancer Research*. 2018; 78 (supp 13):4339-4339.

Perioperative FLOT+ anti-PD-L1 avelumab (FLOT-A) chemo-immunotherapy in resectable oesophagogastric adenocarcinoma (OGA): Safety and biomarker data from the ICONIC trial safety run-in. Davidson M, **Mansukhani S**, Starling N, Chau I, Watkins D, Cunningham D et al. *Annals of Oncology.* 2019; 30 (5_suppl).

Genomic and transcriptomic determinants of therapy resistance and immune landscape evolution during anti-EGFR treatment in colorectal cancer. Woolston A, Khan K, Spain G, Barber LJ, Griffiths B, Gonzalez-Exposito R, Hornsteiner L, Punta M, Patil Y, Newey A, **Mansukhani S**, Davies MN, Furness A et al. *Cancer Cell.* 2019; 36(1):35-50.

Detecting and tracking circulating tumour DNA copy number profiles during first line chemotherapy in oesophagogastric adenocarcinoma. Davidson M, Barber LJ, Woolston A, Cafferkey C, **Mansukhani S** et al. *Cancers*; 2019;11(5). pii: E736.

INDEX

INDEX OF FIGURES

INDEX OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Colorectal cancer (CRC) overview

1.1.1. Epidemiology, aetiology and natural history

Colorectal cancer (CRC) is the third most common cancer in the world with 41,804 new cases in the United Kingdom in 2015 [1]. It accounts for 8-9% of cancer-related mortality worldwide [2] with 16,384 deaths in the UK, in 2016 [1]. CRC develops as a result of an accumulation of genetic and/or epigenetic mutations, which lead to activation of oncogenes and inactivation of tumour suppressor genes. The classical genetic 'Vogelgram' model of CRC depicts that majority of sporadic CRCs are initiated by mutation or loss of the *APC* gene causing dysplasia and adenoma formation, followed by mutations in other key driver genes such as *KRAS* and *TP53* fostering the development of metastases [3]. Since then, large scale genomic characterisation efforts have included multiple omics studies performed on primary and metastatic CRC tissue by The Cancer Genome Atlas (TCGA) and The International Cancer Genome Consortium (ICGC) [4, 5], which have identified numerous recurrently mutated genes (Table 1.1) that are linked to biologically and clinically distinct subsets of mCRC. However, the timing and evolution of clonal expansion and other genome alterations such as copy number changes remain to be elucidated. A proportion of sporadic tumours arise by microsatellite instability through methylation induced silencing of the *MLH1* gene creating genomic instability which leads to a hypermutator phenotype [6] [7]. Furthermore, epigenetic processes have been described as an alternative mechanism of CRC carcinogenesis. Epigenetic silencing is recognised in the Knudson model of colorectal tumourigenesis [8] whereby genetic instability is created either through DNA methylation of cytosine

bases in CpG islands [9] or histone acetylation.

Gene name	Frequency (%)	Hotspot codons	Nucleotide alteration	Protein alteration
APC	61.28	Multiple mutations throughout the gene		
TP53	55.58	Multiple mutations throughout the gene		
KRAS	40.2	12 13 59 61	c.35G>A c.35G>T c.35G>C c.38G>A c.175G>A c.176C>G c.183A>C and c.183A>T c.182A>T c.181C>A	G12D G12V G12A G13D A59T A59G Q61H Q61L Q61K
TTN	18	Multiple mutations throughout the gene		
PIK3CA	16.86	542 545 546 1047	c.1642G>A c.1633G>A c.1634A>C c.1635G>T c.1634A>G c.1633G>C c.1638G>T c.1636C>G c.1636C>A c.1637A>C c.1637A>G c.3139C>T c.3140A>G c.3140A>T	E542K E545K E545A E545D E545G E545Q Q546H Q546E Q546K Q546P Q546R H1047Y H1047R H1047L
SMAD4	10.85	351 361 386 406 445 537	c.1051G>C c.1051G>A c.1051G>T c.1052A>G c.1052A>C c.1082G>A c.1082G>T c.1081C>A c.1081C>G c.1081C>T c.1082G>C c.1156G>A c.1157G>C c.1156G>T c.1156G>C c.1157G>T c.1157G>A c.1216G>A c.1217C>T c.1333C>T c.1334G>A c.1909G>C c.1611C>A c.1610A>T c.1610A>C c.1610A>G c.1609G>T c.1611C>G	D351H D351N D351Y D351G D351A R361H R361L R361S R361G R361C R361P G386S G386A G386C G386R G386V G386D A406T A406V R445* R445Q D537H D537E D537V D537A D537G D537Y D537E
FBXW7	10.15	465 479 505 582	c.1394G>T c.1394G>A c.1393_1394delinsGT c.1393C>G c.1393C>A c.1393C>T c.1435C>T	R465L R465H R465V R465G R465S R465C R479*

			c.1435C>G c.1436G>C c.1436G>A c.1436G>T c.1514G>T c.1513C>G c.1513C>A c.1513C>T c.1514G>C c.1514G>A c.1744T>A c.1745C>A c.1744T>C c.1745C>T	R479G R479P R479Q R479L R505L R505G R505S R505C R505P R505H S582T S582* S582P S582L
TCF7L2	7.65	Multiple mutations throughout the gene		
FAT4	6.48	Multiple mutations throughout the gene		
AMER1	6.4	Multiple mutations throughout the gene		
LRP1B	6.17	Multiple mutations throughout the gene		
BRAF	6.17	594 596	c.1780G>A c.1781A>T c.1782T>A c.1781A>C c.1781A>G c.1790T>C c.1789C>G c.1790T>A c.1790T>G c.1789_1790delinsTC	D594N D594V D594E D594A D594G L597P L597V L597Q L597R L597S
SOX9	5.39	Multiple mutations throughout the gene		
ATM	5.15	Multiple mutations throughout the gene		
NRAS	4.92	12 13 61	c.35G>A c.35G>T c.35G>A c.35G>A c.34G>T c.35_36delinsAG c.37G>A c.38G>T c.37G>T c.38G>A c.38G>C c.37G>C c.183A>T and c.183A>C c.182A>G c.181C>A c.183A>C and c.183A>T c.182A>T c.181C>A c.182A>T c.182A>C	G12D G12V G12S G12D G12C G12E G13S G13V G13C G13D G13A G13R Q61H Q61R Q61K Q61H Q61L Q61K Q61L Q61P
ARID1A	4.06	Multiple mutations throughout the gene		
KMT2C	3.83	Multiple mutations throughout the gene		
PCBP1	2.81	100 102	c.299T>A c.299T>G	L100Q L100R

			c.299T>C c.305T>C c.305T>A c.305T>G	L100P L102P L102Q L102R
EPHA3	2.73	Multiple mutations throughout the gene		
BCL9L	2.65	Multiple mutations throughout the gene		

Table 1.1: Recurrently mutated genes in mCRC.
[5, 10, 11]

Fortunately the clinical outcomes for patients with metastatic colorectal cancer (mCRC) have improved over the past decade which can be attributed to multiple factors including improved adjuvant surveillance programs for earlier detection of metastatic disease, biomarker based treatment selection and improvements in multi-disciplinary treatment regimens [12].

1.1.2. Diagnosis, staging and monitoring

A clinical or biological suspicion of CRC should always be confirmed by radiological imaging, to characterise the extent of disease and a biopsy, to confirm the histopathology of the tumour before commencement of therapy. Histology of the primary tumour or metastases is always necessary before chemotherapy is started. Metachronous lesions presenting atypically or late after the initial diagnosis require histopathological or cytological analysis to confirm source of the primary and to guide treatment choices.

The overall 5-year survival depends on the stage of the disease at the time of resection. Staging for CRC is based on the Tumour, Node and Metastases (TNM) staging system [13] with reference to the size of the tumour and its penetration of the bowel wall (T), spread to regional lymph nodes (N) and distant metastases (M) (Table 1.2). Stage I tumours which are limited to the submucosa and fully resectable have a 5 year survival of 90%, which decreases to 50% in

oligometastatic potentially resectable stage IV (metastatic) cancer and 8% in unresectable stage IV disease [1].

Primary tumour (T)

Tx	Primary tumour cannot be assessed
T0	No evidence of primary tumour
Tis	Carcinoma in situ: limited to intraepithelial or invasive lamina propria
T1	Tumour invades submucosa
T2	Tumour invades muscularis propria
T3	Tumour penetrates muscularis propria and arrives at colorectal fat tissue
T4	Tumour directly invades other organs or structures
T4a	Tumour penetrates visceral peritoneum
T4b	Tumour directly invades or adheres to other organs or structures

Regional lymph nodes (N)

NX	Regional lymph node(s) cannot be assessed
N0	No regional lymph node metastasis and no tumour deposits
N1	Metastasis in 1-3 regional lymph nodes
N1a	1 lymph node metastasis
N1b	Metastasis in 2-3 lymph nodes
N1c	Tumour deposits are submucosal, mesangial or peritoneum covered para-colorectal tissue
N2	Four or more lymph node metastases
N2a	Metastasis in 4-6 regional lymph nodes
N2b	Metastasis in seven or more regional lymph nodes

Distant metastasis (M)

M0	No distant metastasis
M1	Distant lymph node metastasis
M1a	Metastasis is limited to one organ site (e.g. liver, lung, ovary and extra-regional lymph node metastases)
M1b	Transfer more than one organ or site
M1c	Peritoneal metastases with or without metastasis of other organs

Clinical staging

Stage	T	N	M
0	Tis	N0	M0
I	T1-T2	N0	M0
IIA	T3	N0	M0
IIB	T4a	N0	M0
IIC	T4b	N0	M0
IIIA	T1-2	N1/N1c	M0
	T1	N2a	M0
IIIB	T3-T4a	N1/N1c	M0
	T2-3	N2a	M0
	T1-2	N2b	M0
IIIC	T4a	N2a	M0
	T3-T4a	N2b	M0
	T4b	N1-N2	M0
IVA	Any T	Any N	M1a
IVB	Any T	Any N	M1b
IVC	Any T	Any N	M1c

Table 1.2 TNM classification for colorectal cancer [14]

1.2 Pathophysiology

1.2.1. Locoregional and distant metastasis

The majority of CRC related deaths are due to metastatic tumours infiltrating distant organs rather than the primary disease itself. The process of metastasis occurs in sequentially occurring steps: through local invasion of the basement membrane and extracellular matrix of the surrounding tissue and intravasation into the blood and/or lymphatics, survival and dissemination in the bloodstream

to microvessels of distant tissues, and adaptation to the microenvironment of the distant site to enable proliferation and formation of a secondary tumour [15-17].

1.2.2. Molecular pathology of CRCs

Cancers arise from DNA mutations which lead to uncontrolled cell proliferation. These mutations can be detected in cancer specimens by DNA sequencing technologies. The identification of driver mutations in an individual cancer, i.e. those mutations that have been positively selected during the evolution of cancer and confer a growth advantage [18], should provide insights into its biological characteristics. In turn, this could enable predictions about specific therapeutic vulnerabilities of the tumour.

There are a number of gene mutations involved in the molecular repertoire of CRC; *APC*, *TP53* and *KRAS* are the most common somatic mutations in CRC with incidence of 81%, 60% and 43% respectively, as per TCGA Network exome sequencing effort of 276 CRC tumour normal pairs [19]. The IntOGen-mutations platform of 1281 CRC samples reports 72 recurrently mutated genes [11], some of which can serve as predictive biomarkers (those that measure the likelihood of response and allow identification of patients most likely to benefit from a given treatment) and/or prognostic biomarkers (those that are associated with the clinical outcome) [20].

1.2.3. *RAS/RAF* mutation status

Of the three human *RAS* isoforms, *KRAS* is the most frequently altered gene with mutations occurring 30%–40% of CRCs [19, 21]. The most frequent mutation in this gene include point substitution in codons 12 (p.G12D c.35G>A; p.G12V c.35G>T; p.G12A c.35G>C and p.G12C c.34G>T), 13 (p.G13D c.38G>A), 59

(p.A59T c.175G>A; p.A59T c.175G>A; p.A59G c.176C>G) and 61 (p.Q61H c.183A>C and c.183A>T; p.Q61L c.182A>T; p.Q61K c.181C>A) [10]. The presence of a *RAS* mutation is a negative predictive marker of anti-*EGFR* monoclonal antibody treatment outcomes in patients with mCRC and therefore expanded *RAS* analysis testing is recommended for patients who are being considered for systemic therapy with anti-*EGFR* antibodies [22]. Retrospective analysis of pivotal clinical trials for anti-*EGFR* therapies have shown that activating *KRAS* and *NRAS* mutations have been associated with primary resistance to *EGFR* targeting monoclonal antibody therapy and in *RAS* mutant cancers these therapies have a detrimental effect [23-29] and should not be administered.

BRAF mutations, majority of which are p.V600E (c.1799T>A), are found in between 8-12% of mCRC tumours and are almost exclusively non-overlapping with *RAS* mutations [30]. *BRAF* mutations are a negative prognostic marker with a median overall survival (OS) of 10.4 months compared with 34.7 months for patients with *BRAF* wild type tumours [30]. Some *BRAF* V600E mutations have been associated with CpG island methylator phenotype (CIMP) giving rise to sporadic, non-germline microsatellite instability [22]. More recently however, the <1% of mCRC patients with *BRAF* mutations in codons 594 and 596 have shown to have markedly longer OS (62.0 months) than patients with *BRAF* V600E mutations [31]. *BRAF* V600E mutated tumours have certain phenotypic features seen in retrospective analyses for example two thirds of these primary tumours are located on the right side of the colon and associated with mucinous features with increased lymphocyte infiltration, an increased incidence of peritoneal and distant lymph node metastases [30]. Conversely, cancer carrying mutations in *BRAF* codons 594 (p.D594N c.1780G>A; p.D594V c.1781A>T; p.D594E

c.1782T>A; p.D594A c.1781A>C; p.D594G c.1781A>G) or 596 (p.L597P c.1790T>C; p.L597V c.1789C>G; p.L597Q c.1790T>A; p.L597R c.1790T>G; p.L597S c.1789_1790delinsTC) were more frequently rectal, non-mucinous with no peritoneal spread.

BRAF mutation is also recognised as a negative predictive biomarker for anti-EGFR antibody therapy [32]. Meta analyses have shown that there is greater benefit of anti-EGFR antibody therapies in patients with *RAS*-wild type/*BRAF*-wild type tumours than in patients with *RAS*-wild type/*BRAF*-mutant tumours [32, 33]. Thus *BRAF* mutation testing forms part of the diagnostic molecular pathology work up of mCRC to guide systemic therapy and for prognostic assessment.

1.2.4. Microsatellite Instability

Tumours with genetic defects in mismatch repair pathways are known to harbour numerous somatic mutations, especially in regions of repetitive DNA known as microsatellites. Microsatellite regions are susceptible to replicative errors caused by strand slippage and stalling of DNA polymerase. Defects in the mismatch repair system are caused either by germline mutations affecting MMR genes (*MLH1*, *MSH2*, *MSH6* and *PMS2*) in Lynch syndrome, or sporadically as a result of epigenetic silencing of *MLH1* through CIMP, with a minority caused by double somatic mutations [34]. The accumulation of mutations in these regions is known as microsatellite instability (MSI). MSI tumours represent 4-5% of all mCRC.

A phase II study in 2015 demonstrated evidence that MMR status can be used as a predictive biomarker for programmed cell death-1 (PD-1) blockade therapy [35], based on the hypothesis that the mismatch repair-deficient (MMRd) tumour microenvironment strongly expresses immune checkpoint ligands including LAG-

3 and PD-L1. The MMRd status increases the number of mutation-associated predicted neoantigens which is believed to be the basis for enhanced anti-PD-1 responsiveness in MSI cancers [36].

1.2.5. Circulating tumour biomarkers

A cancer biomarker is an objective characteristic that can measure the risk of cancer development or progression, or potential response to therapy. A biomarker can be diagnostic, predictive or prognostic and ideally should serve at least one of these purposes [37]. The ideal cancer biomarker for diagnosis is one that is a well characterised, validated, non-invasive test, with 100% sensitivity and specificity at all disease stages and can be correlated with stage.

Currently the most widely used circulating tumour antigen in clinical practice for CRC is the carcinoembryonic antigen (CEA), however CEA levels can increase with various gastro-intestinal cancers hence its low sensitivity limits its use to patient follow up during assessment of response to systemic treatment [38] and post surgery surveillance. Therefore the need remains to identify a highly sensitive and specific, minimally invasive biomarker that can be used in diagnosis and follow up. A number of research studies into circulating DNA have identified it as a potential biomarker which may be useful diagnostically and prognostically (further described in section 1.5).

1.3 Management of mCRC

The mainstay of treatment for early CRCs is definitive surgery followed by chemotherapy in high risk stage II and stage III cancers to eliminate micrometastatic disease [12]. The prognostic histological features of high risk stage II disease are T4 tumour, perforation or obstruction, poorly differentiated

histology, lymphovascular invasion, fewer than 12 sampled lymph nodes and a high pre-operative CEA level [39]. The use of radiotherapy in colon cancers is generally reserved for palliative settings and in a small proportion of patients where surgical resection margins are positive. Rectal cancers are routinely treated with neoadjuvant chemoradiation due to the local anatomy and relatively higher risk of local recurrence [12].

At the time of diagnosis of mCRC several factors need to be taken into consideration for treatment planning. These include clinical presentation, patterns of tumour metastases, symptoms, prognostic molecular or biochemical markers, as well as patient related factors including comorbidities and patient expectations. Oligometastatic disease, characterised by the existence of metastatic lesions at two or three predominantly visceral sites, is primarily treated with the aim of achieving complete response of all tumour masses. This can be achieved through a multi-modal approach of systemic therapy, surgical resection and/or locally ablative therapies. For patients with low volume oligometastatic disease in the liver or lung, long term survival or cure can be achieved in 20-50% of patients who undergo complete R0 resection of their metastases [40].

For patients with more extensive metastatic disease, the aim is to achieve long term disease control, to improve disease related symptoms and survival with systemic chemotherapy. The systemic therapy options for mCRC comprise three main chemotherapy drugs (fluoropyrimidine, irinotecan and oxaliplatin), in addition to targeted therapies in various combinations (5FU + irinotecan- FOLFIRI; 5FU + oxaliplatin- FOLFOX; capecitabine + oxaliplatin- CAPOX; 5FU + irinotecan + oxaliplatin- FOLFOXIRI) and schedules (see Table 1.3).

1.3.1. First line systemic therapy

Typically, first line chemotherapy backbone comprises a fluoropyrimidine (intravenous 5-FU or oral capecitabine) used in combination with oxaliplatin or irinotecan. Combination chemotherapy provides higher response rates and better progression free survival (PFS) and overall survival (OS) times than a fluoropyrimidine alone [41-43]. FOLFOX or FOLFIRI as chemotherapy regimens have similar activity and can both be the backbone for biologicals, however have different toxicity profiles which may be favoured [44]. Monoclonal antibodies bevacizumab (anti-VEGF) and cetuximab/panitumumab (anti-EGFR) have been shown to improve the clinical outcome of patients with mCRC in combination with chemotherapy [45-52]. First line chemotherapy is usually given for 6 months. There is some thought on continuing patients until disease progression or unacceptable toxicity due to funding restrictions on targeted therapies in the NHS. However consideration needs to be paid to the effects of continuing specifically oxaliplatin containing regimens due to cumulative and potentially irreversible toxicities such as peripheral sensory neuropathy. As a result, discontinuation and/or intermittent combination chemotherapy is considered the mainstay of treatment for mCRC patients with evidence of disease response or stable disease [53-56].

1.3.2. Targeted therapy

There has been a treatment shift from non-selective cytotoxic chemotherapy agents to combination therapies that include targeted therapies. The monoclonal antibodies cetuximab and panitumumab (anti-EGFR) and bevacizumab (anti-VEGF) have been shown to improve clinical outcomes when combined with combination chemotherapy regimens [25, 26, 57, 58].

Expanded *RAS* and *RAF* analysis should be carried out at diagnosis, as a prerequisite for the use of anti-EGFR antibodies cetuximab and panitumumab. Cetuximab has been shown to improve PFS and OS in first line setting in mCRC, in combination with FOLFIRI compared with FOLFIRI alone in *RAS* wild type mCRCs [26, 51]. Both cetuximab and panitumumab also increase the activity of the doublet FOLFOX in *RAS* wild type mCRC tumours [24, 25, 57, 59-61]. The PRIME study used FOLFOX with or without the addition of panitumumab in *KRAS* wild type cancers, and reported a statistically significant increase in median OS from 19.7 to 23.9 months (HR 0.88; 95% CI 0.73-1.06; p=0.17) [24]. However, the addition of anti-EGFR antibodies to oxaliplatin based regimens without infusional fluoropyrimidine e.g. bolus administration, capecitabine alone or CAPOX has not resulted in any benefit [61, 62].

Bevacizumab is a monoclonal antibody that binds circulating VEGF-A and has shown to improve outcomes when combined with chemotherapy in the first line setting of mCRC [50, 63, 64]. The CALGB 80405 trial of physician choice chemotherapy backbone with bevacizumab showed no difference in outcomes between chemotherapy backbones of FOLFOX or FOLFIRI [65]. A further phase III TRIBE study of bevacizumab plus triplet chemotherapy backbone FOLFOXIRI vs. FOLFIRI plus bevacizumab, has demonstrated a superior median OS with triplet compared with doublet chemotherapy, 31.0 vs. 25.8 months respectively which is one of the longest median OS recorded in mCRC setting irrespective of baseline clinical characteristics and *RAS*/*RAF* status [46]. The anti-angiogenic aflibercept has been shown to confer a survival advantage in combination with FOLFIRI [66]. A benefit has also been observed for patients who had received prior bevacizumab therapy in a subgroup analysis of the VELOUR study [67]. The

best reported outcomes in large scale clinical trials in untreated mCRC have generally combined cytotoxic chemotherapy agents with a targeted agent.

1.3.3. Second line systemic therapy and beyond

Second line therapy is proposed after failure of first line strategy in patients with good performance status and adequate organ function and is dependent on the first line therapy choice. In patients in whom the initial chemotherapy backbone has failed, the chemotherapy regimen should be changed. Furthermore, biologicals should be considered in the second line setting although currently these agents are unavailable in the UK on the cancer drugs fund in the second line setting. In the absence of funding restrictions, anti-angiogenics bevacizumab and aflibercept could be considered in the second line setting if they were not used in the first line and have been shown to improve OS when used in combination with FOLFOX [48] and FOLFIRI [68, 69].

Both anti-EGFR agents cetuximab and panitumumab have been shown to increase PFS, but not OS, when combined with irinotecan in the second line setting. Therefore there is value in using these agents if not previously used in *RAS* wild type patients [32, 70, 71]. Cetuximab and panitumumab have also shown efficacy in the third line or salvage therapy setting and are both active as single agents [72, 73]. The combination of cetuximab with irinotecan is more active than cetuximab alone, in irinotecan refractory patients [13].

In patients with treatment refractory mCRC a newer oral agent that combines trifluridine and tipiracil hydrochloride, Lonsurf/ TAS.102, has been shown to be efficacious with a response rate of approximately 5% [74]. It is a well-tolerated agent and therefore is a potential option in the third line setting and beyond.

Regorafenib is an orally available multi-targeted kinase inhibitor, inhibiting several targets including angiogenesis. Regorafenib has shown activity vs. best supportive care in two phase III studies, CORRECT [75] and CONCUR [76]. The CORRECT study confirmed that regorafenib improves OS in heavily pre-treated patients including those that have been treated with all available cytotoxics. The median OS was 6.4 months in the regorafenib group vs. 5.0 months in the placebo group (HR 0.77; 95% CI 0.64-0.94; p=0.0052) [75] and can therefore be used beyond third line setting. However its toxicity profile limits the benefit to patients with a good performance status and adequate organ function.

Population	Pivotal trial	Number of patients	Regimen Name	Efficacy endpoints	Reference
1st line *RAS* unknown Or *RAS* negative	deGramont-FOLFOX4	420	FOLFOX	**PFS**: 9.0 months vs. 6.2 months for 5FU + leucovorin control arm (p= 0.0003) **RR**: 50.7% vs. 22.3% for 5FU + leucovorin control arm (p= 0.0001) **OS**: 16.2 months vs. 14.7 months for 5FU + leucovorin control arm- did not reach statistical significance	De Gramont et al, 2000
	Irinotecan Study Group	387 683	FOLFIRI	**RR**: 49% vs. 31% for 5FU + calcium folinate (p<0.001) **OS**:17.4 vs. 14.1 months (p=0.031) **PFS**: 7.0 months vs. 4.3 for 5FU + leucovorin alone (p=0.004)	Douillard et al, 2000 Saltz et al, 2000
	N016966	1401	Oxaliplatin based doublet + Bevacizumab	**PFS**: 9.4 months vs. 8.0 for oxaliplatin based doublet + placebo (HR 0.83; P=0.0023) **OS**: 21.3 months vs. 19.9 for oxaliplatin based doublet + placebo (HR 0.89; P=0.077)- did not reach statistical significance	Saltz et al, 2008
	AVF2107g	813	FOLFIRI+ Bevacizumab	**RR**: 44.8% vs. 34.8% (p=0.004) **PFS**: 10.6 months vs. 6.2 months for FOLFIRI alone (HR 0.54; p<0.001) **OS**: 20.3 months vs. 15.6 for FOLFIRI alone (HR 0.66; p <0.001)	Hurwitz et al, 2004
	HORG group GONO group	283 244	FOLFOXIRI vs. FOLFIRI FOLFOXIRI vs. FOLFIRI	**OS**: 21.5 months vs. 19.5 months for FOLIRI (p=0.337)- did not reach statistical significance **RR**: 60% vs. 34% for FOLFIRI (p<0.0001) **PFS**: 9.8 months vs. 6.9 months for FOLFIRI; HR 0.63 p=0.0006 **OS**: 22.6 months vs. 16.7 for FOLFIRI (HR 0.70 p=0.032)	Souglakos et al, 2006 Falcone et al, 2007
	TRIBE	508	Bevacizumab plus FOLFOXIRI or FOLFIRI	**PFS**: 12.1 months vs. 9.7 months for FOLFIRI + Bevacizumab (HR 0.75; p=0.003)	Cremolini et al, 2015

				OS: 31.0 vs. 25.8 months for FOLFIRI + Bevacizumab (HR 0.79; p= 0.054)- did not meet statistical significance.	
1st line *KRAS* wild type mCRC	COIN (*negative study)*	1630	Oxaliplatin based chemotherapy (CAPOX or FOLFOX) + Cetuximab or placebo	**OS**: 17.0 months vs. 17.9 months in the placebo group (HR 1.04, p=0.67) **PFS**: 8.6 months vs. 8.6 months in the placebo group (HR 0.96, p=0.60)	Maughan et al, 2011
	OPUS	337	FOLFOX +/- Cetuximab		Bokemeyer et al, 2009
	CRYSTAL	1198	FOLFIRI +/- Cetuximab	**ORR:** 46% vs. 36% in FOLFOX only group **PFS:** 7.7 months vs. 7.2 months in FOLFOX only group (HR 0.57, 0.163) **PFS:** *KRAS* WT group: 9.9 months vs. 8.7 months for FOLFIRI alone (HR 0.68, p= 0.02) **OS**: 24.9 months vs. 21.0 months for FOLFIRI alone (HR 0.84, p= 0.31)	Van Cutsem et al, 2009
	PRIME	1183	FOLFOX+/- Panitumumab	**PFS**: 10.0 months vs. 8.6 months for FOLFOX alone (HR 0.80, p=0.01) **OS**: 23.9 months vs. 19.7 months (HR 0.88, p= 0.17)	Douillard et al, 2014
2nd line	E3200	829	FOLFOX +/- Bevacizumab vs. Bevacizumab alone after 1st line FOLFIRI	**OS**: 12.9 months for FOLFOX + Bev group vs. 10.8 months for FOLFOX alone (HR 0.75, p= 0.0011) **PFS**: 7.3 months for FOLFOX + Bev group vs. 4.7 months for FOLFOX alone (HR 0.61, p< 0.0001) and 2.7 months for bevacizumab alone	Giantonio et al, 2007
	VELOUR	1226	FOLFIRI +/- Aflibercept	**OS**: 13.5 months vs. 12.06 months for FOLFIRI alone (HR 0.817, p= 0.0032) **PFS**: 6.90 months vs. 4.67 months for FOLFIRI alone (HR 0.758, p<0.0001).	Van Cutsem et al, 2012
	RAISE	1072	FOLFIRI+/- Ramiucirumab	OS: 13.3 months vs. 11.7 months for placebo group (HR 0.844, p= 0.0219)	Taverners et al, 2015
3rd line	RACECOURSE	800	TAS.102 vs. placebo	**OS**: 7.1 months vs. 5.3 months for placebo (HR 0.68, p <0.001)	Mayer et al, 2015

	CORRECT	753	Regorafenib + BSC	**OS**: 6.4 months vs. 5.0 months for placebo + BSC (HR 0.77, p= 0.0052)	Grothey et al, 2013
	CONCUR	204	Regorafenib vs. placebo	**OS**: 8.8 months vs. 6.3 months for placebo (HR 0.55 p= 0.0016)	Li et al, 2015

Table 1.3: Pivotal studies in metastatic colorectal cancer
[41, 42, 46-48, 50-52, 59, 61, 63, 74-79]

1.3.4. Therapy resistance

Activating mutations in the *RAS* pathway are responsible for primary and acquired resistance to anti-EGFR antibodies. However despite biomarker based patient selection 65-70% of patients progress within 3 to 12 months of starting therapy [80]. Studies have shown that patients with undetectable *RAS* mutations at baseline can go on to develop treatment resistance through emergence of *RAS*-mutant subclones that have undergone clonal expansion under selective pressure of anti-*EGFR* therapy. More recently recognised pathways of resistance include *BRAF* V600E [81], *MAP2K1/ERK1 [82]* or *PIK3CA* [83] mutations, amplifications of *KRAS* [84] and of the receptor tyrosine kinases (RTKs) *ERBB2*, *MET* and *FGFR1* [82] have been suggested as further drivers of primary resistance.

Epigenetic mechanisms may also be responsible for drug resistance. For example, in vitro studies have demonstrated the presence of *EGFR* gene promoter hypermethylation which silences *EGFR* gene expression [85, 86], thus reducing the effect of anti-*EGFR* therapy in patients with mCRC. The presence of *EGFR* gene promoter methylation in the primary tumour was negatively associated with clinical outcome in a study of 52 patients with mCRC treated with cetuximab and irinotecan in the second line setting [87].

The importance of optimising treatment regimens to escape the emergence of resistance relies strongly on the ability to understand the course of disease in individual patients in order to anticipate and detect resistance and tumour progression early on to guide future clinical decisions. Treatment resistance is driven by intratumour heterogeneity (ITH) (see section 1.4.1) that generates cancer evolution over time and space, thus multiple biopsies in time and space are critical to understanding the complexity of evolving resistance [88, 89].

1.4 Molecular profiling and personalised medicine

The future of personalised medicine lies in being able to molecularly characterise and understand each individual's disease and to thereafter tailor therapy based on findings in order to predict and avoid treatment resistance. It is likely that resistant subclones are present early on in the disease process, however their low prevalence in treatment-naïve cancer cell populations prevents them from being detected by standard assays [90]. Tumour tissue biopsies have served as the foundation of basing therapeutic decisions, however biopsies are often inadequate at capturing the complete biological portrait of a malignancy and its evolution due to spatial and temporal tumour heterogeneity [89].

1.4.1. Tumour heterogeneity in colorectal cancer

Tumours acquire genomic aberrations during cancer progression, which can vary from cell to cell. This generates intratumour heterogeneity (ITH) i.e. subclonal diversity through acquisition of new and diverse mutations as a result of genomic instability [89]. ITH was first described at the cytogenetic level in the 1950's [91], however only recently the full extent and the functional implications of ITH have been appreciated [89].

The changing dynamics of tumour subclonal architecture over space and time may result in previously sub-dominant clones, becoming dominant and generating new cancer phenotypes [92]. This genomic plasticity fosters ongoing evolutionary adaptation to changing environments. Importantly, stochastic processes including mutation generation and genetic drift play a major role in evolutionary processes and varying outcomes, thus affecting long term predictability [90].

Single timepoint genomic assessments are unlikely to be sufficient for treatment personalisation approaches due to dynamic genomic changes within the tumour, resulting from cancer evolution and response/resistance to treatment. Repeat tumour biopsies are clinically difficult and can suffer from sampling biases, furthermore they can be uncomfortable for patients or may even be unsafe if the cancer has spread to places in the body which are difficult to reach. The inability of tumour biopsy approaches to temporally trace solid tumour evolution at the necessary level of detail also precludes thorough scientific testing of cancer evolution theories. For example, it remains unknown whether high heterogeneity confers poor outcomes by increasing the probability that multiple drug resistant clones evolve or whether heterogeneous cancers are simply more robust against a range of adversities, including genomic instability and cytotoxic agents. Thus there is an urgent need to develop an ultrasensitive, minimally invasive assay that can detect low level mutations that can impact treatment efficacy.

1.5 Circulating cell free DNA

1.5.1 Origins of circulating DNA

Circulating cell free DNA (cfDNA) is genetic material found within peripheral blood. cfDNA in the human blood stream was first described in 1948 [93]. These

nucleic acid fragments are released into the bloodstream from dying cells, likely through apoptosis or cell necrosis and can originate from non- malignant somatic tissues e.g. a foetus, or from primary and metastatic tumour deposits [94]. Presence of cfDNA is also being investigated in other biological samples including urine, synovial fluids, saliva and sputum for cancer diagnosis [95].

In patients with cancer, a proportion of cfDNA is composed of tumour-derived DNA which contains tumour-specific alterations including genetic or epigenetic changes, known as circulating tumour DNA (ctDNA) [96, 97]. The fraction of ctDNA within the total circulating DNA varies widely between <0.1% to over 90% and can be dependent on a number of factors including tumour histology, tumour location, vascularisation and metastatic burden [98].

Circulating tumour cells (CTCs) represent intact, often viable cancer cells that can be purified from the blood based on cell surface markers that distinguish them from normal blood cells [96]. However the low abundance of CTCs often makes detection, quantification and characterisation unfeasible in clinical practice. Comparative analyses of CTCs with ctDNA demonstrated that ctDNA is detected in most patients with advanced cancer and showed a greater dynamic range and greater correlation with changes in tumour burden than tumour markers or CTCs [96].

1.5.2. Properties of cfDNA

Tumour derived cfDNA is usually highly fragmented double-stranded DNA with lower molecular weight than genomic DNA. It averages lengths of 170 bp and its sampling offers a minimally invasive method of tumour derived DNA acquisition [95]. cfDNA has a half life of approximately 1-2 hours [99]. As it is cleared rapidly

from the circulation, the ctDNA is likely to reflect a real-time picture of the released fragments from the dying tumour cells, thus making it an appealing real-time biomarker for assessment of molecular tumour genotype.

The discovery that ctDNA is released into the blood by many tumours has opened novel opportunities to perform genetic analyses through minimally invasive 'liquid biopsies' [100, 101]. Furthermore, as ctDNA can be shed from multiple lesions in an individual patient, it can sample DNA from subclones scattered across the primary and multiple metastases [102, 103].

1.5.3 cfDNA as a diagnostic and predictive marker

Early detection of cancer increases the likelihood of curative intervention and improves survival [104]. Detection rates of ctDNA are low, varying between 49 and 78% in patients with localised disease in a study of breast, colon and pancreatic cancer [96]. Furthermore ctDNA analysis can distinguish between residual disease and treatment-related changes through highly sensitive ctDNA detection techniques [105]. During follow up of patients in complete remission following surgery, ctDNA can allow for detection of minimal residual disease (MRD) and ctDNA rise can precede radiologically diagnosed cancer relapse by several months [105, 106]. The reliable identification of patients with subclinical metastases may permit stratification of adjuvant treatments thus, avoiding unnecessary treatments and toxicities in those who would never relapse. In addition, extensive imaging for surveillance can be adapted to individual risk.

Patients presenting with metastatic disease do not have suitable diagnostic tissue from surgical resection and would therefore require a fresh tumour biopsy for molecular analysis prior to treatment initiation [107]. In patients presenting with

early-stage disease and subsequently developing metachronous disease, archival biopsy specimens may not always correspond to the molecular features of the disease at the time of development of metastases and treatment initiation. Even if tumour tissue biopsy remains the gold-standard specimen for diagnosis, liquid biopsies appear as an attractive alternative for tumour molecular assessment [96].

Intratumoural heterogeneity and the evolution of drug resistant subclones in patients with mCRC have been examined by liquid biopsy approaches [96, 108, 109]. Longitudinal monitoring is feasible and can be used to track and understand the subclonal composition of cancers, clonal evolution over the course of disease and the early detection of drug resistant subclones, in real time [97, 110]. This can provide opportunities for early therapeutic changes targeting evolving resistant subclones, which may still be small and treatable. A recent study has shown that frequent (four weekly) longitudinal cfDNA samples can be utilised to model cancer evolutionary dynamics and predict treatment resistance [80]. Promising results have also been seen from studying the concordance of tumour biopsy vs. cfDNA profiling [96, 97, 110-112], demonstrating the clinical utility of this technology for identifying tumour mutations in the blood of mCRC patients.

1.5.4 Techniques for cfDNA analysis

A number of techniques have emerged for cfDNA analysis, with varying degrees of sensitivity and specificity. Broadly, polymerase chain reaction (PCR) based techniques are designed to detect hotspot mutations at frequencies below 0.1% in cfDNA but are restricted to the analysis of a limited number of genomic loci [113, 114]. PCR-based techniques are relatively time-saving and economical. NGS is a sequencing approach used for analysis of larger genomic regions and

target regions can be customised with target gene panels. NGS offers an accurate assessment of ctDNA with information on fragment length, small alterations such as single nucleotide variants, indels, copy number aberrations and genomic rearrangements thus allowing identification of novel alterations without pre-existing knowledge. However, NGS error rates rapidly increase when attempting to call mutations with variant allele frequencies (VAFs) below 5% [115]. Error correction approaches have been incorporated into NGS cfDNA assays to reduce this error rate, for example through background error correction models that remove recurrent false positives occurring at specific positions and more recently by including random unique molecular identifiers/barcodes (MBC) combined with redundant over-sequencing [116, 117]. The latter approach has been shown to enable accurate mutation calling with VAFs ≤0.1%.

1.5.4.1 Beads, Emulsion, Amplification, Magnetics (BEAMing)

BEAMing is a first generation digital PCR method that combines emulsion PCR and flow cytometry to identify and quantify somatic mutations [114]. BEAMing has shown a highly sensitive detection rate down to 0.02% [118].

1.5.4.2 Droplet digital polymerase chain reaction (ddPCR)

ddPCR is an emulsion-based PCR technology that partitions a sample into ~20,000 droplets and PCR reactions occur individually within each droplet. After PCR amplification, nucleic acids may be quantified by counting positive reactions that are tagged by mutation specific Taqman probes. ddPCR is extremely sensitive, able to detect one mutant event in 100,000 wild-type events. However, it is only able to evaluate a limited number of base pair alterations within a single assay. Therefore, this limits its use to detect hotspot mutations and cannot be applied for mutation discovery [119].

1.5.4.3 Quantitative Polymerase Chain Reaction (qPCR)

qPCR can be used for gene expression analysis, copy number analysis as well as quantification of target mutations. This system relies on amplification data of target templates to be measured during the early exponential phase. It is a fast and relatively inexpensive technique however can only detect variants with allele frequencies greater than 10% [120].

1.5.4.4 Amplicon sequencing

Amplicon sequencing is a highly targeted approach that uses a pair of oligonucleotide probes designed to capture regions of interest but larger genomic regions are prone to errors [121]. Coverage of whole genes requires overlapping amplicons and multiplex designs which may require different PCR conditions thus leading to drop out of certain target regions. Significant developments in amplicon sequencing have permitted cancer mutations to be detected down to 2% allele frequency. These methods have been used to demonstrate tumour heterogeneity and clonal evolution in ovarian cancers and breast cancers [100, 103]. Overall amplicon sequencing is easy to setup for low input DNA quantities and permits discovery approaches however, multiplex primer designs can be difficult due to low coverage uniformity and thus higher sequencing error rates.

1.5.4.5 Error-corrected targeted cfDNA sequencing

An improvement in error rates can be achieved by capture-base approaches, where NGS is applied to a target gene panel for sensitive detection of ctDNA mutations within a specific region of interest. Targeted DNA sequencing is more cost effective, as it facilitates higher sample throughput and improves accuracy by optimising the read depth coverage in the regions of interest [122]. In addition to this, error correction approaches have been used to further improve sensitivity.

This can be through digital error suppression used in the CAncer Personalized Profiling deep sequencing (CAPP-Seq) assay [117], which can achieve sensitivities down to 0.1% however requires a large number of healthy donor cfDNA data for digital error suppression.

An alternative error correction approach is based on random molecular barcode (MBC) tagging to suppress the error rates which can otherwise lead to false positive mutation calls that obscure true mutations at VAFs below 5%. Such error correction requires substantial over-sequencing, which limits ultra-sensitive ctDNA analysis to small target regions.

A recent development of duplex sequencing takes advantage of the double-stranded nature of DNA as endogenous barcodes, whereby true mutations only show up in each amplified copy of each DNA strand and single copies of mutations are reflective of errors introduced by PCR or sequencing [123]. This approach reduces error rates without the need for over-sequencing with molecular barcodes.

1.5.4.6 Whole exome and whole genome sequencing

Although targeted panels have a high sensitivity with relative low cost, they are limited to detecting mutations within the target region. Whole exome sequencing (WES) can be used to sequence the entire coding region of the human genome and such comprehensive analyses has several advantages including discovery of novel mutations, identification of mutational signatures and assessment of tumour mutational burden. However, due to the large target region sensitivity is lower than targeted sequencing approaches [124].

Whole genome sequencing (WGS) allows the most comprehensive analysis of cancer genomes, including analysis of rare mutations and novel oncogenes. However, several challenges limit the application of WGS for cfDNA in the clinical setting including the requirement for very high sequencing depths to profile subclonality, or often low tumour content would make this prohibitively expensive. Both WGS and WES require high input of sample DNA, hindering their application in screening and early-stage diagnosis when the concentration of ctDNA is considerably low.

Overall, larger panels such as whole exomes are hence crucial to dissect whole cancer population evolution. Based on the above considerations, these need to incorporate error correction to avoid false positive mutation calls as these can impair subclone identification which depends on accurate measurements of mutations with low VAFs.

1.6 Overview of Gastric and Oesophageal Adenocarcinomas (GOA)

1.6.1 Epidemiology, aetiology and natural history

Gastro-oesophageal adenocarcinomas (GOAs) are the third most common cause of cancer related death in the Western world and together are responsible for more than 1 million deaths per year worldwide [125]. There is a significant variation in incidence worldwide with the highest incidence seen in Eastern Asia. In non-Asian countries, there have been significant changes in the incidence of GOAs, whereby there is a falling rate of gastric adenocarcinoma and a rapid increase in the rate of oesophageal adenocarcinoma [126]. Environmental risk factors which have led to a decrease in the incidence of gastric adenocarcinoma include diets containing less salt and fresher food and a reduction in the prevalence of *Helicobacter pylori* infection [127]. Factors responsible for the

ongoing rise in the incidence of oesophageal junctional adenocarcinoma include increased rates of central obesity and reflux leading to Barrett's oesophagus, the precursor lesion for oesophageal adenocarcinoma [128]. The focus of this thesis is on gastric and oesophageal adenocarcinoma including oesophageal junctional adenocarcinoma, but not oesophageal squamous cell carcinoma which has an epidemiology and treatment paradigm distinct from adenocarcinoma.

Primary tumour (T)

Tx	Primary tumour cannot be assessed
T0	No evidence of primary tumour
Tis	Carcinoma in situ: intraepithelial tumour without invasion of the lamina propria
T1	Tumour invades lamina propria, muscularis mucosae, or submucosa
T1a	Tumour invades lamina propria or muscularis mucosae
T1b	Tumour invades submucosa
T2	Tumour invades muscularis propria
T3	Tumour penetrates subserosal connective tissue without invasion of visceral peritoneum or adjacent structures
T4	Tumour invades serosa (visceral peritoneum) or adjacent structures
T4a	Tumour invades serosa (visceral peritoneum)
T4b	Tumour invades adjacent structures

Regional lymph nodes (N)

NX	Regional lymph node(s) cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in 1-2 regional lymph nodes
N2	Metastasis in 3-6 regional lymph nodes
N3	Metastasis in seven or more regional lymph nodes
N3a	Metastasis in 7-15 regional lymph nodes
N3b	Metastasis in 16 or more regional lymph nodes

Distant metastasis (M)

M0	No distant metastasis
M1	Distant metastasis

Clinical staging

Stage	T	N	M
0	Tis	N0	M0
I	T1	N0	M0
	T2	N0	M0
IIA	T1	N1, N2, N3	M0
	T2	N1, N2, N3	M0
IIB	T3	N0	M0
	T4a	N0	M0
III	T3	N1, N2, N3	M0
	T4a	N1, N2, N3	M0
IVA	Any T	Any N	M0
IVB	Any T	Any N	M1

Table 1.4 TNM classification for gastric cancer [14]

Primary tumour (T)

Tx	Primary tumour cannot be assessed
T0	No evidence of primary tumour
Tis	High grade dysplasia, defined as malignant cells confined by the basement membrane
T1	Tumour invades lamina propria, muscularis mucosae, or submucosa
T1a	Tumour invades lamina propria or muscularis mucosae
T1b	Tumour invades submucosa
T2	Tumour invades muscularis propria
T3	Tumour invades adventitia
T4	Tumour invades adjacent structures
T4a	Tumour invades the pleura, pericardium, azygous vein, diaphragm, or peritoneum
T4b	Tumour invades other adjacent structures, such as aorta, vertebral body, or trachea

Regional lymph nodes (N)

NX	Regional lymph node(s) cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in 1-2 regional lymph nodes
N2	Metastasis in 3-6 regional lymph nodes
N3	Metastasis in seven or more regional lymph nodes
N3a	Metastasis in 7-15 regional lymph nodes
N3b	Metastasis in 16 or more regional lymph nodes

Distant metastasis (M)

M0	No distant metastasis
M1	Distant metastasis

Clinical staging

Stage	T	N	M
0	Tis	N0	M0
I	T1	N0	M0
IIA	T1	N1	M0
IIB	T2	N0	M0
III	T2	N1	M0
	T3-T4a	N0-N1	M0
IVA	T1-T4a	N2	M0
	T4b	N0-N2	M0
	Any T	N3	M0
IVB	Any T	Any N	M1

Table 1.5 TNM staging classification for adenocarcinoma of the oesophagus and oesophagogastric junction [129]

1.6.2 Established treatments for resectable GOA

Approximately 25% of GOAs are resectable at presentation, however relapse following surgery is common even for patients with early stage GOA (e.g. OS is approximately 50% for resected stage IIA disease (see Tables 1.4 and 1.5). Therefore (neo)adjuvant therapy in addition to surgery is routinely recommended for patients. Since an absolute benefit in overall survival of 13-15% has been demonstrated in randomised trials [130-132] (see Table 1.6) perioperative chemotherapy has become a standard of care in operable GOAs.

The first landmark trial was the Medical Research Council Adjuvant Gastric Infusional Chemotherapy (MAGIC) study, which included approximately 500 patients with potentially resectable gastric, gastroesophageal junction (GOJ), or distal oesophageal adenocarcinoma with stage T2 or higher. Participants were randomised to receive surgery alone or surgery plus neoadjuvant and adjuvant (i.e. perioperative) chemotherapy. Chemotherapy consisted of 3 cycles each of preoperative and postoperative epirubicin 50 mg/m^2 3-weekly, cisplatin 60 mg/m^2

3-weekly and continuous intravenous infusion of fluorouracil 200 mg/m^2 for 21 days. Patients who received chemotherapy in combination with radical surgery had a higher rate of curative resection (79. %) than participants receiving radical surgery alone (70%; p=0.03 by the chi square test), with improved 5-year overall survival rates in the perioperative group compared with surgery alone (36% versus 23%, HR for death 0.75, 95% CI 0.60 to 0.93; p=0.009). This study provided the first evidence from a large randomised controlled study that perioperative chemotherapy improves OS in this population.

The next landmark trial was the Fédération Francophone de Cancérologie Digestive (FFCD) 9703 study, a multi-centre randomised controlled study performed in France [131]. The study enrolled 224 participants with gastric, GOJ, or distal oesophageal adenocarcinoma amenable to curative intent resection. Patients were randomised to receive surgery alone or surgery plus neoadjuvant and adjuvant chemotherapy (three or four cycles of cisplatin 100 mg/m^2 + 5-fluorouracil 800 mg/m^2 on days 1-5 as a continuous infusion, every 28 days [FP]). However, only half of the participants received adjuvant therapy due to poor tolerance or decline in performance status post-operatively. Overall, patients receiving perioperative chemotherapy had a higher R0 resection rate (microscopically margin negative) compared with those receiving surgery alone (84% versus 73%, p=0.04), with no evident difference in postoperative complications. The mOS was improved in the perioperative chemotherapy group with HR of death 0.69 (95% CI 0.50-0.95, p=0.02) with 5-year survival rates of 38% (95% CI 29-47%) versus 24% (95% CI 17-33%).

The most recent landmark study trial was the FLOT4-AIO trial, which was a randomised, phase II/III study in patients with resectable gastric and gastro-

oesophageal adenocarcinomas [133]. Patients were assigned (1:1) to either 4 preoperative and 4 postoperative 2-week cycles of FLOT (docetaxel 50 mg/m², intravenous oxaliplatin 85 mg/m², intravenous leucovorin 200 mg/m², and fluorouracil 2600 mg/m² as a 24-hour infusion all on day 1), or 3 preoperative and 3 postoperative 3-week cycles of intravenous ECF/ECX (epirubicin, cisplatin and 5FU or epirubicin, cisplatin and capecitabine). FLOT was associated with significantly higher proportions of patients achieving pathological complete regression compared with ECF/ECX 16% (95% CI 10–23%) versus 6% (95% CI 3–11%; p= 0·02). In the phase III part of the study, FLOT also improved OS with a median OS of 50 months vs. 35 months for ECF/ECX (HR 0.77 [0.63-0.94]; p=0.012) [134]. Based on these results, the FLOT regimen is widely considered a new systemic perioperative treatment standard of localised GOAs.

Trial	Number of patients	Patient population	Treatment	Efficacy endpoints	Reference
MAGIC	503	Resectable adenocarcinoma of the stomach, oesophagogastric junction or lower oesophagus	Perioperative chemotherapy (3 cycles of i.v. epirubicin 50 mg/m^2 on day 1+ cisplatin 60 mg/m^2 on day 1 and continuous i.v. infusion of fluorouracil 200 mg/m^2 for 21 days, q3w) plus surgery Or surgery alone	**OS**: 36% vs. 23% 5 year survival (HR 0.75; 95% CI 0.60 – 0.93, p=0.009) **R0 resection rate**: 79.3% vs. 70.3% (p=0.03)	Cunningham et al, 2006
ACCORD07-FFCD 9703	224	Resectable adenocarcinoma of the stomach and lower oesophagus	Perioperative chemotherapy (2-3 cycles of i.v. cisplatin 100 mg/m^2 and continuous i.v. infusion of fluorouracil 800 mg/m^2 on days 1-5 q4w) plus surgery Or surgery alone	**OS**: 38% vs. 24% 5 year survival (HR 0.69; CI 0.50-0.95, p=0.02) **R0 resection rate**: 84% vs. 73% (p=0.04)	Ychou et al, 2011
FLOT4-AIO	716	Resectable gastric and gastro-oesophageal junction adenocarcinoma	Perioperative chemotherapy with 3x ECF/ECX (epirubicin 50 mg/m^2 and cisplatin 60mg/m^2 on day 1 plus either fluorouracil 200mg/m^2 as continuous i.v. infusion or 1250 mg/m^2 capecitabine orally on days 1 to 21, q3w) Or perioperative chemotherapy with 4x FLOT (docetaxel 50 mg/m^2, oxaliplatin 85 mg/m^2, leucovorin 200 mg/m^2 and fluorouracil 2600 mg/m^2 as 24 hour infusion on day 1, q2w)	**OS**: 35 months vs. 50 months (HR 0.77, CI 0.63-0.94; p=0.012) **Pathological complete response rate**: 6% vs. 16% (p=0.02)	Al-Batran et al, 2019

Table 1.6: Pivotal studies in resectable gastro-oesophageal adenocarcinoma [130, 131, 135]

1.6.3 Future directions to improve response rates to perioperative treatment

Cancers can develop resistance mechanisms to therapy through genetic or epigenetic changes, leading to disease progression and overall reduced survival. Thus, there remains an urgent need to further improve poor survival rates by developing innovative therapeutic combinations, ideally using non-cross resistant therapies that can act synergistically to hinder the evolution of drug resistance. Translating the findings of basic science research and clinical trials from the advanced disease setting to the perioperative setting has been relatively unsuccessful and to date, no targeted therapy has reported a survival improvement in patients with operable gastroesophageal cancer (see Chapter 6 Introduction section for details), thus there is currently no role for these agents in the perioperative setting. However several studies are currently recruiting to combination therapy including the FLOT6 and INNOVATION trials, combining trastuzumab and pertuzumab to perioperative chemotherapy in HER2-overexpressing GOAs. The FLOT7 study is currently recruiting patients with operable HER2-negative GOAs to address the value of ramucirumab an anti-VEGFR-2 antibody in addition to perioperative FLOT [136].

The newest class of systemic anti-cancer therapies which have recently entered the field are immunotherapies including immune checkpoint blockade and adoptive cell therapy. The ultimate goal of immunotherapy is to establish a durable population of highly active, tumour-specific B-cells and T-cells that can work towards lysing tumour cells and eradicating cancers [137]. Strategically combining immunotherapies with other systemic therapies to harness potential synergies can be crucial for maximising their clinical activity in order to achieve

the greatest benefits for patients [138-141]. This formed the basis for developing the ICONIC study detailed in Chapter 6.

1.7 Overview of previous ctDNA work in GOA and CRC

Intratumour heterogeneity is recognised as a major challenge in the delivery of effective treatment in CRC and GOAs [142, 143]. Application of targeted genomic sequencing to cfDNA analysis has been shown to allow the detection of mutations that are heterogeneous within GOAs [144, 145]. For example, 15% of advanced gastric cancer patients overexpress *HER2* and these patients would benefit from targeted therapy, however there is high intratumour heterogeneity in advanced gastric cancer leading to discordant results between immunohistochemistry (IHC) and fluorescence in situ hybridisation (FISH), this can result in the need for multiple biopsies to confirm *HER2* status. ctDNA can partly overcome this issue such that ddPCR has been utilised to demonstrate 91% concordance of *HER2* status with tumour tissue and thus could be used to complement clinical diagnosis where tumour biopsies fail to confirm *HER2* status [144].

Longitudinal ctDNA analysis can serve as a potential biomarker of response to relatively novel treatments, such as PD-1 antibodies in advanced GOA. This was demonstrated in a prospective study of 46 patients with advanced gastric cancer who were treated with PD-1 antibody and underwent longitudinal ctDNA NGS. Patients who achieved a decline of >25% maximal somatic variant allele frequency (maxVAF) had a longer PFS than patients who did not achieve maxVAF decline (7.3 months; 95% CI 2.4-4.8 months vs. 3.6 months; 95% CI 4.6-10.0 months, p= 0.0011). Additionally, a decline was suggestive of a trend

towards improved overall response rates (53.3% vs. 13.3%), though this was not statistically significant (p=0.06) [146].

In early stage GOAs maxVAF has also been used as a surrogate for disease burden [147]. Patients who had a detectable mutation (ctDNA positive) after curative-intent surgery had a poorer disease free survival compared with those who were ctDNA negative (median disease free survival of 12.5 months vs. unreached; HR= 0.1; 95% CI 0.01 – 1.1 p=0.03). However careful analysis is required as highlighted in this study, where three apparently ctDNA-positive patients did not relapse following surgery and it was subsequently realised that these false positive results were due to clonal hematopoiesis, thus ctDNA analysis requires analysis of matched germline samples to accurately determine the origin of mutations detected [147].

Another area where liquid biopsies techniques have been long awaited is in the setting of early stage CRC, where the recommendation is for adjuvant chemotherapy (in high risk stage II and stage III CRCs) to reduce the risk of recurrence. Some patients would be cured by surgery alone and some will recur despite adjuvant therapy as a result of failure to eradicate micrometastatic residual disease, therefore some patients will have been exposed to the risk of treatment related toxicities with no overall benefit. Post-operative ctDNA analysis has been shown to be prognostic in two studies, such that 79% of patients with resected stage II CRC with detectable ctDNA after surgery developed disease recurrence. 10% of patients with negative ctDNA had disease recurrence (HR 18, 95% CI 7.9 to 40; p<0.001) at median follow up of 27 months [148]. In patients with stage III colon cancer, a significant difference in 3-year recurrence free interval was observed in patients with detectable versus undetectable levels of

ctDNA after surgery (47% vs. 76%) and after completion of standard adjuvant chemotherapy (30% vs. 77%; HR 6.8; 95% CI 11.0 – 157.0; p<0.001) [149]. Together these studies demonstrate that ctDNA analysis after surgery is a prognostic marker and serial analysis may define a subset of patients that remain high risk even after completion of standard adjuvant therapy and may require additional therapeutic approaches.

ctDNA has been investigated as an early indicator of treatment response in mCRC [150]. In a prospective study of 53 patients, there was a statistically significant correlation of early changes in ctDNA mutant allele fraction following initiation of systemic therapy (after one cycle) and radiological response at eight weeks, demonstrating that ctDNA could be used as a potential predictive biomarker of non-response to systemic therapy, sooner than radiological assessment. Furthermore, this study demonstrated ctDNA-based response assessment may potentially complement RECIST assessment, especially in patients with non-measurable disease by RECIST criteria.

ctDNA can also be used as tool to predict disease evolution in mCRC, which can improve planning of targeted personalised therapy. In a prospective phase II trial of *RAS* wild-type mCRC patients treated with single agent anti-EGFR antibody cetuximab or panitumumab [80], serial ctDNA profiling combined with matched tissue biopsies, CEA measurements, RECIST measurements and mathematical modelling of tumour evolution enabled predictions of clonal evolution, emergence of therapy resistance and time to treatment failure in individual patients. This study highlights novel application of ctDNA profiling to predict the time to relapse and guide clinical decision making for adaptive therapy personalistion.

1.8 Outline of thesis

Metastatic CRCs are associated with poor outcomes, in part due to intratumoural heterogeneity being a major barrier to predicting the evolution of mCRCs and successful treatment. I aimed to address this through utilising a minimally invasive ctDNA assay to sequence commonly mutated genes and detect subclonal variants with a high degree of sensitivity.

In the first results chapter (Chapter 3), I describe the development of a novel fully customisable error corrected capture based ctDNA targeted sequencing assay to interrogate 32 genes that can be recurrently mutated in mCRC or implicated in chemotherapy resistance. I apply this assay to a series of mCRC patients to detect driver mutations and CNAs and investigate the clonal and subclonal genomic landscape (Chapter 4). I then utilise this assay to study the genetics of chemotherapy resistance (Chapter 5) and develop whole exome ctDNA sequencing for identification of candidate mechanisms of resistance.

I have also designed a translational clinical trial (ICONIC- Chapter 6) to improve treatment approaches in localised GOAs, utilising a novel combination of chemo-immunotherapy as perioperative treatment in early stage GOAs. The study has a strong translational component to investigate candidate predictive biomarkers of tumour response through longitudinal ctDNA and tumour sampling.

CHAPTER 2: MATERIALS AND METHODS

2.1 Materials

2.1.1 3D organoid culture reagents

- Advanced DMEM:F12 (Life Technologies 12634-010)
- DMEM:F12 (Sigma D6421) supplemented with 10% FCS, Penicillin/Streptomycin and Glutamax
- Foetal bovine serum (Life Technologies 10270106)
- Penicillin/Streptomycin X100 stock (Life Technologies 15140122)
- Glutamax X100 stock - Life Technologies (A1286001)
- Matrigel- Growth factor reduced, phenol red free. (Corning 356231) suppliers: Scientific Lab Supplies or Fisher.
- Tryple Express (Life Technologies 2605028)
- PBS (Sigma D8537)
- Fungizone (Life Technologies 15290018)
- Nystatin (Life Technologies 15340029)

Culture Media: Modified from Cleevers lab's formulation for colorectal organoid cultures using Advanced DMEM:F12.

2.2 Methods

2.2.1 Associated clinical trials and sample processing

2.2.1.1 FOrMAT clinical trial

The Feasibility of a Molecular Characterisation Approach to Treatment (FOrMAT) study (ClinicalTrials.gov identifier NCT02112357) was a single-centre translational study conducted at the Royal Marsden Hospital, in patients with

gastrointestinal tumours to assess the feasibility of NGS to guide treatment personalisation in routine clinical practice [151]. Samples were selected from patients who were about to commence a new line of systemic therapy, either because they progressed on prior therapies or because they were recently diagnosed with metastatic disease.

222 patients with locally advanced or metastatic gastro-oesophageal, pancreatic, biliary tract or colorectal cancers were recruited to the study between February 2014 and November 2015. There were no exclusion criteria however patients for whom a tissue sample could not be obtained were replaced (Figure 2.1). Written informed consent was obtained from all patients. The study was carried out in accordance with the Declaration of Helsinki and approved by the national UK ethics committee (UK Research Ethics Committee approval: 13/LO/1274). The primary endpoint of the trial involved targeted capture-based DNA sequencing using archival FFPE specimens in order to detect mutations, copy number aberrations and translocations in up to 46 genes which had prognostic or predictive significance, or were potential targets in existing or upcoming clinical trials (Table 4.1). Non-cancerous control DNA was obtained from whole blood. Mutation calls, BAM files of tumour and blood germline sequencing data as well as clinical data were available for my analysis by the Centre for Molecular Pathology, ICR/RMH. As part of the translational tissue collection component of the trial, blood samples were also taken from enrolled patients at trial entry and at the timepoint of response assessment CT scans during treatment.

Sequencing results were discussed at a sequencing tumour board where recommendations were made on clinical significance and potential actionability

of any variants. Any changes to patient's management were at the discretion of the treating physician.

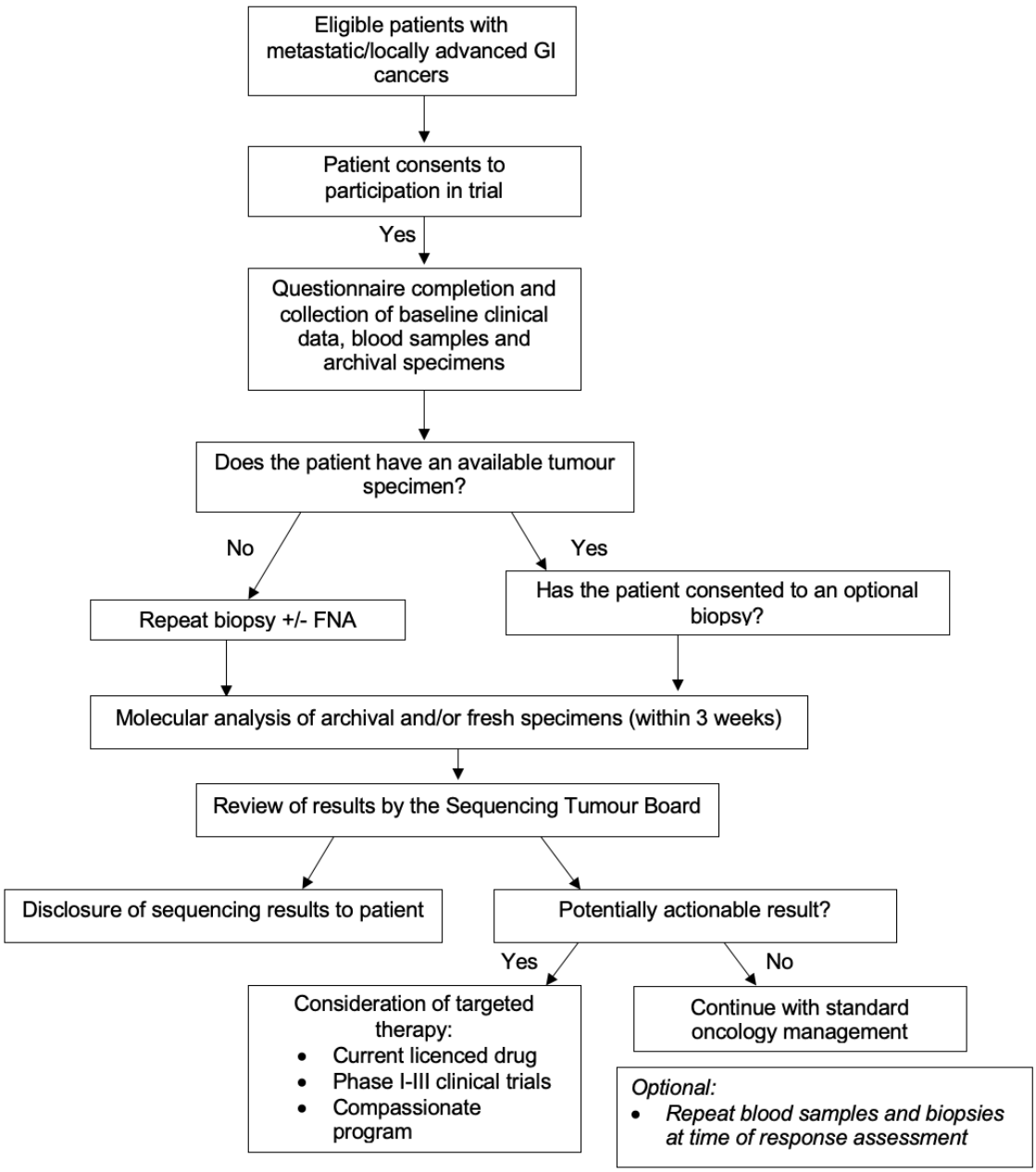


Figure 2.1: FOrMAT study design
Reproduced with permission from the FOrMAT trial protocol

2.2.1.2 PROSPECT-C clinical trial

The PROSPECT-C trial (ClinicalTrials.gov identifier NCT02994888) was a phase II, open-label, non-randomised study of anti-EGFR monoclonal antibodies in patients with *RAS* wild type mCRC. Patients were considered eligible for this

study if they fulfilled all the following criteria: (i) chemorefractory (at least two lines of chemotherapy) metastatic disease; (ii) *KRAS/NRAS* WT (on archival material according to hospital policy); (iii) measurable disease; and (iv) metastatic sites amenable to biopsy. Patients received cetuximab/panitumumab through the Cancer Drug Fund.

Written informed consent was obtained from all patients. The study was carried out in accordance with the Declaration of Helsinki and approved by the national UK ethics committee (UK Research Ethics Committee approval: 12/LO/0914). All participants were required to have mandatory pretreatment biopsies, biopsies at 3 months (if there was evidence of partial response by RECIST v1.1 criteria) and at the time of disease progression and regular plasma collection for ctDNA analysis.

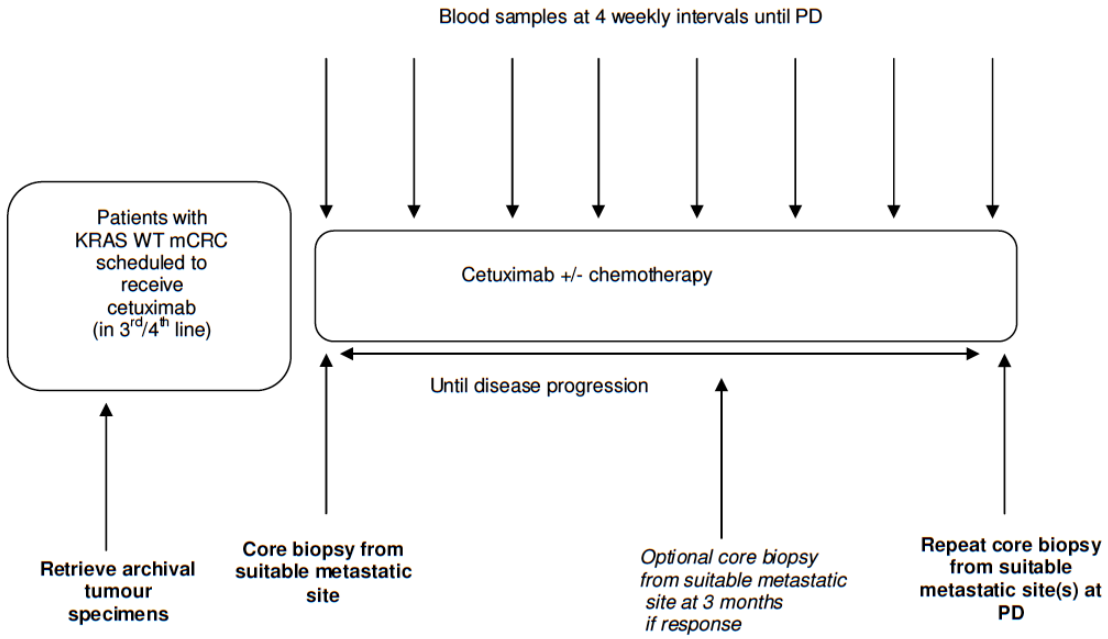


Figure 2.2: PROSPECT-C trial design
Reproduced with permission from the PROSPECT-C trial protocol

2.2.1.3 ICONIC clinical trial

The ICONIC trial (ClinicalTrials.gov identifier: NCT03399071), is a phase IB/II trial utilising a novel combination of chemo-immunotherapy in the peri-operative setting for operable gastro-oesophageal adenocarcinomas (GOA). This trial is designed to evaluate the safety and efficacy of avelumab with cytotoxic FLOT. A translational sub-study has been designed to investigate biomarkers of response/resistance and to identify potentially targetable mechanisms of resistance to improve future combination immunotherapy trials. Patients who were at least 18 years old and had a performance status of 0-1 were eligible if they had histologically confirmed operable GOA with no evidence of distant metastases.

Written informed consent was obtained from all patients. The study was carried out in accordance with the Declaration of Helsinki and approved by the national UK ethics committee. All participants were required to have mandatory endoscopic tumour biopsies at baseline and between cycle 2 and 3 of pre-operative chemo-immunotherapy. Image-guided tumour biopsy is also performed at progression or recurrence after surgery. Longitudinal blood sample collection is performed before each treatment cycle and at each follow-up visit. The study design is described in Chapter 6.

2.2.1.4 Healthy donor samples

cfDNA samples from healthy donors were collected from consenting volunteers through the 'Improving Outcomes in Cancer' Biobanking Protocol at Barts Cancer Centre Hospital. Written informed consent was obtained from all patients. The study was carried out in accordance with the Declaration of Helsinki and approved by the national UK ethics committee (UK Research Ethics Committee

approval: 13/EM/0327). Donors donated 27 ml whole blood, which was collected in EDTA tubes.

2.2.2 Plasma extraction

Whole blood was collected in EDTA (Ethylenediaminetetraacetic acid) tubes and centrifuged within two hours, in two sequential centrifugation steps, one at low speed (1600g at room temperature, for ten minutes) and subsequently at high speed (16000g at 4ºC, for ten minutes) to isolate the plasma for cryostorage at -80°C and subsequent cfDNA extraction.

2.2.3 cfDNA extraction protocol

cfDNA was extracted from 4 ml plasma per mCRC patient and from 2x4 ml from healthy donors using the QIAamp Circulating Nucleic Acid Kit (Qiagen), as per the manufacturer's protocol. cfDNA was eluted in 30 μl 10 mM Tris Ultrapure buffer + 0.1 mM EDTA (low EDTA TE) (Sigma-Aldrich) and stored at -20ºC.

cfDNA fragments in the range 100-700 bp were quantified on a Bioanalyzer High Sensitivity DNA chip (Agilent), and if this showed a high concentration of cfDNA the sample was subsequently quantified on a 7500 DNA chip.

2.2.4 cfDNA quantification

DNA concentrations were quantified using 1 μl of the extracted cfDNA in low EDTA TE, on an Agilent Bioanalyzer High Sensitivity chip as per the manufacturer's protocol. If this indicated a concentration of >15 ng DNA/ml plasma, a second quantification was performed on the standard 7500 chip. DNA fragment size distribution was assessed simultaneously. cfDNA was stored at -20ºC before library preparation. Bioanalyzer quantification was also used following pre-hybridisation PCR and post-hybridisation PCR products during

sequencing library preparation.

2.2.5 ddPCR protocol

ddPCR was performed for independent validation of discordant mutations i.e. those called in cfDNA but not detected in tumour tissue. Pre-validated commercially available ddPCR SNP Genotyping Assays were used for *BRAF, KRAS* and *PIK3CA* variants (Life Technologies; Assay ID A44177 *BRAF*_476, A44177 *KRAS*_555, A44177 *PIK3CA*_763) and the remaining probes were custom designed with ThermoFisher Scientific (Table 2.1).

Assay Name	Target Sequence	Amplicon length (bp)	Target Location
ATM_G449A	CCCAACAGCGACATG **[G/C]** GGAACGTACACCA	125	11: 108121538 G>C
ATM_R2993	GACCAAGAATGCAAA **[C/T]** GAAATCTCAGGTG	139	11: 108235935 C>T
CTNNB1_V676I	AAGAAACGGCTTTCA **[G/A]** TTGAGCTGACCAG	139	3: 41278150 G>A
MAP2K2_E328K	ATCAAGCACAAACCT **[C/T]** GTTCACAATATAG	163	19: 4097279 C>T
TP53_A355S	CTGGCTCCTTCCCAG **[C/A]** CTGGGCATCCTTG	177	17: 7573964 C>A
TP53_E298K	TCCCTGGGGGCAGCT **[C/T]** GTGGTGAGGCTCC	159	17: 7577046 C>T
TP53_H178P	GAGCAGCGCTCATGG **[T/G]** GGGGGCAGCGCCT	163	17: 7578397 T>G
TP53_I254V	CTTCCAGTGTGATGA **[T/C]** GGTGAGGATGGGC	149	17: 7577521 T>C

Table 2.1: Custom designed primers and probes for ddPCR validation

Input cfDNA for each ddPCR experiment varied, depending on amount of residual material available (17 ng for case 10, 20 ng for cases 8 and 14 and >25 ng for all other cases). cfDNA was added to a ddPCR reaction containing 11 μl mastermix (10 μl 2x ddPCR Supermix for Probes and 1 μl 20x target primer/probe mix for both mutant and wild type alleles) and made up to a total volume of 21 μl with nuclease-free water as per the Bio-Rad manufacturer's protocol. At least four positive control (patient cfDNA for *KRAS* Q61H, *BRAF* V600E, *PIK3CA* H1047R or gBlock controls- donated by Turner Lab) and negative control wells were run to verify assay performance and facilitate thresholding in fluorescence values. For each patient, plasma was analysed in triplicate and ddPCR results are based on the merged data of these wells. Erroneous wells were not included in the merge.

The reaction mixture was placed into the sample well of DG8 cartridge (Bio-Rad). A volume of 70 μl of droplet generation oil was loaded into the oil well, and the reaction was partitioned into a median of 17,676 droplets per sample in a Bio-Rad QX-200 droplet generator according to the manufacturer's protocol. Emulsified PCR reactions were transferred to a 96-well PCR plate (Eppendorf) and run on a G-Storm GS4 thermal cycler incubating the plates at 95^{o}C for 10 minutes followed by 40 cycles of 95^{o}C for 15 sec and 60^{o}C for 1 minute. Plates were read on the QX200 droplet reader using QuantaSoft analysis software (Bio-Rad) to acquire and analyse data. ddPCR analysis of gBlock controls and no DNA template controls were included as standard in all experiments. Wild type (WT) and mutant (MUT) targets are detected using different fluorophore conjugations (VIC/HEX for WT and FAM for MUT).

Allele fraction was calculated as:

A.F. = Nmut /(Nmut+Nwt)

where Nmut is the number of mutant events and Nwt is the number of wild type events per reaction.

2.2.6 cfDNA-seq Assay

2.2.6.1 Library preparation protocol

I modified the Agilent SureSelect$^{XT\text{-}HS}$ protocol in order to assure a reliable performance with 25 ng cfDNA input. All PCR steps were performed on an Eppendorf Mastercycler nexus GSX1/SX1e. 8 cycles of pre-hybridisation PCR were optimal for 25 ng of input cfDNA to generate the amount of product required (500–1000 ng) for in-solution capture. The entire product was used as input for hybridisation to the custom-designed Agilent SureSelect capture bait library (Table 3.1). Target fragments were captured using streptavidin beads after modifying the manufacturer's protocol to increase the stringency of washes to reduce off-target reads (see Chapter 3).

To minimise bias during library preparation, identical cfDNA input amounts and master mixes of reagents were used. All other reagents were added according to the manufacturer's protocol and 60 cycles fast hybridisation were performed, taking ~1.5h. Capture was started immediately after the final hybridisation cycle and proceeded for 30 minutes at room temperature.

Next, post-capture washes of 6 incubations x5 minutes each at 70^{o}C were carried out. 10 PCR cycles were used for post-capture amplification and this was followed by 2 rounds of 1x Ampure XP bead cleanup to remove un-incorporated

primers.

The final prepared sequencing libraries were profiled on a Bioanalyzer High Sensitivity DNA chip and quantified by qPCR using the Kapa Library Quantification kit before pooling. Pooled libraries were clustered using the Illumina cBot and sequenced with paired-end 75 reads on an Illumina HiSeq2500 in rapid output mode.

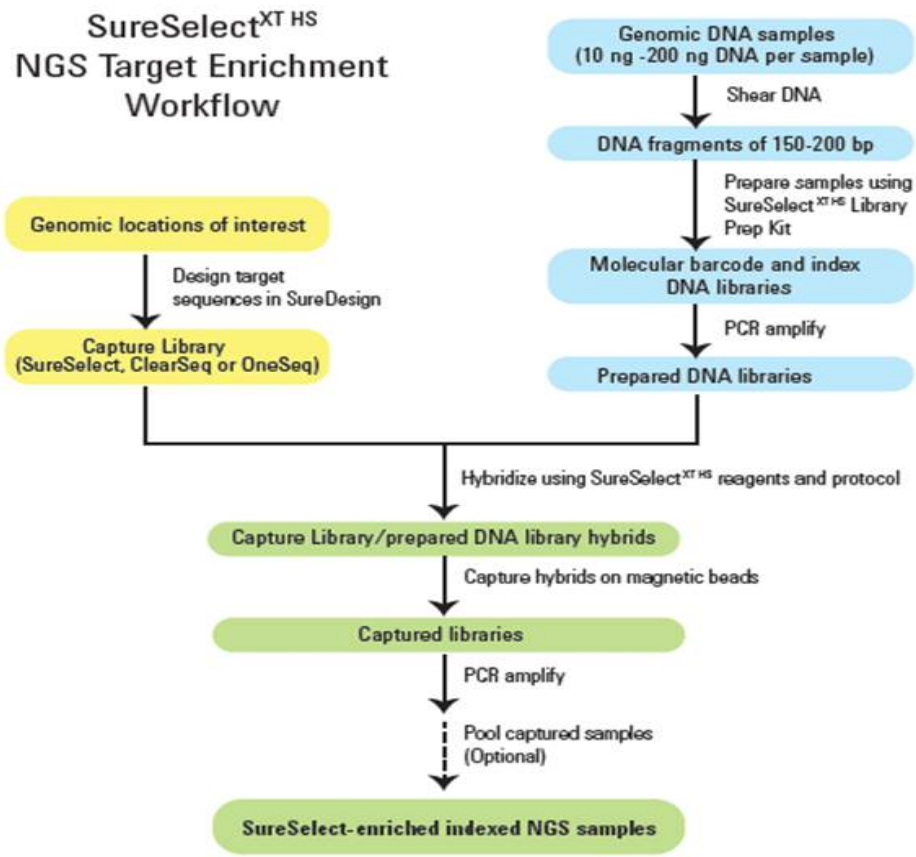


Figure 2.3: Workflow illustrating target enrichment capture process of library preparation with sequencer specific adaptors and indexes.

The library is hybridised with biotinylated RNA library baits. Target regions are pulled-down using magnetic streptavidin beads, which can be amplified and sequenced. A number of experiments were performed to optimise factors such as bait formulation, hybridisation and capture washes in order to improve the sequencing efficiency (see Chapter 3).

2.2.6.2 Sequencing on Illumina 2000 HiSeq

DNA libraries were sequenced by the Tumour Profiling Unit (TPU). Individual libraries were first quantified on a tapestation (Agilent) followed by quantification on QPCR to confirm accurate library molarity prior to pooling in equimolar amounts, or in the desired percentages according to depth required. The pooled libraries were once again requantified on both tapestation and QPCR to ensure the final molarity agrees with what was expected but also to get a reading as accurate as possible, which could affect subsequent clustering.

Based on the pool molarity, the libraries are denatured and diluted in hybridisation buffer to the desired loading molarity (in the case of the Rapid runs- 14pM). 120 μl is loaded on the cBot for subsequent cluster generation and sequenced with 75 bp paired-end reads on an Illumina HiSeq 2500 platform. This is sufficient to sequence the short cfDNA fragments of ~170 bp length.

2.2.6.3 Variant calling

NGS data read alignment, deduplication and quality control checks of BAM files were performed by bioinformaticians in the Gerlinger lab: Dr Matthew Davies, Dr Andrew Woolston and Dr Louise Barber. Thereafter, variant calling and visualisation of mutations on IGV was performed by myself. The Agilent SureCall software (version 4.0.1.45) was used to trim and align fastq reads to the hg19 reference genome with default parameters. Raw on-target depths and on-target fraction were assessed before de-duplication. MBC de-duplication was performed with SureCall permitting one base mismatch within each MBC and consensus reads that were only supported by a single read were removed as MBC error correction is ineffective on such ‘lone reads’. The on-target depth after de-duplication was assessed after single read families were removed.

Variant calling was performed using the SureCall SNPPET SNP Caller with the following parameters: Variant score threshold =0.01; Minimum quality for base =30; Variant call quality threshold =40 (manufacturer default setting: 100); Minimum Allele Frequency =0.001; Minimum number of reads per barcode =2; no region padding; and masked overlap between reads. After calls were generated by the bioinformatics team (Dr Matthew Davies, Dr Andrew Woolston and Dr Louise Barber), I then performed further filtering of the primary calls by visualising sequencing data on IGV [152].

Variants predominantly located in reads with an alignment score of zero or in reads with multiple non-contiguous non-reference bases were removed as these are usually indicative of misalignment. Where the insert size was < 74 and all 74 bases of the read are aligned, any variants occurring at the end of the read were excluded as they represent erroneous mapping of the adapter sequence. Thereafter, Dr Louise Barber cross-checked all variant positions identified in patient cfDNA were with the dataset of six HD samples using bam-readcount tool (https://github.com/genome/bam-readcount). The majority of called variants were completely absent in the six HD samples however, mutations whose VAF was not at least double that of an identical variant in a healthy donor sample were discounted in order to filter out recurrent false positives.

2.2.6.4 Bioinformatic identification of duplexes

Variants only passed filtering if supported by a) at least one duplex pair of consensus families in which the variant was not the last base in the read and b) at least two reads with different alignment positions, this was initially performed manually by myself using IGV. Once I defined the conditions required to confirm a duplex, the ‘duplexCaller’ bioinformatics tool was developed by a staff

bioinformatician in the Centre for Evolution and Cancer (Dr Dimitrios Kleftogiannis), which is freely available for academic use at https://github.com/dkleftogi/duplexFiltering. Dr Dimitrios Kleftogiannis designed this tool using the SureCall variant calls (in a VCF-like format) and the corresponding MBC de-duplicated sample BAM file as input. SAMtools [153] is used to identify only the consensus reads that span the genomic positions of interest. The list of consensus reads is parsed in the SAM format, selecting and saving the start position of each consensus read, the start position of the next read in the template, as well as the length of the template. By applying memory-efficient indexing of the relevant information, a data structure is generated for all consensus reads supporting variants that could potentially form duplexes (i.e. a candidate set of duplexes). In addition, this technique allows discarding of the non-relevant information for every input variant, and thus it improves the execution time.

Next, the data structure is used to process all fragments that start and stop in the same genomic positions. By parsing the list of fragments with the same start and stop positions, the consensus reads being first in the template and second in the template for each fragment can be identified. Thus, in a given set of fragments with the same start and stop positions, the number of duplexes is the minimum number of pairs of fragments having their leftmost consensus read as first in the template, and second in the template respectively. As an output, the duplexCaller reports the total number of duplexes found for every variant call that passed the initial filtering.

2.2.6.5 Standard de-duplication

To compare our MBC analysis with current methods, standard de-duplication and subsequent calling was performed by Dr Louise Barber, with SureCall using the same parameters, except that de-duplication was solely based on genome alignment and insert size. As the post-call filtering of MBC de-duplicated data essentially required a variant to be present in a minimum of three reads (two that form a duplex and one additional read with a different alignment position), we also mandated a minimum of three reads with a specific variant to support a mutation call in the standard-duplicated data to enable a fair comparison.

2.2.6.6 Copy number analysis

Copy number analysis was performed by Dr Louise Barber using CNVkit [154], which analyses both, on-target and off-target reads to reconstruct genome wide copy number profiles. BAM files resulting from MBC de-duplication, before removal of single-read consensus families were used to generate copy number profiles. The six sequenced healthy donor cfDNA samples were used as a pooled normal reference dataset. Antitarget average size was set to 30,000 bp to improve resolution. Seg files were created from the .cns segmented output for visualisation of the detected copy number aberrations as a heatmap on IGV.

2.2.7 Low coverage whole genome sequencing

Libraries were prepared with 10 ng of input DNA using the NEBNext Ultra II DNA Library Prep kit (New England Biolabs) as per manufacturer's protocol. Libraries were quantified using the Bioanalyzer (Agilent), pooled and sequenced using an Illumina HiSeq2500. Genome wide copy number analysis was subsequently performed by viewing seg files as a heatmap using the Integrated Genome Viewer software (Broad Institute; v.2.3.97), allowing comparison of genomic

somatic CNA profiles across multiple samples with the ability to zoom in to areas of interest in order to investigate genes located within this genomic region.

2.2.8 AVENIO sequencing

AVENIO sequencing was carried out by the Hubank Team in the Centre for Molecular Pathology. DNA sequencing libraries were prepared using the AVENIO ctDNA Analysis Kit (Roche) with 25 ng input DNA where available. Library preparation was performed as per manufacturer's protocol and hybridised to the AVENIO Expanded capture kit to enrich for a panel of 77 target genes. Sequencing was performed on an Illumina NextSeq 500 with 150 paired end reads in High Output mode. Data was analysed on the Roche AVENIO Custom App via a locally installed Roche server to generate mutation calls. The Roche analysis pipeline is reported to call variants down to 0.5% VAF and targeted indels and fusions to 1%.

2.2.9 Whole Exome Sequencing on Novaseq

Whole exome sequencing was performed on cfDNA extracted from FOrMAT patient plasma, sheared FFPE DNA and germline DNA from blood cell pellets. Exome sequencing libraries were prepared with 25 – 50 ng cfDNA input, using the Agilent optimised SureSelect XT-HS protocol with the Human All Exon v5 baits, according to manufacturer's protocol. Matched normal germline DNA libraries from peripheral blood were prepared and sequenced using less sequencing capacity than cfDNA samples (1% of the run) to help identify somatic variants from germline variants or SNPs during mutation calling.

Paired end sequencing was performed on the Illumina Novaseq S2 flow cell using a read length of 2 x 100 bp.

2.2.10 ICONIC Trial sample analysis

2.2.10.1 Multiparametric immunohistochemistry

Multiparametric analysis of ICONIC matched pre-treatment and on-treatment biopsies were performed by Dr. K. von Loga, Consultant Molecular Pathologist, RMH using the Perkin Elmer OPAL assay, as per manufacturer's protocol. The Opal workflow allows for simultaneous detection of multiple T-cell biomarkers, within a single tissue selection. Combining Opal with multispectral imaging and analysis enables simultaneous, quantitative results for biomarkers in fluorescence (CD4 stain for T helper cells; CD8 stain for cytotoxic T cell; PD-L1 stain in tumour or immune cells; nuclear FOXP3 stain for regulatory T cell and CD45RO stain for T memory cells).

Image analysis was performed using Perkin Elmer Vectra 3.0 Quantitative Pathology Imaging System and Perkin Elmer InForm v2.2 tissue finder to segment specific tissue compartments, phenotype cells based on their biomarker expression and quantify immune cell subtypes.

2.2.10.2 Patient derived organoid (PDO) culture

At the time of the mandatory baseline and week 3 (early on-treatment) endoscopies, one fresh biopsy core is collected into a vial containing phosphate-buffered saline (PBS) and transported to the laboratory on ice where it is transferred to a sterile tissue culture hood.

The tumour core is transferred to 10cm dish and coarsely chopped with a scalpel. Sections are transferred to the centre of 1 well of a 24 well plate and embedded into Matrigel, with sufficient volume to cover the tissue and attach it firmly to the

plastic. The plate is transferred to a 37ºC incubator to allow contents to set for 10 minutes. Once set, culture media and additional antifungal agents (Fungizone (1/100) and Nystatin (1/100)) are added, due to the high risk of fungal contamination. PDO cultures are incubated at 37ºC, 5% CO2 for approximately one to two weeks. Cultures are passaged into a new 24 well when good growth is visible, or after two weeks. PDOs were established and maintained by Beatrice Griffiths, Scientific Officer, Gerlinger Lab.

2.2.11 Statistical analysis

cfDNA quantities were correlated with clinical characteristics using a two-sided Student's t-Test with Excel to determine p-values and hence significance of there being a true difference between two groups. $p < 0.05$ was considered statistically significant

Progression free and overall survival (PFS, OS) were calculated as the time from the date of initiating therapy to the date of reported event. Median PFS and OS were calculated using the Kaplan-Meier method using GraphPad Prism software version 7.0 and compared using the log-rank test.

2.2.12 Tumour purity corrected VAF of subclonal mutations

To calculate purity corrected VAF of subclonal mutations arising in DNA damage repair (DDR) genes, I divided the VAF of the DDR mutation by the VAF of a truncal mutation in the same sample (either *TP53* or *BRAF*).

A variant was defined as subclonal if the purity corrected VAF was less than 25% of the highest VAF in the sample (Figure 5.1) and was defined as clonal if it was above this threshold, as previously described in mCRCs [155].

CHAPTER 3: DEVELOPMENT OF AN ERROR-CORRECTED ULTRA-DEEP CIRCULATING TUMOUR DNA SEQUENCING TECHNOLOGY FOR COLORECTAL CANCER

3.1 Introduction

Many cancers release cfDNA into the blood and numerous publications have demonstrated the feasibility to detect tumour somatic mutations, copy number changes and genomic rearrangements from cfDNA [96, 112, 156-159]. Despite clear advantages over invasive biopsies, the low tumour-derived cfDNA fractions found in many cancers, coupled with the often extremely low abundance of DNA molecules harbouring mutations that drive acquired drug resistance, complicates the detection of mutations [89, 92, 160-162]. Detection techniques with high specificity and low false positive error rates are hence crucial for reliable cfDNA analysis.

ddPCR and Beads, Emulsions, Amplifications and Magnetic (BEAMing) assays can accurately detect point mutations present at frequencies below 0.1% in cfDNA but they are restricted to the analysis of a limited number of genomic loci [113, 114]. This limits their use for mutation detection and discovery without prior knowledge of genetic alterations in the tumour. In contrast, amplicon sequencing can be applied to larger regions of interest, enabling *de novo* mutation discovery. However covering an entire gene can be difficult as this requires overlapping amplicons and multiplex designs. Additionally, variable GC contents of distinct DNA target regions lead to different melting temperatures of amplification primers. This sometimes makes multiplex PCR impossible or causes drop out of

difficult to amplify target regions. Consequently there is loss of coverage for poorly performing amplicons and overall low coverage uniformity [163]. In comparison, solution hybrid capture target enrichment technology (Figure 3.1) can interrogate larger DNA regions such as all exons of thousands of genes simultaneously. The high levels of coverage that can be achieved for entire genes makes this one of the target enrichment technologies of choice for tumour suppressor genes (TSG), such as the commonly mutated drivers of colorectal cancer *TP53* and *APC*, which are not only mutated at specific hotspots but across most of the gene length, need to be sequenced. Target gene panels can generate deeper sequencing and improved mutation detection sensitivity compared with whole exome or whole genome sequencing, however next generation sequencing (NGS) error rates rapidly increase when attempting to call mutations with variant allele frequencies (VAFs) below 5% [115].

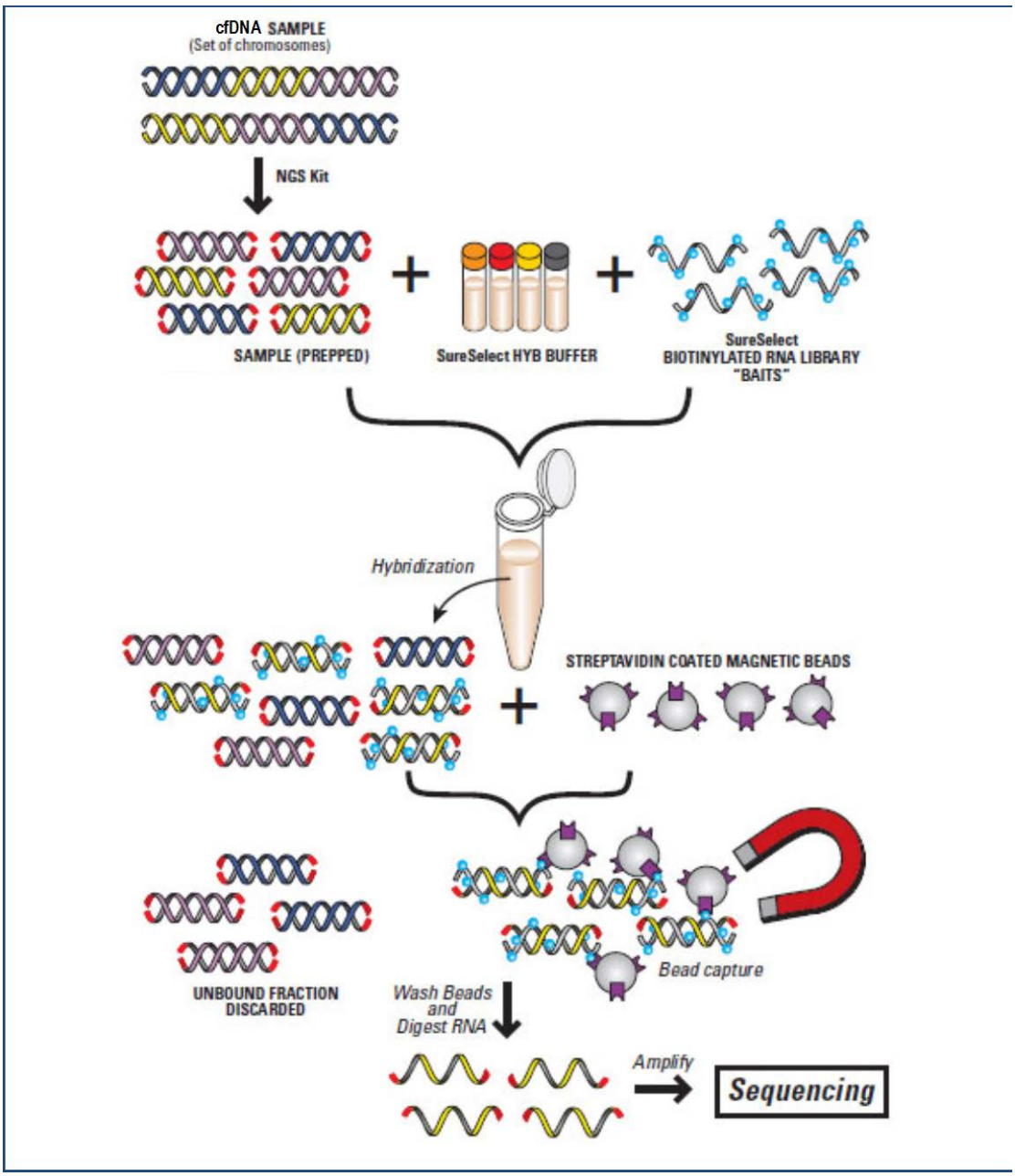


Figure 3.1: Agilent SureSelect solution hybrid capture method of target enrichment. © Agilent Technologies, Inc. Image reproduced with permission, courtesy of Agilent Technologies, Inc.

Target enrichment allows isolation of specific fragments of DNA for sequencing using a library of complimentary oligonucleotide “baits” to retrieve target DNA. The target DNA hybridises well with the baits but other DNA does not, which forms the basis of target enrichment technology for sequencing [164].

Error correction approaches have been developed and incorporated into NGS cfDNA assays to reduce this error rate [116, 117]. One approach is through building a background error correction model by sequencing multiple samples

with known mutation profiles, usually from healthy donor (HD) cfDNA, to identify positions that frequently show false positives. This information can be used to either define a minimum VAF below which a call is likely to be an error, or not call such loci. However this approach requires a lot of training data and many positions in the target region may not be callable in the end [117].

Another main approach is incorporation of molecular barcodes (MBCs) during sequencing library preparation for subsequent computational error correction. MBCs are short and random nucleotide sequences that act as a unique tag for each original DNA fragment (Figure 3.2). The simplest method of MBC tagging involves incorporation of random MBCs into one of the index read positions on the sequencing adapter. Ligation of the MBC-containing adapter to a cfDNA fragment results in a specific identifier being assigned separately to each strand of the input DNA molecule. Thereafter, polymerase chain reaction (PCR) amplification generates multiple copies of this molecule which can subsequently be used for error correction.

Errors may occur during PCR amplification or during sequencing, however in most instances errors will be confined to the minority of copies generated from the original DNA molecule, whereas a true mutation present in the starting template DNA will be present in every duplicate containing the same MBC. MBCs can be used for error correction by computationally grouping reads that have identical MBCs and the same genomic alignment position and have therefore originated from the same input cfDNA molecule, into ‘consensus families’ (Figure 3.2). Thus, if a mutation is detected in all or the majority of reads in a consensus family then it is likely that the original DNA template also carried that mutation.

However if a variant is only detected in a minority of the reads in the family, it will be disregarded and the consensus read sequence will be wild-type [116, 117].

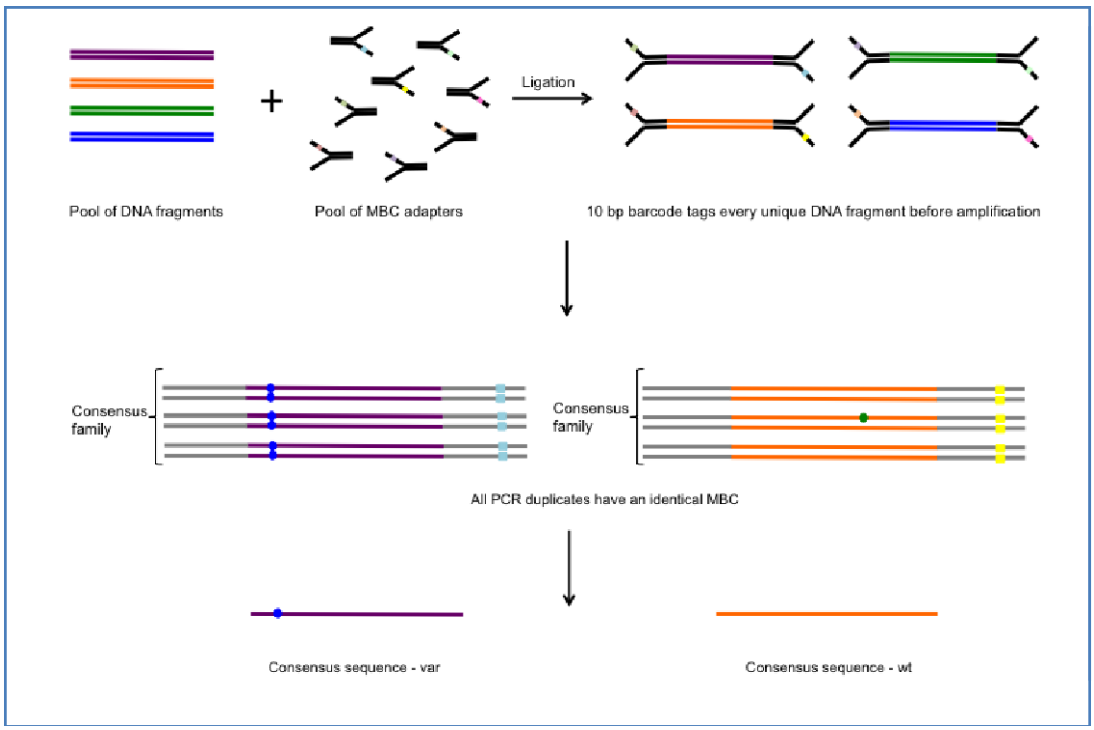


Figure 3.2: Random molecular barcodes with 10 bp length are incorporated into adapters to enhance error correction
Error correction is performed by enabling the identification of false positive calls. The MBC located in one of the index read positions in the adapter (depicted as a coloured box within the black adapter, with the other index read containing the multiplex barcode (not shown). Collapsing the reads of an entire MBC family into a single high confidence consensus read enables substantial reduction of the false positive mutation calls generated through sequencing and PCR errors. Blue circle = variant likely to be real; Green circle = variant likely to be sequencing error.

The third main approach for error correction in cfDNA-NGS is duplex sequencing [165] which entails tagging both strands of each double stranded DNA (dsDNA) with the same MBC such that the resulting forward and reverse consensus reads can be traced back to the same initial dsDNA molecule. This is achieved by incorporating a random yet complementary double stranded nucleotide MBC sequence into the adapter at the point where it is ligated to the DNA molecule. Every PCR duplicate that arises from either strand of the original dsDNA will carry

the same tag sequence therefore the two strands of a double stranded input DNA molecule can be re-identified after sequencing. True mutations can be distinguished from sequencing errors by comparison of the tag sequence on each of the two strands in the duplex, where true mutations will be present on both strands and PCR or sequencing errors will only be seen on a single strand [123]. Error correction by duplex molecules is however limited by the complicated sequencing adaptor generation, the limited adapter stability arising from the random DNA sequence and the recovery of sufficient duplex molecules to detect rare variants [166, 167].

3.2 Hypothesis and Objectives

As described in Chapter 1, the main limitations of cfDNA sequencing technologies at the start of this PhD project included the inability to target multiple whole genes with low cfDNA input amounts with simultaneous incorporation of error correction tools to accurately define subclonal mutations with VAFs <0.5%. In particular, there was no solution hybrid capture technology that incorporated MBC error correction for cfDNA sequencing. My aim was to develop an ultra-deep cfDNA sequencing (cfDNA-seq) assay which is fully customisable with commercially available sequencing reagents. Customisability would allow for straightforward adaptation of the target panel, for example to incorporate genes of interest for academic research such as candidate driver genes associated with therapy resistance. In order to develop the technology, the following objectives were set:

i. Extract, quantify and investigate the characteristics of cfDNA obtained from healthy donors (HD) and patients with metastatic colorectal cancer (mCRC) and to define the optimal input for cfDNA sequencing.

ii. Test plasma cfDNA from a pilot mCRC cohort to determine the optimal library preparation conditions from low input DNA quantities.

iii. Optimise error correction techniques for accurate detection of low frequency variants.

iv. Test the sensitivity and specificity of the optimised capture-based sequencing technique.

v. Assess its research and clinical utility in a series of mCRC patients (Chapter 4).

3.3 Properties of circulating free DNA in patients with metastatic colorectal cancer and healthy donors

3.3.1. Quantification of cfDNA

cfDNA was extracted from plasma of 25 HDs and 58 patients with mCRC enrolled in the FOrMAT trial at the Royal Marsden Hospital (ClinicalTrials.gov NCT02112357) [168]. I selected plasma samples from 3 groups of patients: i) treatment naïve (n=19); ii) patients whose tumours started to progress following first line palliative systemic therapy (n=17) and iii) patients whose tumours started to progress following two or more prior lines of palliative systemic therapy (n=22) (Table 3.1).

Case ID	Gender	Age at sample collection	Site of Primary	Histology	RAS/RAF mutation present	Number of lines of systemic therapy received	Timepoint at which ctDNA was sampled
F01	M	52	Rectum	Moderately differentiated adenocarcinoma	KRAS: c.35G>A p.Gly12Asp	2	Post second line
F02	F	58	Caecum	Moderately differentiated adenocarcinoma	Nil	2	Treatment naïve
F03	F	68	Transverse colon	Moderately differentiated adenocarcinoma	BRAF: c.1799T>A p.Val600Glu	2	Post first line
F04	M	65	Caecum	Moderately differentiated adenocarcinoma	KRAS: c.35G>T p.Gly12Val	4	Post third line
F05	M	79	Caecum	Moderately differentiated adenocarcinoma	KRAS: c.30_35dupAGCTGG p.Gly12_Gly13insAlaGly	2	Post first line
F06	M	74	Caecum	Moderately differentiated adenocarcinoma	KRAS: c.38G>A p.Gly13Asp	5	Post fourth line
F07	M	76	Sigmoid	Moderately differentiated adenocarcinoma	Nil	5	Post second line
F08	M	45	Transverse colon	Moderately to poorly differentiated adenocarcinoma	BRAF: c.1799T>A p.Val600Glu	4	Post first line
F09	M	70	Rectal	Moderately differentiated adenocarcinoma	Nil	3	Post second line
F10	M	51	Rectal	Moderately differentiated adenocarcinoma	Nil	4	Post third line
F11	M	66	Rectal	Moderately differentiated	Nil	4	Treatment naïve
F12	F	62	Rectal	Moderately differentiated adenocarcinoma	NRAS: c.182A>T p.Gln61Leu	2	Treatment naïve
F13	F	70	Rectal	Moderately to poorly differentiated adenocarcinoma	KRAS: c.34G>A p.Gly12Ser	2	Post first line
F14	M	61	Sigmoid and caecum (synchronous primaries)	Moderately differentiated adenocarcinoma	KRAS: c.35G>A p.Gly12Asp	2	Post second line
F15	M	53	Rectal	Moderately differentiated adenocarcinoma	KRAS: c.35G>A p.Gly12Asp	6	Post fifth line
F16	F	37	Caecum	Moderately differentiated adenocarcinoma	NRAS: c.181C>A p.Gln61Lys	3	Post first line
F17	M	59	Right sided (site not stated)	Moderately differentiated adenocarcinoma	BRAF: c.1799T>A p.Val600Glu	1	Treatment naïve
F18	M	63	Rectal	Moderately differentiated adenocarcinoma	Nil	3	Treatment naïve
F19	F	38	Sigmoid	Poorly differentiated carcinoma	Nil	1	Treatment naïve
F20	M	67	Hepatic flexure	Poorly differentiated carcinoma	Nil	3	Treatment naïve
F21	F	66	Caecal	Moderately differentiated adenocarcinoma	KRAS: c.35G>A p.Gly12Asp	3	Post first line
F22	F	71	Caecum	Moderately differentiated adenocarcinoma	Insufficient tumour for analysis	1	Treatment naïve
F23	M	74	Caecum	Moderately differentiated adenocarcinoma	Nil	3	Post first line
F24	M	55	Sigmoid	Moderately differentiated mucinous adenocarcinoma	KRAS: c.183A>C p.Gln61His	4	Treatment naïve
F25	F	51	Sigmoid	Moderately differentiated adenocarcinoma	KRAS: c.35G>A p.Gly12Asp	2	Post first line
F26	M	68	Caecum	Moderately differentiated adenocarcinoma	KRAS: c.35G>T p.Gly12Val	3	Post second line
F27	M	81	Caecum	Moderately differentiated adenocarcinoma	KRAS: c.35G>A p.Gly12Asp	2	Post first line
F28	F	38	Caecum	Moderately differentiated adenocarcinoma	Insufficient tumour for analysis	1	Treatment naïve
F29	F	50	Rectal	Poorly differentiated desmoplastic carcinoma	Nil	2	Post first line
F30	M	74	Sigmoid	Moderately differentiatied adenocarcinoma	Nil	4	Treatment naïve
F31	M	72	Splenic Flexure	Moderately to poorly differentiated mucinous adenocarcinoma	KRAS: c.35G>A p.Gly12Asp	4	Treatment naïve
F32	M	67	Rectal	Poorly differentiated carcinoma	KRAS: c.Unknown p.Codon12/13.	2	Post first line
F33	M	66	Sigmoid	Moderately differentiated adenocarcinoma	Nil	5	Post first line
F34	M	75	Hepatic flexure	Moderately differentiated adenocarcinoma	Nil	4	Post third line
F35	F	64	Caecal	Moderately differentiated adenocarcinoma	NRAS: c.182A>G p.Gln61Arg	3	Post first line
F36	F	35	Hepatic flexure	Moderately differentiated mucinous adenocarcinoma	KRAS: c.35G>A p.Gly12Asp	3	Post second line
F37	M	63	Sigmoid	Moderately differentiated adenocarcinoma	Nil	4	Post second line
F38	F	84	Sigmoid	Moderately differentiated adenocarcinoma	Nil	4	Post second line
F39	M	79	Transverse colon	Differentiation not stated	Nil	2	Treatment naïve
F40	M	62	Hepatic flexure	Moderately to poorly differentiated adenocarcinoma	KRAS: c.182_183delAAinsTG p.Gln61Leu	3	Treatment naïve
F41	M	83	Caecal	Moderately differentiated adenocarcinoma	BRAF: c.1799T>A p.Val600Glu	2	Treatment naïve
F42	M	71	Caecal	Moderately differentiated mucinous adenocarcinoma	KRAS: c.35G>A p.Gly12Asp	7	Post sixth line
F43	M	52	Splenic flexure	Poorly differentiated carcinoma	BRAF: c.1799T>A p.Val600Glu	3	Post third line
F44	M	69	Caecal	Moderately differentiated adenocarcinoma	KRAS: c.Unknown p.Codon12/13.	3	Post second line
F45	M	60	Caecal	Poorly differentiated carcinoma	KRAS: c.Unknown p.codon 61	3	Post second line
F46	M	74	Rectal	Poorly differentiated carcinoma	Nil	2	Treatment naïve
F47	M	33	Splenic Flexure	Poorly differentiated carcinoma	Nil	4	Post first line
F48	M	66	Rectal	Moderately differentiated adenocarcinoma	Nil	3	Post first line
F49	M	49	Sigmoid	Moderately differentiated adenocarcinoma	KRAS: c.35G>A p.Gly12Asp	2	Treatment naïve
F50	F	37	Transverse colon	Poorly differentiated signet ring cell type	BRAF: c.1799T>A p.V600E	2	Treatment naïve
F51	M	66	Rectal	Moderately to poorly differentiated adenocarcinoma	Nil	5	Post third line
F52	F	43	Rectal	Poorly differentiated carcinoma	Nil	3	Post second line
F53	F	69.00	Ascending colon	Moderately differentiated adenocarcinoma	BRAF: c.1799T>A p.Val600Glu	2	Post first line
F54	M	79	Sigmoid	Moderately differentiated adenocarcinoma	KRAS: c.38G>A p.Gly13Asp	5	Post fourth line
F55	M	72	Transverse colon	Moderately differentiated adenocarcinoma	KRAS: c.35G>T p.Gly12Val	2	Treatment naïve
F56	M	72	Caecum	Moderately differentiated adenocarcinoma	NRAS: c.181C>A p.Gln61Lys	2	Post first line
F57	F	36	Rectal	Moderately differentiated adenocarcinoma	Nil	3	Post second line
F58	F	79	Sigmoid	Moderately differentiated adenocarcinoma	KRAS: c.unknown; p.codon 12/13	3	Post second line

Table 3.1: Table of patient characteristics of 58 FOrMAT trial patients from whom cfDNA was extracted.

Blood was collected in EDTA tubes and processed as soon as possible following blood draw to prevent lysis of white blood cells which can dilute the cfDNA with normal leucocyte genomic DNA. If plasma separation was not possible within 6 hours of collection, cell stabilisation (Streck tubes) were used, which can store peripheral blood for up to 48 hours without compromising cfDNA quality [169]. The blood was processed through two sequential centrifugation steps, one at low speed (1600g at room temperature, for ten minutes) and thereafter at high speed (16000g at 4^{o}C, for ten minutes) to isolate the plasma for cryostorage and subsequent DNA extraction. These steps have been incorporated into standard operating procedures (SOPs) at the Royal Marsden Hospital (see methods section 2.2.2).

Following cfDNA extraction, I next assessed the concentration of cfDNA that can be extracted from a fixed volume of plasma from mCRC patients and from HDs (for use as negative control cfDNA), where cfDNA levels are known to be lower than in cancer patients [170]. cfDNA was extracted from 4 millilitres (ml) of patient plasma and 8 ml of HD plasma and quantified on the Agilent Bioanalyzer. cfDNA is naturally fragmented with a modal fragment size of 160 to 180 base pairs (bp), corresponding to the length of DNA that wraps around a mono-nucleosome, as *in vivo* cleavage during apoptosis occurs between nucelosomes [171, 172]. DNA fragments with multiples of this size (nucleosome dimer, trimer etc.) can also be found but these are less abundant [173]. cfDNA within a size range of 100 to 700 bp, encompassing the three predominant fragment peaks, was quantified using a Bioanalyzer High Sensitivity chip (Figure 3.3A). DNA concentration across all 3

peaks was calculated based on the area under the curve method. Samples with DNA concentrations above 15 nanograms per millilitre (ng/ml) plasma exceeded the accurate analysis range of the Bioanalyzer High Sensitivity chip (Figure 3.3C) and required repeat quantification using the standard sensitivity DNA 7500 chip (Figure 3.3D).

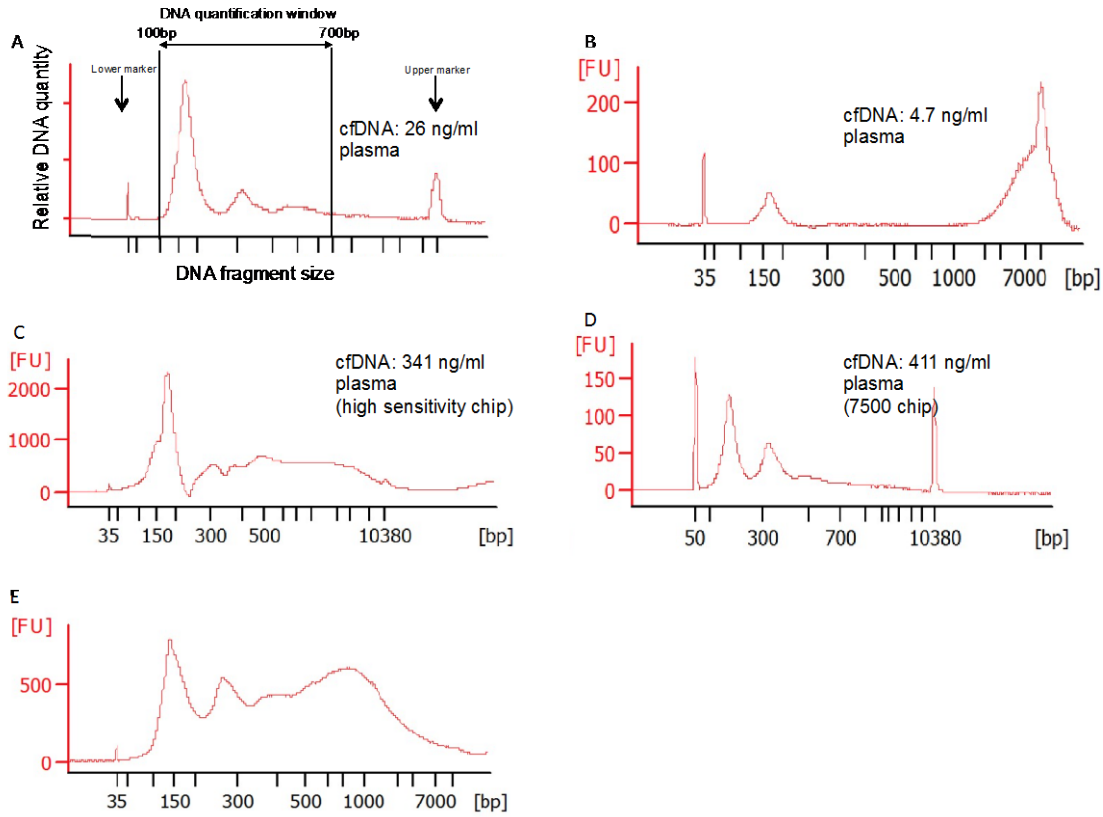


Figure 3.3: Bioanalyzer profiles of fragmented cfDNA extracted by Qiagen circulating nucleic acid extraction protocol and quantified using the Agilent Bioanalyzer.
(A) a Bioanalyzer profile of good quality cfDNA with characteristic three-peak profile.

(B) a very low yield cfDNA sample from a patient with mCRC. This sample also displays high molecular weight (HMW) DNA contamination skewing the upper marker peak

(C) a very high yield cfDNA trace on a high sensitivity chip with little/no visibility of the marker peaks, requiring re-quantification on a DNA 7500 chip, shown in panel (D)

(E) a profile of poor quality cfDNA extracted from 4 ml plasma from a patient with mCRC, likely contaminated with HMW DNA, causing loss of the upper marker peak.

cfDNA sample quality was assessed by reviewing the baseline of the trace, the height, width and clarity of each peak, including the marker peaks. For accurate analysis, reference markers need to be identified correctly as they are used for fragment sizing and specifically the upper marker is used for quantification. cfDNA was quantified across a window from 100 bp to 700 bp in all samples. A trace with an elevated baseline (Figure 3.3E) or a high level of fragments larger than 1 kilobase (kb) indicates contamination by HMW DNA, likely arising from lysis of nucleated blood cells during phlebotomy or blood/plasma processing which dilutes the fragmented cfDNA. High fractions of HMW DNA can deviate the baseline and distort the upper marker preventing accurate quantification of the cfDNA and hence affecting the precision of library preparation through overestimation of the input cfDNA amount. Where Bioanalyzer sample quantification was deemed unreliable, the sample was not carried forward for library preparation until it had been requantified on the Bioanalyzer and by an alternative method i.e. Qubit. If the requantification demonstrated very poor profiles or vastly inconsistent cfDNA concentrations, samples were excluded. Overall, two out of 58 samples were excluded to avoid imprecision for library preparation.

How pre-analytical variables affect the quality of cfDNA, including types of collection tube, time to preparation, time to shipping, transportation and storage conditions has previously been assessed [169]. To assess if these factors consistently affected cfDNA quality and to understand the reasons for potential poor sample quality, I assessed and categorised the degree of haemolysis visually based on the colour of plasma: i) straw-coloured plasma- indicating no haemolysis, ii) orange- indicating moderate haemolysis and iii) red- severe haemolysis. The degree of haemolysis was recorded for correlation with amount

of HMW DNA present in extracted cfDNA Bioanalyzer traces (Figure 3.4). Of the 58 plasma samples, 36 out of 58 (62%) had no evidence of haemolysis on inspection of the plasma colour, yet following cfDNA extraction nine of these 36 cases demonstrated evidence of HMW DNA on the Bioanalyzer trace. 19 out of 58 (33%) samples showed moderate haemolysis and two of these 19 cfDNA Bioanalyzer traces demonstrated HMW DNA contamination. Finally, the remaining three cases had red coloured plasma indicating severe haemolysis yet none of these samples showed HMW DNA contamination.

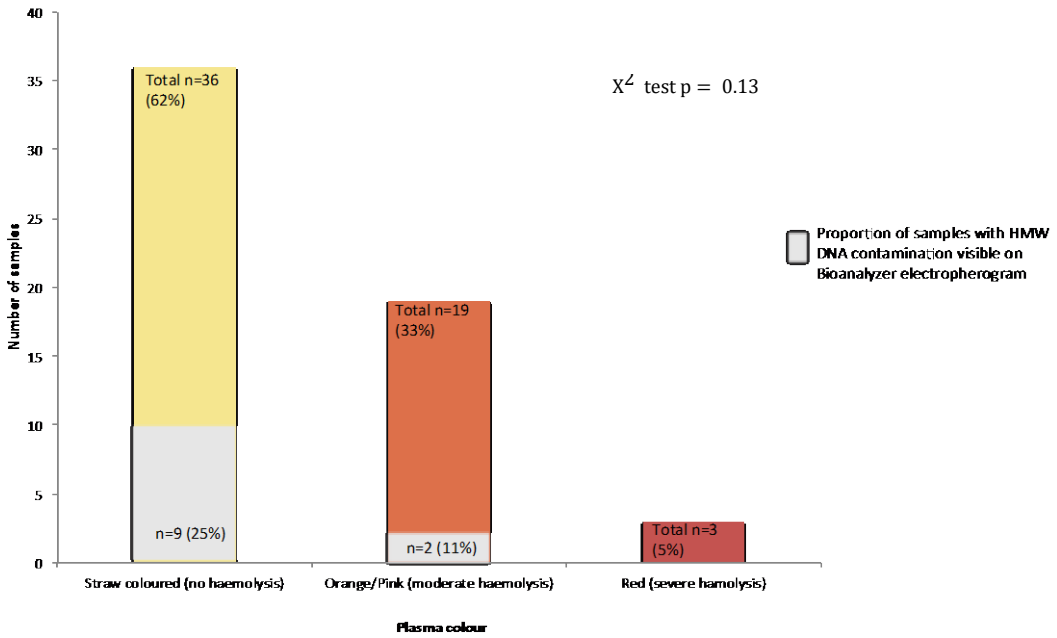


Figure 3.4: Plasma colour and degree of HMW DNA contamination within each of the three groups.

Overall, 11 of 58 (19%) of the extracted cfDNA Bioanalyzer traces displayed high amounts of large fragment size HMW DNA contamination and nine of these cfDNA samples were extracted from plasma with no evidence of haemolysis. Chi squared test for trend demonstrated a nonsignificant p value of 0.13, thus there is no association of plasma colour with the degree of HMW DNA contamination, suggesting that lysis of red blood cells is not proportional to that of white blood

cells. In view of these observations, samples were not selected based on plasma colour and HMW DNA contamination was only assessed on the Bioanalyzer trace.

3.3.2. cfDNA yields in metastatic colorectal cancer

A median of 20.8 ng cfDNA per ml of plasma were extracted from the 58 mCRC patients (25^{th} percentile: 7.1 ng/ml; 75^{th} percentile: 49.1 ng/ml) (Figure 3.5). HDs had significantly lower median cfDNA yield (median 4.6 ng/ml) than patients with mCRC (median 20.8 ng/ml; $p<0.001$, Student's t-test). There was no statistically significant difference in the cfDNA yield for untreated mCRC patients (median: 19.18 ng/ml) and for patients who had received one prior line of therapy (median: 23.98 ng/ml; $p=0.36$, Student's t-test). The yields for those who received 1 vs. ≥ 2 lines of prior therapy was also similar (median 23.98 ng/ml vs. 23.02 ng/ml; $p=0.35$ Student's t-test). Thus mCRC patients have a higher cfDNA concentration than HDs, consistent with several studies [96] [148, 156]. In addition, there is no difference in cfDNA yields in mCRC patients based on the number of prior therapies.

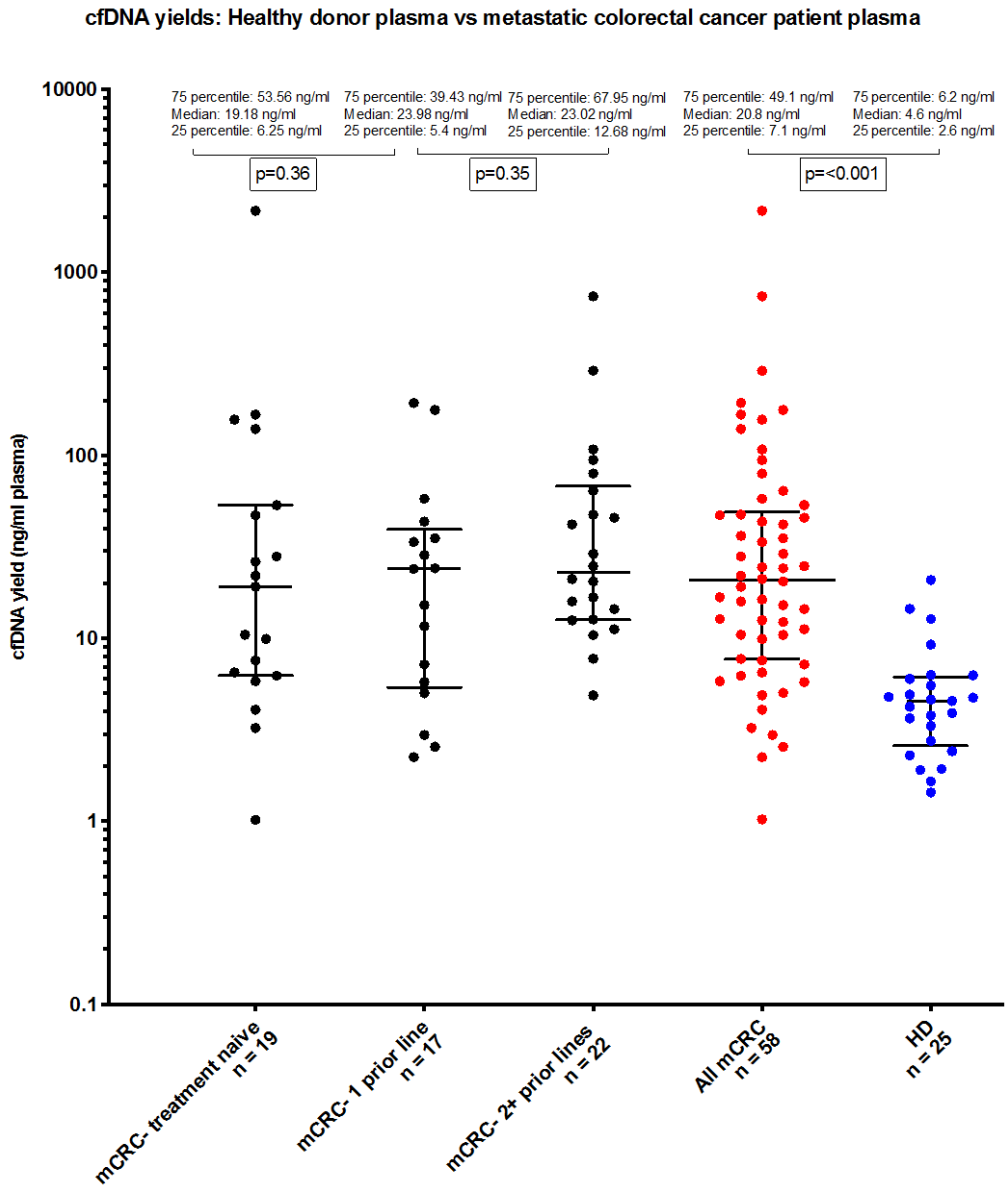


Figure 3.5 cfDNA yields (ng per ml plasma) obtained from 25 healthy donors and 58 patients with mCRC.
Horizontal lines represent median with interquartile ranges. P-values were calculated using Student's t-test demonstrating no difference in median cfDNA yields between consecutive groups.

3.3.3 Defining optimal input cfDNA quantity for sequencing

cfDNA yields obtained from the 58 patients were used to define the cfDNA quantity that can be extracted from >95% of mCRC patients collected through a standard blood draw of 20-30 ml, which should allow the reliable separation of 10 ml plasma. I used the cfDNA yield data to model how the plasma volume

influences the fraction of patients from which at least 15, 20, 25 or 30 ng of total cfDNA can be obtained (Figure 3.6). This model predicts that 25 ng cfDNA can be obtained from over 95% of mCRC patients similar to our cohort, if 10 ml plasma is available.

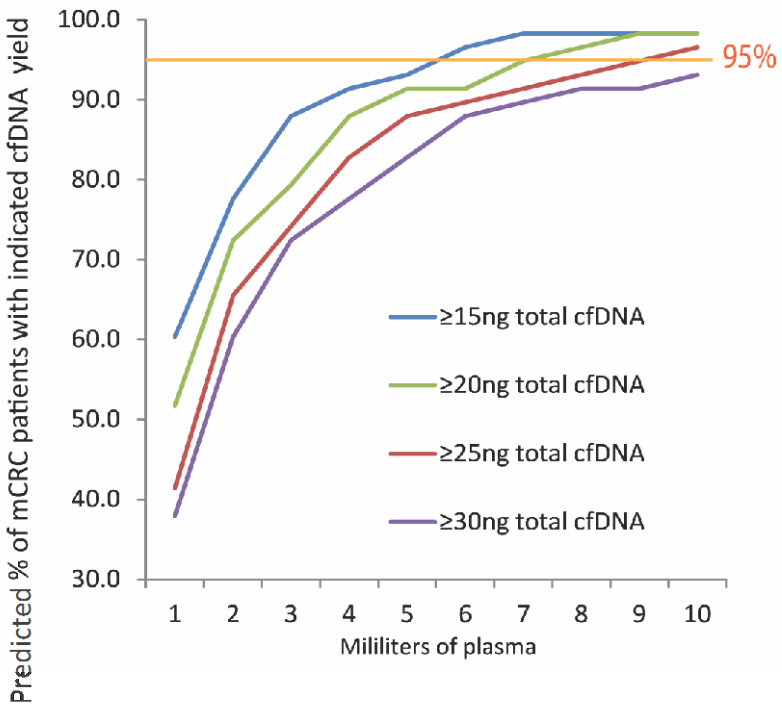


Figure 3.6 cfDNA obtained from 58 patients with mCRC to calculate the fraction of patients that achieved the indicated total cfDNA yield based on the plasma volume available from each patient.

With a haploid human genome mass of ~3.3 picograms (pg), 25 ng of cfDNA should contain >7500 genome equivalents (GE) which is theoretically sufficient for the detection of mutations with frequencies around 0.1% VAF, even if only 20% of the cfDNA fragments (i.e. 20% of 7500 = 1500 fragments) are incorporated into sequencing libraries [94]. Based on these results, we chose 25 ng as the standard input quantity for cfDNA library construction.

3.4 Solution hybrid capture based cfDNA sequencing optimisation

3.4.1. Colorectal cancer specific target capture panel

The first step of optimising library preparation for mCRC patients was to select the genes that should be included into the targeted sequencing assay. The key considerations in so doing were to incorporate all genes of interest, yet maintain a small target size in order to achieve high sequencing depths at reasonable cost.

I selected CRC genes which were observed to be recurrently mutated in non-hypermutated CRCs in TCGA exome sequencing dataset of 276 patients and the Broad Institute CRC analysis of 489 patients [39, 79]. I included the eight most commonly mutated genes in both cohorts (*APC, TP53, KRAS, PIK3CA, FBXW7, SMAD4, TCF7L2* and *NRAS*) plus those that were mutated with a frequency >5% and known to be driver genes in either TCGA or Broad cohorts (*AMER1, BRAF, SMAD2* and *CTNNB1*). Additionally, I selected genes of interest including *B2M*- a TSG involved in immune evasion that has been found to be mutated in some CRCs [174] and *TYMS*- implicated in driving resistance to fluoropyrimidine based therapy [175]. I also included genes involved in DNA repair that were observed to be mutated in chemotherapy resistant CRCs in studies in my lab and were therefore of ongoing interest including: *REV3L*, *ERCC3, ERCC4, ERCC6*, *FANCM*, *RAD54L*, *RAF1*, *MAPK1* and *MAPK3*. The remaining genes were included as they can be mutated and/or altered at the copy number level in CRC and are also of ongoing interest in my lab: *MET, PTEN, EGFR, ERBB2, ERBB3, MAP2K1, MAP2K2* and *CDH1* [176]. *TTN* is an extremely large gene (103,050 bp) that is mutated in 29% of cases in the TCGA cohort but mutations within this gene have no known functional relevance in these cancers [10]. Therefore *TTN* was excluded in order to keep the total target region size relatively small and focused. In total, 32 genes were incorporated into the panel (Table 3.2). Finally,

40 genomic positions on chromosome 18 of single-nucleotide polymorphisms (SNPs), which are frequently heterozygous in Caucasian populations, were included in the target panel (Table 3.3) to assist in tumour content estimation and as a trial to assess whether these could be successfully used for future copy number analysis.

Gene Name	Type of alteration	Frequency (%) TCGA	Frequency (%) Broad study	Stringency	Predicted coverage of the gene based on Agilent design tool
APC	Mutation	81	76	Moderate	99.5
TP53	Mutation	60	52	Moderate	94.2
KRAS	Mutation	43	43	Most	100
PIK3CA	Mutation	18	19	Most	100
FBXW7	Mutation	11	16	Most	100
SMAD4	Mutation	12	12	Most	98
AMER1	Mutation	11	Not stated	Most	100
BRAF	Mutation	10	9	Moderate	100
TCF7L2	Mutation	9	9	Most	100
NRAS	Mutation	9	9	Most	100
SMAD2	Mutation	6	6	Most	100
CTNNB1	Mutation	5	4	Most	100
PTEN	Mutation	4	3	Most	100
ERBB2	Mutation	4	4	Most	95.9
EGFR	Mutation	4	4	Most	100
B2M	Mutation	2.4	2	Most	100
REV3L	Mutation	6	Not stated	Moderate	100
ATM	Mutation	32	32	Most	100
FANCM	Mutation	5.7	Not stated	Most	100
ERCC6	Mutation	5	7	Moderate	100
RAF1	Mutation	3	3	Most	100
ERCC3	Mutation	3	Not stated	Most	100
ERCC4	Mutation	2	Not stated	Most	98.9
MAPK3	Mutation	1.4	Not stated	Most	100
MAPK1	Mutation	1	Not stated	Moderate	100

RAD54L	Mutation	1	Not stated	Most	100
ERBB3	CNA	7	6	Most	100
CDH1	CNA	2.4	4	Most	100
MET	CNA	2	10	Most	98.7
TYMS	CNA	1	Not stated	Most	100
MAP2K1	CNA	1	Not stated	Most	100
MAP2K2	CNA	1	Not stated	Most	100

Table 3.2: 32 genes selected for CRC target panel.
Genes are listed in order frequency of alteration based on TCGA and Broad Institute cohorts [10, 19, 21, 177-179]. Recurrently mutated genes in TCGA/Broad Institute and potential CRC drivers in are indicated in orange, genes involved in DNA repair or recurrently mutated in chemotherapy resistant CRCs (identified in ongoing studies in the lab) are indicated in blue and copy number altered genes indicated in purple.

Target ID	Interval	Size	Coverage
rs11082523	chr18:43923841-	21	100
rs11152375	chr18:60906674-	21	100
rs1132845	chr18:72912459-	21	100
rs11659760	chr18:55350449-	21	100
rs11876282	chr18:50905853-	21	100
rs1317694	chr18:46401460-	21	100
rs1431583	chr18:59660533-	21	100
rs1450428	chr18:44086672-	21	100
rs1511155	chr18:47717232-	21	100
rs17227021	chr18:69474209-	21	100
rs17383821	chr18:50269489-	21	100
rs17707448	chr18:44346214-	21	100
rs17801621	chr18:47509829-	21	100
rs1787578	chr18:47419742-	21	100
rs1866987	chr18:51170927-	21	100
rs2075405	chr18:56037839-	21	100
rs216558	chr18:54051331-	21	100
rs2849372	chr18:61013895-	21	100
rs2850544	chr18:48422889-	21	100
rs2886018	chr18:67141441-	21	100
rs3744865	chr18:56034177-	21	100
rs4493166	chr18:61849637-	21	100
rs4510128	chr18:61631322-	21	100
rs4891217	chr18:72702101-	21	100
rs4986228	chr18:44595249-	21	100
rs521663	chr18:57822230-	21	100
rs568069	chr18:66216246-	21	100
rs640401	chr18:56966822-	21	100
rs6566386	chr18:66485539-	21	100
rs6567365	chr18:61164592-	21	100
rs7233258	chr18:57655238-	21	100
rs7240680	chr18:70674918-	21	100
rs8091955	chr18:46358897-	21	100

rs8094161	chr18:56743355-	21	100
rs8097619	chr18:55720210-	21	100
rs8098217	chr18:63496813-	21	100
rs833503	chr18:46625058-	21	100
rs899103	chr18:53327360-	21	100
rs9953959	chr18:57600956-	21	100
rs9961742	chr18:72709872-	21	100

Table 3.3**:** 40 SNP probes included within target panel.
Each probe was 21 base pairs in size with 100% coverage. Total size of SNP targets was 8.64 kilo base pairs (kbp).

Following identification of the targets for the library design, various probe parameters were adjusted to ensure uniformity of coverage and target enrichment efficiency. Across all target regions 5x tiling density was selected (i.e. every base in the target region, including SNP targets, is covered by 5 overlapping probes) so that even regions with a high GC content, which can be difficult to capture and sequence, would be adequately enriched for sequencing. In order to optimise the capture of DNA fragments with variable GC-content, which are under-represented in sequencing I also maximised the boosting (i.e. the concentration) of GC-rich probes. Maximising boosting can increase the overall target panel size, however this was considered an acceptable compromise in view of the predicted improvement of capture and sequencing efficiency. Most genes within the library were designed with maximum stringency in order to reduce off-target binding. However the stringency had to be lowered for 6 genes (*APC, BRAF, ERCC6, MAPK1, REV3L, TP53*) in order to achieve coverage of most of the coding exons in these genes. The total target size of the final panel was 163.3 kbp.

3.4.2. Optimisation of MBC error corrected cfDNA sequencing technology

At the time of starting this project, there were no commercially available technologies for MBC error correction in combination with solution hybrid capture sequencing. My lab had pre-marketing access to Agilent's new SureSelect$^{XT\text{-}HS}$ library preparation kit and my aim was to optimise this cfDNA sequencing technology to detect mutations present in ~0.1% of DNA molecules (i.e. the ability to detect 1 mutant DNA molecule in ~1000 unique DNA molecules). To do this, each of the following critical steps were individually attempted to be improved:

1. conversion of cfDNA into the sequencing library: the DNA conversion rate describes what proportion of input DNA molecules are converted into sequenceable fragments. I aimed to generate a library that incorporates 1000-1500 unique double stranded DNA fragments covering each nucleotide position in the target region to allow sufficient sensitivity for accurate detection of low abundance variants with a frequency close to 0.1%.
2. reduction of off-target reads: off-target reads are those that map outside of the targeted region. A degree of off-target sequencing is necessary for copy number variation detection, but too much is detrimental as it reduces effective sequencing capacity per unit cost. Low input DNA and small target panels have historically resulted in high off-target rates [180, 181].
3. error correction: reduction of false positive errors generated through PCR and sequencing errors is required to detect true variants down to 0.1% VAF.

I performed experiments modifying individual conditions of the library preparation and sequencing pipeline in order to assess their impact on the on-target read number, sequencing depth and MBC error correction of this assay. The individual parameters assessed were i) bait preparation ii) hybridisation conditions, iii)

library capture conditions and iv) number of reads sequenced. This experiment utilised 25 ng from each of 6 different cfDNA samples (HD36, HD37, HD38 F30, F32, F33) a mix of HD and patient samples were used so as not to consume FOrMAT trial samples which had limited amount of cfDNA available. Samples were prepared in duplicate or triplicate, varying one library preparation condition at a time, in total 15 libraries were prepared. The final prepared libraries were checked on a Bioanalyzer High Sensitivity DNA chip and quantified by qPCR using the Kapa Library Quantification kit before pooling. Pools were sequenced with paired end 75 reads on an Illumina Hiseq2500 in Rapid output mode. After demultiplexing, Agilent SureCall software was used to align the reads to the hg19 human genome and the data was deduplicated using the MBCs (i.e. bioinformatic removal of reads with identical length and sequence identity) for subsequent MBC error-corrected variant calling.

3.4.2.1 Improving the on-target read number

3.4.2.1.1 Effect of bait preparation

Baits are biotinylated oligonucleotides which are reverse complementary to the target DNA regions and are hybridised to the input DNA in order to then retrieve DNA fragments of interest for sequencing (Figure 3.1). I tested two different Agilent SureSelect$^{XT\text{-}HS}$ bait formulations on 3 cfDNA samples that were each prepared with both baits A and baits B: bait formulation A was the standard formulation and bait formulation B (available to my host lab through developer access program) was a modified propriety formulation developed by Agilent for enrichment of small target regions. Sample preparation continued as described in the manufacturer's protocol (Methods section 2.2.6.1).

The median total reads per sample was similar for baits A (47,660,748 reads, with an average read length of 73 bp) and baits B (median of 46,814,226 reads, average read length 73 bp). I anticipated that baits B would improve on-target rates due to improved target enrichment, however comparison of the 3 samples that were each prepared with both baits A and baits B, showed a similar fraction of on-target reads with no statistically significant differences between baits (Figure 3.7; baits A median on-target= 45.46%; baits B median on-target= 46.76% ANOVA p=0.49, further analysis revealed similar post deduplication sequencing depths of 1176X median for baits A and 998X median for baits B ANOVA p=0.44). Based on these results, I used bait formulation A for all future experiments as these are readily available from the manufacturer and performance is equivalent to baits B.

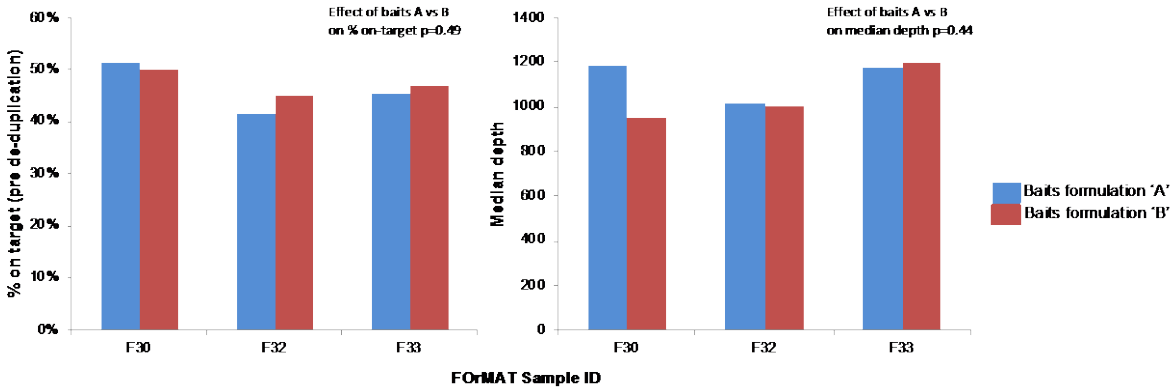


Figure 3.7: Histogram showing the effect of baits formulation on percentage on-target reads and median depth
Three patient samples (F30, F32, F33) are demonstrated above. Suffix –A and –B denote the formulation of Agilent baits (baits A blue bars- standard SureSelect XT formulation; baits B red bars- modified formulation for targeting small <100kb regions). Two-way ANOVA was used to compare the variance between baits A and baits B on % on-target (p=0.49) and median depth (p=0.44), there was no statistically significant difference between either bait formulation.

3.4.2.1.2 Effect of hybridisation conditions

I next conducted a comparison between the new 'fast-hybridisation' protocol, that cycles the hybridisation temperature between 65^{o}C and 37^{o}C for 60 cycles over 1.5 hours, to the standard-hybridisation protocol of 65^{o}C constant temperature over 16 hours in two mCRC patient samples (F32 and F33). The 4 libraries were simultaneously prepared and hybridised.

The total reads were comparable across samples (F32 standard-hybridisation 49,543,772, vs. fast-hybridisation 44,071,036; F33 standard-hybridisation 46,814,226 vs. fast-hybridisation 48,248,050, with an average read length of 73 bp for all samples) and the on-target rate was similar for sample F32 (45% standard-hybridisation vs. 44% fast-hybridisation) and slightly improved in sample F33 (47% vs. 51%), Figure 3.8. Overall, though the total reads and on-target rates were similar, the decision was taken to use the fast-hybridisation protocol because of the time advantage allowing for a more efficient workflow.

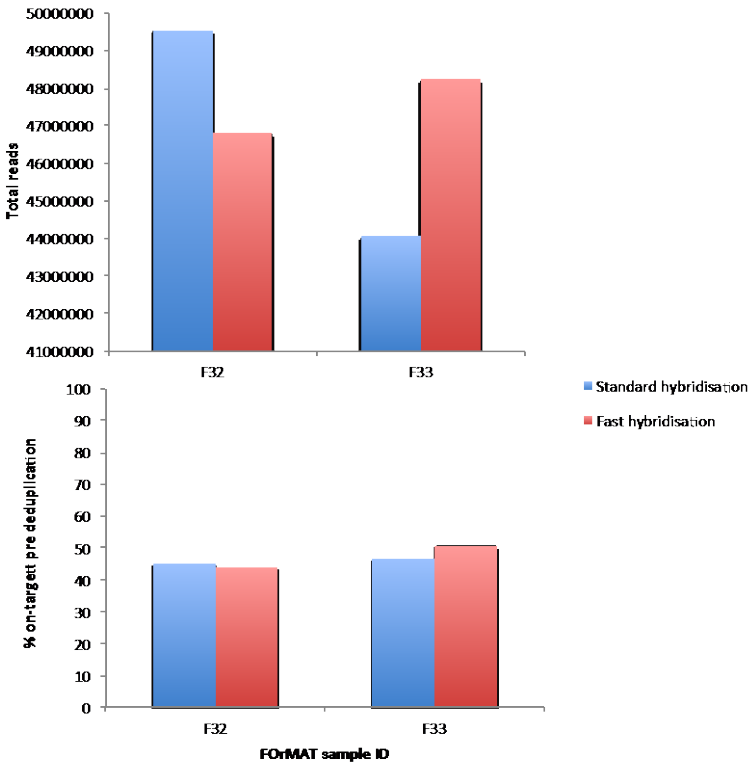


Figure 3.8: Histogram showing a comparison between standard-hybridisation and fast-hybridisation protocol.
Resulting total reads (top) and on-target percentage after deduplication (bottom) from two patient samples (F32, F33). Each library was prepared in duplicate, the only difference being hybridisation conditions.

3.4.2.1.3 Effect of capture wash conditions

I attempted to improve the on-target fraction by modifying the capture wash conditions, through washing the DNA bound to capture beads after hybridisation more stringently with a temperature of 70°C instead of 65°C. I hypothesised that the higher temperature would relax the target/bait-bond (DNA:RNA bond) in the hybridised molecules and a higher number of mismatched fragments would be removed. This should lead to a reduction in the non-specific carry over of DNA fragments into the library and thus improve the on-target rate. To test this, two library preparations were started in parallel from each of four cfDNA samples and

the post-capture wash was performed at 65°C in one set of samples and at 70°C in the other. After sequencing with similar read numbers (65°C: median of 92,820,887 per sample; 70°C: median of 102,582,694 per sample, average read length of 73 bp; p=0.46 paired t-test) the on-target fraction increased from 30-35% to 71-74% (p=0.0011) with the 70°C protocol (Figure 3.9). Furthermore the on-target depth tended to be greater with the higher temperature across all samples, with a median of 1345X with 65ºC to 1775X with 70ºC capture wash (p=0.08) (Figure 3.9). The trend towards higher depth is thought to be through removal of the non-specific off-target DNA during capture, which in turn allows the on-target DNA to have more sequencing capacity and hence improved read depth.

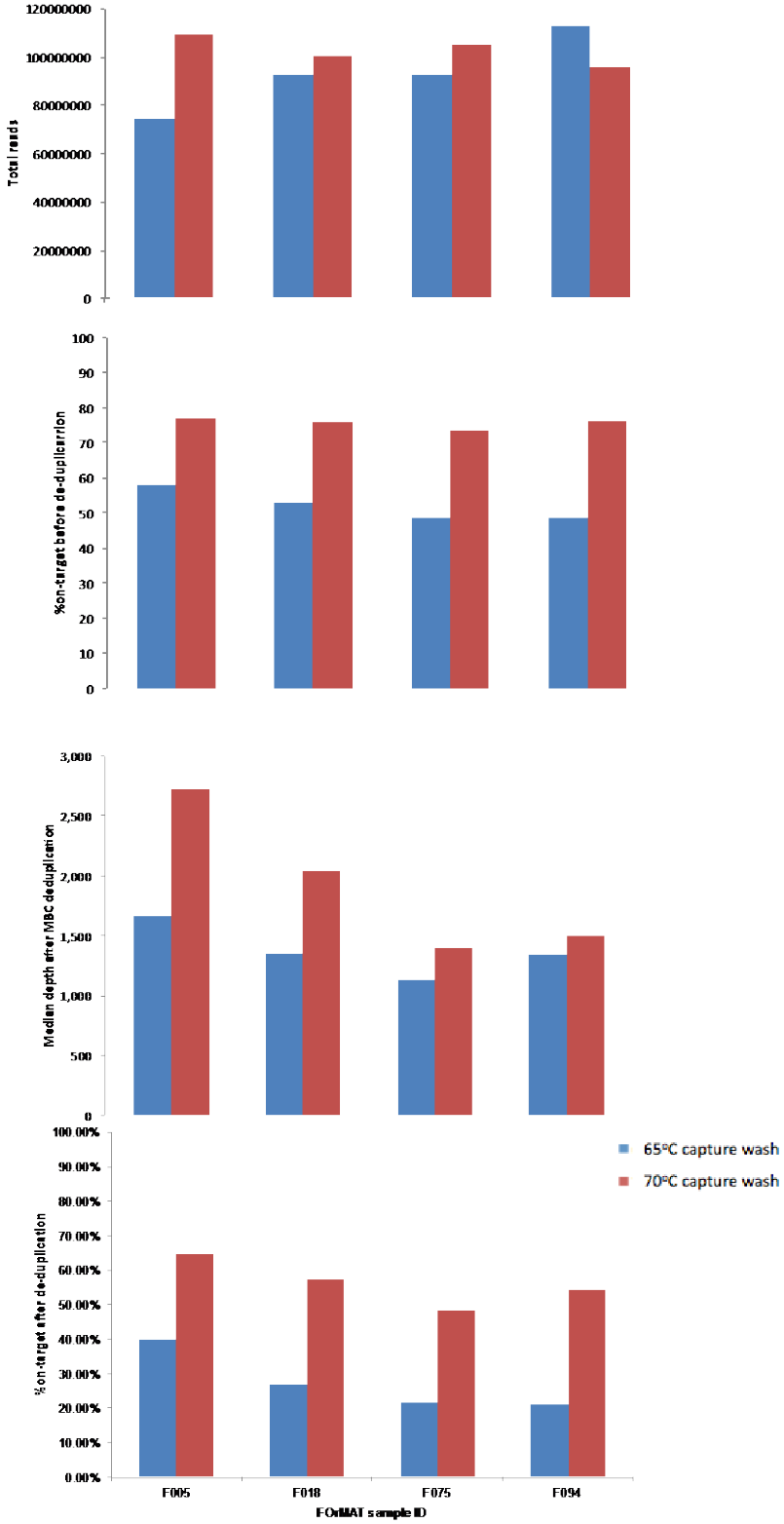


Figure 3.9: Histograms summarising the comparison of 65°C and 70°C capture washes. The effect of changing capture wash conditions on total reads, median depth and percentage on-target rate on four patient samples is demonstrated. Each library was prepared in duplicate, the only difference being capture wash conditions.

As the off-target percentages decreased with the higher temperature washes, the manufacturer recommended an increase from 9 post-capture PCR cycles with 65^{o}C protocol to 12 post-capture PCR cycles to compensate for the lower amount of non-specific DNA carryover and to generate sufficient yields for sequencing. However, this increased the final library molarity to a median of 5027 pmol/l, approximately 5-fold higher than the minimum required for sequencing. The higher PCR cycle could lead to increased PCR bias and hence uneven coverage but also a higher final library concentration can introduce sampling bias as only a small aliquot is taken when pooling samples for sequencing. Therefore I made the decision to reduce the post-capture PCR cycle number for subsequent library preparation to 10 cycles, which still reliably produced sufficient yield (~1000 pmol/l) for sequencing without an excessively high duplicate rate.

3.4.2.2 Improving the sequencing depth and molecular barcode analysis

3.4.2.2.1 Effect of the number of reads sequenced on median depth

I next investigated the effect of increasing the number of reads sequenced on sequencing depth. I performed 2 sequencing runs using the same set of prepared libraries, the first run multiplexed 15 libraries and the second run multiplexed 9 libraries in order to dedicate a greater sequencing effort to each sample and assess the impact of increasing the number of sequencing reads.

Run 1 achieved an average of 47,544,706 reads per sample and run 2 reached an average of 99,369,347 reads per sample, with an average read length of 73

bp in both runs. For subsequent analyses of sequencing depth, I excluded eight samples (HD 36, HD 38 and FOrMAT samples suffixes 1, 2 and 3 in Figure 3.10) as these were prepared with an earlier version of the SureSelect Kit which could confound the results of the depth analysis. After deduplication using MBCs, the median on-target read depth was 1113X in samples in run 1 (Figure 3.10, blue bars), compared with 1671X for run 2 (Figure 3.10, red bars) and calculated to be a median of 2853X for the combined dataset (Figure 3.10, green bars). Thus, increasing the sequencing effort by 2.5X the median depth increases by only 1.5 fold. This indicates that the libraries have been sequenced close to saturation somewhere between 6 and 9 samples per lane, as per the calculations derived from the combined datatset i.e. between 1671X and 2853X depth.

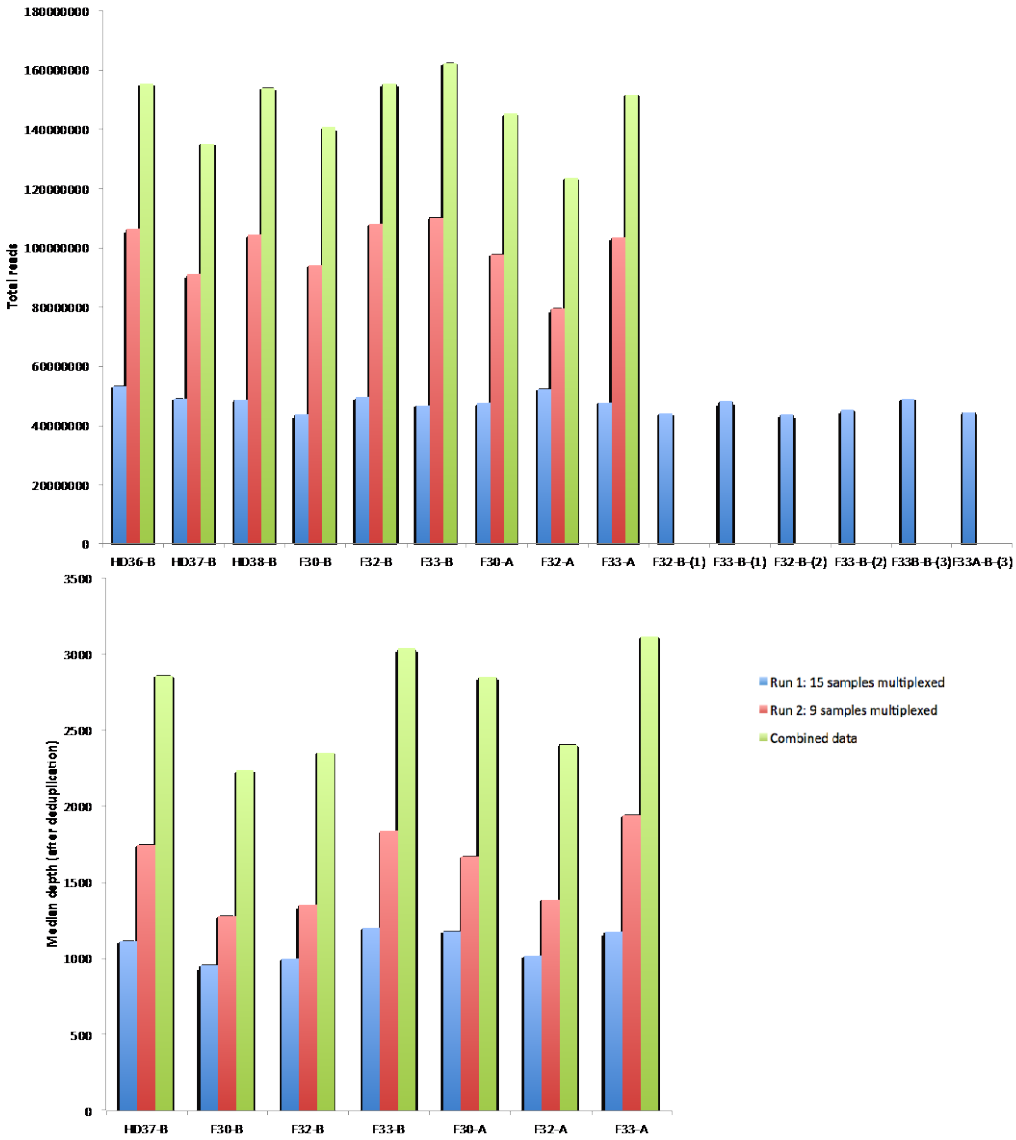


Figure 3.10: Histogram showing total reads and median depth after reducing the number of samples multiplexed in the sequencing run.
Blue bars represent a sequencing run with 15 samples multiplexed (achieving an average 47,544,706 reads per sample), red bars represent a sequencing run where nine samples were multiplexed (achieving an average 99,369,347 reads per sample), and green bars represent the combined data equivalent to six samples multiplexed on a run (achieving an average 147,038,465 reads per sample). For median depth analysis, samples prepared with a previous version of the SureSelect kit were removed.

Despite the additional sequencing effort, the median on-target rate was only 45% i.e. over half the reads were non-specific off-target, which is only slightly higher than the on-target rate of run 1 (median 37%). Thus optimal sequencing depth can likely be achieved somewhere between 6 and 9 samples per run (~110 million reads, per sample) in order to achieve adequate depth without over-

sequencing, which is associated with excessive costs and analysis efforts, for diminishing returns.

3.4.2.2.2 Effect of the number of reads sequenced on consensus family distribution

I next used molecular barcode sequencing data generated for runs 1 and 2 to assess the effect of sequencing depth on ‘consensus family’ distribution (Figure 3.2) and hence error correction. The variant calling software (SureCall) groups sequencing reads, that have identical MBCs and map to the same genomic alignment position, into consensus families. For a family of reads that do not all have identical sequences, the software computes a score for each candidate sequence based on the mapping and base qualities in that read, which is then multiplied by the number of reads with identical sequences. The sequence with the highest resulting score is used as the consensus sequence in the merged read. If a variant occurs in all reads then the consensus sequence will be variant for that base, indicating that a detected variant is likely a true mutation. However, if a variant is only detected in a small fraction of reads in the consensus family, it will be disregarded and the consensus read sequence will be wild-type at that position [116, 117].

A key factor in error correction is the median size of consensus families. Multiple family members are necessary to perform this error correction as small families cannot definitively distinguish a true variant from false positive errors. The optimal family size is a sufficient number of reads which maximises the efficiency of MBC error correction yet not too high such that the additional reads consume sequencing capacity without yielding utilisable data. A recent publication on error corrected ultralow frequency mutation detection [123] defined a family size of six

maximises the efficiency of MBC error correction, data yield continues to increase for family sizes up to 16 members, although the overall efficiency decreases as the peak family size approaches 16. Once the family size exceeds this upper limit nearly all additional reads are redundant and consume sequencing capacity without yielding further meaningful data [123].

I compared the consensus family size distribution (XI) between run 1 and run 2 to assess whether increased sequencing effort would result in more reads per family (Figure 3.11).

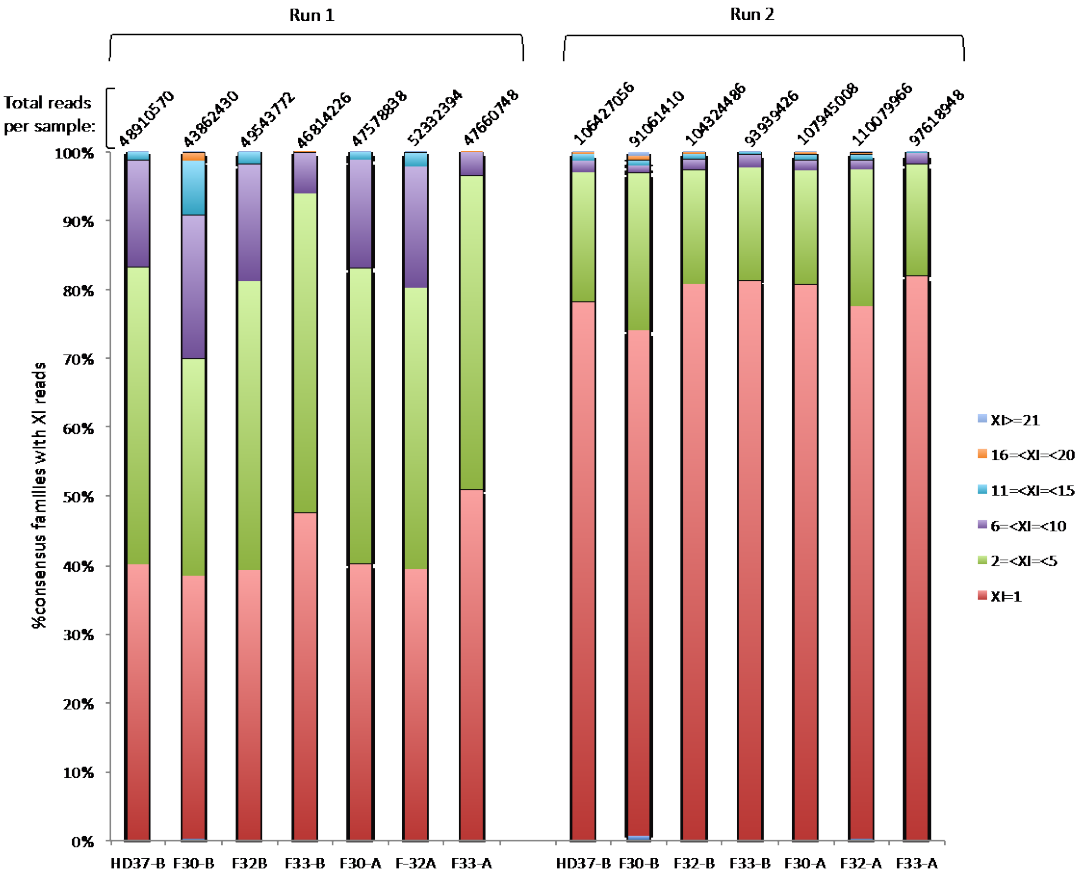


Figure 3.11: Histogram showing consensus family sizes
Left: 15 samples multiplexed in run 1 and right: nine samples multiplexed in run 2. XI= consensus family size.

The increased sequencing effort (run 2 samples) showed a considerably higher proportion of families containing lone reads i.e. a family comprising a single

consensus read supporting a particular base. These lone reads do not contribute towards error correction and thus would not improve the accuracy of variant calling. For *de novo* mutation detection, two reads are required to apply MBC error correction and exclude false positive reads arising from PCR or sequencing errors. Therefore families comprising of at least two consensus reads (XI ≥2) were used in onward analysis for variant calling. This approach however only reduces false positives. False negatives and sensitivity were tested in mixing experiments (section 3.4.6).

Based on these experiments, six to nine samples were combined for all subsequent sequencing runs to improve the total number of reads sequenced, the depth and the number of on-target reads.

3.4.3 Index hopping as a source of false-positive calls

Libraries are usually multiplexed together by tagging each DNA fragment with a sample-specific index primer prior to pooling libraries and loading onto the flow cell. Prepared libraries can then be multiplexed and sequenced on the same flow cell and subsequently demultiplexed computationally, resulting in significant time and cost saving and experimental scalability [182]. As discussed above, the ideal number of samples for multiplexing is approximately six to nine per sequencing run to achieve an adequate read count (~110 million reads average per sample) and depth (~1500X average per sample). Multiplex sequencing on the same flow cell relies on precise reassignment of each DNA sequencing read to the appropriate sample after sequencing [183].

During sequencing data analysis of two consecutive runs, I became aware of unexpected identical calls arising from different samples sequenced on the same

lane of a flow cell. This became noticeable in two sequencing runs where depth was increased stepwise by reducing the number of samples to six per run, compared with 8-9 samples in prior runs. I investigated the origin of these calls to understand how they arise, whether they interfere with variant calling and how to mitigate their effects on variant calling.

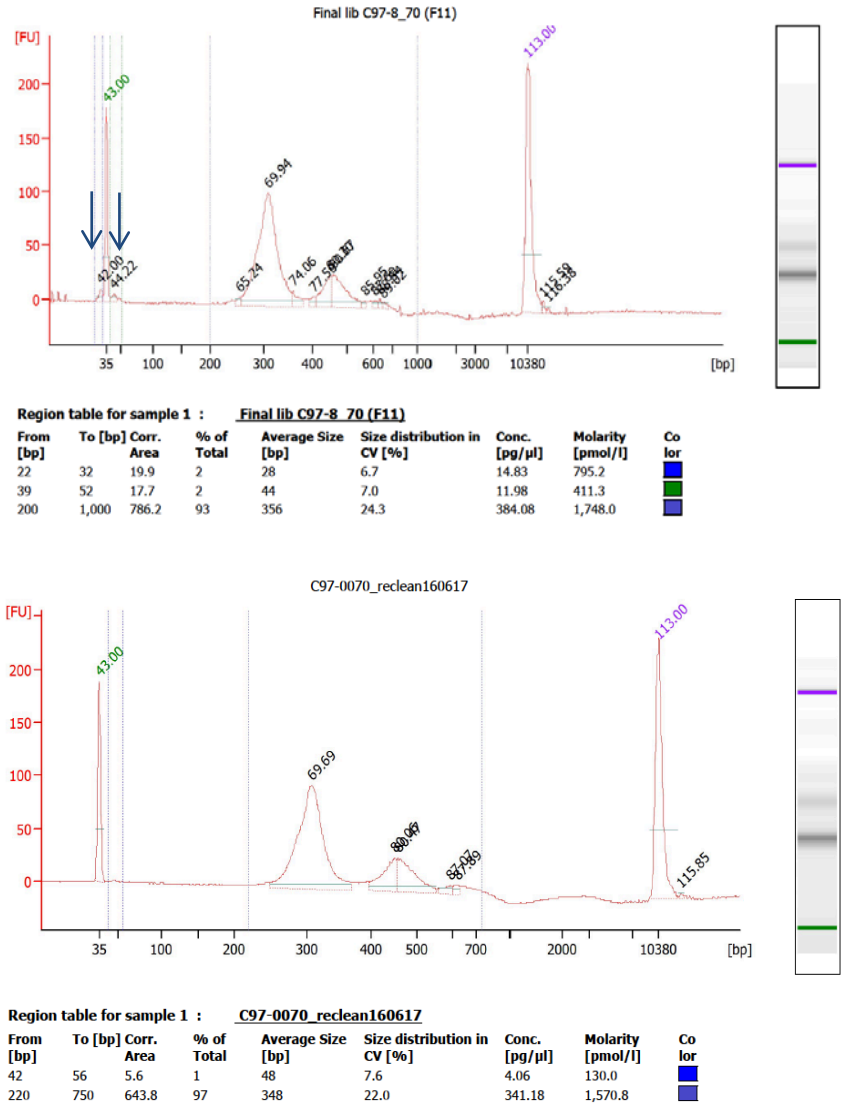
First, I identified that these calls were always of low frequency (median 0.53% VAF in the first run and 0.06% VAF in the next run), which could be potentially detrimental when aiming to detect subclonal tumour mutations with low VAFs. The vast majority of the likely false positive calls were present in consensus families containing only 1 or 2 reads. Second, these low level variants were always identical to a heterozygous or homozygous SNPs called in one or more of the unrelated samples sequenced on the same lane (referred to hereafter as the donor sample). There was no evidence of cross-contamination between lanes and many of the samples that were affected by this cross-contamination had not been prepared together, excluding accidental contamination during library preparation. Therefore we reasoned that the cross-contamination between samples multiplexed together on a lane occurs during cluster generation on the flow cell.

Coincidentally shortly after we noticed this issue, similar problems had been reported in a BioRxiv pre-print article [183] that described up to 5-10% of sequencing reads that were incorrectly assigned from a given (donor) sample to other multiplexed samples that had been sequenced on Illumina platforms. The authors hypothesised that these events arise from low levels of free index primers present in the multiplexed sample pool, which can prime pooled library fragments

to create a new library molecule. This finding was subsequently acknowledged as ‘index hopping’ in Illumina’s White Paper released in April 2017 [183-185].

Our sequencing data on the HiSeq 2500 highlighted that the issue can cause data misinterpretation, above what can be accepted as background noise in a high sensitivity cfDNA sequencing assay. I wanted to ensure that the level of index hopping remained below the limit of detection of our assay and hence implemented some changes to the library construction and sequencing workflow.

As recommended by Illumina to address the problem, I performed two rounds of Ampure clean up on the final sequencing libraries to remove residual free index primers. I retrospectively quantified the amount of unbound index primers and primer dimers in the final libraries of all samples in the two affected runs. These can be quantified on the Bioanalyzer trace between 23-30 bp and 39-52 bp (Figure 3.12- top panel). For the two sequencing runs where index hopping had been detected, I performed a second round of 1x Ampure bead clean up on the final libraries and reassessed the library profile (Figure 3.12- bottom panel) on the Bioanalyzer. The unbound index primer peaks were no longer visible after the second round of clean up.



Region table for sample 1 : Final lib C97-8_70 (F11)

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Molarity [pmol/l]	Color
22	32	19.9	2	28	6.7	14.83	795.2	
39	52	17.7	2	44	7.0	11.98	411.3	
200	1,000	786.2	93	356	24.3	384.08	1,748.0	

Region table for sample 1 : C97-0070_reclean160617

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Molarity [pmol/l]	Color
42	56	5.6	1	48	7.6	4.06	130.0	
220	750	643.8	97	348	22.0	341.18	1,570.8	

Figure 3.12: Unbound index primer visible on Bioanalyzer trace
Top panel: Retrospective quantification of the amount of unbound index primers and primer dimers in the library of a sample affected by index hopping. Leftover free index primers are visible after a single round of Ampure clean up (blue arrows) with a concentration of 14.83 pg/µl between 22-32 bp and 11.98 pg/µl between 39 and 52 bp.
Bottom panel: A second round of Ampure clean up was performed on the same library, thereafter no residual index primer was visible on the Bioanalyzer trace.

Additionally, since high cluster density has also been reported to be a source of introducing mis-priming events [183], an automated cluster generator (Illumina cBot) was introduced as an additional step prior to sequencing. This would theoretically reduce the risk of mis-priming events as cBot optimises the cluster density [186].

With the combination of 2 rounds of Ampure clean up, introduction of the Illumina cBot for optimal cluster density, I re-sequenced the 12 samples multiplexing the same samples on each flowcell as before. I carried out repeat mutation calling and manual BAM file review at SNP positions where the false positive calls due to index hopping had previously been seen (Figure 3.13 A and B, top panels). The implementation of these measures reduced the index hopping related false positive calls from 0.93% to 0.12% VAF at position Y1078C in *REV3L* gene and from 2.08% to 0.00% VAF at position R521K in *EGFR* gene (Table 3.4). Across all 12 samples there was a significant reduction in mis-assigned reads from a mean VAF of 0.15% to 0.03%, p=0.003 paired t-test. The VAF of these remaining index hopping events was below the limit of detection of our assay. Furthermore, review of the BAM file showed that the erroneous reads were all lone reads (consensus family size =1) and hence would be filtered out by the MBC error correction tool. All samples that were multiplexed together were checked for index hopping events in subsequent experiments, by assessing if SNPs from one sample became detectable in any of the other as low frequency variants.

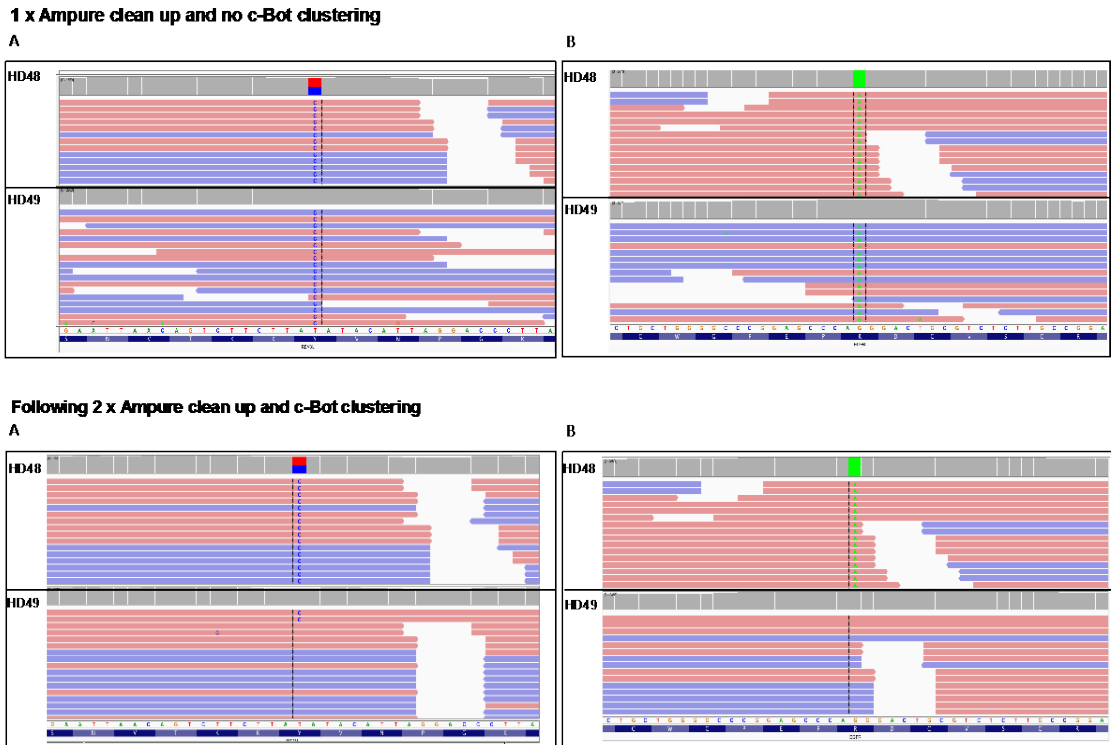


Figure 3.13: IGV screenshot demonstrating false positive calls emerging from mis-assignment of reads.
Index misassignment is demonstrated from sample HD48 to sample HD49 at two genomic positions (A and B). These errors were present in consensus families comprising 1 read. Following 2 sequential rounds of Ampure clean up and cBot clustering the false positive calls were below the limit of detection.

A

REV3L Y1078C		1x Ampure clean	2x Ampure clean up
HD49 (donor	Total reads	1514	1556
	Ref allele (T)	769	797
	Variant allele	745	759
HD48	Total reads	2036	1731
	Ref allele (T)	2015	1721
	Variant allele	19	2

B

EGFR R521K		1x Ampure clean	2x Ampure clean up
HD48 (donor	Total reads	2346	2346
	Ref allele (G)	0	0
	Variant allele	2346	2346
HD49	Total reads	3071	2485
	Ref allele (G)	3005	2485
	Variant allele	64	0

Table 3.4: Table depicting number of reads arising from misassignment.
Reference reads and variant reads at SNP positions of the donor sample are given, with false positive reads arising from index misassignment before and after 2x rounds of Ampure clean up and cBot clustering

Using the same process for all subsequent sequencing runs I pooled three samples per lane using cBot clustering, instead of six samples across two lanes with on-board clustering and ensured that longitudinal samples from the same patient were not pooled together. In the subsequent cfDNA libraries sequenced, no further index hop false positives were called by SureCall or through manual BAM review at SNP positions, confirming that the combination of these measures reduced the frequency to below the limit of detection.

3.4.4 Summary of optimisation results

In summary, the optimised protocol using Agilent's SureSelect$^{XT\text{-}HS}$ kit included the fast-hybridisation protocol with 70ºC capture wash temperature and 10 post-capture PCR cycles. In addition 2x Ampure bead clean-ups were performed. Overall, 7 cfDNA libraries were prepared from FOrMAT mCRC patients with the manufacturer's initial library preparation protocol (standard-hybridisation and post-capture wash temperature of 65°C) and 21 FOrMAT cfDNA libraries with the optimised protocol, for the final cohort of patient samples presented here and in subsequent chapters. These were sequenced with 75 bp paired-end reads on an Illumina HiSeq2500 in rapid output mode with overall improved quality control metrics with the optimised protocol (Table 3.5).

	Initial protocol	Optimised protocol
Total read count	80,453,088	109,347,960
Duplicates (%)	83	91
Depth after deduplication	1205	2087
On target (%) pre deduplication	68	75
On target (%) after deduplication	45	57

Table 3.5: Median quality control (QC) metrics across all FOrMAT patient samples prepared with either the initial protocol (7 samples) or the optimised protocol (21 samples)

The average total read count is higher in the samples prepared with the optimised protocol as fewer samples were multiplexed per run, but this alone was not found

to affect depth or on-target rates. A key factor responsible for improved depth and on-target rates was the improved capture wash temperature which removes non-specific off-target reads, enabling improved capture and subsequent sequencing of target regions. After MBC deduplication, the median on-target depth for the 70^{o}C washed samples was 2087X. This is theoretically sufficient to achieve a detection limit of approximately 1 mutated DNA fragment in 2087 molecules (0.048%). However, the sensitivity for the *de novo* detection of mutations may be lower in practice since more than one read is required to support robust bioinformatic variant calls. Two independent reads are required thus the detection limit is closer to 0.09% (2 reads in 2087).

3.4.5 Removal of false positive errors with molecular barcodes

As detailed in section 3.1, MBCs are incorporated into SureSelect$^{XT\text{-}HS}$ library prep kit to enhance error correction (Figure 3.2). Reads with identical mapping positions and MBCs are grouped by the SureCall software and every nucleotide position is analysed for the most frequently represented base to build a consensus read.

Identification of low frequency variants requires over-sequencing in order to obtain sufficient reads that can then be grouped and collapsed into the consensus read. With the optimised protocol, I sequenced a median of 109 million reads per sample. MBC-based deduplication reduced the median depth from 31734X pre deduplication to 2087X. Of the 21 samples sequenced with the optimised protocol, consensus families contained a median of 8 to 15 supporting reads (Figure 3.14), which is within the range that has been deemed optimal for barcode error correction [123].

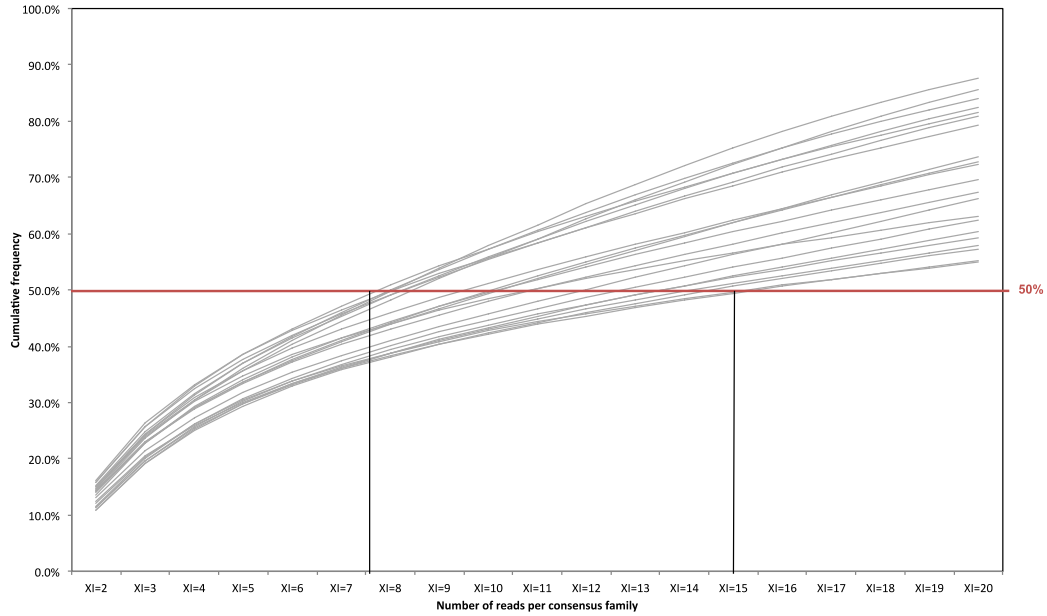


Figure 3.14: Cumulative consensus family size distribution.

The median number of reads (XI) per consensus family varied between 8 and 15 across all 21 samples sequenced with the optimised protocol.

3.4.5.1 Benefit of molecular barcode deduplication over standard deduplication to reduce error rates

I next compared MBC deduplication with standard deduplication (bioinformatic removal of duplicated reads, without the use of MBCs) of sequencing data from Horizon cfDNA reference standards to assess the added benefit of MBCs in error correction. Horizon reference standard cfDNA contains six engineered single nucleotide variants (SNVs) located within our target region that have a pre-defined VAF. These reference samples can be used to assess the limit of detection and to analyse sensitivity and specificity of sequencing assays.

I sequenced the Horizon cfDNA reference standard with VAFs around 1% and called mutations with both standard deduplication and MBC deduplication. I

compared the number of variants called below 2% VAF to assess sequencing errors as this VAF range is error prone in Illumina sequencing [187] but is important for the detection of subclonal mutations and acquired resistance mutations [176, 188, 189]. Utilisation of the MBCs significantly reduced the number of called variants from 218 to 18, with all true variants accurately called (Figure 3.15).

A false positive is a variant detected in cfDNA when one is not present at this position in the reference cfDNA, and a false negative is when a variant was not detected in cfDNA but is known to be present at that position. There was a 92% reduction of false positives through MBCs, illustrating the ability of this error correction approach to reduce errors arising from PCR or sequencing errors. The false negative rate was 0 regardless of whether barcodes were used for error correction or not, as all six known mutations were detected.

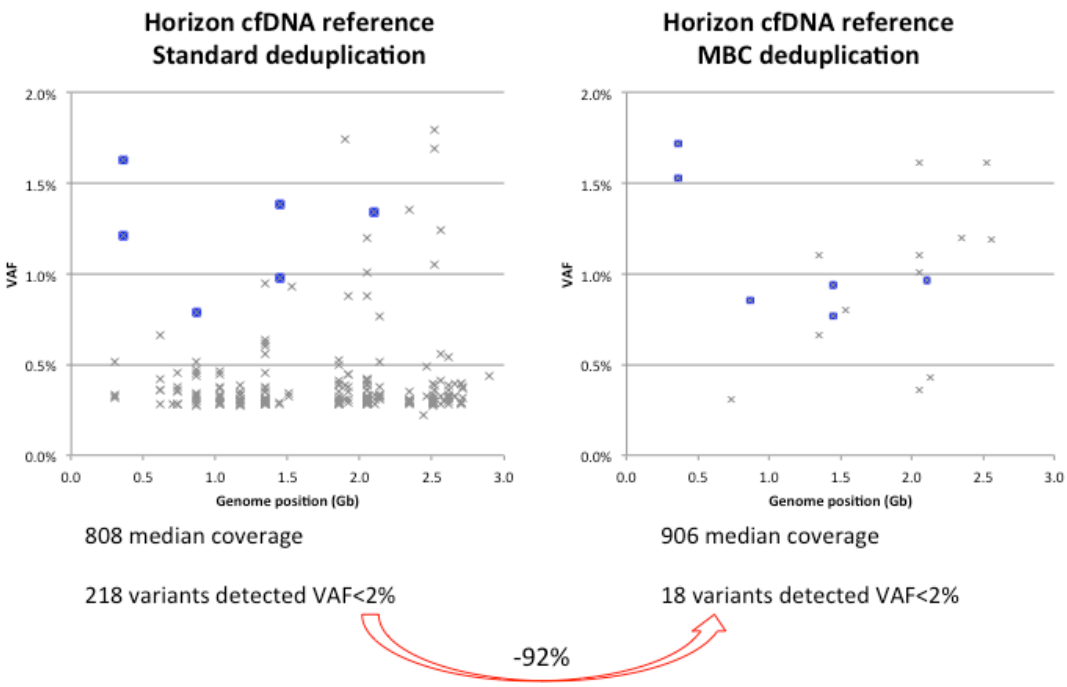


Figure 3.15: Variant calls using standard deduplication vs. SureCall MBC deduplication
This figure demonstrates utility of MBC error correction in reducing rare variant false positive calls in Horizon reference standard DNA. The aligned raw reads show that the expected variants (blue dots) are obscured by false positive noise (grey dots).

3.4.6 Sensitivity of cfDNA-seq assay

I next wanted to investigate the sensitivity of the cfDNA-seq assay. A mixing experiment was designed to test the limit of detection of cfDNA-seq assay. Based on the achieved sequencing depth, this was expected to be close to or below 0.1%.

I utilised cfDNA from two previously sequenced healthy donors that differed in 16 homozygous SNPs within the target region. A dilution series was prepared of 0.15%, 0.075% and 0.0375% cfDNA from donor A spiked into the cfDNA from donor B. 25 ng of each mixture were then prepared for sequencing using the optimised protocol. A median of 74,030,118 reads were sequenced per sample, with an average read length of 73 bp and a median on-target depth of 21651X before deduplication. Sequencing data from each sample was processed with MBC deduplication. Standard deduplication was performed in parallel to compare the number of false positives between the two approaches (Figure 3.16).

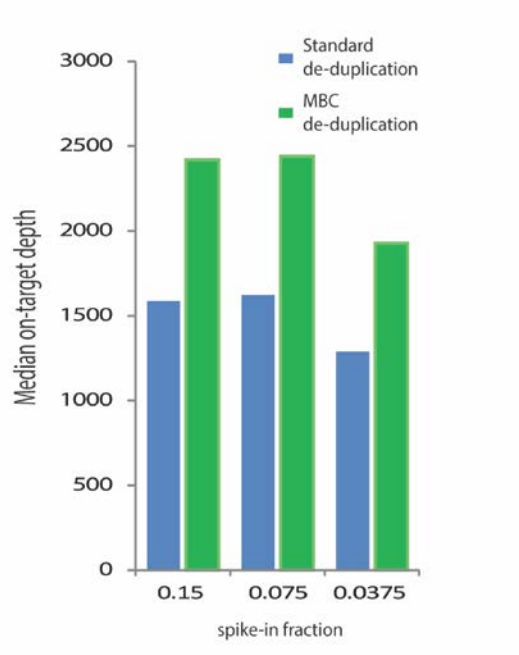


Figure 3.16: cfDNA mixing experiment.
25 ng mixes of donor A spiked into donor B at 0.15%, 0.075% and 0.0375%. Library preparation was performed with the optimised protocol.

The median on-target depth was higher after MBC deduplication (2420X versus 1587X with standard deduplication; Figure 3.16). This was anticipated as DNA fragments in the cfDNA sample that map to the same genomic location will be tagged with distinct MBCs, as will the forward and the reverse strand of each double stranded DNA molecule (so-called ‘duplex’ DNA) in the original sample. These reads are retained as independent units of information when MBCs are used in NGS and this over-sequencing facilitates the identification and correction of randomly introduced errors [190]. Standard deduplication removes all but one of these types of reads as it cannot distinguish them from PCR duplicates.

I investigated whether the spiked-in SNPs could be re-identified in the MBC deduplicated BAM files using IGV. All 16 SNPs were detected with at least one read showing the variant in the 0.15% mix, 14/16 at 0.075% and 11/16 at 0.0375% mixing ratios (Table 3.5). The observed VAFs varied between the SNPs within each sample, likely as a result of stochastic sampling at such low VAFs. Overall, this ultra-deep cfDNA sequencing assay allowed detection of SNPs at 0.15% and still showed a high sensitivity for detection at 0.075% (i.e. one mutated DNA molecule out of 1333 DNA molecules).

		Expected VAF in sample		
Hg19 position		0.150%	0.075%	0.038%
chr5	112162854	0.095%	0.060%	0.039%
chr5	112164561	0.074%	0.163%	0.000%
chr5	112175770	0.183%	0.078%	0.245%
chr5	112176325	0.055%	0.056%	0.000%
chr5	112176559	0.323%	0.097%	0.041%
chr5	112177171	0.479%	0.071%	0.087%
chr6	111695268	0.198%	0.236%	0.295%
chr6	111696852	0.339%	0.126%	0.102%
chr7	55268916	0.052%	0.000%	0.000%
chr7	116339282	0.110%	0.027%	0.035%
chr7	116435768	0.123%	0.072%	0.061%
chr7	116436022	0.315%	0.105%	0.034%
chr7	116436097	0.161%	0.028%	0.000%
chr18	44346224	0.131%	0.043%	0.192%
chr18	51170937	0.211%	0.000%	0.000%
chr18	61164602	0.050%	0.126%	0.034%
Median observed VAF		0.146%	0.071%	0.037%

Table 3.6: cfDNA mixing experiment.
cfDNA from donor A was spiked into cfDNA from donor B at 3 different spiking ratios: 0.15%, 0.075% and 0.0375%. Expected and observed VAFs and genomic positions for 16 SNP positions are shown.

The next aim was to assess whether MBC error correction improved the bioinformatics calling accuracy of ultra-low frequency variants, which is technically more challenging than re-identification of known variants. The SureCall software was used to assess how MBCs influence the ability to detect any of the nine homozygous variant SNPs from sample A that were present at 0.15% in the cfDNA mixture. SNPs that are variant in sample B and wild type in the spiked in donor A could not be assessed as they were present in 99.85% of DNA molecules in the cfDNA mix and were hence always called accurately against the reference genome.

Calling of the mixing experiment was first performed using MBC deduplicated data and then the same data was processed with standard deduplication. The same variant calling parameters were used for both. Mutation calling in the entire target region using standard deduplication and a low stringency variant call quality threshold of 40 (VCQT- thresholding used to bioinformatically trim low

quality reads) detected five out of the nine spiked-in homozygous SNPs but also generated 156 additional mutation calls. These additional variants are likely false positives as they were not identified by deep sequencing of the individual cfDNA samples. Dr Louise Barber then performed stepwise increase of the VCQT which first reduced the likely false positive calls to 147 while retaining all five true positive calls, but a further reduction in false positives (to only three at the highest tested VCQT of 80) was accompanied by a loss of sensitivity with only a single spiked-in SNP called (Figure 3.17). When the same data were called using MBCs and a low stringency VCQT of 40, four of the spiked-in SNPs that had been found above were called but with only two likely false positive variants. I assessed why calling with MBC error correction failed to pick up the five other spiked SNPs. Each of these had a VAF <0.1% when observed in the BAM files and the SureCall software requires a minimum threshold of 0.1% VAF to make a call. I also compared the number of false positive calls after adjusting the VCQT threshold for calling from standard deduplicated data so that this identified the same four true positive variants detected by MBC supported calling. 81 likely false positives were called compared to just two using MBC.

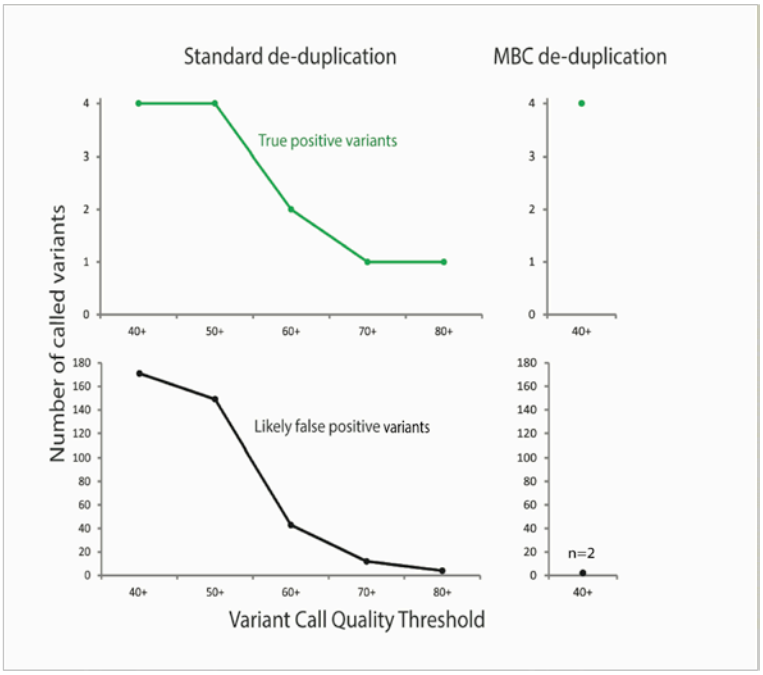


Figure 3.17: Impact of MBC error correction on true positive and false positive calls.
The top panels show the number of true positive variants (expected SNPs) that were bioinformatically called in the mixing experiment with standard deduplication (left) and MBC deduplication (right) using different variant call quality thresholds. The lower panel shows the number of likely false positive variant calls (not observed in the deep sequencing of either cfDNA sample used in the mix) for standard deduplication (left) and MBC deduplication (right).

Hence variant calling with a VCQT of 40 showed the same sensitivity for both approaches but deduplication using the MBC increased the specificity by 97.5%. This demonstrates that MBC error correction leads to a dramatic decrease of false positive mutation calls and enabled the identification of variants with a sensitivity of 0.15% with high precision.

3.4.7 Duplex DNA-molecule identification for enhanced error correction

I next visualised sequencing data from the above mixing experiment on IGV to understand patterns associated with true positive variants. I observed that all true variants were supported by at least two consensus families that mapped to the same genomic position but differed in whether the variant was seen in read 1 or read 2 in paired-end sequencing. Read 1 and read 2 originate from the same cfDNA fragment such that in paired-end sequencing the cfDNA fragment is first

read from one end to the other (read 1) and then the process starts again in the other direction (read 2) (Figure 3.18).

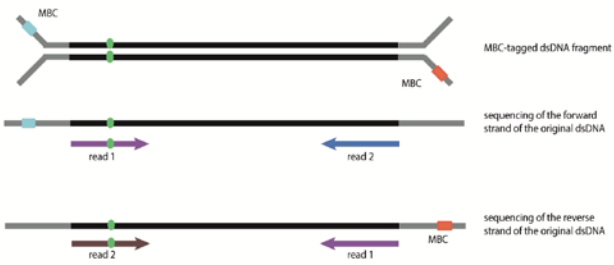


Figure 3.18: Illustration of duplex read pair detection.

The original forward strand of the cfDNA duplex will result in read pairs where the forward read (relative to the reference genome) will be read 1 and the reverse read will be read 2. The reverse strand of the original cfDNA duplex will result in read pairs with the opposite orientation: i.e. the reverse read is read 1 and the forward read is read 2. Hence the order of the read pair sequences with respect to the reference genome alignment can be used to detect duplex DNA strands.

These were highly likely to represent the forward and reverse strand of the double stranded input cfDNA, as when MBCs are ligated to the cfDNA fragment the orientation of the MBC will be different for the two strands. This strategy allows for reconstruction of parental double-stranded DNA duplexes from sequencing data, which is an important approach that can be used to re-identify reads from the forward and the reverse strand of the initial duplex molecules and should hence enable further error correction (Figure 3.18). This is a similar approach used by Schmitt et al [165], however their approach of duplex sequencing relies on complex and often unstable duplex adaptors, which tag both strands of the original dsDNA compared to the tagging of a single strand in our technique which is technically more straightforward.

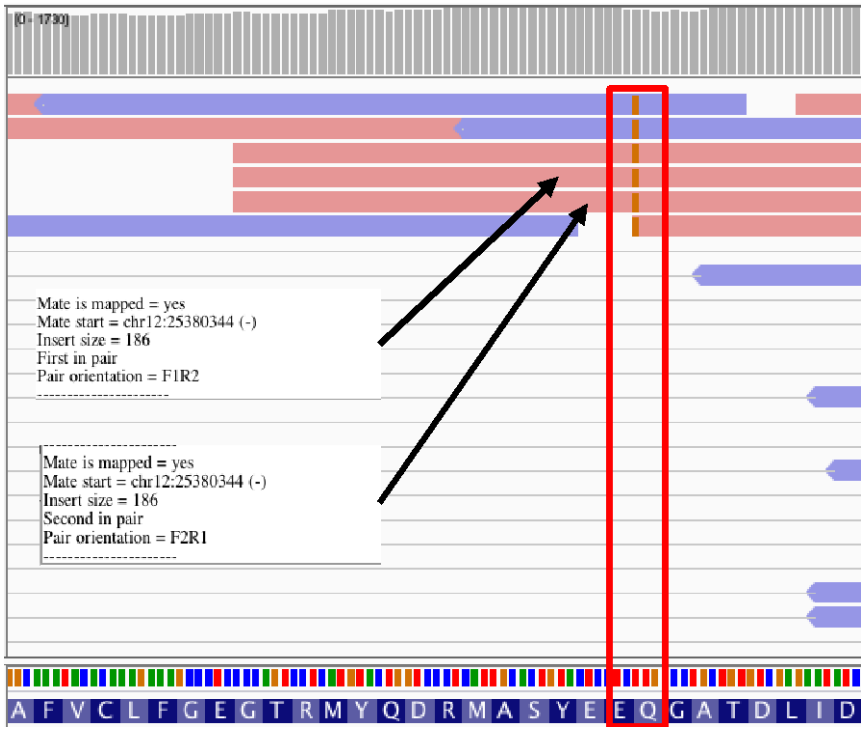


Figure 3.19: IGV screenshot illustrating an example of a subclonal mutation supported by duplex reads.
Variant bases are present in both, read 1 (F1R2) and read 2 (F2R1) of paired end sequencing and supported by five consensus families.

Based on this observation, the 'duplexCaller' bioinformatics tool was developed by a staff bioinformatician in the Centre for Evolution and Cancer (Dr Dimitrios Kleftogiannis –Methods section 2.2.6.4; available at https://github.com/dkleftogi/duplexFiltering). This tool identifies variants supported by duplex reads representing parent double-stranded cfDNA. First, duplexCaller identifies the consensus reads that span a mutation that has been called by SureCall software. As an output, the duplexCaller tool reports the total number of duplexes found for every variant detected by the SureCall software.

From the mixing experiment above, we know that MBC error correction on its own is not sufficient to reduce the false positive error rate to zero, therefore I wanted to assess the additional advantage of duplex calling for error correction. I analysed six healthy donor samples by first applying the SureCall software to

MBC deduplicated data. This called homozygous and heterozygous SNPs with VAFs of 50% and 100%, respectively, and a median of 29 variants with a VAF <2% in each of the six healthy donors. After applying the duplexCaller the number of calls with VAF <2% decreased to a median of three calls (9.6 fold decrease) per sample (Figure 3.20).

Some of the low frequency variants that remained after duplexCaller filtering were found to be recurrent error positions in our dataset including *PIK3CA* E707K and *MAP2K2* V8M (indicated in Figure 3.20 in green and blue, respectively), which were identified by reviewing these positions manually on IGV. Following removal of calls from recurrent error positions, three of the six healthy donors had had no further low frequency variants but the remaining three still had a total of 11 likely false positive calls. I manually reviewed these calls in IGV to establish the nature of these (marked in red in Figure 3.20). Six of the 11 false positives calls occurred at the same genomic position in all three HD samples and are alignment errors with a mapping quality score of zero, always occurring in the *PTEN* gene. These variants were subsequently bioinformatically filtered out of the patient cfDNA datasets and thereafter filtered out of the duplexCaller tool.

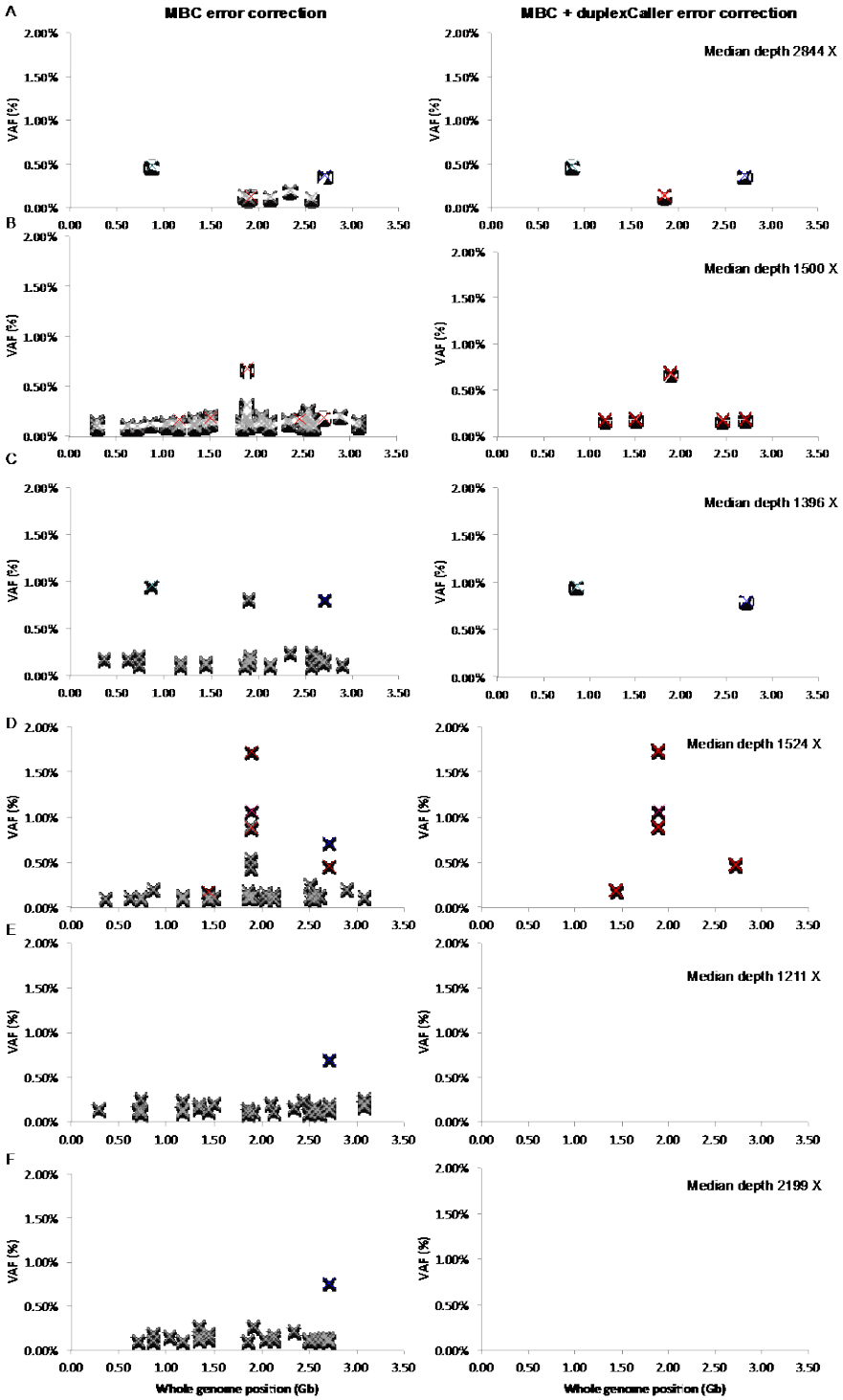


Figure 3.20: Error correction with MBC deduplication using SureCall and additional error correction with duplexCaller tool in six healthy donor samples.

Whole genome position is represented on the x-axis to show VAF by genome position on the y-axis. All low level variant calls (<2%VAF) are shown. Variants highlighted in red are common to both MBC error corrected and MBC plus duplexCaller error corrected calls and result from alignment errors with a mapping quality of 0 in the *PTEN* gene. Recurrent *PIK3CA* E707K variant is highlighted in green and recurrent *MAP2K2* V8M variant is highlighted in blue. Other base

changes detected in the target region relative to the hg19 reference genome are marked as grey crosses.

The remaining five false positive calls found in the HD's were observed to have a common pattern such that the variant base was always located at the end of a read in fragments that had insert sizes less than 74 bp (Figure 3.21). The sequencing adapter starts at this position of the read and with an intended sequencing read length of 75 bp, those reads sequenced with a length below 75 bp will incorporate erroneous mapping of the sequencing adapter to the human reference genome.

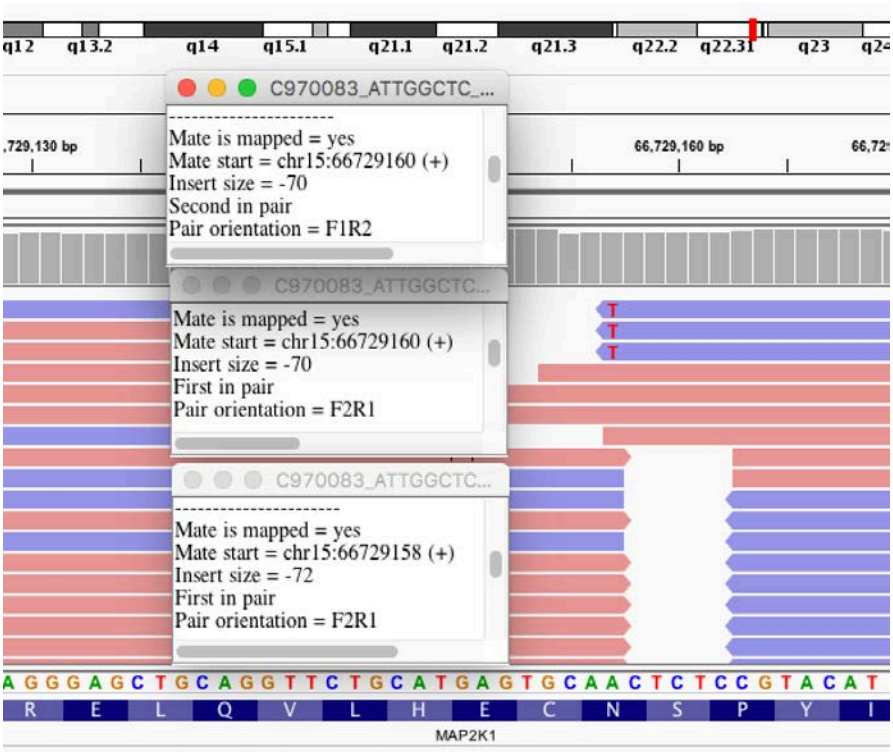


Figure 3.21: An illustration of an adapter error.

First and second read form the duplex (identical insert size and start position but differ in whether the variant is present in the forward read (F1R2), relative to the reference genome, or the reverse read (F2R1). However, the read length is <74 bp and variant is present at the end of the read and thus represents erroneous mapping of the adapter.

As reported above, most sequencing runs had an average sequencing read length of 73 bp, thus these would acquire one or two bases of the adapter at the

end of the read. Based on this observation the duplexCaller tool was upgraded to eliminate such errors.

3.5 Summary and Discussion

The low tumour-derived cfDNA fraction and technical challenges associated with ultra-deep NGS techniques complicates the analysis of cfDNA with high sensitivity and specificity. The above results show proof of principle that the novel solution hybrid capture cfDNA-seq assay, which incorporates MBC error correction and the duplexCaller bioinformatic tool, can significantly reduce false positive mutation calls by 97.5% while maintaining true positive calls and enable the identification of variants with a sensitivity of 0.15%.

Through cfDNA extraction from 58 FOrMAT trial mCRC patients and 25 healthy donors I demonstrated that the amount of cfDNA is almost five-fold higher in patients with mCRC than in HDs (median 20.8 ng/ml vs. 4.6 ng/ml), which is in keeping with studies performed in different tumour types including lung, breast and colorectal cancers [96, 191, 192]. Modelling based on these yields confirmed that 25 ng cfDNA could be extracted from 10 ml plasma from the majority (>95%) of mCRC patients. 25 ng was therefore chosen as the standard library input for the cfDNA-seq assay, with a long term view of optimising this assay for clinical use where cfDNA quantities can be low.

A number of improvements have been described in this chapter to optimise the SureSelect $^{XT\text{-}HS}$ cfDNA library preparation protocol for low input amounts and development of a bioinformatics pipeline to improve error correction. The modifications that have been combined in the standard library preparation and sequencing workflow for sequencing experiments in later chapters are:

- Incorporation of MBC technology which reduces false positive mutation calls
- The fast-hybridisation protocol which improves the efficiency of the workflow through an overall reduced protocol time to eight hours.
- Increased stringency washes during the capture to reduce off-target reads by ~40%
- Double Ampure clean up to reduce free index primers and cBot clustering to reduce index hopping events
- Based on my data which showed that duplex strands can be identified, the duplexCaller tool was developed. This is applied following MBC error correction and mutation calling to further improve error correction. This tool is freely available to use in the public domain and can be integrated into any bioinformatics pipeline.

I furthermore showed that these optimisations have yielded an assay which can achieve de-duplicated sequencing depths of 2029X median from 25 ng cfDNA. At this sequencing depth, subsequent error correction with MBCs enabled a sensitivity for point mutation detection down to 0.15% VAF and a specificity of 97.5%, which were calculated through cfDNA mixing experiments using two healthy donors with known heterozygous SNPs.

One further optimisation that could be considered for future deep sequencing experiments is the use of unique dual multiplex indexing. This would help remove the potential interference of index hopping, which is more common on newer Illumina sequencers that use ExAmp and patterened flowcell technology which has been shown to further aggravate index hopping compared to the HiSeq 2500

instrument which was used for sequencing above [183]. Using unique dual 8-plex indexes and avoiding the use of any index more than once per sample pool creates sample-specific index pairs and therefore any unexpected index combinations are filtered out as undetermined reads [185]. Illumina are also developing an enzymatic treatment to block the free 3' ends so the residual adapter cannot act as a primer to produce an index hopped strand [185].

Overall, this cfDNA-seq has advantages over similar cfDNA sequencing technologies which have been published while this work was ongoing such as CAPP-Seq which achieves a specificity of 96% and detection limit down to ~0.02% [117]. CAPP-Seq is dependent on a background error correction model that is constructed from a large number of healthy donor samples that have to be sequenced for digital error suppression [193]. CAPP-seq technology has been developed into the commercially available Avenio ctDNA Expanded Kit which requires higher cfDNA input amounts of 50 ng to achieve a sensitivity of 100% for all variants down to 0.18% [194, 195]. However when input levels are reduced to 10 ng input, the limit of detection is 0.6% with a sensitivity of 92% [194]. Finally, the clinically validated Guardant360 assay has a high specificity of >99.99%, with a high analytic sensitivity limit of detection down to 0.06% in the 54-gene panel with 5-30 ng cfDNA input [116]. However these assays require proprietary reagents and bioinformatics pipelines and do not have customisable target panels, hence limiting their use for research. Our cfDNA sequencing approach was developed with a customisable gene panel and off-the-shelf reagents and demonstrated a similar sensitivity and specificity compared to published cfDNA sequencing techniques that use bespoke reagents.

One limitation of the design of the target capture panel arose as a result of the inclusion of 40 heterozygous SNP targets on chromosome 18. These SNP probes were initially included with the intention of identifying chromosomal regions with copy number variation, however the SNP probes were not used for this purpose and genome wide CNVs were detected using the CNVkit algorithm. Thus the target panel size was unduly larger than it needed to be, which may have affected overall sequencing performance since smaller panels can be sequenced to a higher depth. The other main limitation of the work in this chapter is that I was not able to test the entire error correction duplexCaller pipeline on a second set of HD samples, which would have confirmed elimination of the recurrent error positions. This is being performed with ongoing sequencing experiments in the lab.

CHAPTER 4: MINIMALLY INVASIVE MUTATION DETECTION AND GENOME WIDE DNA COPY NUMBER RECONSTRUCTION OF METASTATIC COLORECTAL CANCERS THROUGH CIRCULATING FREE DNA SEQUENCING

4.1 Introduction

Sequential interrogation of cancer genomes from patients with mCRC enables identification of driver mutations relevant to therapy resistance and treatment outcome. For example, in metastatic colorectal cancer (mCRC) the presence of *KRAS* and *BRAF* mutations predicts resistance to anti-EGFR therapy [196]. The presence of *BRAF* V600E mutation, or microsatellite instability may guide therapeutic options towards novel combined RAF/MEK/EGFR inhibitor therapy [197] or immunotherapy agents, respectively [198]. *ERBB2* amplifications also predict benefit from anti-HER2 therapy herceptin [199].

Conventional invasive tumour biopsies pose difficulties due to fine needle or core biopsies yielding insufficient quantity and quality of DNA. Additionally, tissue biopsies can be risky depending on the anatomical location of the tumour deposits, cause discomfort and are expensive. Furthermore ITH may lead to spatial biases and resistance mutations may change prior to treatment initiation or during progression [200]. Thus in clinical practice, longitudinal clonal evolution should ideally be studied prior to making any treatment decisions and this may be achieved by utilising liquid biopsies for minimally invasive tumour monitoring over time. ctDNA analysis can sample mutations more comprehensively across metastatic sites than invasive tumour biopsies [96]. Minimally invasive ctDNA

based liquid biopsies are hence a promising tool to genotype tumours for specific biomarker identification and also to track the evolution of drug resistance during therapy.

4.1.1 Hypothesis and Objectives

I analysed the circulating free DNA sequencing (cfDNA-seq) data from 28 mCRC patients enrolled in the FOrMAT clinical trial described in Chapter 2 (section 2.2.1.1) in order to:

- assess the concordance of cfDNA-detected mutations with mutations and copy number changes identified by tumour biopsy sequencing from the FOrMAT trial.
- Analyse the clinical relevance of mutations identified in cfDNA but not by sequencing of matched tumour samples
- define the correlation of plasma cfDNA concentration with clinicopathological features (site of disease, tumour metastatic burden, histopathology and number of lines of systemic therapy) to identify patients where sufficient cfDNA is available for liquid biopsies, to guide future selection of patients from who cfDNA can provide useful insights into tumour genomic profiling.

4.2 Mutation analysis and concordance with tumour biopsy sequencing

cfDNA from 28 cases from the FOrMAT clinical trial was subjected to ultra-deep targeted cfDNA-seq (as described in Chapter 3). Seven cases were sequenced with the original library preparation protocol using 65°C wash temperature and standard hybridisation, and 21 cases were sequenced with the optimised protocol using 70ºC washes and fast-hybridisation (detailed in Chapter 3 section 3.4.2.).

Mutation calling was performed using the duplexCaller error correction tool (Chapter 3 section 3.4.7).

DNA was obtained from formalin fixed paraffin embedded (FFPE) archival specimens from 23 cases and matched peripheral blood from all patients in the FOrMAT trial had been sequenced with solution hybrid capture enrichment of 46 cancer driver genes relevant for gastrointestinal cancers [151], 16 of these genes were also present in our cfDNA sequencing assay (Table 4.1). Tumour samples from four further cases had been analysed with an amplicon sequencing panel used in routine clinical diagnostics that included five genes (*BRAF*, *KRAS*, *NRAS*, *PIK3CA* and *TP53*), which were all present in our cfDNA sequencing panel. One case had failed any tissue sequencing and no mutations were known. Mutation calls, from FOrMAT trial tumour and blood sequencing data were available for my analyses.

FOrMAT Tumour NGS Target Panel		cfDNA Sequencing Target Panel	
Gene Name	**Exons**	**Gene Name**	**Exons**
APC	1-16	*APC*	1-16
TP53	2-11	*TP53*	2-11
KRAS	2-4	*KRAS*	2-5
SMAD4	2-12	*SMAD4*	2-12
FBXW7	1-11	*FBXW7*	1-14
CTNNB1	2-15	*CTNNB1*	2-16
TCF7L2	1-14	*TCF7L2*	1-14
ATM	2-63	*ATM*	2-64
NRAS	2-4	*NRAS*	2-5
MAP2K1	1-11	*MAP2K1*	1-11
MAP2K2	1-11	*MAP2K2*	1-11
PIK3CA	2, 5-6, 10, 21	*PIK3CA*	2-21
PTEN	1-9	*PTEN*	1-9
BRAF	11, 15	*BRAF*	1-18
EGFR	18-21	*EGFR*	1-22
ERBB2	8-11	*ERBB2*	1-27
		AMER1	2
AKT1	2-14		
ARID1A	1-20		
		B2M	1-3
		CDH1	1-16
CDK4	2-8		
CDKN2A	1-3		
CDKN2B	1-2		
DOCK2	1-52		
		ERBB3	1-28
		ERCC3	1-15
		ERCC4	1-11
		ERCC6	2-21
ELMO1	2-22		
ERBB4	1-28		
		FANCM	1-23
HRAS	2-5		
IDH1	3-10		
IDH2	1-11		
JAK3	2-24		
KIT	9, 11, 13-14, 17-18		
		MAPK1	1-8
		MAPK3	1-7
		MET	2-21
NOTCH1	1-34		
NOTCH2	1-34		
NOTCH3	1-33		
PDGFRA	12, 14, 18		
		RAF1	2-17
RET	2-20		
		REV3L	2-35
		RAD54L	3-19
ROS1	32-43		
		SMAD2	2-10
		TYMS	1-7
VHL	1-3		
UGT1A1	1		
FGFR2	Copy number		
IGF2	Copy number		
TRIM44	Copy number		
CCND1	Copy number		
IGF1	Copy number		
ALK	Copy number		
MST1R	Copy number		
CDK6	Copy number		
MET	Copy number		

Table 4.1: Comparison of genes in FOrMAT Tumour NGS target panel and customised 32-gene SureSelect target panel for cfDNA sequencing.
16 genes overlapped both panels where grey fields indicate that a gene was not covered. The final nine genes listed in the FOrMAT tumour NGS panel were included as they can often be copy number altered in gastrointestinal cancers [151].

cfDNA-seq in this cohort of mCRC patients demonstrated that the *TP53* gene was the most commonly mutated gene in 24 out of 28 (86%) cases, higher than the 60% mutation rate (p=0.008) found in CRC primary tumours in the TCGA cohort [19], followed by *APC* mutations affecting 22 out of 28 (79%) cases, comparable to the 81% reported in TCGA [19]. A high proportion of patients (20/28) showed aberrations in *RAS*/*RAF* pathway genes (Figure 4.1), which confers resistance to anti-EGFR therapy [201].

For each of the 28 cases, I analysed the concordance of mutation calls within the target regions common to the sequencing assay applied to tumour tissue within the FOrMAT trial and the cfDNA sequencing target panel (Table 4.1). Overall 80 out of 91 (88%) of all mutations that had been found by tumour sequencing were called in the cfDNA (Figure 4.1).

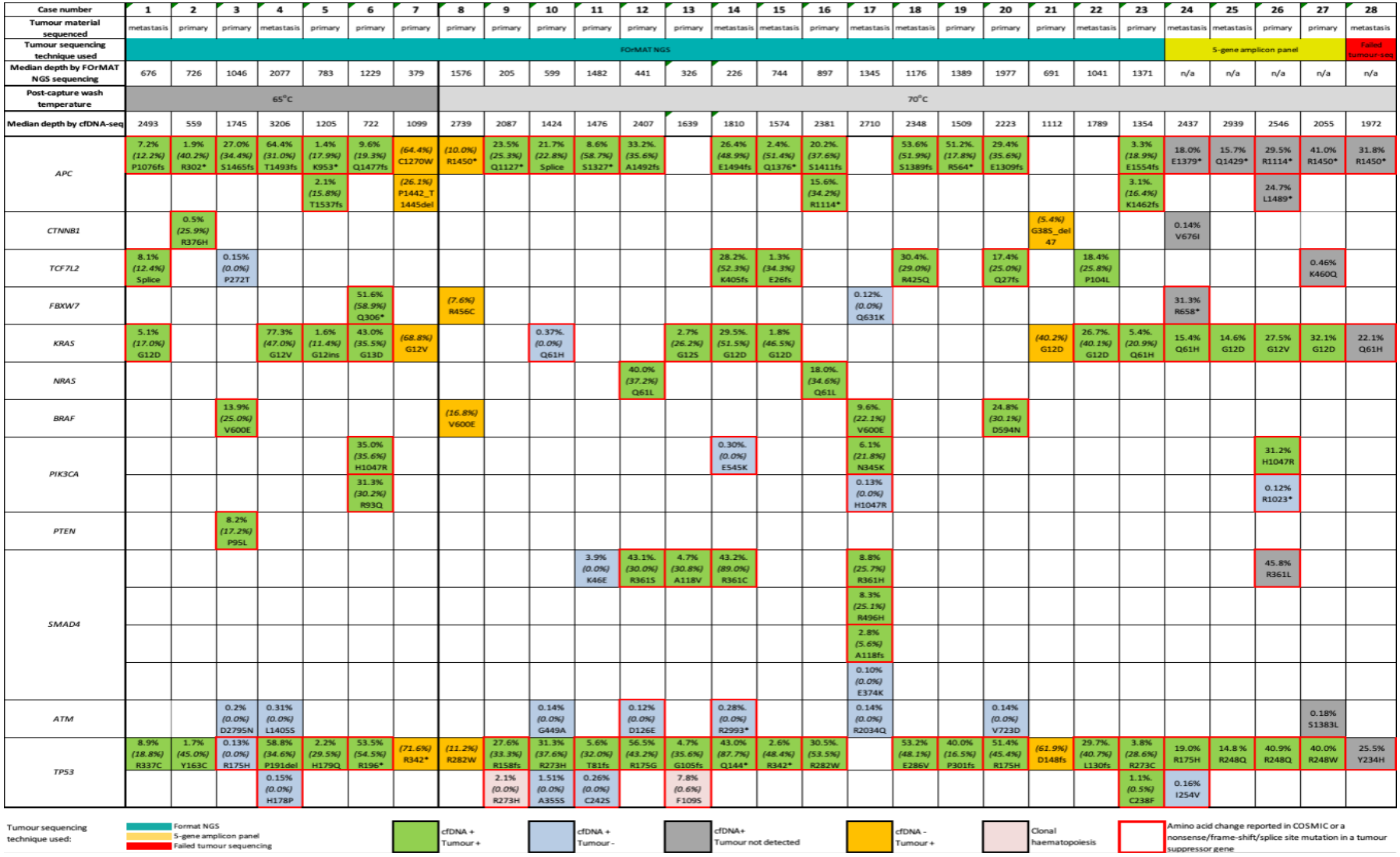


Figure 4.1: Concordance of mutations identified by cfDNA-seq and by sequencing of tumour material.
Within each cell, percentages indicate mutation VAFs acquired by cfDNA-seq and percentages in parantheses indicate mutation VAF acquired by FOrMAT trial sequencing of solid tumour tissue. Mutations identified in both cfDNA-seq and tumour sequencing are coloured green. Novel variants called by cfDNA-seq and not by tumour sequencing are coloured blue. Variants not detected by cfDNA-seq that were detected in tumour sequencing are coloured orange. Pink indicates clonal haematopoiesis. Red borders indicate mutations reported as tumourigenic in COSMIC. Variants in grey have been identified in the cfDNA of patients that either had been sequenced using the limited 5-gene amplicon panel or failed FOrMAT sequencing.

In twenty four out of the 27 cases (89%) where tumour tissue and cfDNA had been analysed, all mutations detected in tissue were also called in the cfDNA. The 11 mutations that were not called in cfDNA (Figure 4.1 orange boxes) were all from three of the 28 cases (11%). Review of the sequencing BAM files on the integrated genomics viewer (IGV) showed that five out of the 11 mutations had been detected in only one or two sequencing reads each, with a mean VAF of 0.035% i.e. one mutated DNA fragment in ~3000, which was below the limit of detection of our assay (Table 4.2). The other six mutations that were not visible in any sequencing reads were likely absent from cfDNA altogether due to very low/no tumour content in these samples. The median read depth of 2031 at these 11 positions demonstrates that sequencing was satisfactory but that the tumour derived DNA content within these three cfDNA samples was very low.

Case Number	Gene	Protein change	Variant reads in cfDNA	Total reads in cfDNA	VAF (%) by cfDNA-seq	Tumour sequencing depth	Tumour sequencing VAF
7	*APC*	C1270W	0	1083	0.0%	413	64.4%
	APC	P1442_T144	0	2031	0.0%	272	26.1%
	KRAS	G12V	1	794	0.13%	452	68.8
	TP53	R342*	1	2143	0.05%	654	71.6%
8	*APC*	R1450*	1	3702	0.03%	1693	10.04%
	BRAF	V600E	2	3050	0.07%	1711	16.8%
	FBXW	R465C	0	2686	0.0%	1342	7.6%
	TP53	R282W	0	4456	0.0%	1031	11.2%
21	*CTNN*	G38S_del47	0	1363	0.0%	1074	5.4%
	KRAS	G12D	1	990	0.10%	860	40.2%
	TP53	D148fs	0	1396	0.0%	139	61.9%

Table 4.2: 11 discordant variants that were not called in cfDNA but had previously been detected in tumour tissue.
Read depth and number of consensus reads supporting each of the 11 discordant variants in cases 7, 8, and 21 are detailed.

In order to confirm that the failure to detect the known mutations was due to a very low tumour-derived cfDNA fraction, I applied highly sensitive ddPCR to case 8, from which sufficient cfDNA was remaining for orthogonal assessment of the mutation abundance. I used an established and validated ddPCR probe for detection of the *BRAF* V600E mutation and included case 3 as a positive control as a *BRAF* V600E mutation had been detected by cfDNA-seq with a VAF of 13.9%. The mutant allele fraction in this sample by ddPCR was 16.7%. A total of 21 ng from case 8 was used as input material, generating 2,830 wild-type droplets but no mutant droplets (Figure 4.2). The 95% confidence interval for a binomial distribution indicates that the true allele frequency is likely between 0% and 0.11%. This orthogonal validation confirms a very low tumour-derived cfDNA fraction and hence a biological cause rather than technical assay failure explained the inability to detect this variant by cfDNA-seq.

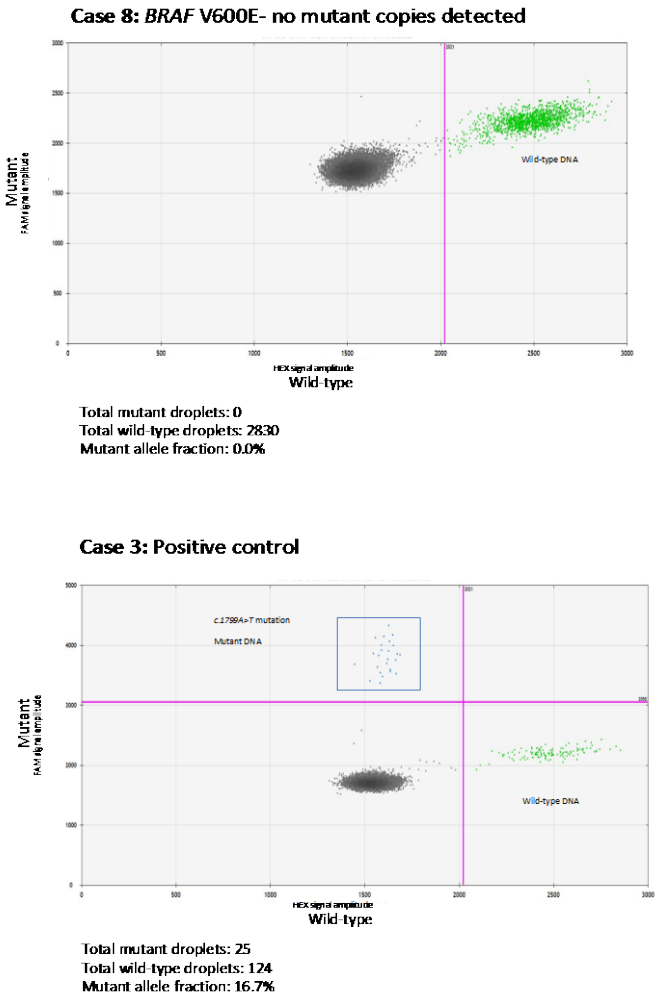


Figure 4.2: ddPCR to investigate *BRAF* mutation in a discordant call.
ddPCR of *BRAF* mutation was performed on case 8 to confirm absence of mutant DNA (top panel), pooled data from 7 wells each containing 3 ng cfDNA. Total wild-type droplets =2830, total mutant droplets =0. Case 3 (bottom panel) with known *BRAF* c.1799A>T (p.V600E) mutation confirmed by cfDNA-seq (VAF 13.9%) was used as a positive control. Total wild-type droplets =124, total mutant droplets =25, fractional allele frequency =16.7%. In each plot green dots represent wild-type DNA, blue dots represent mutant DNA and black dots represent droplets with no DNA incorporated.

I next assessed mutations called by cfDNA-seq in genes that had not been sequenced in the corresponding tumour tissue (highlighted in grey in Figure 4.1). cfDNA sequencing detected *APC* mutations in each of 4 cases whose tumours had only been analysed by the 5-gene amplicon panel that did not cover the *APC* gene. Furthermore, one mutation was found in each of *FBXW7*, *CTNNB1,*

TCF7L2, ATM and *SMAD4*, which were not present in the panel. The cfDNA-seq assay furthermore detected mutations in *APC*, *TP53* and *KRAS* in case 28 that had previously failed all tumour tissue sequencing attempts. 11 out of these 13 mutations (85%) encoded protein changes previously observed in cancer samples and reported in the COSMIC cancer mutation database [10]. Further review of the mutations identified in the tumour suppressor genes *APC* (case 24: E1379*, case 25: Q1429*, case 26: R1114* and L1289*, case 27: R1450*, case 28: R1450*), *FBXW7* (case 24: R658*), *SMAD4* (case 26: R361L) and *TP53* (case 28: Y234H) showed that all *APC* and *FBXW7* mutations were disrupting mutations as they were either premature stop codons, terminating the gene prematurely, or frame shift mutations. The mutation in *SMAD4* is an inactivating mutation in a hotspot region in the MH2 domain of the gene which is responsible for the transcriptional activation of the SMAD protein [202], hence resulting in loss of function of the gene [203]. The *TP53* Y234H mutation is a common hotspot mutation occurring in the core DNA binding domain, causing disruption of specific DNA binding and sequential transactivation [204] inferring this is a likely a driver mutation. This substantiates that the cfDNA-seq assay can detect biologically important cancer driver mutations directly from cfDNA *de novo*, without prior knowledge from tumour biopsy sequencing.

I next analysed mutations called in cfDNA that had not been detected when the same gene had been analysed in tumour tissue. 22 such mutations were detected (highlighted in blue or pink in Figure 4.1): seven in *TP53*, seven in *ATM*, three in *PIK3CA*, two in *SMAD4* and one each in *KRAS*, *FBXW7* and *TCF7L2*. The four mutations that were called in oncogenes (one in *KRAS*, three in *PIK3CA*) were canonical activating cancer driver mutations. The other 18 mutations were located in tumour suppressor genes and 8/18 (44%) of these encoded for amino

acid changes that had previously been detected in cancer according to the COSMIC database (outlined with a red border in Figure 4.1). One of these mutations was a premature stop codon (*ATM* R2993*) and the other seven were missense point mutations (case 3: *TP53* R175H, case 4: *TP53* H178P, case 9: *TP53* R273H, case 11: *TP53* C242S, case 12: *ATM* D126E, case 13: *TP53* F109S, case 24: *TP53* I254V). The remaining ten mutations are likely to be passenger mutations. Overall, this suggests that a considerable fraction of the mutations detected in cfDNA but absent in tumour are likely functionally relevant.

The VAFs of mutations called in cfDNA but not in matched tumour tissue were on average 125-fold lower than the VAF of the most abundant mutation detected in the same cfDNA sample (Figure 4.3). The dramatically lower VAFs suggest that most of these mutations would have to be confined to small subclones in these tumours.

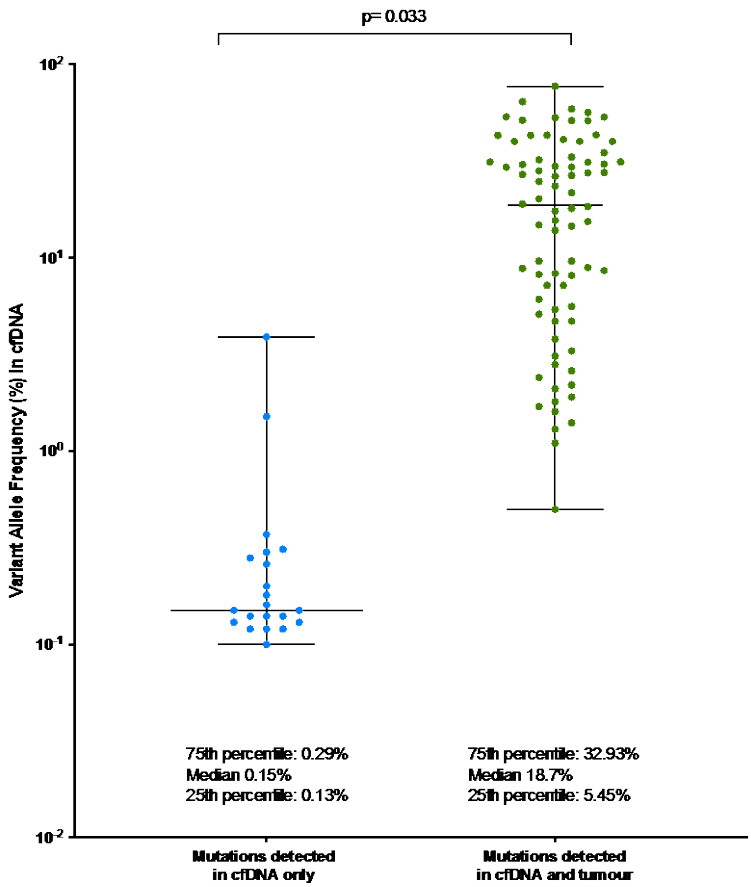


Figure 4.3: VAF comparison of variants called by cfDNA-seq only, with variants identified by tumour tissue and cfDNA-seq.
Blue dots: Variants detected by cfDNA-seq only; green dots: concordant variants detected by both tumour tissue and cfDNA-seq. The median is shown as a long horizontal line and whiskers indicate the range.

Similar to data showing that subclonal *APC*, *KRAS*, *NRAS* and *BRAF* mutations are extremely rare in pre-treated mCRCs [205], only a single divergent alteration between cfDNA and tumour tissue sequencing was detected among these genes: an activating mutation in *KRAS* gene (Q61H) had a VAF of 0.37% in cfDNA of case 10 and none of 633 reads from the matched tumour showed the mutation. Assessing the clinical data of our cohort showed that this was the only case that had been treated with anti-*EGFR* antibody cetuximab prior to blood collection. This subclonal *KRAS* mutation was a likely driver of acquired therapy resistance that had evolved during cetuximab therapy [201]. Orthogonal validation of this mutation by ddPCR (Table 4.4 and Figure 4.4), demonstrated that cfDNA-seq technology is suitable for the detection of clinically relevant subclonal alterations.

4.2.1 ddPCR for orthogonal validation of discordant mutations

ddPCR is the gold standard for point mutation detection with low variant frequencies (<1%) [206]. I performed ddPCR for orthogonal validation of discordant mutations that were called by cfDNA-seq but not called by tumour tissue sequencing (cfDNA+/tumour-). Of the 11 cases where such mutations had been identified with low frequency in cfDNA and were hence likely subclonal, sufficient cfDNA for ddPCR was available from four cases (4, 10, 14, 24). Nine discordant cfDNA+/tumour- variants from these cases were selected for independent validation (Table 4.3).

Case Number	Gene	cDNA base change	Protein change	VAF by cfDNA-seq
4	*ATM*	c.4213 T>C	p.L1405S	0.31%
	TP53	c.533 A>C	p.H178P	0.15%
10	*KRAS*	c.183 A>C	p.Q61H	0.37%
	ATM	c.1346G>C	p.G449A	0.14%
	TP53	c.1063 C>A	p.A355S	1.51%
14	*ATM*	c.8977 C>T	p.R2993*	0.28%
	PIK3CA	c.1633 G>A	p.E545K	0.30%
24	*CTNNB1*	c.2025 G>A	p.V676I	0.14%
	TP53	c.760 T>C	p.I254V	0.16%

Table 4.3: Nine discordant (cfDNA+/tumour-) low frequency variants selected for orthogonal validation with ddPCR

Pre-validated commercially available ddPCR SNP Genotyping Assays were used where available (for *KRAS* and *PIK3CA* variants), and the remaining probes were custom designed by ThermoFisher (Sequence designs are listed in methods section 2.2.5).

Input cfDNA quantity for ddPCR experiments varied, depending on the amount of remaining DNA available (17 ng for case 10, 20 ng for case 14 and >25 ng for all other cases). The reactions were partitioned into a median of 17,676 droplets per well. Overall, 7/9 ddPCR assays successfully detected and quantified the low frequency mutations (Figure 4.4). All 7 detected variants showed VAFs by ddPCR that were similar to those found by cfDNA-seq, confirming that these calls were reliable calls of true mutations.

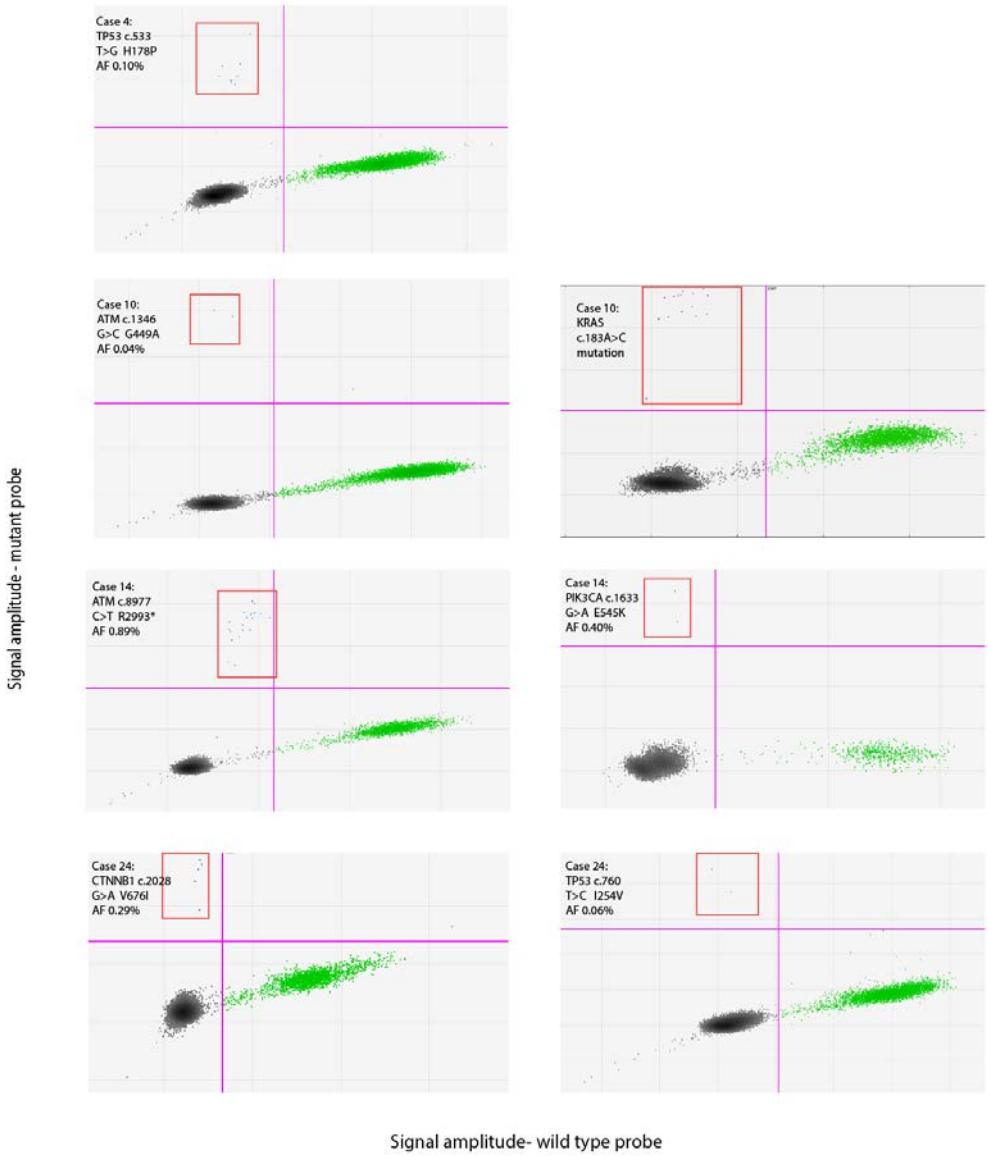


Figure 4.4: ddPCR validation of seven low frequency mutations detected in cfDNA but not in tumour tissue.

In each plot green dots represent wild-type DNA, blue dots represent mutant DNA and black dots represent droplets with no DNA incorporated.

Case	cfDNA conc (ng/ul)	Variant	VAF by cfDNA-seq	AF by ddPCR
4	98.96	TP53 H178P	0.15%	0.10%
	98.96	ATM L1405S	0.31%	Assay design
10	8.54	ATM G449A	0.14%	0.04%
	8.54	KRAS Q61H	0.37%	0.48%
	1.5	TP53 A355S	1.51%	Assay design
14	2.73	ATM R2993*	0.28%	0.89%
	2.73	PIK3CA E545K	0.30%	0.40%
24	16.76	CTNNB1 V676I	0.14%	0.29%
	16.76	TP53 I254V	0.16%	0.06%

Table 4.4: ddPCR validation of subclonal mutations detected in cfDNA but not in tumour tissue.

Two of the custom assays failed- *ATM* L1405S and *TP53* A355S. I performed multiple experiments with thermal gradient optimisations to improve annealing, yet the assays failed to detect wild-type and mutant DNA molecules. This indicated that either the custom self-designed probe was unsuccessful or that the genomic DNA (control sample) and the clinical cfDNA sample were both degraded, which is unlikely as the same control genomic DNA was used successfully for all other ddPCR experiments.

Overall, orthogonal validation with ddPCR demonstrated that cfDNA-seq assay can detect *de novo* subclonal variants that are not detectable in matched tumour tissue using conventional sequencing methods. These low frequency variants are a feature of intratumoural heterogeneity [160, 162] and may represent the evolution of drug resistance [176] for example in case 10 where *KRAS* Q61H mutation was detected, a known driver of anti-*EGFR* resistance [201], eight months following disease progression on cetuximab therapy. cfDNA-seq also identified two distinct *PIK3CA* mutations in the same patient (case 17 N345K and H1047R), indicative of parallel evolution which is as signal of Darwinian selection [89]. Only one of these mutations had been detected in the baseline tumour

biopsy and the other may have either been present in a clone that was not sampled or emerged as an acquired mutation during systemic therapy. Both mutations are activating mutations with prior studies showing a functional role in targeted therapy resistance [207]. *PIK3CA* H1047R furthermore has a role in promoting epithelial-mesenchymal transition and cell invasion [208] which is a necessary step for tumour metastasis.

4.2.2 Validation of variant allele frequencies of mutations detected by cfDNA-seq

I wanted to compare variant frequencies obtained with cfDNA-seq with ddPCR as an independent method to validate the reliability of the VAFs. Validation of hotspot *KRAS* variants were chosen as they have a high rate of mutation in mCRCs (15/28 patients in this cohort) and off-the-shelf commercial ddPCR probes were readily available.

The VAFs determined by cfDNA-seq and ddPCR were compared in nine of the 15 *KRAS* mutant cases where sufficient cfDNA was remaining, and in addition I included the 7 subclonal mutations described above plus the control *BRAF* V600E ddPCR result for this analysis (Figure 4.5). Linear regression analysis showed an excellent concordance between VAFs measured by cfDNA-seq and ddPCR (R^2=0.98). This illustrates that the abundance of mutant DNA detected by cfDNA-seq was highly consistent with that obtained by ddPCR. Of note, ddPCR was able to detect the *KRAS* G12V variant (with an allele frequency of 0.16%- 1 mutant molecule in 623 wild type molecules), which was a false negative result from cfDNA-seq in case 7.

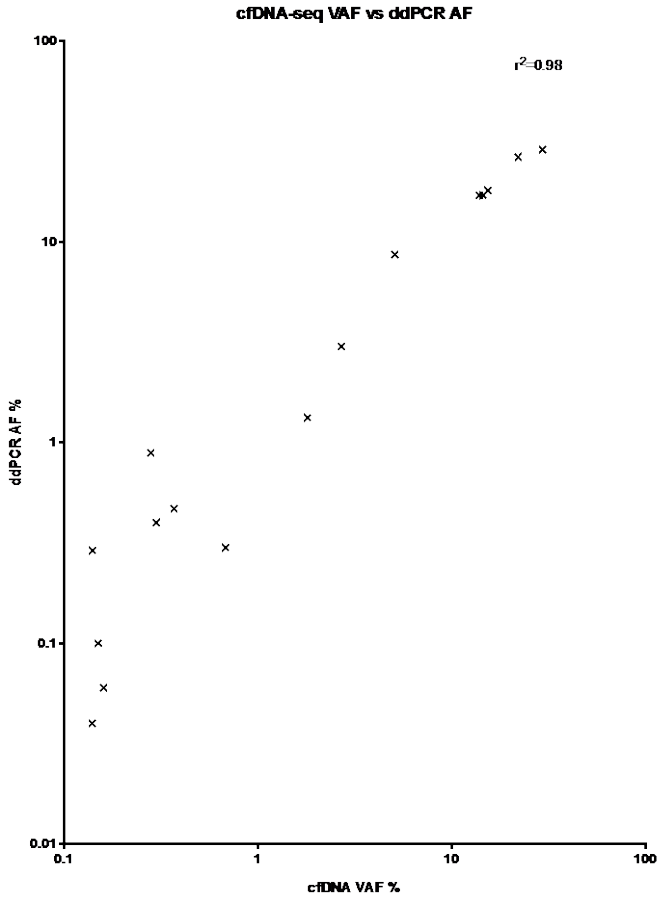


Figure 4.5: Concordance of mutation allele frequencies between cfDNA-seq and ddPCR. The detected allele frequency between ddPCR and cfDNA-seq shows an excellent correlation with based on the Pearson correlation coefficient r2=0.98, p=<0.0001.

This concordance analysis was limited to testing known mutations detected by cfDNA-seq but does not include analysis of mutations that may have been missed by cfDNA-seq.

4.3 cfDNA-seq mutation comparisons with matched tumour and blood genomic DNA

During comparison of cfDNA-seq with tumour tissue sequencing and matched genomic DNA extracted from blood cells that had been sequenced with the FOrMAT panel, I noticed that two out of seven *TP53* mutations detected in cfDNA but not in tumour tissue (cases 9 and 13), were present in matched normal germline DNA from peripheral blood with VAFs of 2.1% and 7.8%, respectively.

TP53 mutations in cfDNA do not necessarily originate from the tumour itself but can be the result of a clonal expansion of blood cells [209], termed clonal haematopoiesis [210, 211]. Clonal haematopoiesis describes somatic acquisition of genomic alterations in haematopoietic stem cells leading to clonal expansion that may progress to malignant or symptomatic disease [210-212]. Clonal haematopoiesis is rare in individuals under 40 years of age and the prevalence rises with age and is reported to occur at a frequency of 10% in over 70 year olds. Both clonal haematopoiesis events in this cohort were noted in patients aged 70, older than the cohort median of 62.5 years. The most common genes affected by clonal haematopoiesis are *DNMT3A*, *ASXL1*, *TET2*, *PPMD1*, *JAK2* and *TP53*, of which *TP53* was the only gene in our panel [213].

The observation of possible clonal haematopoiesis in two out of 28 (7%) cases in this cohort demonstrates the importance of sequencing matched normal germline DNA from peripheral blood to avoid interpreting such variants as cancer associated mutations. Although *TP53* mutation status is currently not used for treatment/prognostic decision making in current clinical practice, stratification by *TP53* mutation status is being used for early phase clinical trials (NCT03149679 and NCT03144804), hence accurate identification of the origin of *TP53* variants is imperative.

4.4 Orthogonal validation with a commercially available high sensitivity cfDNA assay (AVENIO)

In order to compare the performance of our cfDNA-seq assay against a commercial targeted cfDNA sequencing assay that had been launched in 2018, cases with remaining cfDNA that harboured mutations with VAFs <2% (cases 3, 15 and 23) were selected for re-sequencing with the Roche AVENIO Expanded

ctDNA sequencing kit by Dr. Mike Hubank's laboratory in the Centre for Molecular Pathology, ICR Sutton. This technique employs integrated digital error suppression combining molecular barcodes with *in silico* error suppression techniques [195]. 25 ng, 17 ng and 25 ng were used as input cfDNA amounts, respectively. Libraries were generated as per the manufacturer's protocol and sequenced on an Illumina NextSeq500. Samples were sequenced to depths between 2689X and 6420X after duplicate removal.

Ten out of 13 (77%) mutations in genes targeted by both panels (Table 4.4) were called by both cfDNA-seq and the AVENIO assay. Three frame shift mutations detected by cfDNA-seq in *APC* (in cases 3 and 23), with VAFs of 27%, 3.1% and 3.3% were not detected by the AVENIO assay, despite a reported limit of detection of 1% for insertions and deletions (indels) for the commercial assay [195]. The highest frequency *APC* S1465fs variant was visible by manual review of the BAM files but not called, although the two lower frequency variants were not detectable in the BAM files. Further, the low frequency *TP53* R175H variant in case 3 was not called by the AVENIO software, but was seen to be present upon manual review of the BAM file at a frequency of 0.17% which is below the reported limit of detection of 0.5% for single nucleotide variants. All mutations called by the AVENIO assay were also detected by our bespoke assay.

Case Number	Gene	Variant	cfDNA seq VAF (%)	AVENIO variant reads/total reads	AVENIO seq VAF (%)
3	*APC*	S1465fs	26.99	not called	
	ATM	D2795N	0.20	not covered	
	BRAF	V600E	13.91	1214/7264	16.71
	PTEN	P95L	8.19	779/5665	13.75
	TCF7L2	P272T	0.15	not covered	
	TP53	R175H	0.13	7/4001	0.17
15	*APC*	Q1367*	2.4	43/2277	1.89
	KRAS	G12D	1.7	45/2599	1.73
	TCF7L2	G26fs	1.34	not covered	
	TP53	R342*	2.6	46/1973	2.33
	APC	Q1367*	2.4	43/2277	1.89
23	*APC*	E1554ins	3.30	not called	
	APC	K1462ins	3.13	not called	
	KRAS	Q61H	5.44	188/4704	4
	TP53	R273C	3.81	159/3726	4.27
	TP53	C238F	1.14	97/9312	1.04

Table 4.5: Comparison of cfDNA sequencing of 3 cases by AVENIO ctDNA Expanded Kit with cfDNA-seq

This concordance analysis of our customisable cfDNA-seq assay against a validated and commercially available assay demonstrates a superior performance of our assay for which I showed a sensitivity for point mutation detection down to a VAF of 0.10%. Furthermore, cfDNA-seq was able to call three indels, all of which remained undetected by the AVENIO assay despite of good read depths at these positions (median 4872X).

4.5 Genome wide DNA copy number aberration analysis from targeted cfDNA-seq

Cancer genomic aberrations are not confined to mutations but also include structural DNA aberrations and somatic copy number aberrations (CNAs). CNAs are defined as gains (amplifications) or losses (deletions) of whole chromosomes, chromosome arms, or fragments from the normal number of two copies per cell. Specific CNAs may be characteristic of different tumour types and are important to analyse during genomic profiling as some CNAs are drivers for instance *TP53* loss is associated with formation of CRC [214]. CNAs can also be associated with development of resistance to therapy and hence contribute to tumour evolution, for instance amplifications in *KRAS*, *ERBB2*, *MET* or *FGF10* can play a role in the development of resistance to anti-EGFR therapy [215]. Other key CNAs such as *ERBB2*/*HER2* can targetable in mCRC clinical trials. Moreover, copy number profiling is important for accurate subclonality analyses as CNAs and subclonality both influence the VAF of a mutation, hence subclonality cannot be determined without knowing the local copy number state.

CNAs have historically been analysed through array comparative genomic hybridization [216], although more recently improved sensitivity has been achieved through whole genome or exome based sequencing. I aimed to investigate if the genomic characterisation of individual cancers could be maximised from a single assay by reconstructing genome-wide CNA profiles from cfDNA-seq. Dr Louise Barber applied the CNVkit [154] pipeline for CNA detection which utilises both on-target reads and the non-specifically captured off-target sequencing reads for copy number calling across the genome. This toolkit was initially developed using targeted sequencing cell line data and has subsequently

been applied to targeted sequencing data from melanoma tumour tissue, however until now has not been applied to cfDNA sequencing data.

The sequencing BAM files are input into the CNVkit programme, along with reference BAM files from sequencing of six healthy donors with the same assay. Log_2 copy ratios are then calculated across the genome from both the on-target reads and the non-specifically captured off-target reads, which are constructed into target bins and off-target bins. Bins are individual segments of data used to plot copy number signal, using the reference file to estimate the expected read depth of each on- and off-target bin across the control samples. At each bin, the read depths from the reference file are used to normalise and correct for biases in the bin coverage in order to assign corrected copy number ratios to the samples. Narrower bins are required to assess focal genomic amplifications, whereas large bins assess copy number states of whole chromosomes or chromosome segments [154]. 30 kb bins were used to generate genome-wide scatter plots (Figure 4.6A) per sample or heatmaps of multiple samples (Figure 4.6).

The median off-target read fraction in the cfDNA-seq data from the 28 patients was 43.5%. However only 20 cases (71%) revealed informative CNA heatmaps with appreciable copy number changes. The eight cases where CNA profiling was not possible had a median off-target rate of 48.8%, which is above the cohort median, and therefore this is unlikely to be the reason for the inability to generate a CNA profile. On further review, these eight samples had maximum mutation VAFs ≤5.6 % (median VAF 1.8%) and thus a low tumour content was the most probable reason for the failure to define CNA profiles. CNVkit analysis was

possible for all cases where the maximum mutation VAFs was ≥8.6% (examples in Figure 4.6A).

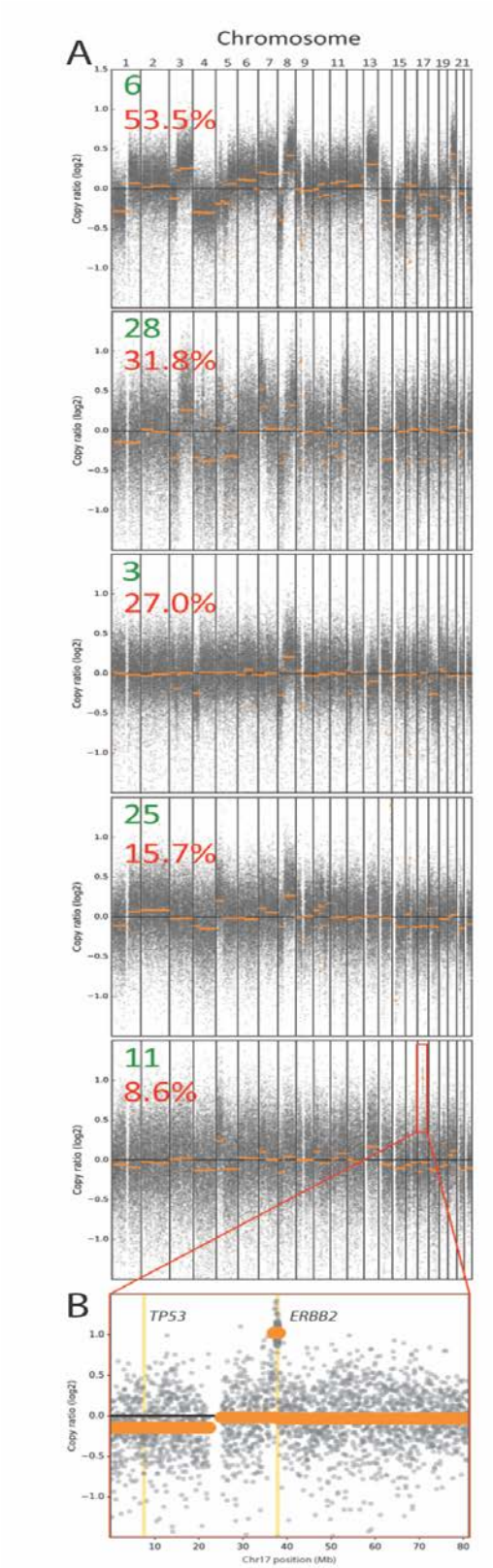


Figure 4.6:Genome wide copy number aberrations can be detected from targeted cfDNA-seq using CNVkit.
(A): Representative genome wide log copy ratio scatter plots for five cases (green number) in our cohort with tumour content ranging from 53.5% to 8.6% (red number indicates max VAF) are shown. (B) Focused log copy ratio plot of chromosome 17 for case 11 which had a high level amplification of ERBB2.

I next wanted to assess whether any actionable somatic copy number aberrations (SCNAs) were detected in those samples where tumour content was low. A high-level amplification was detected in case 11 despite a low fraction of tumour-derived cfDNA (8.6%), this is a potentially targetable amplification involving the *ERBB2* oncogene (Figure 4.6, panel B), which has been described in CRC, breast and gastro-oesophageal tumours [19]. This amplification had also been detected in the matched tumour, validating the ability to profile CNA with our targeted cfDNA-seq technology to detect actionable copy number aberrations. No other focal amplifications were detected in the tumour biopsies with the FOrMAT NGS panel and no further targetable focal amplifications were apparent in the remaining cfDNA samples.

Upon review of the heatmaps for the cohort, the most frequent gene copy loss was seen in chromosome arms 17p (13 of 20, 65%) and 18q (14 of 20, 70%), which are regions harbouring the key tumour suppressor genes *TP53* and *SMAD4*, respectively. This is consistent with TCGA data on non-hypermutated tumours where *TP53* loss occurred in 56% of tumours and *SMAD4* loss in 66% [19]. There were several chromosome arms with copy number gains: 7p (14 of 20, 70%), 8q (16 of 20, 80%) which contains the *MYC* gene, 13q (12 of 20, 60%) which harbours the mCRC candidate oncogene *CDK8*, 20p (9 of 20, 45%) and 20q (16 of 20, 80%) were present, which are frequently detected in CRC (Figure 4.7) [19, 177, 217].

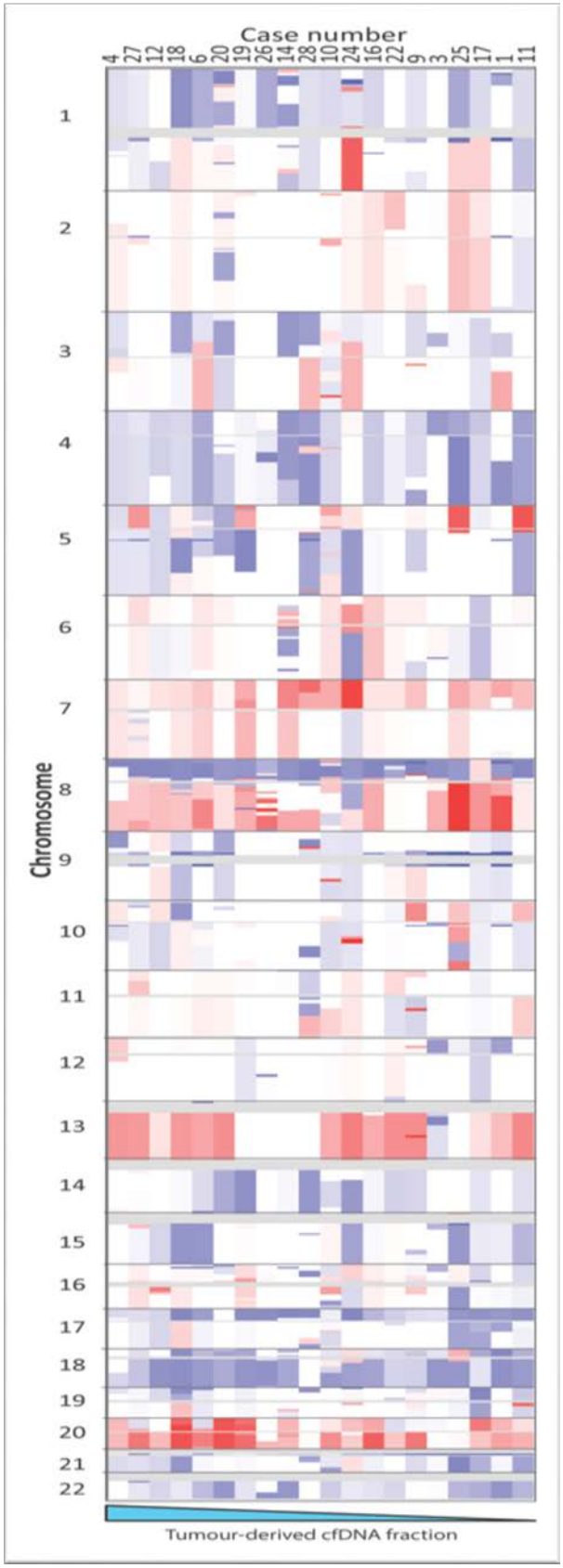


Figure 4.7: Genome wide heatmap of segmented copy number raw log ratio data after amplitude normalisation.
Gains are red and losses are blue. Profiles are ordered (left to right) from highest to lowest tumour content (based on maximum VAF) for all 20 cases that had a visible CNA profile.

4.6 Validation of SCNAs by low coverage whole genome sequencing (lcWGS)

To validate the use of off-target reads from targeted cfDNA sequencing for the reconstruction of genome-wide CNA profiles, I analysed 18 of these cases that had sufficient cfDNA material remaining, by low coverage whole genome sequencing (lcWGS). lcWGS is an established and cost effective approach for genome wide copy number profiling [218]. Libraries were sequenced to a median depth of 0.42X i.e. only approximately half of the genome is covered by sequencing reads. Copy number profiles were generated (by Dr Louise Barber) with 50 kb bins for broad assessment of genome wide CNA. Copy number profiles visually showed high concordance across these 18 cfDNA samples (Figure 4.8). This independent validation showed a median Spearman correlation of 0.886 with the profiles generated from cfDNA-seq using CNVkit (by Dr Louise Barber), although this is limited by tumour content of >8.6%. The tumour content threshold for lcWGS is reportedly similar at 5-10% [219].

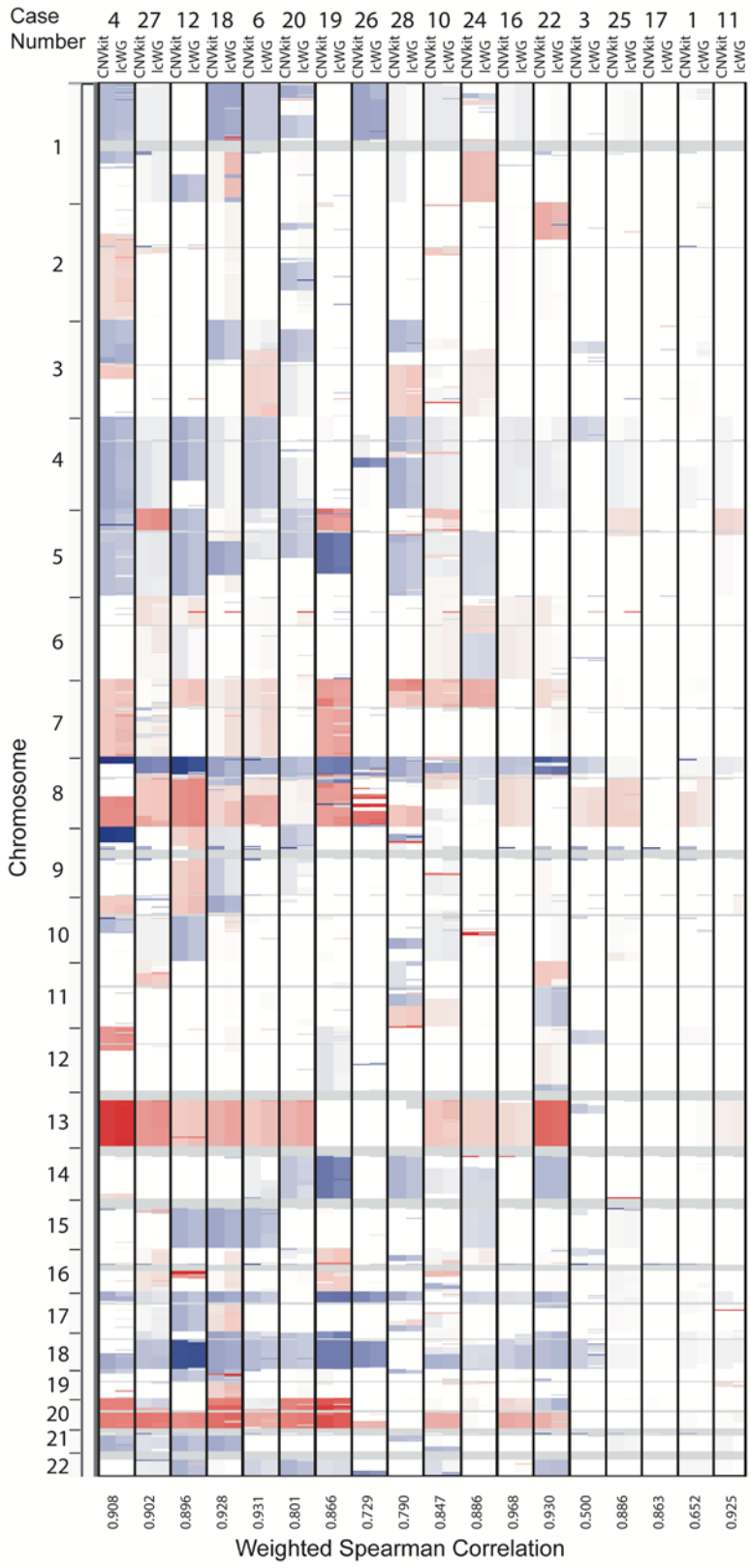


Figure 4.8: Comparison of genome wide heatmap of segmented copy number data using off-target reads from cfDNA-seq (CNVkit) and lcWGS.

lcWGS copy number profiling shows good correlation with profiles generated from cfDNA-seq using CNVkit. Case numbers are labelled at the top with matched CNVkit and lcWGS heatmaps aligned sided by side for each case.

Though lcWGS is a low cost method of obtaining CNA data, the cfDNA sample requires a different library preparation method, whereas CNVkit utilises sequencing data from the same target cfDNA-seq assay. Furthermore, lcWGS missed focal losses e.g. in case 27 chromosome 1 and 2, and generated some gains in chromosome 6p in most samples, though this was not present in CNVkit data.

4.7 Correlation of plasma cfDNA concentration and tumour derived cfDNA with clinical features in patients enrolled in the FOrMAT clinical trial

Clinical correlation was initially performed using data from all 58 patients from whom cfDNA was originally extracted, further analysis was performed on the 28 consecutive cases which were sequenced. Consecutive cases were selected for sequencing, unless there was <25 ng cfDNA available. I first investigated the correlation of cfDNA concentration from 58 patients with clinicopathological characteristics, including CEA secretor status (patients with a CEA level greater than the upper limit of normal by the lab reference range of 5 μg/ml were classified as CEA secretors), sites of metastatic disease, tumour histology and number of prior lines of systemic therapy (Table 4.6).

Variable	Patients (n)	Average cfDNA concentration (ng/ml) +/- SD	P-value (unpaired T-test)	Fold change
CEA status				
- **Secretor** **- Non-secretor**	47 11	68.35+/-127.5 12.8 +/- 14.9	0.16 (95% CI 30.9 – 105.8)	5.33
Liver metastases				
- Present - Absent	44 14	71.1+/-131.4 16.10 +/- 15.6	0.12	4.42
Lymph Node metastases				
- Present - Absent	39 19	68.17 +/- 138.12 35.59 +/- 47.3	0.34	1.86
Pulmonary metastases				
- Present - Absent	29 29	62.45 +/-141.87 53.20 +/- 87.26	0.77	1.17
Number of sites involved				
- ≥3 - <3	46 12	66.78 +/- 128.11 23.50 +/- 22.28	0.26 (95% CI 105.8 – 30.9)	2.84
Previous treatment				
- Treatment naive - Prior treatment exposure	46 12	63.30 +/- 103.69 55.36 +/- 123.46	0.81	1.14
Differentiation*				
- Moderately **-** Poorly	42 15	46.68 +/- 48.95 39.80 +/- 104.31	0.76	1.72
Location of primary**				
- Right sided - Left sided	29 28	78.01 +/- 159.21 37.44 +/- 40.66	0.2	2.08

Table 4.6: Average plasma cfDNA concentrations obtained from 58 patients with mCRC and correlation to independent clinicopathological features.
* Tumour differentiation not recorded for all patients. ** One patient had synchronous right and left sided primary tumours, hence excluded

Although no statistically significant difference or correlation was observed with any clinical characteristics, the largest numerical changes were observed in patients who secreted CEA- 5.3 fold higher yield than those who were CEA non-secretors (average: 68.35 ng/ml vs. 12.8 ng/ml plasma, p=0.16, 95% CI 30.9-105.8, correlation coefficient of r= 0.02) and those who had liver metastatic disease- 4.4 fold higher yield than those who did not have liver metastases

(average: 71.1 ng/ml vs. 16.1 ng/ml plasma, p=0.12, correlation coefficient of r= 0.20).

Based on published data, several mCRC features are associated with greater ctDNA concentration, including stage IV disease, higher CEA levels, greater tumour burden and moderately differentiated histology [220-223]. A higher ctDNA concentration is associated with improved concordance of cfDNA mutation calls with biopsy tissue calls [96], providing a rationale for pre-selection of patients who are particularly suitable for genotyping by cfDNA sequencing based on biomarkers that indicate high mutant ctDNA fraction. I investigated whether ctDNA concentration correlated with any of these features. Of the 28 cases that underwent cfDNA-seq*,* I used the variant frequency of predicted truncal *TP53* or *BRAF* somatic mutations [205] to estimate ctDNA concentration by multiplying VAF of *BRAF*/*TP53* mutation by cfDNA concentration. No statistically significant association was observed between the estimated mutant ctDNA concentration and candidate clinical predictors (Table 4.7). Though the largest numeric difference was seen with CEA secretors and non-secretors, the imbalanced subgroup size (27 vs. 1) makes this comparison futile. The next largest fold change was observed in histological differentiation, where the average ctDNA concentration was 12.7 fold higher in patients with moderately differentiated tumours than in patients with poorly differentiated tumours (average 22.11 ng/ml vs. 1.74 ng/ml plasma, p=0.37). Many of the subgroup sizes, including those with poorly differentiated tumours, were very small thus precluding statistically significant results.

Variable	Patients (n)	Estimated mutant ctDNA concentration (ng/ml) +/- SD	P-value (unpaired T-test)	Fold change
CEA status				
- Secretor - Non-secretor	27 1	20.67+/-37.1 0 +/- n/a	n/a	20.67
Liver metastases				
- Present - Absent	25 3	20.85+/-38.3 12.21 +/- 20.5	0.71	1.71
Lymph Node metastases				
- Present - Absent	17 11	26.06 +/- 45.48 10.45 +/- 12.37	0.29	2.49
Pulmonary metastases				
- Present - Absent	15 13	15.43 +/-17.01 25.13 +/- 51.27	0.49	0.61
Number of sites involved				
- ≥4 - <4	16 12	27.59 +/- 45.90 10.50 +/- 15.31	0.25	2.62
Previous treatment				
- Treatment naive - Prior treatment exposure	10 18	32.47 +/- 56.84 12.96 +/- 16.67	0.18	2.5
Differentiation				
- Poorly - Moderately	3 25	1.74 +/- 3.02 22.11 +/- 38.26	0.37	12.69
Location of primary*				
- Left sided - Right sided	15 12	15.24 +/- 19.27 26.20 +/- 52.42	0.46	1.72

Table 4.7: Correlation of estimated mutant ctDNA concentration and clinicopathological features in patients studied in this cohort.
Note imbalanced group sizes in CEA secretor status and differentiation are likely responsible for the large numerical differences in fold change but not statistically significant results. *One patient had synchronous right and left sided primary tumours hence excluded from analysis.

An important observation in the three patients that had no tumour derived DNA (cases 7, 8 and 21) was that they all had their primary tumour resected and all three patients had <4 sites of metastatic disease, specifically absent pulmonary and bilobar liver metastasis, although only case 7 had single liver lobe

involvement. Cases 8 and 21 had oligometastatic disease limited to the peritoneum, muscle and lymph nodes. By linear regression of the entire cohort these were the two factors that had the highest correlation coefficient of r =0.26 for number of sites of metastatic disease and r =0.21 for primary disease in situ. This implies that the presence of primary lesion *in situ* and metastases in a higher number of well-perfused organs e.g. the lungs and liver are likely to contribute to the ctDNA load through dissemination of tumour DNA into the plasma.

4.8 Correlation of tumour derived cfDNA content with survival

I next investigated the utility of estimated somatic mutation burden (defined as median mutant allele frequency of a truncal mutation) in predicting progression free (PFS) and overall survival (OS). Time to tumour progression and time to death were calculated as months from the date of initiating the current line of therapy to the date of event. Patients with a high somatic mutant allele frequency (defined as greater than the median VAF of a truncal mutation of 18%), had a significantly shorter median OS (3.5 months) than patients with a lower mutant allele frequency (11 months p=0.0143) (Figure 4.9). This was not significant when evaluating PFS.

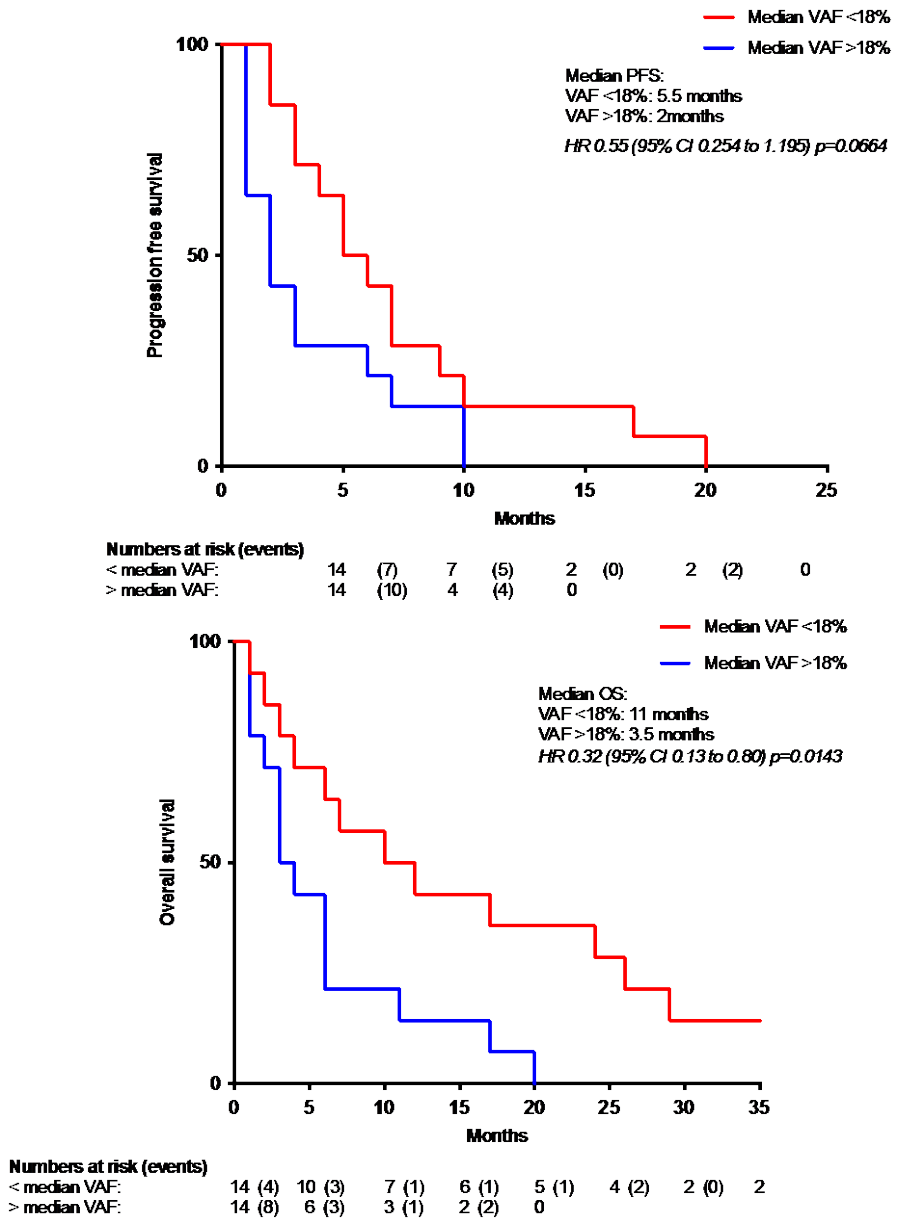


Figure 4.9: Kaplan Meier curve depicting impact of variant frequency of somatic mutations detected by cfDNA-seq on PFS and OS.

The cohort median VAF (18%) was used to define groups. Groups were compared using the log-rank test.

Tumour mutation VAF analysis may be associated with OS in mCRC whereby patients with low somatic mutation burden have a significantly greater OS than those with a high somatic mutation burden (11 months vs. 3.5 months, p=0.01), irrespective of prior and current treatments received [224]. This should be confirmed in a larger cohort by multivariable analysis.

4.9 Summary and Discussion

This chapter describes cfDNA detection and feasibility of our novel cfDNA-seq assay in mCRC patients. I applied this technology for sequencing analysis of cfDNA from a cohort of FOrMAT mCRC patients. This demonstrated that cfDNA-seq captures the overall genomic landscape, calling 88% of the cancer mutations that were detected by clinical grade tumour tissue sequencing. Small indels can be more difficult to call than point mutations, particularly when present at low abundance. cfDNA-seq assay called 23 of 26 indels (88.5%) that were known based on tumour tissue sequencing, showing a similar performance to point mutation detection (58 of 66 called; 87.9%). Further analysis using ddPCR of cases where mutations were not called revealed that the reason for the false-negative calls was due to low tumour content within the cfDNA sample, rather than technical failure of the assay. This sensitivity is similar to that reported for a 5-gene amplicon cfDNA assay applied to CRC patients (87.2%) [96] and the commercially available Guardant 360 54-gene assay applied to multiple tumour types including CRC (85%) [116, 225]. Although the Guardant 360 assay contains a larger number of genes, the assay is not customisable hence additional genes of interest cannot be added for research or clinical use in other tumour types. I tested the sensitivity of the cfDNA-seq assay against the commercially available AVENIO ctDNA 17-gene target kit. This is marketed as a highly sensitive assay with >99% sensitivity for detection of mutations down to 0.5% VAF. The cfDNA-seq assay I developed detected all point mutations found by the AVENIO assay in three cases, however cfDNA-seq is superior to the AVENIO kit which failed to call any of the known indels with VAFs up to 26%.

Besides sensitivity, the specificity of results posed some additional challenges. cfDNA-seq detected several additional cancer driver mutations not reported by

tumour sequencing. Eight of these variants were observed in *TP53*, all but one of which were reported in COSMIC as being mutated in CRCs. Two of these were also observed in the matched blood samples, suggesting that they arise as a consequence of clonal haematopoiesis. The discovery of clonal haematopoiesis in 7% (2 of 28) of the cohort demonstrates the importance of sequencing DNA from blood cells in parallel with cfDNA, in order to assess whether mutations in cfDNA are of cancer origin. One of the eight *TP53* mutations was detected in a case where tumour tissue had failed sequencing, this is a recognised driver mutation and hence unlikely to be a false positive cfDNA call. Three of the remaining 5 cfDNA+/tumour- *TP53* mutations had been validated orthogonally using either ddPCR or the AVENIO assay, concluding that they are true mutations rather than PCR or sequencing errors. The remaining 2 could not be validated orthogonally due to ddPCR assay design failure and insufficient cfDNA material available, thus the source of these 2 mutations remains undetermined.

In 39% (11 of 28) of cases cfDNA-seq detected mutations that were not found in archival tumour tissue. These cfDNA+/tumour- discordant calls can result from tumour heterogeneity or technical factors such as differing read depths. One patient who had previously been treated and subsequently progressed on cetuximab therapy, cfDNA-seq demonstrated a *KRAS* Q61H variant that was absent from the matched tumour. It is likely that this represents a drug resistant subclonal mutation that evolved during cetuximab therapy and remained detectable at a low level eight months after cetuximab exposure. Two activating *PIK3CA* mutations were detected in each of two cases (case 17 H1047R, VAF= 0.13% and case 26 R1023*, VAF= 0.12% both reported in COSMIC cancer database) had subclonal VAFs compared to likely truncal mutations in the same sample. These are likely to represent parallel evolution events. Though parallel

evolution commonly occurs in response to selective pressure by targeted therapies [201, 226], neither of the two patients received treatment with targeted agents. This suggests that cfDNA-seq detected distinct somatic mutations which were present in separate tumour regions and thus missed by the tumour biopsy due to spatial or temporal heterogeneity or differing read depths. Overall, mutations that were only called in cfDNA but absent in tissue sequencing had significantly lower VAFs than concordant mutations in the same sample. The subclonal VAFs supports the notion that many of these are evolving mutations that could not be detected by the single time point tumour biopsies. To confirm the absence of these subclonal mutations from tumour biopsy tissue, deeper sequencing would be required to reach the same depths that cfDNA-seq reached, or gold standard ddPCR could be run on the samples to investigate the presence/absence of these specific variants in tumour tissue.

In 11% (3 of 28) cases, mutations present within tumour tissue were not detected by cfDNA-seq. One potential cause for tumour tissue:cfDNA discordance is the variable shed of tumour DNA into the plasma causing false negative tumour genotyping from cfDNA. This is supported by evidence from ddPCR analysis which confirmed an absence of *BRAF* V600E mutant alleles, concluding that there was an absence of detectable tumour derived DNA within the cfDNA. One possible reason for this is the potential effect of tumour burden on the ability to detect mutations in ctDNA, such that all 3 patients had <4 sites of metastatic disease, where majority of the patients in our cohort had ≥4 sites of disease, suggesting that a lower tumour burden may contribute to a higher likelihood of discordant ctDNA calls as seen in previous studies of matched tumour/ctDNA analysis [227]. Furthermore each of these cases had at least one mutation in *TP53* and *BRAF*, which are usually acquired as truncal events in CRCs and the

mutational status of these truncal variants are highly stable throughout the course of this the disease [228], thus it is highly unlikely that the mutant clones detected in tumour tissue sequencing had been lost.

Together, these data show that this ultrasensitive cfDNA sequencing technology can address the subset of 20% or more of CRC patients who cannot be molecularly profiled due to unobtainable or inadequate biopsy tissue [168, 229]. This is in addition to the ~2% of patients that require repeat invasive tumour tissue biopsies to undergo tumour genotyping [151] and cfDNA-seq can spare these patients intervention that can be associated with inherent complication risks. As the number of targeted therapies increases, small custom target enrichment panels which can be readily adapted and scaled for the tumour type and therapeutic agent in question, could be used to investigate the genomic landscape. This would facilitate both the identification of novel resistance mechanisms and help to further characterise cancer evolution.

While copy number calling did not reveal aberrations in patients with low tumour derived cfDNA, the copy number landscape in patients with tumour variants of VAF ≥8.6% revealed characteristic losses of mCRC, including recurrent chromosome arm losses of 17p and 18q, regions where tumour suppressor genes *TP53* and *SMAD4* are located, and a potentially actionable *ERBB2* amplification.

Tumour derived cfDNA can be used as a surrogate marker of tumour burden, such that patients with metastatic disease have a higher ctDNA fraction than those with localised disease [105, 148]. In our cohort, ctDNA concentrations estimated by the variant frequency of a truncal mutation varied from 0 to 189.38 ng/ml plasma. There tended to be a 2.5 fold higher ctDNA content in patients who

were treatment naïve compared to those who had prior treatment exposure (32.47 ng/ml vs. 12.96 ng/ml respectively, p=0.18). This is thought to be because exposure to systemic therapy induces clonal selection, causing the ratio of wild-type and mutated DNA to change over time [176]. Therefore following first line treatment in those who respond, the decrease in ctDNA load is likely to result potentially from reduction in tumour burden through selective pressure of systemic therapy and persistence/progression of only chemotherapy resistant tumour cells. Analysis of serial cfDNA samples should confirm this as disease burden may be a potential confounding factor.

Consistent with reports from lung cancer, melanoma and gastrointestinal cancers [230, 231] our data, from a pooled analysis of 28 patients, confirms that low mutation frequency detected by ctDNA is associated with a longer overall survival and a non-significant trend towards longer time to treatment failure. This is thought to be because patients with a higher tumour burden are likely to have more active or resistant disease shedding tumour derived cfDNA into the blood. This analysis includes pre-treated patients, some of which have had good responses to prior systemic therapy, with subsequent progression. Clinically this correlates with an improved prognosis which is in keeping with the prognostic value of lower median amount of mutated cfDNA and prognosis. The prognostic value of cfDNA could be utilised adjunctively in the clinical setting to improve decision making for treatment initiation, whereby risk of progression may be discernible by the analysis of tumour-derived cfDNA. For example, where patients with a low mutation frequency are predicted to have a longer PFS and OS, as per our cohort, they may benefit from longer time off treatment. As there are only a limited number of lines of systemic therapy available for mCRC, this could avoid patients developing resistance to treatment early on.

Furthermore with this cfDNA-seq technology the mutation allele frequency can be monitored with longitudinal cfDNA sampling for early detection of resistance acquisition, which is reported to precede radiological disease progression by six months or more [232].

In conclusion, this cfDNA sequencing approach with customisable and off-the-shelf reagents showed a similar or better sensitivity as published cfDNA analysis methods such as 5-gene assay using amplicons with a sensitivity of 87.2% [96], NGS assay CAPP-seq with a sensitivity 92% [117] and the 54-gene target capture cfDNA assay (Guardant 360) with a sensitivity of 85% for mutation detection [116]. Our assay has important advantages over other targeted cfDNA-seq techniques such as CAPP-Seq, which has recently been marketed as AVENIO ctDNA assay, in that it doesn't require complex background digital error correction models which is only possible with large numbers of healthy donors. This can prove impractical for applications requiring frequently changing custom gene panels. The Guardant 360 assay does not simultaneously assess genome wide CNAs, furthermore it utilises bespoke reagents and is hence not customisable. Through analysis of the FOrMAT clinical trial cohort of matched tumour tissue and cfDNA, I have shown that cfDNA-seq assay has the ability to genotype mCRCs where prior tumour tissue sequencing has failed. Performance from small cfDNA quantities that can be obtained from most patients with mCRC enables minimally invasive longitudinal studies for monitoring and detection of evolving mutations and CNAs simultaneously from a single assay.

CHAPTER 5: CFDNA-SEQ AS A TOOL TO INVESTIGATE THE GENETICS OF CHEMOTHERAPY-RESISTANT MCRCS

5.1 Introduction

Cytotoxic chemotherapy remains the most effective treatment option in mCRC. The mainstay of treatment includes combination of 5FU (5-fluorouracil) and leucovorin with oxaliplatin (FOLFOX) or irinotecan (FOLFIRI). However, nearly tumours eventually acquire resistance to chemotherapy. Under selective pressure of treatment, cancers can evolve over time, gaining and losing genetic alterations including mutations and gene copy number variations such as the amplification of oncogenes or deletion of tumour suppressor genes. To date, the genetic alterations and molecular mechanisms that confer acquired resistance to cytotoxic chemotherapy in mCRC remain largely unknown and their identification is an ongoing priority in my host lab. Exome sequencing of biopsies from 55 mCRCs that were refractory to 5FU, irinotecan and oxaliplatin had identified mutations in the DNA damage repair (DDR) genes *FAMCM*, *RAD54L*, *ERCC6* and *REV3L* (unpublished data) each occurring in 2-4 progression biopsies and were not present in the primary tumour. Furthermore the DDR genes *ERCC3* and *ERCC4* harboured one mutation each, however no obvious driver aberrations have been identified in any one gene. DDR genes are integral to maintaining genome integrity. DNA repair pathways can enable tumour cells to survive DNA damage induced by chemotherapeutic treatments [233] and aberrations in DDR genes can promote the accumulation of DNA errors and affect the response of cells to DNA damaging anti-cancer treatment [234].

Both lost or increased DNA repair fidelity can affect the response of cells to anti-cancer treatment and have been implicated in chemotherapy resistance in CRC [235-238]. For instance, downregulation in mismatch repair (MMR) genes contribute to resistance to platinum and antimetabolite drugs [239] and overexpression of *ERCC* genes (specifically *ERCC1*, *ERCC3*, *ERCC4* and *ERCC6*) which are key members of the nucleotide excision repair (NER) pathway have been associated with higher levels of resistance to platinum drugs [238, 240, 241].

RAD54L is located on chromosome 1p32 which is a region of frequent loss of heterozygosity in breast tumours [242]. Deletions and mutations in *RAD54L* can lead to tumour formation and progression [243]. It has been implicated as a tumour suppressor in lymphoma, breast and colorectal cancers however the literature around this is relatively sparse [242-244].

REV3L gene is located on chromosome 6q21 and encodes the catalytic subunit of DNA polymerase (pol) ζ, which functions in the DNA damage tolerance mechanism of translesion DNA synthesis and protects against mitochondrial DNA damage [245, 246]. *REV3L* is ubiquitously expressed in normal and malignant cells and expression levels can affect resistance to platinum chemotherapy [247]. *In vitro* studies in cervical cancer have shown that cells deficient in *REV3L* demonstrate a higher sensitivity to platinum containing chemotherapy regimens [248], yet *REV3L* deficient oesophageal cancer cells demonstrate increased sensitivity to 5FU therapy through G1 phase arrest and apoptosis [249]. In mCRC the consensus is that there is reduced *REV3L* expression, independent of carcinoma stage [250] suggesting that downregulation may occur early during tumourigenesis.

ATM is a large gene on chromosome 11q that encodes a 3057 amino acid *PI3K*-related serine/threonine protein kinase that helps to maintain genomic integrity [10] through repair of DNA double stranded breaks. It is one of the most commonly aberrant genes in sporadic cancers and has been revisited here as a candidate driver of resistance. *ATM* mutations are predicted to result in increased sensitivity to platinum chemotherapy and topoisomerase inhibitors [251]. Finally, the thymidylate synthase gene (*TYMS*) has been found to be amplified in 5FU resistant cells [252]. Together, this indicates a potential role of the aforementioned genes in chemoresistance.

As discussed in earlier chapters, liquid biopsies have several potential advantages for the analysis of the genetic causes of chemotherapy resistance in cancer. They can theoretically capture subclonal genetic alterations present in a cancer cell population which is spatially dispersed across multiple metastases and hence improve inference of a tumour's clonal architecture [103]. Furthermore, cfDNA-seq can be applied longitudinally throughout the disease course and crucially at the time of disease progression to investigate the emergence of therapy resistance.

Results presented in Chapters 3 and 4 demonstrate that the cfDNA-seq technology I have optimised can detect mutations down to a VAF of <0.15% with a concordance of 89% of cfDNA mutation calls with tissue biopsy calls, such that invasive and more costly tumour biopsies may not be necessary. As drivers of acquired chemotherapy resistance are often subclonal with VAFs below 5% [96], the ability of this assay to identify variants down to <0.15% should theoretically enable detection of chemotherapy resistance drivers.

5.2 Hypothesis and Objectives

I wanted assess the performance of the optimised ultra-deep target sequencing assay to identify novel drivers of chemotherapy resistance in mCRC by performing cfDNA-seq on a larger cohort of mCRC patients. I also wanted to develop the next generation of whole exome ctDNA sequencing for resistance driver identification. Overall, to establish whether specific genes are recurrently mutated in mCRC the following objectives were set:

i. Apply cfDNA-seq to evaluate the genetics of treatment-refractory mCRCs
ii. Evaluate mutations in a panel of candidate genes to investigate mCRC resistance
iii. Apply cfDNA-seq technology for whole exome sequencing (cfDNA-WES) to identify recurrently mutated genes in longitudinal samples
iv. Refine error correction approaches for cfDNA-WES.

5.3 Application of cfDNA-seq to identify potential drivers of chemotherapy resistance in metastatic colorectal cancer

I included seven DDR genes: *ATM*, *ERCC3*, *ERCC4*, *ERCC6*, *FANCM*, *RAD54L,* *REV3L* and *TYMS* into the 32-gene 164 kb CRC target panel (Chapter 3, Table 3.2). My aim was to test whether these genes are recurrently mutated in the cfDNA of an independent cohort of mCRC patients treated with up to six lines of prior systemic anti-cancer treatment. Plasma samples were obtained from 13 patients from the PROSPECT-C trial (Table 5.1), a single arm phase II prospective translational study investigating predictive biomarkers for the anti-*EGFR* antibody cetuximab in the third line setting in patients with *RAS* wild type chemo-refractory mCRC (ClinicalTrials.gov: NCT02994888 PI: D Cunningham).

Samples were sequenced with the 32-gene CRC target panel described in Chapter 3 (section 3.4.1), in addition to the 28 cases from the FOrMAT study described in section 3.3.1 (Table 5.1). Three FOrMAT samples (cases 7, 8 and 21) and one PROSPECT-C sample (case 30) had low tumour derived cfDNA, thus analysis was performed on the remaining 37 cfDNA samples.

Case Number	Number of prior lines of therapy at time of cfDNA sample	5FU/ Capecitabine	Oxaliplatin	Irinitoecan	Anti-EGFR	Anti-angiogenic	Clinical trial drug
1	1	☐	☐			☐	
2	0						
3	1	☐	☐				
4	≥3	☐	☐	☐		☐	
5	1	☐	☐			☐	
6	≥3	☐	☐	☐		☐	
9	2	☐	☐	☐		☐	
10	≥3	☐	☐	☐	☐	☐	
11	0						
12	0						
13	1	☐	☐			☐	
14	2	☐	☐	☐		☐	
15	≥3	☐	☐	☐		☐	☐
16	1	☐	☐			☐	
17	0						
18	0						
19	0						
20	0						
22	0						
23	1	☐	☐			☐	
24	0						
25	1	☐		☐			
26	2	☐	☐	☐		☐	
27	1	☐				☐	
28	0						
29	2	☐	☐	☐		☐	
31	≥3	☐	☐	☐			
32	≥3	☐	☐	☐		☐	
33	≥3	☐	☐	☐		☐	
34	≥3	☐	☐	☐		☐	
35	≥3	☐	☐				
36	2	☐	☐	☐		☐	
37	≥3	☐	☐	☐		☐	
38	≥3	☐	☐	☐		☐	
39	2	☐	☐	☐		☐	
40	1	☐				☐	
41	2	☐	☐	☐			

Table 5.1: Prior lines of therapy delivered to patients sequenced by cfDNA-seq.
Cases 1-28 were from the FOrMAT study and cases 29 onwards were from the PROSPECT-C study. Cases 7, 8, 21 and 30 were removed due to low tumour-derived cfDNA content.

Within the sequenced cohort, ten patients were treatment naïve and 27 patients had prior exposure to systemic therapy. 100% of pre-treated patients received a fluoropyrimadine agent. Of the nine patients who had received one prior line of therapy, 66% (six of nine) were treated with oxaliplatin and five of these patients also received the anti-VEGF bevacizumab, 11% (one of nine) with irinotecan and the remaining two received capecitabine + bevacizumab. Of the seven patients treated with two prior lines of therapy, 100% received all three standard chemotherapy agents and 86% (six of seven) received anti-angiogenic therapy. One of 11 patients treated with three or more lines received anti-EGFR therapy and two of 11 received other targeted treatment within the context of an early phase clinical trial.

As detailed in Chapter 3 section 3.4.7, all mutation calls were performed after molecular barcode deduplication and duplexCaller error correction (Figure 5.1)

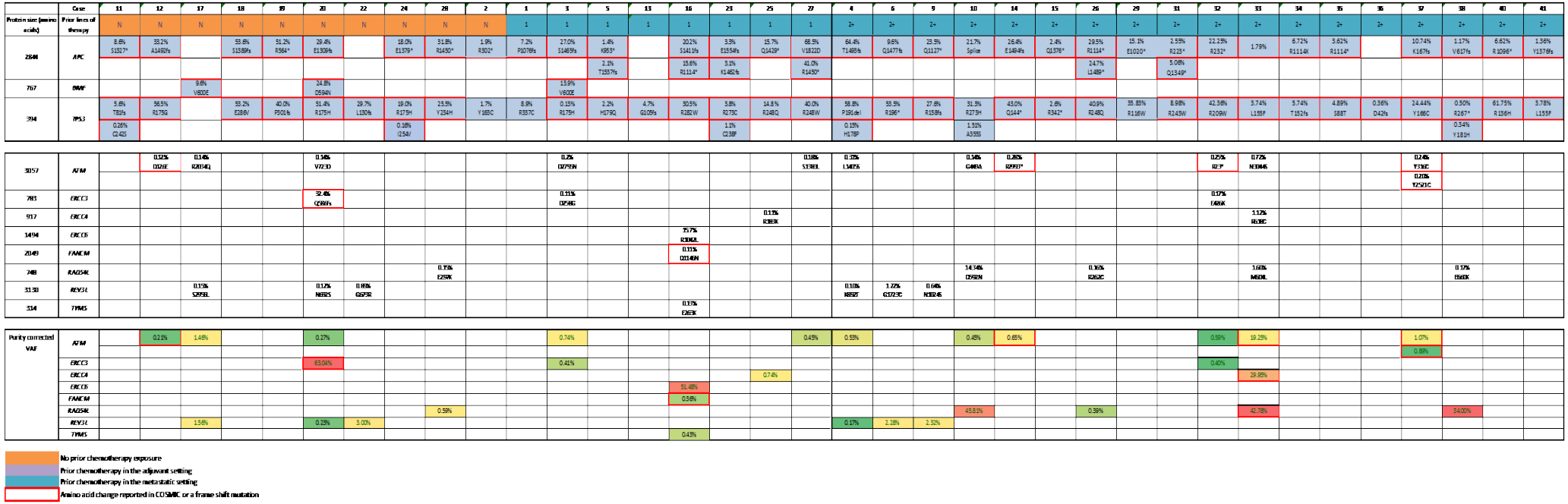


Figure 5.1: Non-silent mutations detected in driver genes, that are known to acquire truncal mutations, and in DNA damage repair genes in 37 mCRC cases.

Cases 1-28 were from the FOrMAT study and cases 29 onwards were from the PROSPECT-C study. Top panel- known clonal variants detected in cfDNA; middle panel- variants detected in DDR genes in cfDNA; bottom panel: purity corrected variant frequencies of mutations in DDR genes. Variants highlighted in red in the bottom panel are likely to be clonal and the rest are likely to be subclonal. VAF= variant allele frequency

Mutations driving acquired resistance can often be subclonal, therefore I aimed to evaluate clonality of these mutations. First, mutations in *TP53* and *BRAF* are usually acquired as early truncal events in CRCs [205] and the mutational status of these truncal variants are highly stable throughout the course of the disease [228]. These mutations are hence present in all cancer cells of a tumour which allows the use of their VAFs to a) demonstrate the presence of tumour derived DNA within the cfDNA sample and b), to benchmark whether other mutations in cfDNA are likely clonal or subclonal [253]. In mCRC, some groups have described subclonal mutations as those where the VAF is less than 25% of the highest VAF in the sample, and mutations are recognised as clonal if their VAF is above this threshold [155]. All 37 patients had one or more truncal mutations in *TP53* or *BRAF* genes with a median VAF of 21.7%, demonstrating presence of tumour derived DNA in all samples. 21 of 37 (57%) cases had two or more truncal mutations and the mutation with highest VAF was used for clonality assessment. *TP53* mutation VAFs (or *BRAF* in cases 3 and 17) were used to calculate the purity corrected VAF of a DDR gene variant as a ratio and if the DDR gene mutation purity corrected VAF was >25% of the highest VAF in the sample it was deemed a clonal mutation (Figure 5.1 lower panel).

15 of 37 (41%) had a mutation within at least one DDR gene. The VAF of these mutations ranged from 0.10% to 32.4% (Figure 5.1, middle panel). Five out of 31 somatic mutations in DDR genes had high purity corrected VAFs that were likely to be clonal (Figure 5.1, lower panel red boxes). In contrast, the VAFs of 26 of 31 were below the 25% threshold used to define a subclonal variant, which may have evolved under selective pressure of cytotoxic agents and would require sequencing of longitudinal samples to confirm this.

5.3.1 Frequently mutated DNA damage repair genes

The two most commonly mutated genes were *REV3L* and *ATM* with non-synonymous mutations in six and 11 cases, respectively. In order to understand whether there is an association of these recurrently mutated genes with chemotherapy resistance I wanted to assess the role of each mutation and whether they were likely to have a functional impact on the gene.

5.3.1.1 *REV3L*

Of the six patients with *REV3L* mutations, three were chemotherapy naïve (cases 17, 20 and 22) and three had received prior therapy (cases 4, 6 and 9) with 5FU + oxaliplatin containing regimens (Table 5.1). Cases 4 and 6 progressed while on treatment with 5FU/oxaliplatin combination regimens and case 9 had an initial response to CAPOX therapy in the first line setting with early progression off treatment. Thus all three patients had either no or only short benefit from 5FU/oxaliplatin treatment.

I next assessed the type and likely function of these mutations. None have previously been reported in the COSMIC cancer mutation database [10]. The DNA polymerase domain occupies the last third of the *REV3L* gene (Figure 5.2) and the central region harbours two adjacent binding domains for MAD2L2. MAD2L2 is necessary for pol ζ activity and functions as a bridge protein for interaction with REV1 protein. REV1 in turn interacts with Y-family DNA polymerases that insert bases opposite sites of DNA damage [254]. The central region of *REV3L* encodes a positively charged domain and a region with strong homology to KIAA2022, which is necessary for the efficient polymerase function of the recombinant protein. Two of the six mutations

observed in pre-treated patients are located in the KIAA2022 domain (K892T case 4 primary-progressor on 2nd line FOLFIRI + bevacizumab, and N1024S case 9 early progression following completion of 2nd line CAPOX treatment). The other pre-treated case (case 6) had a mutation which was not in a known protein coding domain and so the functional impact cannot be interpreted without experimental characterisation. Of the mutations found in treatment-naïve patients two were found in non-protein coding domains and the third was in the polymerase domain (S2956L case 17) (Figure 5.2). Interestingly, case 17 was later treated with capecitabine + irinotecan and had a durable partial response to treatment.

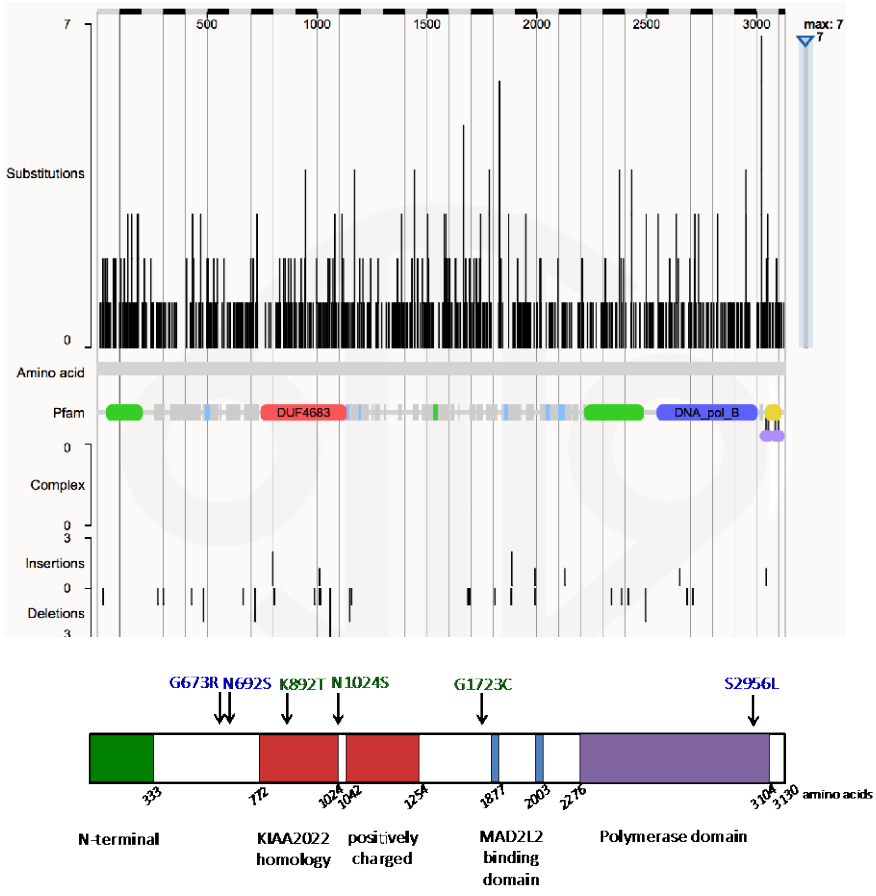


Figure 5.2: Mapping of subclonal mutations arising in mCRC patients in *REV3L* gene. Gene view histogram of mutations across *REV3L* genomic position of known hotspots adapted from COSMIC [19]. Labels beneath the coloured bar indicate domain names, adapted from Lange, 2016 [254]. Variant labels in

blue font are from treatment-naïve patients and green font are pre-treated patients.

Overall, the *REV3L* mutations found in 6/37 patients were not clustered in any single domain and there is no evidence to suggest that they disrupt protein function. Rather, these mutations are likely passenger mutations and the high mutation count in this gene may be explained by the large size of the gene. Contrary to the current knowledge that *REV3L* mutations are expected to lead to sensitivity to 5FU, I found that the *REV3L* mutations were not enriched in chemotherapy resistant cases, where three of 27 (11%) occurred in resistant tumours and three of 10 (30%) in treatment naïve tumours.

5.3.1.2 *ATM*

The *ATM* tumour suppressor gene was the most frequently mutated gene out of the seven DDR genes investigated (with 12 mutations occurring in 11 of 37 patients (30%). This was higher than the previously reported incidence of 20% in mCRC TCGA cohort [251]. *ATM* mutations were seen in three of 10 (30%) treatment naïve patients (cases 12, 17 and 20), two of eight (25%) post 1st line patients (cases 3 and 27) and six of 19 (32%) post ≥2nd line patients (cases 4, 10, 14, 32, 33 and 37, case 37 had 2 *ATM* mutations) (Table 5.2). A χ^2 test was performed which showed no relationship between number of prior lines of therapy and *ATM* mutations (p= 0.94).

Number of prior lines of treatment	Percentage of patients receiving 5FU	Percentage of patients receiving oxaliplatin	Percentage of patients receiving irinotecan	Percentage of patients receiving anti-angiogenic	Percentage of patients receiving anti-EGFR
1	100%	50%		50%	
≥2	100%	100%	100%	100%	16.6%

Table 5.2: Prior lines of therapy and individual drugs received by patients harbouring ATM mutations.
30% of treatment naïve patients harboured an *ATM* mutation, 25% of post 1st line patients and 32% post ≥2nd line patients harboured *ATM* mutations.

Overall, there was no clear enrichment of *ATM* mutations in chemo-naïve or chemo-resistant patients with relatively equal proportions of untreated and pre-treated patients (30% vs. 29%). As majority of the pre-treated patients received standard chemotherapy agents 5FU, oxaliplatin and irinotecan, it is not possible to draw conclusions on the predictive value of *ATM* mutations with individual drugs, and particularly topoisomerase inhibitors, as reported in other tumour types [251]. Similarly there was no enrichment of *ATM* mutations in patients receiving anti-angiogenics (bevacizumab or aflibercept) where the majority (15 of 23 patients = 65%) who received anti-angiogenics were *ATM* wild type and only eight of 23 (35%) had an *ATM* mutation.

I next assessed the type and likely function of these mutations to see if this correlated with treatment outcome. Five of the 12 mutations (42%) have previously been identified in cancer and are reported in the COSMIC cancer mutation database [10] (mutations highlighted with a red border Figure 5.1). Three mutations encode for non-synonymous substitutions (D126E, Y316C and Y2521C) and two for premature stop codons (R2993* and R23*). Recurrence in cancer and the occurrence of disrupting mutations suggests that these *ATM* mutations are not merely passenger mutations but likely have

some functional relevance. Case 14 harbours the R2993* variant which leads to loss of ATM protein function through loss of the FATC domain (Figure 5.3) and can be associated with an increased risk of breast cancer [255]. The FATC domain is critical in regulation of activation of *ATM* through interaction with Tip60, a histone acetyltransferase, that is a key component of the DNA damage-signalling network [256].

The other truncating mutation was R23* found in case 32 which also leads to inactivation of *ATM* through loss of the TAN domain (Figure 5.3) which functions in the DNA damage response and regulation of telomere maintenance [257]. Thus, TAN inactivation reduces the recruitment of ATM to double strand DNA breaks. Case 12 also harboured a mutation within the TAN domain, D126E, which is described in COSMIC [10] with four reported mutations in soft tissue tumours and haematological malignancies. The remaining two missense mutations, which were reported in COSMIC arise in non-coding regions of the gene and have each been reported only once.

The D2795N mutation identified in case 3 has not been reported in COSMIC, but is located within the PIKK regulatory domain (Figure 5.3). This region recognises serine-glutamine and threonine-glutamine motifs in a number of proteins including those involved in cell-cycle checkpoint arrest (e.g. Chk1 and Chk2), DNA repair (BRCA1) and apoptosis (p53) [258]. A mutation that impairs this interaction could perhaps lead to an inability to repair double strand breaks and hence affect genome integrity.

Case 37 was a heavily pre-treated case that harboured two *ATM* mutations. Although both missense mutations were outside of protein-coding domains,

they were both previously reported in COSMIC database. This could represent parallel evolution arising in this patient.

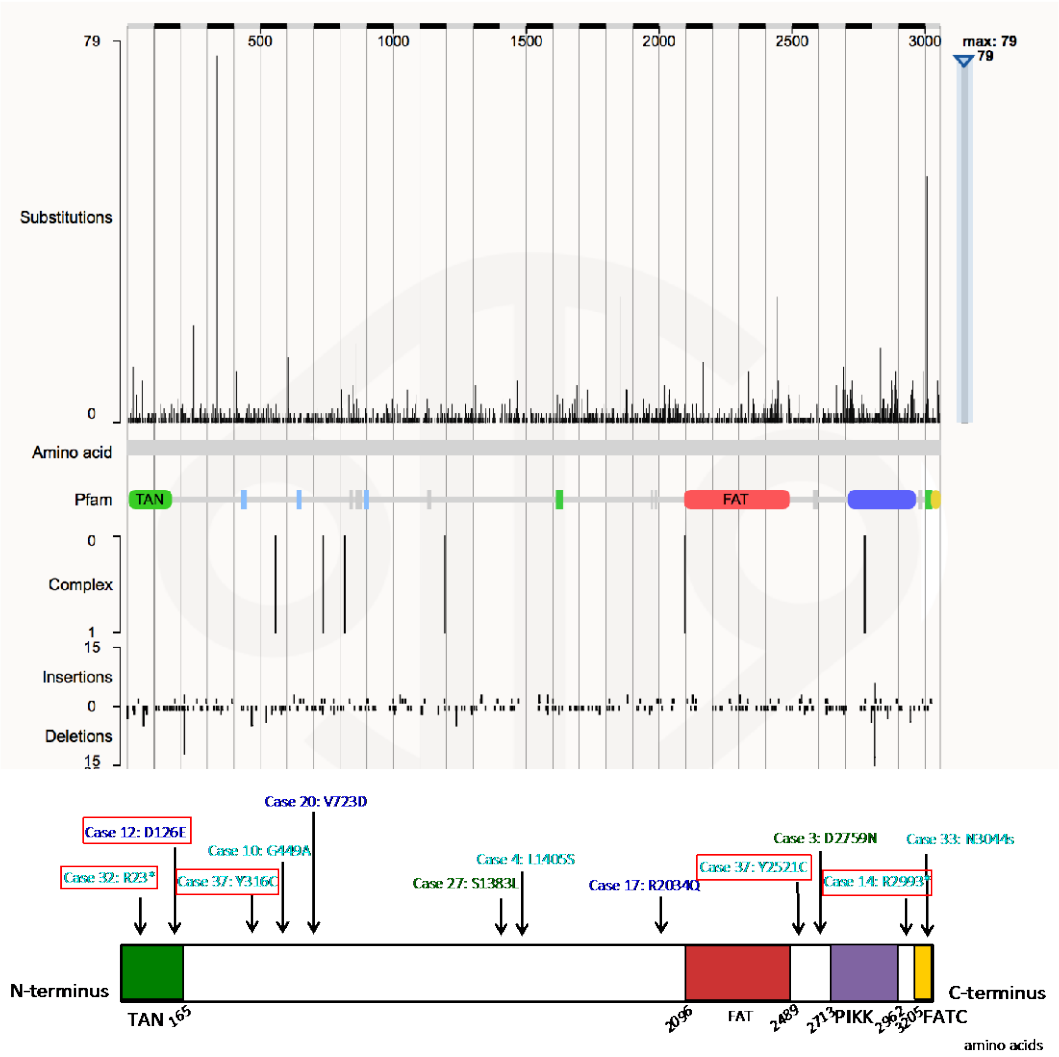


Figure 5.3: Mapping of the 12 *ATM* subclonal mutations detected in 11 patients.
Blue font denotes treatment naïve patients, green font for patients treated with 1 prior line of therapy and cyan font for those treated with 2 or more prior lines. Mutations are scattered across the full length of this large gene of 3057 bp. Gene view histogram of mutations across *ATM* genomic position of known *ATM* hotspots adapted from COSMIC [10]. Mutations identified in COSMIC are outlined in a red box. Labels beneath the coloured bar indicate domain names, adapted from Weber et al [259].

Eleven of the 12 *ATM* mutations had low VAFs with a median of 0.14%, two of these are likely driver mutations leading to inactivation of the gene and two are functional missense mutations found in protein coding domains. *ATM* mutations have been implicated in clonal haematopoiesis (CH) [210] however this is unlikely the source of mutations in this cohort as I excluded CH events

through simultaneously analysing germline blood DNA BAM files and *ATM* variants were absent in all cases. In conclusion, subclonal *ATM* mutations are common and several of these are likely functional but there is no correlation with prior chemotherapy exposure.

5.3.2 Less frequently mutated DNA repair genes

The *ERCC* genes encode a set of proteins involved in DNA repair and polymorphisms of the *ERCC* genes are implicated in cancer development, progression, survival and can influence response to chemotherapy [260]. Specifically, *ERCC6* overexpression has been associated with resistance to 5–fluorouracil [235].

Case 25 was the single case in this cohort that harboured a subclonal *ERCC4* variant. This variant is not reported in the literature as being a cancer associated variant. *ERCC6* mutation was also detected in a single case (case 16- 1 prior line of treatment), however this is likely a clonal mutation based on the purity corrected VAF calculation and thus likely present in the majority of cancer cells. Interestingly, review of clinical details confirmed that this patient progressed on both first line and second line treatment, thus this mutation could be further investigated in the future as a potential resistance driver.

There are two cases harbouring an *ERCC3* mutation, one of which (case 20- frame shift mutation Q586fs) has a VAF similar to that of truncal mutations and is therefore clonal. This variant is likely to affect *ERCC3* function as it results in loss of part of the C-terminal helicase domain (amino acids 496-792). This patient had a favourable response to a single line of

fluoropyrimidine/topoisomerase inhibitor (FOLFIRI) treatment, but did not receive platinum-based therapy. The other case with a variant in *ERCC3* had not received any prior treatment. Overall, mutations in *ERCC* genes were rare and are unlikely to play an important role in acquired resistance.

The single *FANCM* mutation (D1146N VAF 0.11%) observed in case 16, who had been treated with one prior line of systemic therapy has also been reported once in the COSMIC cancer database [10]. It arises in a non-protein coding domain thus functional analysis is required to ascertain its role in resistance.

5.3.3 Copy number analysis of candidate chemotherapy resistance driver genes in cfDNA

I next performed genome wide DNA copy number profiling using CNVkit algorithm for all 37 patients utilising the off-target reads (as described in Chapter 4 section 4.5) (Figure 5.4). Ten cases did not demonstrate gains or losses by visual inspection and had low tumour content based on the VAFs of *TP53* and *BRAF* mutations (median VAF <2.5%). These were excluded as non-informative, leaving 27 cases for analysis. I examined the copy number profiles to assess the copy number status of each DDR gene individually: *ATM*, *ERCC3*, *ERCC4*, *ERCC6*, *REV3L* and *RAD54*. There were no high level amplifications detectable in any of these samples. I also assessed for evidence of allelic loss, which is a pattern characteristic of tumour suppressor genes. In case 20, a copy number loss of the *ERCC3* locus on chromosome 2 was detected in addition to the inactivating Q586 frame shift mutation. No other losses were found concurrently with mutated genes.

It is reported that 5FU resistance is closely related to the expression of its primary target, thymidylate synthase (TS) [261]. TS is encoded by the *TYMS* gene, the level of *TYMS* expression is indicative of prognosis as patients with low expression display a better overall survival (OS) than patients with higher expression [175, 204]. However no *TYMS* amplifications were identified in this cohort, despite all pre-treated patients having prior exposure to 5FU based chemotherapy regimens.

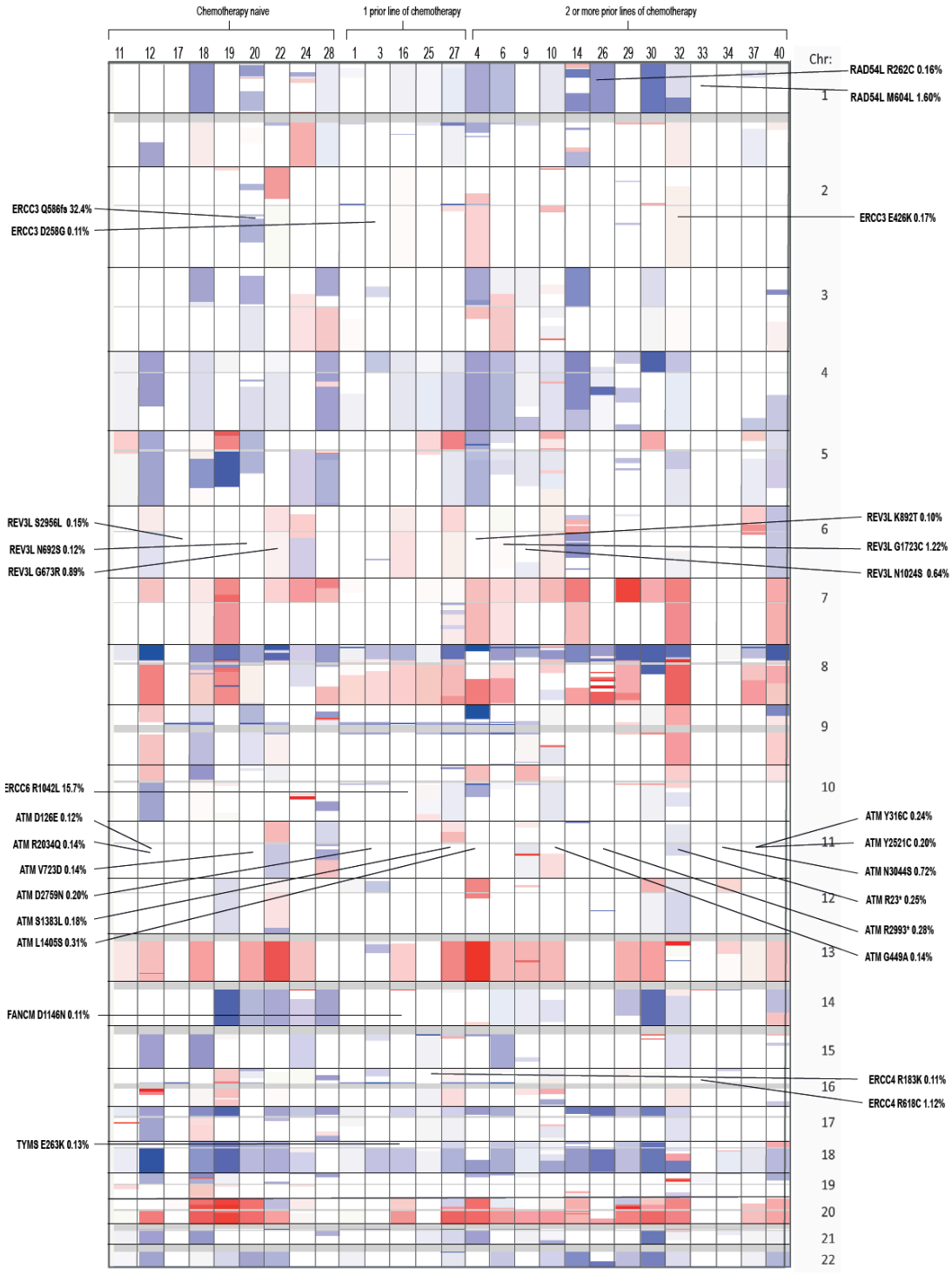


Figure 5.4: Heatmap showing the genome-wide DNA copy number analysis generated from cfDNA-seq to assess copy number status of DDR gene mutations.
Positions of mutations in DDR genes have been indicated. Chemotherapy naïve (left: cases 11 to 28) and prior chemotherapy exposure (cases 1 to 40).

In summary, following the analysis of DDR genes and *TYMS* only 51% of cases harboured a mutation or copy number change. There was no clear enrichment in cases that were refractory to chemotherapy and with the exception of *ATM* mutations, the evidence for functional relevance was

sparse, i.e. only 3/31 mutations were disrupting mutations e.g. early stop and frameshift and there was no clear clustering in specific domains or evidence of recurrence suggesting hotspot mutations based on cancer mutation database. Taken together, the role of the mutations in driving chemotherapy resistance is difficult to ascertain in this small series and ideally further investigation in a larger cohort, with functional analysis is required.

The current understanding of chemotherapy resistance is limited as biopsies are often difficult to obtain. cfDNA has major potential but assays are currently limited to a relatively small number of target genes (usually a few dozen and up 100-200 in the largest assays in use). However, chemotherapy resistance may be complex and better assays are required. Thus, my final project was to apply our cfDNA sequencing technology, which achieves sequencing depths of ~2000X with low cfDNA input quantities, to develop ultra-deep cfDNA whole exome sequencing to analyse all protein coding regions.

5.4 Optimising error correction approaches for cfDNA sequencing

5.4.1 Duplex error correction

As described in Chapter 3, while developing molecular barcode error corrected cfDNA-seq technology, I found that both DNA strands (duplexes) of the input DNA molecules could be identified in the sequencing data. Dr Dimitrios Kleftogiannis used this to build a two-step model of error correction in which MBCs were first used to eliminate some sequencing errors through the formation of consensus families. Secondly detection of duplex DNA molecules supporting a variant was required for high fidelity calling of low frequency mutations. However the use of MBCs requires high depth

sequencing (21,650X pre-deduplication) in order to form consensus families, and this increases sequencing and computational costs. We reasoned that error correction could be applied solely through the detection of duplex molecules, without barcodes and excessive over-sequencing. If duplex error correction could be achieved at shallower sequencing depths it may offer a more economic approach for detection of subclonal mutations with VAFs <5% that could even be applied to whole exome cfDNA-seq.

A novel software package (DuDe- duplex deduplication) was developed by our Dr Dimitrios Kleftogiannis to separately align the F1R2 and F2R1 oriented paired reads in order to distinguish the original strands of each input DNA molecule. After deduplication and mutation calling using SureCall SNPPET, variants supported by duplex reads could be identified with the duplexCaller bioinformatics tool, without the need for MBCs.

To provide proof of principle that duplex error correction can be performed to detect mutations without the need for MBCs, I selected three of the 28 FOrMAT mCRC cfDNA-seq datasets (cases 10, 15 and 23) that each contained mutations with a VAF <5% (ten low frequency mutations in total, median VAF: 3.4%, range: 0.37% to 3.8%) for downsampling. Downsampling is a process by which the total number of reads in a fastq file can be randomly reduced bioinformatically in order to simulate the effect of sequencing to a lower depth. Dr Louise Barber downsampled to 8% of the sequenced reads per sample and performed analysis to determine whether low VAF mutations could still be detected by DuDe and duplexCaller error correction at a sequencing depth of only ~1000X (Figure 5.5).

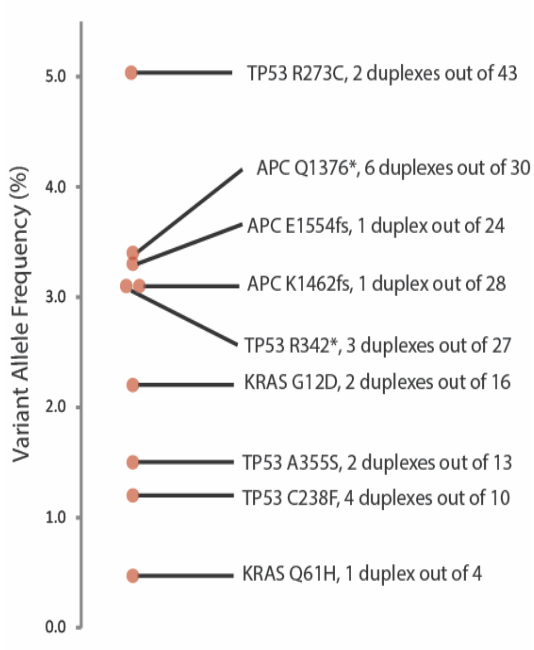


Figure 5.5: Pilot data showing that duplex reads can be identified after targeted sequencing at depths of ~1000X without the need for MBCs.

Of the ten subclonal mutations identified by ultra-deep targeted cfDNA-seq, nine were called and supported by the presence of duplex reads in the down-sampled data. The single mutation that was not supported by duplex reads was a single nucleotide deletion in the *TCF7L2* gene of case 15, with a variant frequency of 1.3%. This is likely because this variant was originally supported by 15 reads and only one duplex family but after downsampling, the reads supporting the duplex were lost. Based on this result, some loss of sensitivity may be expected with lower sequencing output, however overall the data appeared promising that duplex error correction can be performed at lower depths. Two other indels were called accurately, these were of higher variant frequency (3.1% and 3.3%) and one duplex-read was identified out of 28 and 24 reads, respectively.

The ideal would be to have majority of reads supported by duplex molecules, which should improve the calling of low frequency mutations close to the detection limit. However in the downsampled dataset 18% of reads were

supported by duplex molecules which is lower than the original ultra-deep cfDNA-seq data with a maximum 29% duplex reads. This is likely due to some of the duplex reads only being present in the original dataset at a low frequency, with few duplicates, such that many are lost in downsampling. The implication is that if sequencing depth was reduced to ~1000X, fewer variants would have duplex support. Therefore, in order to utilise duplex error correction at lower sequencing output, it will be important to improve duplex recovery.

There are two major factors influencing the effective detection of both strands of a duplex DNA molecule: i) losses of either the forward or the reverse strand of the input DNA molecules could occur during PCR amplification or other steps of sequencing library preparation; and ii) the sequencing depth that was used to generate the targeted cfDNA-seq data was lower than the library complexity, i.e. the number of unique DNA molecules in the library covering each target nucleotide was higher than the sequencing depth, which in turn leads to under-detection of duplex molecules which are present in the sequencing library. Minimising PCR duplicates can help to improve sequencing saturation, but ultimately this is a balance between the cost of sequencing and detecting as many unique molecules as possible.

I therefore wanted to investigate how to optimise duplex molecule retention in order to utilise duplex error correction more effectively at a sequencing depth of ~1000 X, which would be the target depth for future whole exome sequencing.

5.4.2 Methods to improve duplex recovery through library preparation modifications

In order to explore which steps in library preparation can increase the retention of duplex molecules, I reanalysed our existing datasets generated from early library preparation optimisation experiments (Chapter 3 section 3.4.2), where libraries had been constructed from six cfDNA samples in parallel while modifying a number of library preparation conditions (bait preparation, sequencing depth, hybridization and capture wash conditions). I hypothesised that the following modifications in the sequencing library preparation may help to improve duplex retention:

- Input DNA amount- 25 ng, 50 ng or 125 ng: higher input DNA quantity increases the likelihood of more unique molecules being incorporated into the sequencing library and may allow for fewer cycles of PCR
- Choice of library preparation kit: Agilent SureSelect XT-HS end repair and A-tailing vs. Kapa end repair and A-tailing. The Kapa kit could improve duplex recovery as it recommends fewer PCR amplification cycles, which minimises amplification bias and translates into higher molecular complexity
- Baits A vs. B: although direct comparison of A vs. B showed no difference in sequencing depth and on-target %, as detailed in Chapter 3 section 3.4.2.1, it is plausible that the formulation of the target panel could improve retention of duplex reads
- hybridisation conditions: fast hybridisation vs. normal hybridisation.

I first compared the duplex rate achieved with libraries that used varying cfDNA input amounts (Figure 5.6). Libraries were prepared using the Agilent

XT-HS protocol, which is identical to the optimised library preparation protocol (Chapter 3 section 3.4.2) with the exceptions of 65ºC capture wash temperature instead of 70ºC and a single round of post-capture Ampure clean-up.

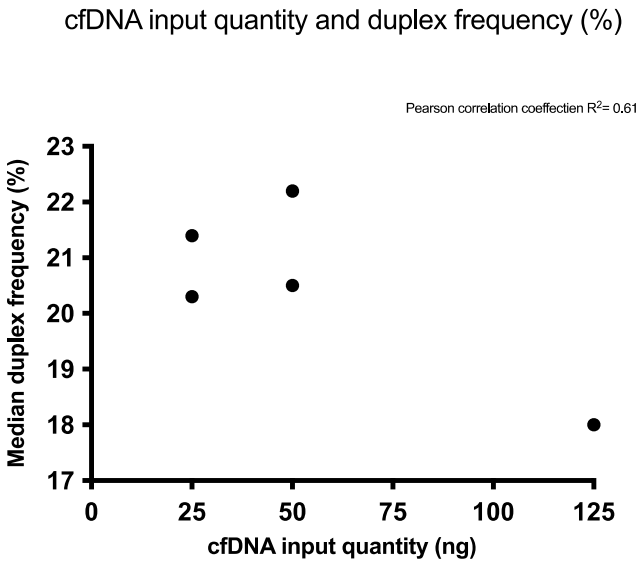


Figure 5.6: Scatter plot demonstrating the duplex frequency in coding regions versus input cfDNA amounts.

Libraries constructed with lower input cfDNA amounts demonstrated some improvement in duplex retention rate, comparing 25 ng input with 125 ng input (20.9% vs. 18%). There were no large differences in duplex frequency between 25 ng to 50 ng input (20.9% vs. 21.4%). Given that library complexity is a measure of unique DNA molecules in the library covering each target nucleotide across the target region, an increase in the cfDNA input should improve complexity [262]. It is possible that duplex recovery requires more sequencing output for a library with higher complexity as there are more DNA molecules to detect. Although this analysis is based on a small number of samples, the drop in duplex rate with higher input cfDNA is likely related to sequencing less and less of the complexity with high input amounts.

Considering these relationships, it was clear that an increase in input DNA would require an increase in sequencing depth. As my aim was to optimise a sequencing assay that only requires ~1000 X depth, an increase of input DNA would be counterproductive.

I next reviewed the effect of an alternative library preparation kit (KAPA) on duplex recovery (Figure 5.7). The KAPA library preparation purports higher adaptor ligation and polymerase efficiency and therefore requires fewer PCR cycles to amplify the adaptor-ligated library [263], even with low input DNA amounts. This should reduce PCR duplicates which can introduce bias and occupy sequencing capacity without generating new data.

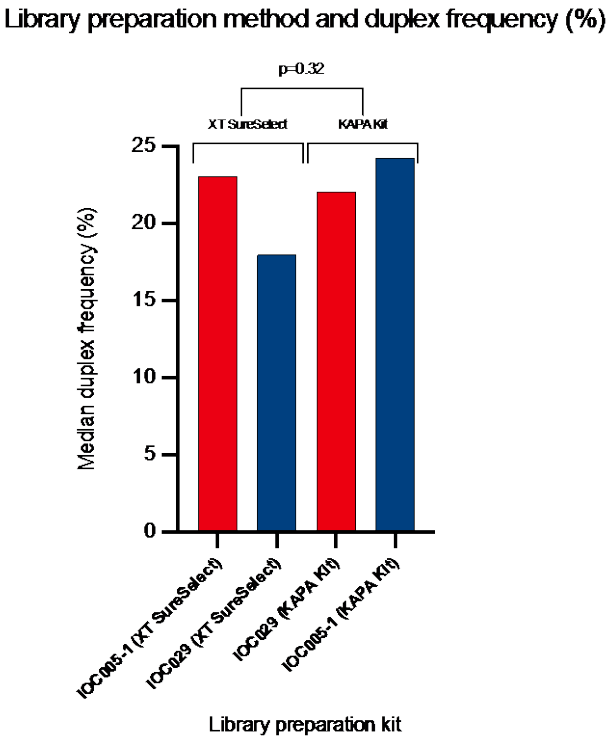


Figure 5.7: Comparison of library preparation kits in two cfDNA samples, to assess their impact on duplex frequency.

Figure 5.7 demonstrates no difference in duplex frequency between the library preparation kits. Overall, sample IOC005-1 had a higher duplex rate than IOC029, irrespective of library preparation method thus inferring that duplex retention is at least in part, sample dependent and not library preparation protocol dependent.

I also investigated the impact of bait preparations on duplex retention (Figure 5.8). Libraries from three samples (F30, F32 and F33) were prepared in duplicate, one set of libraries were hybridised with the standard formulation (baits A) and the other set were hybridised with the proprietary formulation (baits B) of biotinylated oligonucleotides.

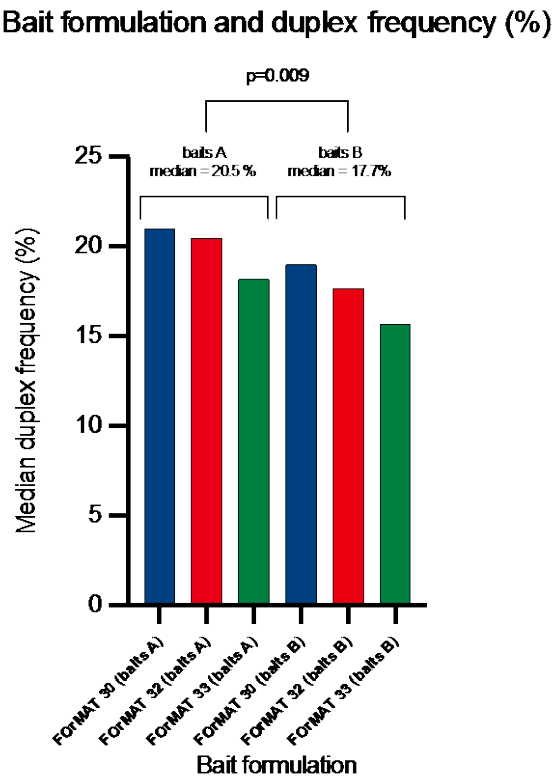


Figure 5.8: Comparison of oligonucleotide bait formulation on duplex frequency.
baits A: off the shelf; baits B propriety formulation for small target panels

Libraries showed a significant improvement in duplex frequency when baits A were used instead of baits B (paired t test p=0.009). However, the

differences in duplex frequency rates were relatively small and baits A were already selected for use in the optimised sequencing protocol.

The final library preparation parameter analysed was fast hybridisation (developed in the optimised library preparation protocol – section 3.4.2.3) vs. standard hybridisation of 65ºC over 16 hours (Figure 5.9).

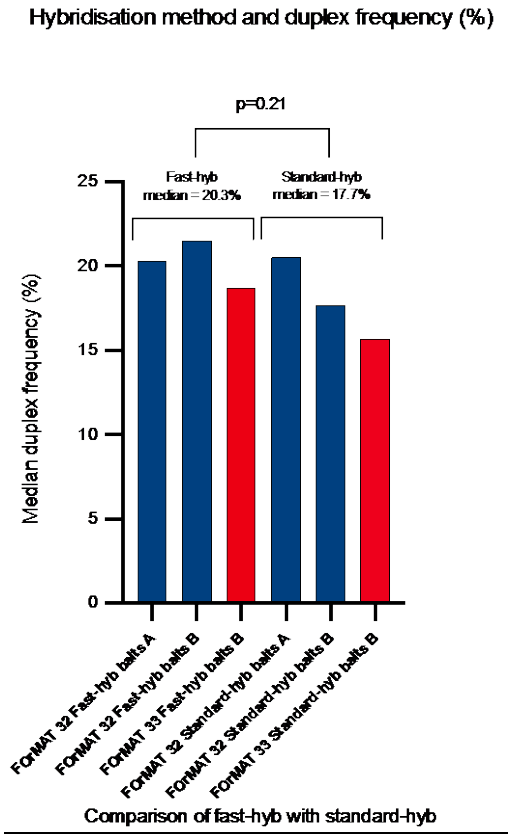


Figure 5.9: Comparison of hybridisation method on duplex frequency.

Across the six libraries there is no significant difference of duplex molecules between fast or standard hybridisation (median 20.3% vs. 17.7% p=0.21). Sample F32 had a duplex rate above the median of 19% in three out of four libraries irrespective of the hybridisation and bait preparation method. Sample F33 had a duplex frequency below the median in both fast and standard hybridisation methods, once again suggesting that duplex retention is probably more influenced by inherent sample characteristics.

Overall, baits A preparation was the only factor associated with a statistically significant improvement. It became increasingly apparent that certain samples always retained duplex reads greater than the cohort median of 19% (F54 and IOC5001) compared with sample F33 which could not recover duplexes for greater than 18% of all reads regardless of the technical changes to library construction condition.

5.4.3 Impact of sequencing depth on duplex retention

Further review of the samples prepared with the final optimised protocol and specifically those samples that had a duplex frequency greater than the cohort median of 19%, there was a trend towards increased duplex frequency in samples that had achieved a higher sequencing depth (Figure 5.10).

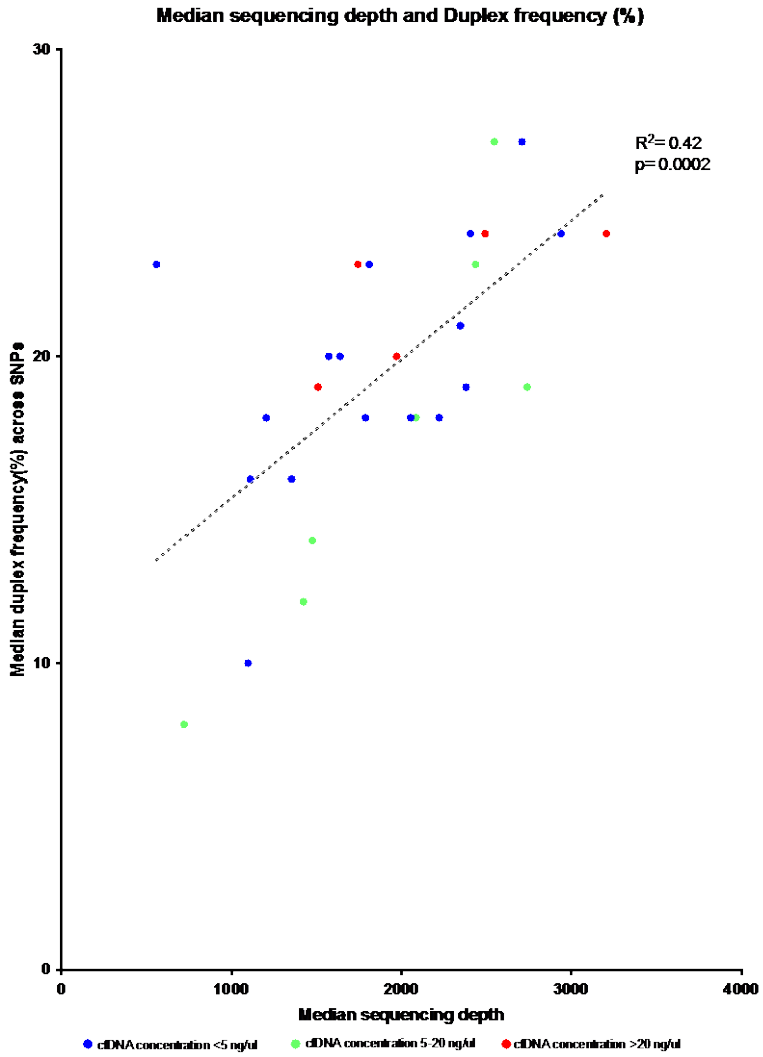


Figure 5.10: Relationship of median depth and median duplex frequency for each sample sequenced using the final optimised protocol.
Data points have been colour coded by concentration of cfDNA in the sample.

The relationship between depth and duplex retention rate is more pronounced in samples reaching the target sequencing depth of >2000 X. The reason for this could be because samples that have an efficient library conversion can capture more of the library complexity, which is important to recover duplex molecules. However, the duplex rate plateaus at depths >2500 X in our cohort, such that even if the sequencing depth exceeds 3000 X the median duplex rate is still only 24%.

5.4.4 Effect of cfDNA concentration and duplex retention

In the search of identifying which inherent sample properties could be responsible for affecting the duplex rate, I investigated the impact of cfDNA concentration of the sample on the duplex retention rate. There was no difference in duplex rates in samples with low and medium cfDNA concentration (<5 ng/µl and 5-20 ng/µl, blue and green circles respectively in Figure 5.10 above). However all five samples with a high cfDNA concentration (>20 ng/µl red circles in Figure 5.10) had a median duplex rate of 23% which was greater than the cohort median of 19%.

5.5 Development of ultra-deep whole exome ctDNA sequencing

Whole genome sequencing allows the most comprehensive analysis of cancer genomes, however the requirement for very high sequencing depths to profile subclonality would make this prohibitively expensive. Analysis of the protein coding exome by WES is a more cost effective option, but WES of cfDNA has only been performed to date with depths of up to 300 X [107, 264, 265]. High error rates of WES without dedicated error correction and the lower sensitivity at 300 X sequencing depth, compared to the much higher depths routinely used for cfDNA target gene panel sequencing, preclude the reliable detection of low frequency mutations. Drivers of chemotherapy resistance are often present in subclones and hence have VAFs below 5% [188, 201, 215]. As MBC error correction requires over-sequencing to make use of the barcodes, it cannot be widely used for WES due to time and financial constraints. Over-sequencing is not always necessary when error correction can be performed using duplex configuration as a means of endogenous barcode error correction for subclonality assessment. Thus, I aimed to apply the expertise and library preparation protocols established for

deep targeted cfDNA-seq to test if cfDNA-WES can be performed at much higher depth than published (~1000 X) from 25 ng cfDNA. I also aimed to apply duplex error correction to reduce false positives.

5.5.1 Sample selection

I selected two patients from the FOrMAT cohort who had a good response to initial chemotherapy but failed to respond to subsequent chemotherapy lines (Figure 5.11) so potentially had acquired genetic drivers of resistance. I hypothesised that selective pressures during chemotherapy would alter the subclonal composition of these tumours and that this should be traceable in the cfDNA.

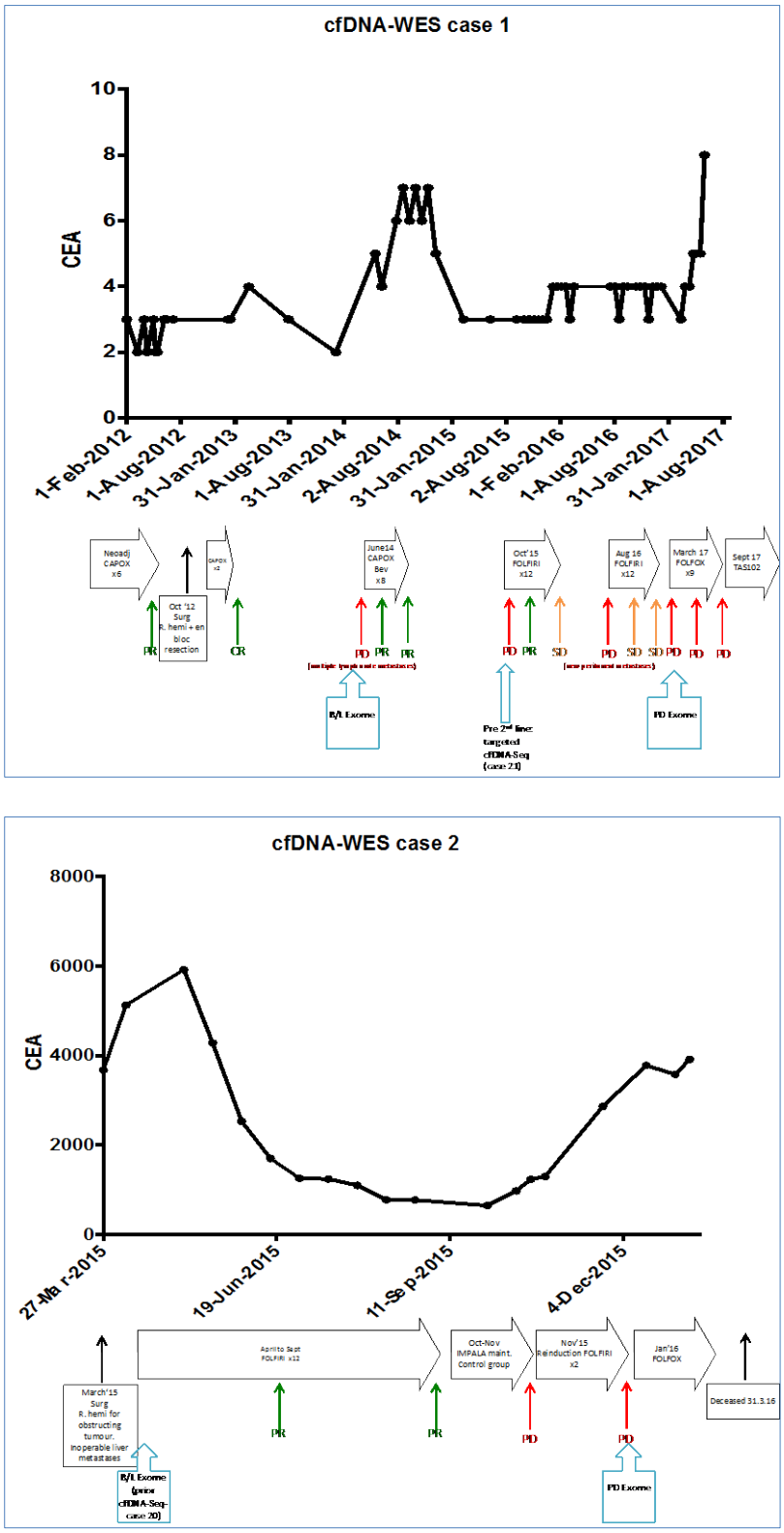
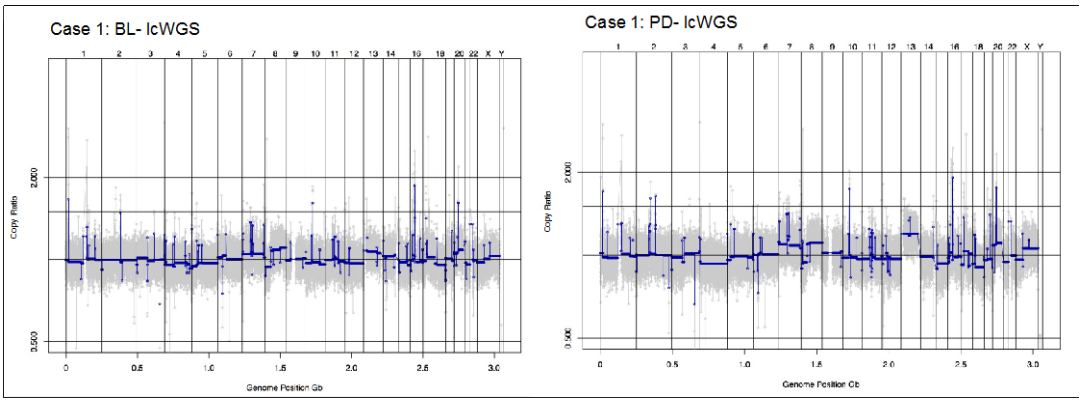


Figure 5.11: Graphical representation of patient treatment history, CEA response and radiological response (by RECIST 1.1) of two patients selected for cfDNA-WES. cfDNA-WES was performed on baseline (BL) and post progression (PD) on 3rd line treatment samples for case 1 and on BL and PD on 2nd line treatment for case 2, as indicated by blue boxes.

I expected that the new cfDNA exome sequencing assay would not be as sensitive as the targeted panel during the initial development, as it may be difficult to achieve uniformly high depths across the whole exome. Therefore a key consideration was to select cases with sufficiently high tumour derived DNA content in the cfDNA (>20%), which should facilitate the detection of multiple subclonal mutations. To estimate the tumour content of cfDNA samples that had not been previously sequenced, I performed low coverage whole genome sequencing in case 1, which had no known mutations, and Sanger sequencing in case 2 where truncal mutations were already known from prior cfDNA targeted sequencing (*BRAF*, *TP53*) (Figure 5.12).



Case 2: BL- targeted cfDNA-Seq variant detection

Mutation	VAF (%)
TP53 R175H	51.4
APC E1309fs	29.4
TCF7L2 Q27fs	17.4
BRAF D594N	24.8

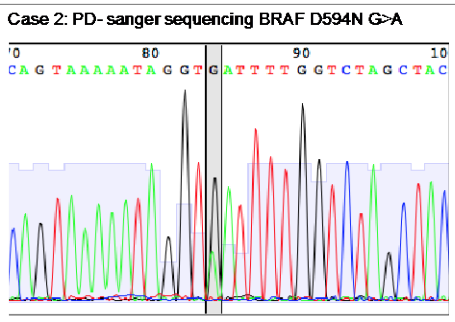


Figure 5.12: Tumour content estimation of two samples selected for cfDNA-WES-seq. Top panel- lcWGS on 5 ng cfDNA from patient 1, at baseline and progression timepoints. Bottom panel left- cfDNA-seq results for patient 2 BL: truncal mutation VAF up to 51%. Right- PD cfDNA sample Sanger sequencing results for BRAF G>A encoding p.D594N.

As this was a pilot experiment of ultra-deep cfDNA-WES-seq with high costs, I wanted to test two different amounts of cfDNA as starting material: 25 ng, which is sufficient for targeted cfDNA-Seq; and 50 ng to determine whether this would increase the number of unique molecules converted into the sequencing library and facilitate adequate sampling of rare subclonal populations.

5.5.2 cfDNA-WES library preparation and sequencing

Library preparation was performed using the optimised XT-HS protocol as described in Chapter 3, followed by whole exome capture, which was performed with Agilent SureSelect Human All Exon Kit v5.0 according to manufacturer's instructions. I planned sequencing of the BL and the PD cfDNA samples from the same patient on separate runs to avoid the risk of potentially false positive mutation calls arising from index hopping. In addition, matched normal germline DNA libraries from peripheral blood mononuclear cells were also prepared and sequenced using less sequencing capacity than BL and PD samples (1% of the run) to help identify somatic variants from germline variants or SNPs during mutation calling.

Each sample was sequenced by the Tumour Profiling Unit, on an Illumina Novaseq S2 flow cell using a read length of 2 x 100 bp. I aimed for a targeting sequencing post deduplication depth of ~1000 X for this pilot experiment, as a compromise between affordability and panel size for cfDNA-WES as we believe that DuDe bioinformatics tool can be used to identify duplexes instead of MBCs. Thus, we did not need to reach the same high depth required for targeted cfDNA-seq assay. Paired end reads were aligned against the human reference genome (hg19) by Dr Louise Barber.

5.5.3 cfDNA-WES QC and error correction results

The median MBC deduplicated sequencing depth of coding gene regions of BL and PD samples was 1150 X and that for germline DNA sample was 56 X. Median on target depth was 86% for all samples. The analysis of the samples collectively is ongoing however case 2 was initially analysed in further detail. The median number of duplex molecules in case 2 PD was 10.9% which is lower than the median achieved for target cfDNA-seq of 19%, although this will require further review with more samples.

I next compared standard deduplication with DuDe error correction and analysed the reduction in false positive calls (Figure 5.13).

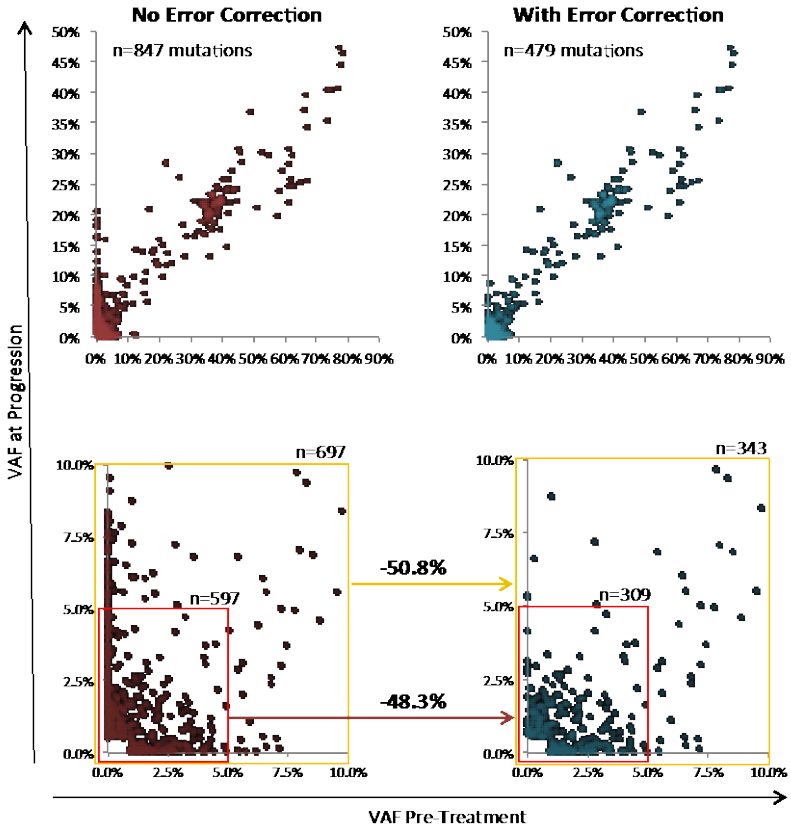


Figure 5.13: Application of error correction using the DuDe tool to cfDNA-WES data from case 2.

The panels at the bottom show a magnification of mutations with VAFs below 10%. Yellow square highlights all calls with ≤10% VAF and the red square highlights calls with ≤5% VAF.

This demonstrates that duplex error correction reduced the number of calls by 43% (847 mutations to 479 mutations). The proportional reduction was similar when only considering mutations with VAFs below 10% or 5%. Of note, mutations that were only present in one of the two samples were removed by error correction (those superimposed on the x or y axis in Figure 5.13). In contrast, those mutations detected in both samples were largely preserved after duplex error correction. As sequencing and PCR errors are usually unique to individual samples and clonal mutations would be present in both, this indicates that duplex error correction indeed removed predominantly the former and confirmed the latter in these samples. Although this provides the first proof of principle that whole exome cfDNA sequencing at a depth of 1000 X is possible and that duplex error correction leads to a strong reduction of calls that are likely false positives, further mixing experiments as described in Chapter 3 will be necessary to define the sensitivity and specificity of this technology. This pilot data has been used towards an European Research Council grant application, securing a ∈2 million grant for ongoing work in cfDNA-seq.

5.6 Summary and Discussion

Although mutations in the seven studied DDR genes were found in 19 of 37 patients (51%), these were not confined to cases with acquired chemotherapy resistance. Five of 19 were detected in patients that had not received any palliative chemotherapy. The most frequently mutated genes were *ATM* (11 cases) and *REV3L* (six cases), however there were no hotspot mutations which are frequently observed in genes that acquire gain of function mutations. Three deleterious mutations were identified in *ATM*, however there was no concomitant copy number loss which is a pattern characteristic for tumour suppressor genes in cancers. Furthermore, two out of 28 pretreated cases harboured loss of function mutations which would be expected to sensitise these tumours to DNA damaging chemotherapy rather than induce resistance. Mutations in *ATM* detected by cfDNA sequencing were hence considered unlikely to drive acquired chemotherapy resistance. Further studies will be required to further investigate the functional role of these mutations in mCRC. It is not possible to ascertain from single timepoint sequencing at which point the mutation was acquired in the pre-treated cohort and whether they were present prior to therapy initiation or arose following exposure to therapy. To confirm the evolution of each mutation, longitudinal sampling would be required.

The majority of the DDR gene mutations were subclonal based on low VAFs, there was one heavily pre-treated case which harboured 2 concurrent *ATM* mutations suggestive of parallel evolution. Although overall definitive evidence of parallel evolution was not seen in the remainder of the pretreated cases, which is another pattern that has been described at the time of

acquired resistance [160]. The low VAFs of these mutations and their presence prior to chemotherapy initiation indicates they are not always the result of selective pressures from chemotherapy. The heavily pretreated patients had a greater degree of mutational heterogeneity such that the majority (86%) harboured subclonal mutations, compared with only 55% of treatment naïve patients. There was also evidence of clonal mutations arising in DDR genes, most commonly occurring in *RAD54L* arising in two heavily pre-treated cases. It is likely that the clonal status suggests that all cancer cells within this patient harbours this mutation.

ATM is among the most commonly aberrant genes in sporadic cancer [266]. Until recently it was thought that alterations in *ATM* confer sensitivity to PARP inhibitors (PARPi) in breast, ovarian, prostate and pancreatic cancers [267]. However recent studies have demonstrated that *ATM* mutations in fact do not exhibit sensitivity to PARPi, but putatively to ATR inhibitors [268, 269]. This is thought to occur through a synthetic lethal interaction i.e. two co-occurring genomic events that are each individually non-lethal but become lethal when they occur together [270, 271]. Resultantly, ATR inhibitors are being investigated in early phase trials in patients with solid tumours as single agents [272] ,in combination with chemotherapy (NCT04535401) and in combination with other inhibitors of core proteins involved in DDR genes e.g. PARPi [273]. Overall however, the very low VAFs and subclonal status in our cohort make it unlikely that these drugs would be clinically effective in a tumour where only small subclones harbour mutations.

Another possibility is that these mutations do not derive from the mCRC but from DNA released from other tissues in the body. I ruled out the origin of

ATM mutations from clonal haematopoiesis but did not further investigate the tissue origin of any of the mutations as the lack of enrichment in chemo-refractory cases and the characteristics of the mutations did not provide strong support for driver status. High incidences of *ATM* mutations have also been reported in treatment refractory prostate cancer [270, 274]. There is no documented history of prostate cancer in any of these 11 patients in this cohort, although majority (9/11) were male patients with a median age of 63 years. *ATM* mutation load is much lower in early prostate cancer and therefore it is not thought that the *ATM* mutations are derived from occult prostate cancers in this cohort.

The main limitation to the work presented in this chapter was the inability to sequence longitudinal samples, in order to confirm the timing of when suspected resistance mutations arose. Serial cfDNA samples pre- and post-treatment would track the emergence of subclonal mutations and whether these developed as a result of selective pressures exerted by cytotoxic therapy. A limitation of targeted sequencing is that the breadth is limited and therefore it is not possible to perform discovery work to identify novel genetic drivers when the candidate resistance driver genes are unknown or where there could be a large number of genes involved. It could be that the genes within the target panel are not major contributors to resistance or that chemotherapy resistance is not due to genetics alone, which would require alternative methods to study e.g. tumour tissue biopsies or cfDNA methylation analysis. For the purpose of detection of novel candidate genes I investigated the use of deep whole exome cfDNA sequencing technology. This will not only be useful to investigate chemotherapy resistance but could be applied broadly to interrogate genetics of acquired resistance for novel

targeted therapies or immunotherapy agents where candidate resistance drivers are unknown.

In order to identify resistance drivers which may be subclonal, the technology needs to be sensitive. DuDe error correction technique previously described increases the sensitivity of cfDNA-seq, however I wanted to investigate factors that influence duplex retention which can be optimised for improved error correction without MBCs. Losses of the forward or the reverse strand of double stranded DNA molecules can occur during library preparation or sequencing as described previously [117]. Therefore I reanalysed data from the optimisation experiments in Chapter 3 in order to improve understanding of how to maximise the fraction of duplex DNA strands retained for sequencing. Through this, I discovered that there is a sample-specific component in determining the duplex retention rate, which may be due to DNA strand damage. Thus future experiments can be performed to optimise sample intrinsic factors such as incorporating the use of DNA repair kits to correct potential nicks or use of increasing concentrations of the molecular crowding agent polyethyleneglycol (PEG) to bring cfDNA strands and enzymes in closer proximity to increase the probability of interaction. Additionally, use of increasing concentrations of DNA ligase or splitting the PCR reaction to overcome saturation of primers and PCR bias may encourage duplex molecule retention.

Results displayed in this chapter provide proof of concept that ultra-deep whole exome sequencing from low input DNA amount is feasible and that error correction can be performed using duplex calling with or without molecular barcodes (DuDe). Further analysis is required to determine the

sensitivity and specificity of this technology, which is ongoing within my host lab as part of a large European Research Council funded project. This should lead to novel insights into the mechanisms of cancer evolution and inform therapeutic strategies that can prevent resistance evolution or treat heterogeneous cancers more effectively.

CHAPTER 6: RATIONALE OF IMMUNOTHERAPY COMBINATION TO IMPROVE OUTCOMES IN EARLY STAGE GASTROESOPHAGEAL ADENOCARCINOMA (GOA): ICONIC CLINICAL TRIAL DEVELOPMENT.

6.1 Introduction

GOAs are the third most common cause of cancer related death in the Western world [125]. In the UK there are around 9,000 new oesophageal cancer cases diagnosed per year and 7,000 new gastric cancer cases diagnosed with 5-year survival rates across all stages of only 15% [275, 276]. Even with early stage resectable disease the prognosis with surgery alone is poor, with a 5-year OS rate of approximately 25% although this can be increased to 35% with perioperative chemotherapy [277]. Thus, multimodality combination therapies are required to improve outcomes, comprising neoadjuvant or perioperative chemotherapy, radiotherapy or chemoradiotherapy for patients with locally advanced disease [278, 279]. Despite the progress in the management of locoregional disease with multimodality treatment strategies, GOAs continue to cause substantial mortality. This reflects the diagnostic and treatment challenges associated with this heterogeneous disease. Advances in treatment of this disease have been limited compared with advances in other common cancers including colon, lung and breast cancers.

Overall GOA is a highly aggressive and heterogeneous disease in which targeted therapies including *EGFR*, *FGFR*, *MET* and *VEGF* tested in phase

II and III trials have not improved outcomes [6-13]. Furthermore resistance develops rapidly to systemic therapies. The most effective treatment has been multimodality therapy with multiagent chemotherapy or chemoradiotherapy, perhaps because combinatorial therapies overcome intratumour heterogeneity to a degree by targeting multiple subclones with distinct drug sensitivities.

6.2 Current treatment guidelines for early stage disease

Surgical resection of early stage GOA is potentially curative, however most patients relapse following resection alone and therefore combined modality treatment involving chemotherapy and surgery are standard of care for stage I, II and III disease (see Tables 1.4 and 1.5, section 1.6.1 staging of GOAs) [278]. Chemotherapy is delivered in a perioperative approach with 3-4 cycles preoperatively followed by surgery and then a further 3-4 cycles of chemotherapy postoperatively in patients have adequate performance status and can tolerate further chemotherapy.

Perioperative chemotherapy for GOAs has been shown to improve survival in two landmark clinical trials: MAGIC [130] and the French FNCLCC/FFCD study showed a similarly improved 5-year OS in patients treated in the perioperative chemotherapy groups [131] (see Table 1.6 for trial and regimen details). Recently, a significant further increase in OS has been shown in a phase III trial for perioperative FLOT compared to ECF/ECX. At a median follow up of 43 months FLOT was associated with a statistically improved OS of 50 vs. 35 months (HR 0.77, 95% CI 0.63–0.94) and 3-year OS rate of 57% vs. 48% compared to ECF [134]. Based on these results, FLOT regimen is

widely considered a new systemic perioperative treatment standard of localised GOAs.

Despite the development of better therapies for early stage GOAs, there remains an urgent need to further improve the poor survival rates by developing innovative therapeutic combinations, ideally using non-cross resistant therapies that hinder the evolution of drug resistance.

6.3 Improving perioperative treatment approach for resectable GOAs- drawing on lessons learned from prior studies

Several recent unsuccessful attempts have been made to improve the efficacy of perioperative systemic treatment in resectable gastroesophageal cancers. For example, the addition of molecularly targeted agents to chemotherapy have failed to improve PFS or OS (Table 6.1). The multi-centre randomised phase II-III ST03 trial failed to show an OS, PFS or disease free survival (DFS) benefit by adding the anti-angiogenic antibody bevacizumab to perioperative epirubicin, cisplatin and capecitabine (ECX) chemotherapy [280]. Another large multi-centre phase II study combining the EGFR-inhibiting monoclonal antibody panitumumab with ECX chemotherapy failed to reach its primary endpoint of down-staging locally advanced GOAs [281]. Likewise, the use of the anti-EGFR monoclonal antibodies have not shown any advantage in the first-line advanced gastric cancer setting. The addition of cetuximab to cisplatin and capecitabine showed no OS benefit in the EXPAND study, and the REAL-3 trial of panitumumab in combination with epirubicin, oxaliplatin and capecitabine (EOX) showed a potential detrimental OS effect of panitumumab compared to the chemotherapy alone arm [282, 283].

The *MET* signaling pathway is aberrantly activated in GOAs via gene amplification, receptor overexpression, or upregulation of stromal ligand production of hepatocyte growth factor (HGF). *MET* copy number gain has been reported in up to 10% of patients [178]. The RILOMET-1 trial investigated rilotumumab (an anti-MET/HGF inhibitor) vs. placebo in combination with chemotherapy ECX as first line therapy in *MET*-positive advanced gastric cancer. OS and PFS were significantly worse in the rilotumumab arm [284]. A further negative trial of the *MET* targeted drug onartuzumab was in the METGastric trial, which showed no difference in OS in the intention to treat population vs. the *MET* 2+/3+ patients [285].

The only monoclonal antibodies that have successfully improved outcomes for patients with advanced GOA are ramucirumab, as monotherapy or in combination with paclitaxel in the second line setting [286, 287] and trastuzumab in combination with cisplatin and 5FU in the first line setting in patients with *HER2* amplified tumours [288]. A subsequent conducted phase II multi-centre study of perioperative oxaliplatin, capecitabine and trastuzumab in *HER2* positive resectable GOAs [289] demonstrated an improvement in 18-month DFS, its primary endpoint [290]. The action of *HER2* targeted agents, trastuzumab and pertuzumab, are currently being investigated in an international randomised phase II trial in the perioperative setting [291]. However, consistent with the disease biology, only patients with *HER2* amplified GOAs are included into this trial, limiting the benefit that may be achieved to ~15% of GOA patients whose tumours overexpress this biomarker.

Trial	Number of patients	Patient population	Treatment	Efficacy endpoints	Reference
STO3 (phase II/III)	1063	Resectable oesophagogastric adenocarcinoma	Perioperative chemotherapy alone vs. perioperative chemotherapy plus bevacizumab (3 pre-operative and 3 post-operative cycles of i.v. epirubicin 50 mg/m² on day 1+ cisplatin 60 mg/m² on day 1 and oral capecitabine 1250 mg/m² on days 1-21 +/- 7.5mg/kg i.v. bevacizumab on day 1 of every cycle of chemotherapy and for 6 further doses q3w following chemotherapy, as maintenance treatment)	**3 year OS**: 50.3% (95% CI 45.5-54.9) in the chemotherapy alone group vs. 48.1% (43.2-52.7) in the chemotherapy plus bevacizumab group (HR 1.08, 95% CI 0.91-1.29; p=0.36)	Cunningham et al, 2017
NEOPECX (phase II)	160	Locally advanced oesophagogastric adenocarcinoma	Perioperative chemotherapy alone vs. perioperative chemotherapy plus panitumumab (i.v. epirubicin 50 mg/m² and cisplatin 60 mg/m² on day 1 plus capecitabine 625 mg/m² orally twice daily on days 1–21, q3w +/- panitumumab 9 mg/kg i.v. on day 1, q3w)	**PFS:** 27.7 in the chemotherapy + panitumumab group vs. 33.5 months for the chemotherapy alone group (HR 1.19, 95% CI:076-1.88; p=0.45) **3 year OS**: 49% in ECX+P group vs. 62% in ECX group (HR 1.37, 95% CI: 0.84-2.25; p=0.2)	Stahl et al, 2018
EXPAND (phase III)	904	Locally advanced unresectable or metastatic adenocarcinoma of the stomach or GOJ oesophagogastric adenocarcinoma	Chemotherapy alone vs. chemotherapy plus cetuximab (twice daily capecitabine 1000mg/m² on days 1-14 and i.v. cisplatin 80mg/m² on day 1 q3w +/- cetuximab 400mg/m² initial infusion on day 1 followed by 250mg/m² per week thereafter)	**PFS**: 5.6 months (95% CI 5.1-5.7) in the chemotherapy alone group vs. 4.4 (4.2-5.5) months in the chemotherapy plus cetuximab group (HR 1.09, 95% CI 0.92-1.29; p=0.32)	Lordik et al, 2013
REAL-3 (phase III)	553	Metastatic or locally advanced oesophagogastric adenocarcinoma	Chemotherapy alone vs. chemotherapy plus cetuximab (eight 21-day cycles of epirubicin 50mg/m² and oxaliplatin 130mg/m² on day 1 and capecitabine 1250mg/m² per day on days 1-21) vs. modified dose EOC plus panitumumab (epirubicin 50mg/m² and oxaliplatin 100mg/m² on day 1, capecitabine 1000mg/m² on days 1-21 and panitumumab 9mg/kg on day 1)	**OS:** 11.3 months (95% CI 9.6-13.0) for EOC group vs. 8.8 months (7.7-9.8) for EOC + P group (HR 1.37, 95% CI 1.07-1.76; p=0.013)	Waddell et al, 2013
RILOMET-1 (phase III)*	609	Unresectable locally advanced or metastatic MET-positive GOJ (MET positive tumours: ≥25% tumour cells with membrane staining of ≥1+ staining intensity) Trial terminated early	Rilotumumab 15mg/kg i.v. or placebo in combination with chemotherapy (epirubicin 50mg/m² and cisplatin 60mg/m² i.v. on day 1, capecitabine 625mg/m² twice daily on days 1-21) q3w for up to 10 cycles.	**OS:** 8·8 months (95% CI 7·7-10·2) in the rilotumumab group vs. 10·7 months (9·6-12·4) in the placebo group (stratified HR 1·34, 95% CI 1·10-1·63; p=0·003)	Catenacci et al, 2017
METGastric (phase III)	562	HER2 negative, MET positive advanced, unresectable GOA (MET positive tumours: ≥50% tumour cells with membrane staining of ≥1+ staining intensity by IHC) Trial terminated early	Chemotherapy + onartuzumab or placebo (oxaliplatin 85 mg/m² i.v.on day 1, fluorouracil 400 mg/m² i.v. bolus followed by 2400 mg/m² over 48 hours starting on day 1, leucovorin 400 mg/m² i.v. day 1 +/- onartuzumab 10 mg/kg i.v. on day 1) q2w	**OS:** 11.3 months for FOLFOX + placebo group vs. 11.0 months for FOLFOX + onartuzumab group (HR 0.82, 95% CI, 0.59-1.15; stratified p=0.24), **PFS:** 6.8 vs 6.7 months (HR 0.90, 95% CI 0.71-1.16; stratified p=0.43)	Shah et al, 2017
REGARD (phase III)	355	Advanced GOA and disease progression after 1st line platinum-containing or fluoropyrimidine-containing chemotherapy	Best supportive care + ramucirumab (8mg/kg i.v. q2w) vs. best supportive care + placebo	**6 month OS:** 41·8% (35·4-48·1) in the ramucirumab group vs. 31·6% (23·2-40·2) in the placebo group (HR 0·767, 95% CI 0·598-0·984; p=0·037) **DFS:** 4·2 months (2·8-8·1) vs 2·9 months (2·7-4·3; p=0·036)	Fuchs et al, 2014
RAINBOW (phase III)	655	Advanced GOA and disease progression on or within 4 months after first-line chemotherapy	Ramucirumab 8 mg/kg or placebo i.v. on days 1 and 15, plus paclitaxel 80 mg/m² i.v. on days 1, 8, and 15 q4w	**OS:** 9·6 months (95% CI 8·5-10·8) in the ramucirumab group vs. 7·4 months (95% CI 6·3-8·4) in the placebo group (HR 0·807, 95% CI 0·678-0·962; p=0·017)	Wilke et al, 2014
ToGA (phase III)	594	HER2 overexpressed (IHC 3+ or FISH positive) advanced GOA	Chemotherapy alone vs. chemotherapy plus trastuzumab (cisplatin 80 mg/m² i.v. on day 1 plus either capecitabine 1000 mg/m² orally twice daily for 14 days, or fluorouracil 800 mg/m² per day continuous i.v. infusion on days 1-5 +/- trastuzumab 8 mg/kg i.v. on day 1 of the first cycle, followed by 6 mg/kg q3w)	**OS:** 13.8 months (95% CI 12-16) in trastuzumab plus chemotherapy group vs. 11.1 months (10-13) in chemotherapy alone group (HR 0.74; 95% CI 0.60-0.91; p=0.0046	Bang et al, 2010
NEOHX (phase II)	36	HER2 overexpressed (IHC 3+ or FISH positive) resectable GOA	Chemotherapy plus trastuzumab (capecitabine 1000 mg/m² twice daily for 14 days, oxaliplatin 130 mg/m² i.v. on day 1, and trastuzumab 8 mg/kg i.v. on day 1 of the first cycle, followed by 6 mg/kg q3w)	**2 year DFS:** 60% (95% CI, 41–75%) **2 year OS:** 76% (95% CI, 58–87%)	Rivera et al, 2021
INNOVATION (phase II)	215	HER2 overexpressed (IHC 3+ or FISH positive) resectable GOA	Perioperative chemotherapy + trastuzumab + pertuzumab	Trial in progress	Wagner et al, 2019

Table 6.1: Pivotal trials for targeted therapies in GOA
[280-283, 292-298]

Taken together, more effective perioperative systemic treatment regimens are needed to improve the poor outcomes in patients with operable GOAs. However oncogene targeted therapy, with the exception of anti-HER2 antibodies, are unlikely to be successful given the lack of efficacy in metastatic disease and the intratumoural heterogeneity of *MET*/*EGFR* amplifications in this disease.

6.4 Checkpoint inhibitors as a promising new systemic treatment for gastric and oesophageal adenocarcinomas.

Targeting immune checkpoint to activate the host immune system against cancer cells is an approach that has rapidly evolved in the recent years [299]. Programmed cell death receptor 1 (PD-1) is an inhibitory immune checkpoint receptor that is expressed on a variety of immune cells, including cytotoxic T-cells of the adaptive immune system. The PD-1 ligand (PD-L1) can be upregulated by cancer cells in response to T-cell recognition and inhibits their cytotoxic activity through binding of PD-1 on the T-cell surface, allowing cancer cells to evade destruction by the immune system. One approach to overcome this inactivation of T-cells by cancers has been to target the PD-1/PD-L1 interaction with monoclonal antibodies [300].

There has been a paradigm shift in the treatment of solid malignancies with the approval of a number of immune checkpoint inhibitors in melanoma, lung, bladder and renal cancers and a number of trials using perioperative immune checkpoint inhibitors are now ongoing in these indications [301-304]. PD-L1 overexpression has been demonstrated on up to 40% of gastric cancers and 43% of oesophageal cancers [305-307] suggesting a role of immune checkpoint activation for immune evasion in this cancer type. A number of studies have reported some efficacy of checkpoint inhibitors in advanced GOAs:

Trial	Number of patients	Patient population	Treatment	Efficacy endpoints	Reference
KEYNOTE-012 (phase Ib)	39	PD-L1-positive (≥1% by IHC), recurrent or metastatic GOA	Pembrolizumab i.v. 10 mg/kg q2w	**ORR:** 21% (95% CI 10-39%) **OS:** 11.4 months (95% CI 5.7 months – not reached)	Muro et al, 2016
KEYNOTE-028 (phase Ib)	23	PD-L1 positive (≥1% by IHC) metastatic squamous cell carcinoma or adenocarcinoma of the oesophagus or gastroesophageal junction who have progressed on >2 prior lines of therapy	Pembrolizumab i.v. 10 mg/kg q2w	**ORR:** 30% (95% CI, 13% to 53%) **PFS:** 1.8 months (95% CI, 1.7 to 2.9 months) **OS:** 7.0 months (95% CI, 4.3 to 17.7 months)	Doi et al, 2018
KEYNOTE-061 (phase III)	395	PD-L1 positive (CPS ≥1) advanced GOA who have progressed on first line therapy	Pembrolizumab 200 mg i.v. q3w or paclitaxel 80 mg/m^2 i.v. on days 1, 8, and 15 of 4-week cycles.	**OS:** 9·1 months (95% CI 6·2–10·7) for pembrolizumab group and 8·3 months (95% CI 7·6–9·0) for paclitaxel group (HR 0·82, 95% CI 0·66–1·03; one-sided p=0·0421) **PFS:** 1·5 months (95% CI 1·4–2·0) for pembrolizumab and 4·1 months (3·1–4·2) for paclitaxel (HR 1·27, 95% CI 1·03–1·57)	Shitara et al, 2018
CHECKMATE-032 (phase I/II)	160	Locally advanced or metastatic chemotherapy–refractory oesophagogastric cancer	Nivolumab 3 mg/kg (NIVO3) i.v. q2w or nivolumab 1 mg/kg plus ipilimumab 3 mg/kg (NIVO1 + IPI3) q3w for four cycles or nivolumab 3 mg/kg plus ipilimumab 1 mg/kg (NIVO3 + IPI1) q3w for four cycles. All combination regimens were followed by NIVO3 every 2 weeks	**ORR:** were 12% (95% CI, 5% to 23%) for NIVO3, 24% (95% CI, 13% to 39%) for NIVO1 + IPI3, and 8% (95% CI, 2% to 19%) for NIVO3 + IPI1 **12-month PFS:** 8%, 17%, and 10%, respectively **12-month OS:** 39%, 35%, and 24%, respectively	Janjigian et al, 2018
ONO-4538-12 ATTRACTION-2 (phase III)	493	Metastatic or recurrent chemo-refractory GOA	Nivolumab 3 mg/kg i.v. q2w or placebo	**OS:** 5.26 months (95% CI 4.60-6.37) in the nivolumab group and 4.14 months (3.42-4.86) in the placebo group (HR 0.63 95% CI 0.50-0.78; p<0.0001). **12 month OS:** 26.2% (95% CI 20.7-32.0) in the nivolumab group and 10.9% (6.2-17.0) in the placebo group	Kang et al, 2017
JAVELIN (phase Ib)	150	Locally advanced or metastatic GOA and prior treatment with >1 line chemotherapy [2 subgroups: those without disease progression received avelumab as 1 L switch maintenance (1 L-mn subgroup), and those with disease progression received avelumab as 2 L treatment (2 L subgroup)]	Avelumab 10 mg/kg i.v. q2w	**PFS:** 2.8 months (95% CI, 2.3–4.1) in 1 L-mn subgroup and 1.4 months in 2 L subgroup (95% CI, 1.3–1.5). **OS:** 11.1 months (95% CI, 8.9–13.7) in 1 L-mn subgroup and 6.6 months (95% CI, 5.4–9.4) in 2 L subgroup	Chung et al, 2019
MEDI4736 multi-arm trial (phase I)	10-20	Metastatic GOA	Durvalumab 10 mg/kg i.v. q2w	**ORR:** 25%	Segal et al, 2014

Table 6.2: Pivotal trials of checkpoint inhibitors in GOA [308-314]

These studies show that immune checkpoint inhibition is a promising option in advanced GOAs and a small proportion of patients can achieve durable disease control from single agent treatment. However it is not entirely clear why some patients respond well and others do not. This may be due to several factors including heterogeneous immunogenicity within tumours and also between tumours and the different immunosuppressive mechanisms that may hamper effective tumour control by the host immune cells. In order to improve the efficacy of checkpoint inhibitors in a larger proportion of GOAs,

we reasoned that consideration of the following factors would together allow improved response rates:

i) combining checkpoint inhibitors with pro-immunogenic agents which should act synergistically
ii) using combination therapies as early as possible when the tumour load is low should improve responses
iii) extrapolating from data in melanoma, where patients with lower baseline disease burden have improved clinical and pathologic responses to immunotherapy and survival compared with higher baseline tumour burden. This is likely because smaller tumours are likely to have had a shorter time to acquire immune evasion mechanisms [315].

6.5 Rationale for combining checkpoint inhibitor with chemotherapy in resectable GOAs

The ultimate goal for cancer immunotherapy is to initiate cancer cell kill and target the rate-limiting step to promote ongoing immune response within the tumour. Melanoma is the prototypical tumour for demonstrating immunogenicity and responses to checkpoint inhibition. A common feature among cancers with a higher probability of response to immunotherapies is the prevalence of somatic mutations in their genomes. GOAs have modest mutational loads compared to melanomas [316], however given that some tumours with low mutation load e.g. renal cancers do respond to checkpoint inhibitors, it remains an open question as to which mechanisms may enable this. Furthermore, GOAs are not immunologically “hot” tumours. The distinction between “hot” and “cold” tumours is characterised by the presence

or absence of infiltrating T-cells, more specifically CD8-positive cells in the tumour microenvironment [317]. This can be due to the lack of tumour antigens or reduced T-cell priming/activation and impaired trafficking of T-cells to the tumour mass due to spatial barriers and/or immunosuppressive factors that avoid T-cell infiltration [318]. Prior studies have demonstrated that checkpoint inhibitor therapy is more efficacious in immunologically hot tumours as the tumour microenvironment already has a repertoire of exhausted PD-1+ T-cells ready to be expanded [319]. A cold tumour has little T-cell infiltration, which is either completely absent or localised to the periphery of the tumour and can be a key factor involved in the initial resistance to checkpoint inhibition. A number of therapeutic approaches are being trialled to overcome the absence of T-cell infiltration in a tumour to transform a cold tumour into an immunologically hot tumour. These include combination treatments with oncolytic viruses or vaccinations and induction chemotherapy or radiotherapy to induce immunogenic cell death (ICD), which is cell death that induces chronic exposure of damage-associated molecular patterns (DAMPs) in the tumour microenvironment and can lead to long term antitumour activity [318].

For an anti-cancer immune response to lead to effective killing of cancer cells a series of steps must be initiated and completed in a process known as the cancer-immunity cycle (Figure 6.1) [320]. Effective progression through all steps of the cycle, from tumour antigen release (1) to uptake and presentation by antigen presenting cells (APCs) (2) which then prime and activate T-cells (3), over trafficking of these activated T-cells to the tumour (4) and their infiltration into the tumour (5), followed by recognition of cancer cells by these T-cells (6) are all essential steps before cancer cell killing (7) can occur.

Therefore failure to spontaneously progress through the cancer immunity cycle may be the key hurdle that precludes better responses to immunotherapies in tumours. Thus, understanding which step(s) of the cancer-immunity cycle is non-functional is crucial.

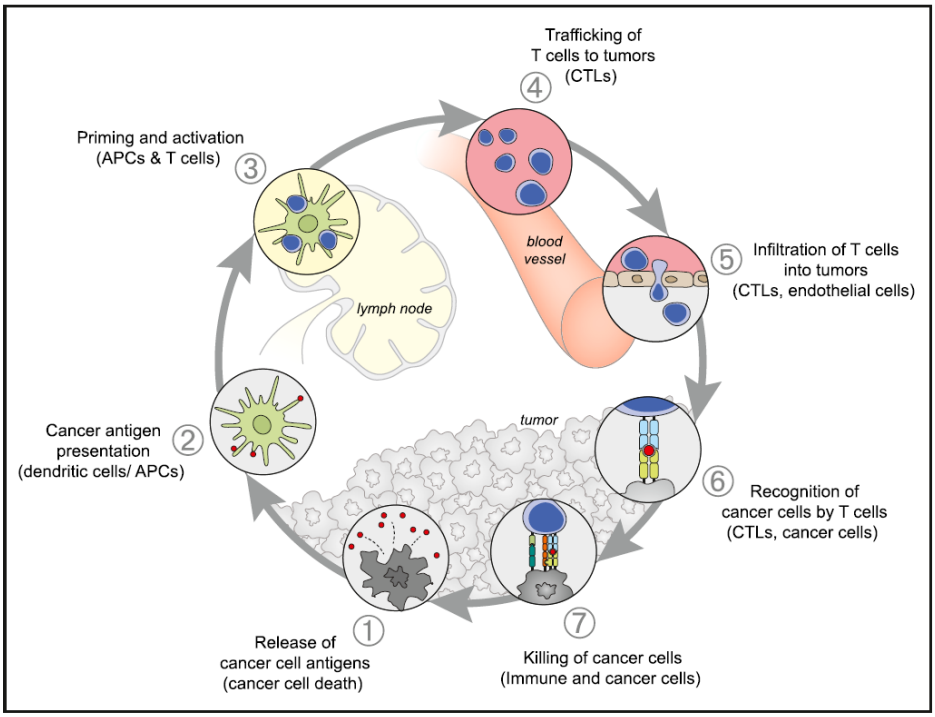


Figure 6.1: The cancer-immunity cycle

[320]. The cancer-immunity cycle is a self-propagating process, leading to an accumulation of immune-stimulatory factors that can amplify T-cell responses. The cycle is characterised by inhibitory factors that lead to immune regulatory feedback mechanisms, which can limit the immune-stimulation. The cycle is divided into seven steps (1-7), starting with the release of antigens from the cancer cell and ending with the killing of cancer cells.

The challenge is to optimally address these steps not singly but as a group. I hypothesised that the cancer-immunity cycle can be initiated and propagated, even in localised tumours with low mutation burden, through combining chemotherapy with immune checkpoint inhibitors to generate synergies, which will generate sufficient immune stimulation to propagate the cycle. In so doing, we wanted to facilitate progression through most of the cancer-immunity cycle steps.

We reasoned that by combining checkpoint blockade with chemotherapy agents that can promote ICD in early disease, we could generate synergies to improve response rates. This should occur through the following mechanisms [321]:

- Once ICD is triggered by chemo-immunotherapy, it initiates a strong antitumour response through a cascade of events [322]. First, cancer cells release antigens through a type of cell death that is particularly immunogenic, based on the release of high-mobility-group box 1 (HMGB1) and ATP [137]
- cells undergoing ICD upregulate the expression of the 'eat me' signal calreticulin (CALR) on the cell surface, which stimulates immune activation and in particular enables the phagocytosis of these dying cells by dendritic cells, which can then more effectively prime T-cells [323]
- increased secretion of interferon gamma (IFN-γ) in response to immune activation in the tumour could also promote antigen processing within cancer cells and the expression of antigen presenting MHC class I molecules on the surface of cancer cells. This should increase the visibility of cancer cells to tumour reactive T-cell. Together, HMGB1, ATP, CALR and IFN-γ are known as DAMPs [321] and failure of cancer cells to emit one or more of these DMAPs completely compromises ICD [324]. DAMP molecules are not accessible by the immune system in physiological conditions but are released or exposed in a coordinated fashion during cell death, leading to recruitment of antigen presenting cells (APCs) to sites of ICD [325].

- cell death will also trigger inflammation and cytokine release in the tumour tissue [326]. An inflammatory environment can promote antigen uptake and presentation by APCs, thus stimulate an adaptive immune response to cancer cells [326, 327]
- the chemokine and cytokine milieu established by an inflammatory response could help to attract cells of the immune system into the cancer mass and generate enough space within these tumours for T-cells to migrate and distribute
- the deposition of anti-PD-L1 antibodies on tumour cells and the subsequent activation of antibody-dependent cell-mediated cytotoxicity (ADCC) *in vitro* [328, 329] may direct immune cells to the cancer cells. Unlike other anti-PD-L1 monoclonal antibodies, avelumab was designed as an IgG1 monoclonal antibody to trigger ADCC against tumour cells and has been shown to be a mechanism to eliminate triple negative breast cancer cells *in vitro* [329].
- finally, by reducing the cancer load and by improving the nutritional status of some patients through the reduction of the degree of luminal obstruction in the oesophagus or stomach, chemotherapy may also reduce systemic immunosuppression.

I aimed to investigate through literature search which chemotherapy agents would be ideal candidates for combination with a checkpoint inhibitor to drive progression through the cancer-immunity cycle and trigger ICD. These agents should promote T-cell priming which may overcome the limited spontaneous T-cell recognition of moderately immunogenic cancers. Combination of ICD-inducing drugs with checkpoint inhibitors is known to enhance cancer-immunotherapy efficacy in mice however it is unknown

whether this confers benefits in patients [327]. I performed a literature search to investigate chemotherapy agents used in GOAs that are either known to induce ICD or be associated with alternative pro-immunogenic effects (Table 6.3):

Drug	Ability to induce immunogenic cell death	Other immune effects	Reference
Cisplatin	N	Elicits cytotoxic T lymphocyte responses *in vivo* M6P receptor upregulation, rendering tumour cells sensitive to granzyme –B killing	Zitvogel et al 2013 Hato et al 2014
Oxaliplatin	Y	Increased CD8+ T cell:T reg ratio Promotes DC maturation and T-cell proliferation	Pfirschke et al 2016 Hato et al 2014
5FU	N	MDSC depletion in the spleen and tumour microenvironment In combination with RT promotes tumour infiltrating cytotoxic T lymphocytes	Vincent et al 2010 Wang et al 2018
Carboplatin	N	Nil recognised	
Paclitaxel	N	Increases expression of MHC class I molecules Promotes differentiation of immunosuppressive MDSCs into DCs Potentiates IFN-induced PD-L1 cell surface expression in breast cancer cells	Zitvogel et al 2013 Michels et al 2012 Zhang et al 2008
Docetaxel	N	MDSC depletion in the spleen and tumour microenvironment Upregulation of calreticulin expression. Sensitisation of tumour cells to the lysis of CD3+ and CD56+ cytokine induced killer cells	Kepp et al, 2014, Sevko et al 2013 Wang et al, 2015
Anthracyclines	Y	Deplete circulating T-regs	Kroemer et al, 2014, Tesniere et al 2010, Zitvogel et al 2013, Bezu et al 2015
Irinotecan	Unknown	Deplete circulating T-regs	Wang et al 2018

Table 6.3: Summary of literature search to identify chemotherapy agents used in GOA that can induce immunogenic cell death and hence candidates for chemoimmunotherapy combination

MDSCs: myeloid derived suppressor cells. M6P: mannose-6-phosphate. DC: dendritic cell. ICD: immunogenic cell death. [321, 324-327, 330-337].

Based on this literature search, the two agents that can induce ICD are oxaliplatin and anthracyclines, and were therefore considered as potentially favourable combinations with immunotherapy. However the efficacy of anthracycline agents remain debated as both MAGIC [130] (anthracycline based regimen) and FNCLCC/FFCD ACCORD [131] (not anthracycline based regimen) trials demonstrated similar results. The MRC OE05 trial [338] showed that treatment with four preoperative cycles of ECX does not improve survival in patients with resectable oesophageal adenocarcinoma compared to two cycles of CF thus making anthracycline-free regimens the treatment of choice. Thus, we reasoned that if a two drug combination (platinum + 5FU) is likely to be as effective as a three drug combination (with the addition of epirubicin), then using a two drug combination will be easier to combine with avelumab as the toxicity profile is likely to be better. Oxaliplatin has shown equivalent efficacy as cisplatin in metastatic GOA and was used widely in early stage preoperative treatment [339]. Thus the two drug combination of 5FU and oxaliplatin (FOLFOX), which included an ICD inducer and had largely non-overlapping toxicities with avelumab was the first combination selected as the chemotherapy backbone for this trial.

6.6 Trial design

Based on the above considerations, I designed the ICONIC (**I**mmuno-**C**hemotherapy in **O**perable oesophageal a**N**d gastr**I**c **C**ancer) trial. This is a single-arm, single centre open-label, phase Ib/II clinical trial of perioperative

chemoimmunotherapy initially designed with combination FOLFOX-A (folinic acid 400 mg i.v. infusion day 1, oxaliplatin 85 mg/m^2 i.v. infusion day 1, fluorouracil 400 mg/m^2 i.v. bolus day 1, fluorouracil 2400 mg/m^2 over 48 hours i.v. with avelumab 10 mg/kg i.v. on day 1) in operable GOA. Each patient receives 4 cycles of down staging FOLFOX-A followed by surgery and those that have adequate performance status will receive a further 4 cycles in the postoperative period.

The primary objectives of the ICONIC trial are to evaluate the safety, efficacy and toxicities of FOLFOX-A combination and to explore biomarkers of perioperative chemoimmunotherapy benefit. As this is a novel combination treatment being used in the potentially curative setting, the study was designed in two phases; the safety run-in phase and the efficacy phase. The safety run-in phase was designed to assess the safety of adding avelumab to FOLFOX chemotherapy in patients with operable GOA and to recommend a maximum administered dose (MAD) of avelumab for use in the efficacy phase of this trial. The safety run-in is conducted using a 3+3 dose finding design. After completion of the safety run-in phase, patients will receive the established MAD of avelumab in combination with FOLFOX. Enrolment into the efficacy phase will continue until a total of 38 patients have been treated at the MAD and undergone surgical resection.

6.7 Endpoints

Primary Endpoint:

Safety run-in phase:

To recommend a MAD for use in the efficacy phase of this trial.

Efficacy phase:

- To assess the pathological complete response (pCR) rate. We aim to increase the pCR rate after perioperative treatment from <5% to 20% through the addition of avelumab to perioperative FOLFOX chemotherapy.

Secondary Endpoints:

- PFS
- OS
- The safety and tolerability profile of perioperative FOLFOX-A in patients with operable GOA in relation to drug toxicity, perioperative immunomodulation effects and surgical safety outcomes in the intra-operative, early and late postoperative phases
- Efficacy endpoints:
  - Mandard pathological regression grading assessed in the resection specimen
  - Radiological response rate assessed at the pre-operative scan using RECIST 1.1 criteria, additional radiological response criteria may also be applied
  - R0 resection rate

6.8 Statistical considerations

A total of 38-44 patients plus additional patients (estimated to be a maximum of 9) to replace those patients who commenced study treatment but did not undergo surgery will be treated within this study, with the aim of increasing pCR rate from <5% to 20%. In the safety run-in phase 3 to 12 patients will be recruited in a 3+3 dose finding design. The efficacy phase will continue recruiting until 38 patients have been treated with FOLFOX-A at the MAD and

have undergone surgical resection. Using an A'hern single stage design to rule out a lower limit of a 5% pCR rate (p0) and to demonstrate an increase to a 20% pCR rate (p1) requires the inclusion of 38 patients and will be deemed a positive trial if 5 or more pCR events are observed.

6.9 Protocol amendment

Following the approval and opening of the ICONIC trial using combination FOLFOX-A in July 2017, the RMH standard of care for perioperative treatment of GOAs was updated shortly after based on the results of the FLOT4 chemotherapy trial data (oral presentations: Al Batran et. al. ASCO 2017 and ESMO 2017) [132, 340]. The data showed that FLOT chemotherapy (5FU, oxaliplatin and docetaxel) was superior for all efficacy endpoints including pathological complete resection rates, PFS and OS compared to previous standard of care chemotherapy without docetaxel.

Given the improved efficacy of the FLOT regimen over ECF/X in all subgroups, the rationale to change the chemotherapy backbone became clear, this also became the new standard of care treatment. The protocol amendment was in line with similar amendments that were being prepared at the time for major clinical trials that were testing the addition of targeted agents to chemotherapy in the perioperative setting e.g. INNOVATION trial of chemotherapy and trastuzumab +/- pertuzumab in the perioperative setting (NCT02205047), KEYNOTE 585 trial of chemotherapy +/- pembrolizumab in the perioperative setting (NCT03221426). Furthermore, docetaxel also has pro-immunogenic effects of depleting immunosuppressive myeloid derived suppressor cells in the spleen and the tumour microenvironment [325, 335, 341] and is known to induce tumour cell death through cytokines from T-cells

and NK-cells [337]. We hypothesised that FLOT chemotherapy, which includes the ICD-inducer oxaliplatin should boost T-cell priming and combination with the PD-L1 inhibitor avelumab (FLOT-A regimen) should unleash these T-cells, leading to synergies and improved outcomes in early stage GOA.

6.9.1 Amended protocol- statistical considerations

Following results of the phase II FLOT 4 trial, where 128 patients were treated with perioperative FLOT, 15.6% patients achieved a pCR (95% CI:10.3%-23.0%) [342]. Based on the lower bound of this confidence interval, the updated version 2 ICONIC trial aims to increase the pCR rate from the minimum expected pCR rate for perioperative FLOT chemotherapy - 10% to a superior rate of 25% through the addition of avelumab to perioperative chemotherapy. Using an A'hern single stage design for efficacy to rule out a lower limit of a 10% pCR rate (p0) and to demonstrate an increase to a 25% pCR rate (p1), requires the inclusion of 40 patients, to achieve a power of 80% and significance level α=0.05. The combination therapy of FLOT-A would be considered promising for further development (e.g. in a randomised phase III trial) if 7 or more pCR events are observed in the 40 patients.

Patients will be recruited until 40 patients have received perioperative FLOT combined with avelumab at the MAD and have undergone surgery. It is likely that more than 40 patients may need to be recruited into the trial at that dose, as not all patients will undergo surgery for any reason e.g. inoperability due to deconditioning or disease progression on treatment, and such cases will not be evaluable for the primary outcome measure (pCR). In total the study will recruit 40-46 evaluable patients in order to account for a possible dose

reduction during the 3+3 safety run-in phase of the trial. Recruitment will continue until 40 patients have been treated, at the maximum tolerated dose identified in the safety run-in phase and undergone surgery.

It is expected that accrual of 40-46 patients, plus any additional patients required to replace patients who do not undergo surgery, would take 2 years given that the Royal Marsden Hospital treats up to 80 patients with operable localised gastro-oesophageal adenocarcinomas per year.

6.10 Trial Schematic

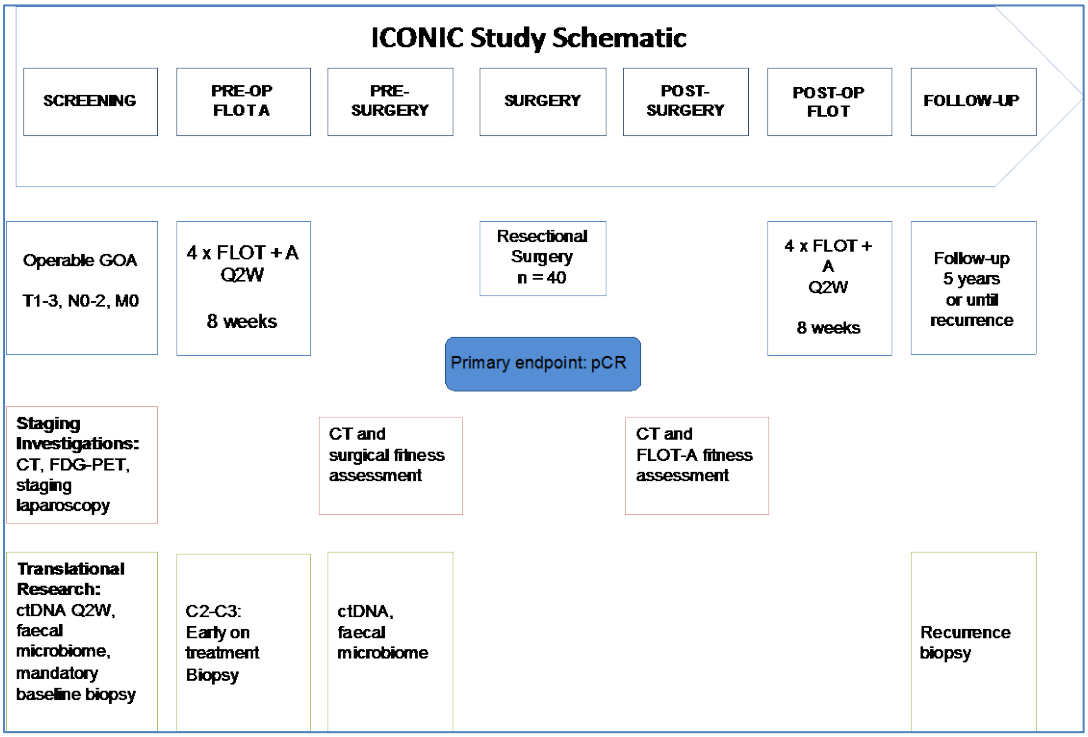


Figure 6.2: ICONIC trial schematic depicting the study design.

All radiological investigations will be performed as per standard of care. An extensive translational research program was built into the trial to make use of the unique opportunities offered in early stage GOAs to obtain biopsies and subsequent availability of the resection specimen, which should enable

detailed molecular studies of baseline characteristics and dynamic changes through chemoimmunotherapy (see section 6.14).

6.11 Rationale for evaluation of pathological tumour response

Pathological complete response, defined as the absence of cancer cells in the surgical resection specimen following neoadjuvant therapy, is strongly correlated with long term survival rates [343-346]. In the MAGIC and ST03 trials, good pathological response (TRG 1-2) scored by Mandard tumour regression grade [347] correlated with good overall survival. For example 3 year OS in the ST03 trial for patients achieving Mandard TRG 1 and 2 was 73%, Mandard TRG 3: 58% and Mandard TRG 4-5: 44% [348, 349]. As a potential surrogate for outcome, pCR could facilitate the reporting of results from phase II trials before OS data mature.

In order for the primary endpoint of the efficacy phase to be assessed systematically and reproducibly, we obtained pathology input from Dr Katharina von Loga (Consultant Molecular Pathologist, Biomedical Research Centre) and Dr Andy Wotherspoon (Consultant Histopathologist, GI Unit Royal Marsden Hospital) to develop a standardised protocol. The macroscopically visible tumour is measured and representative 5 mm thick slices of tumour and lymph nodes are obtained and embedded. A pathologist systematically examines all slides to assess for complete histopathologic response. Complete histopathologic response is defined by no vital tumour cells in the oesophagus, the stomach or in the regional lymph nodes. In cases of residual tumour, the response assessment follows criteria described by Mandard et. Al. [347].

6.12 Interim analysis to assess the safety of FLOT-A in combination with surgery

As this is the first trial assessing perioperative FLOT + avelumab treatment and one of the first trials assessing PD1/PD-L1 targeting immunotherapies in the perioperative setting in GOA, safety assessments are performed on a frequent schedule. A pre-planned interim safety analysis was built into the protocol to assess accumulating safety and complication data for novel perioperative FLOT-A treatment combination. This is carried out by an independent data monitoring committee (IDMC) and occurs after the 15th patient (40% of the planned cohort for efficacy assessment) has reached the 30th post-operative day. This interim analysis aimed to evaluate any unexpected toxicities of FLOT-A and the safety of FLOT-A in the perioperative setting.

6.13 Patient and Public Involvement of ICONIC Trial

During trial development we invited patient and public involvement (PPI) via RMH Biomedical Research Centre for Cancer (BRC). PPI involvement was sought for review of the trial patient information sheet (PIS), for oversight of the procedures, potential impact on individual patients and wider patient communities. The PPI panel was supportive of the study, which includes mandatory research biopsies and regular research blood samples. The constructive comments about trial documents were addressed to optimise the design of a comprehensible and patient-friendly PIS, and successfully gained ethical approval.

6.14 ICONIC Exploratory Translational sub-study

Novel combinations of chemo-immunotherapy represent a new paradigm of systemic anti-cancer treatment. However, predictive biomarkers for patient stratification and the molecular mechanisms regulating chemo-immunotherapy responses or resistance have not been identified in GOAs. The current limited understanding hinders the development of better combination treatments with activity against a larger proportion of GOAs.

The key questions we wanted to address through the translational study are:

i) to understand which patients benefit from FLOT-A combination

ii) why a subset of patients do not benefit from chemo-immunotherapies and

iii) identify mechanisms of resistance that are potentially targetable in order to improve combination immunotherapy trials in the future. Together this can inform more effective combination partners for immunotherapies and effective treatment allocation to patients who are likely to derive benefit.

As part of the translational research protocol tumour tissue samples, blood samples and faecal samples are collected from all patients registered into the trial (Figure 6.3). Through the longitudinal sample collection, we aim to:

1. identify candidate biomarkers in pre-treatment biopsies through assessing the mutational load (including clonal and subclonal mutations) and molecular markers (including IHC PD-L1 expression, immune cell infiltrates and RNA sequencing gene signatures) differing between those who achieve pCR vs. suboptimal response vs. no response to neoadjuvant FLOT-A

2. assess whether ICD induction (using immunofluorescence) and whether dynamic changes of immune contexture in early on-treatment biopsies correlate with pCR and whether this can identify immune inhibitory mechanisms that hinder pCR
3. understand mechanisms of acquired resistance through tumour tissue biopsies obtained at recurrence
4. perform serial ctDNA analysis to assess whether responses can be predicted early and whether we can identify those patients with very good long term survival based on clearance of ctDNA following surgery
5. assess the faecal microbiome as a biomarker of immunotherapy response.

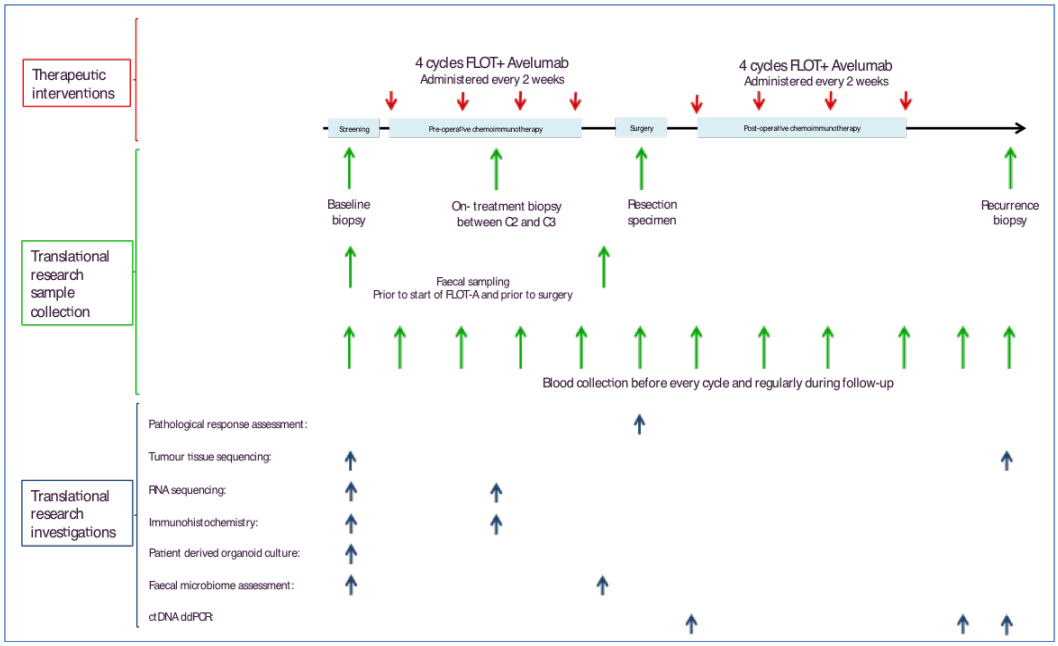


Figure 6.3: ICONIC translational sub-study schematic detailing the sample collection throughout the trial.

6.14.1 Tumour tissue analysis

Some of the key factors considered during design of the tumour tissue sample collection were:

i) Timing of the biopsies: 3 mandatory timepoints were selected for tumour tissue analysis, in addition to surgical specimen analysis

- pre-treatment: to analyse the treatment naïve tumour
- early on treatment: after 2 cycles/3 weeks of FLOT-A chemo-immunotherapy to generate patient derived organoid (PDO) cultures as *in vitro* models to assess residual viable tumour cells, which should still be present at this timepoint based on data from other trials with checkpoint inhibitors [317]. This should enable evaluation of ICD and tumour cell-immune cell interactions
- at the time of resection: to assess the changes of immune infiltrates and molecular tumour characteristics induced by combined chemotherapy and PD-L1 blockade and
- at the time of tumour recurrence after resection: an endoscopic or percutaneous image guided biopsy will be taken to interrogate mechanisms of acquired resistance after peri-operative FLOT-A.

i) Location of the serial biopsies was a consideration to minimise biological variability. In addition to intratumoural heterogeneity, further variability in tumour immune landscape can exist if the pre- and on-treatment biopsies are not derived from the same tumour lesion in the same region. This can confound interpretation of results from biopsy pairs and therefore all biopsies were taken from the invasive margin to sample active tumour and the site of the biopsy was recorded as a reference for future biopsies.

6.14.2: Candidate biomarkers for PD1/PD-L1 inhibitors

Biomarkers can be used to stratify patients predicted to not respond to current checkpoint immunotherapies in order to avoid unnecessary toxicities and instead use combination treatment strategies which could have a major

impact on patient care. So far, multiple biomarker strategies have emerged [350, 351] as being associated with response rates:

- IHC assessment of PD-L1 positive tumour cells, immune cells or both such that a higher proportion of PD-L1 positivity is associated with improved clinical outcomes in lung cancer [352, 353] and melanoma [354, 355]
- Tumour infiltrating lymphocytes (TILs) assessed by IHC, such that an increased proportion of CD8+ TILs density pre-treatment predicts for an improved survival in melanoma and lung cancers [356, 357]
- Tumour mutational burden and microsatellite instability can be established by sequencing. Tumours with high mutational burdens are more responsive to immunotherapies and indeed, melanoma, non small cell lung cancer (NSCLC) and MSI-high CRCs have the greatest number of neoantigens and are responsive to checkpoint immunotherapies [35, 358-361]. Further support for the role of somatic mutations in phase 2 data of immunotherapy in mismatch repair-deficient colorectal cancer, where the mutational burden was >20 times higher than in mismatch proficient cancers and the response rates was significantly improved [35]
- Immune gene signature assessment from the tumour and microenvironment can give insights into presence of T-cell inflamed profiles or interferon γ, which predict for improved clinical outcomes in melanoma [362, 363]. An analysis of 200 patients with non-small cell lung cancer revealed that *IFNG* gene was most correlated with response to treatment [364]

- Multiplex IHC to assess multiple protein markers on tumour cells and immune cells and their spatial relationships can be predictive of response to immunotherapy [365]
- Gut microbiome can be assessed by NGS or polymerase chain reaction (PCR). Four independent studies have shown an association between specific gut bacteria and checkpoint response in melanoma, non-small cell lung cancer, renal cell cancer and bladder cancer. The microbiome can affect anti-tumour immunity and enhance therapy induced anticancer response by increasing reactive oxygen species (ROS) production from tumour infiltrating myeloid cells or by generating bacteria-specific T-helper 17 (Th-17) responses [366]. Two *in vivo* studies have demonstrated that an intact commensal microbiota is critical for therapeutic efficacy of oxaliplatin-based [367] or cyclophosphamide-based chemotherapies [368]. The presence of *Bifidobacterium* species, *Collinsella aerofaciens* and *Enterococcus faecium* is associated with improved response to checkpoint inhibition in patients with melanoma, whereas *Bacteroides* species were enriched in non-responding patients [369]. Overall, increased microbiota diversity is associated with improved response rates [370-372]
- T-cell receptor clonality assessed by next generation sequencing of T-cell receptor β chain. A more clonal, less diverse population correlates with improved response rates [365].

The ICONIC trial collects suitable samples to interrogate these biomarkers by exome-seq, RNA-seq, IHC and stool cultures. This should identify the

most promising molecular markers for further development and validation in larger and ideally randomised trials. How samples will be analysed in the trial has been detailed below.

6.14.2.1 Exome sequencing

Pre-treatment baseline and early on treatment biopsies will be analysed by exome sequencing to investigate immunotherapy biomarkers including mutation and neoantigen load and subclonality. Furthermore, the correlation of genetic GOA subtypes such as microsatellite unstable, chromosomally unstable and genomically stable GOAs [373, 374] and presence of immune evasion mechanisms (e.g. *B2M* or *JAK1/2* mutations [375, 376]) on response will be studied.

6.14.2.2 RNA-sequencing

Tumour RNA-sequencing will be performed on biopsies to investigate the abundance of 28 immune cell subtypes based on single sample Gene Set Enrichment Analysis (GSEA). GSEA calculates separate enrichment scores for each paired sample (pre-treatment biopsy and surgical resection specimen) to represent the degree to which genes are coordinately up- or down-regulated and thus the activation or repression of corresponding biological process [377]. This will be used to test if specific immune cell types or patterns of immune cell infiltrates in the pre-treatment sample correlate with pCR or failure to respond. Immune-suppressive gene signatures are expected to be enriched or depleted through treatment in tumours with suboptimal response or pCR respectively [378]. How these dynamic changes differ between responders vs. progressors will identify potential immune-suppressive mechanisms (e.g. immunosuppressive cell accumulation or alternative checkpoint upregulation) for co-targeting in future clinical trials.

6.14.2.3 Immunohistochemical analysis of PD-L1 expression

Pre-treatment and on-treatment (week 3) biopsies will be stained for PD-L1 expression on tumour and immune cells. We will use the combined positive score (CPS), which is calculated by summing the number of PD-L1 stained cells (tumour cells and immune cells) as a percentage of the total number of viable cells, with a cut off of 1 or more confirming PD-L1 positivity [379]. This score has been utilised as a novel scoring method by the US Food and Drug Administration in approving pembrolizumab as a therapy for GOAs.

Although patients are not selected into the ICONIC study based on PD-L1 status, I hypothesise that PD-L1 positive tumours will have improved response to FLOT-A treatment and thus be positively correlated with pathological response characteristics. However, PD-L1 negative gastric tumours can also respond to checkpoint inhibitors, inferring that other tumour related factors have a role in driving the cancer-immunity cycle to generate a treatment response.

6.14.2.4 Immunohistochemical analysis of immune cell infiltrates

IHC performed on baseline and on-treatment biopsies will reveal dynamic changes of immune cell infiltrates induced by combination chemotherapy and PD-L1 blockade. This will show if FLOT-A converts immunologically cold into hot tumours through treatment [380].

6.14.2.5 Immunogenic cancer cell death with FLOT-A

Concurrent chemotherapy and immunotherapy in ICONIC allows correlation of ICD with immune responses and pathological response. ICD will be measured in pre- and on-treatment biopsies on FLOT treated patient derived

organoids (PDOs) by quantifying levels of ICD-associated signals conveyed by DAMPs. The DAMP signal that will be primarily assessed in the ICONIC trial is cell surface calreticulin exposure. This will be analysed using an ICD marker panel to corroborate ICD induction *in vitro* through treatment of PDOs with FLOT chemotherapy. Overall, correlation of ICD induction *in vitro*, with immune responses in matched biopsies and improved outcomes would substantiate the clinical relevance of ICD for immune responses.

6.14.3 Circulating tumour DNA analysis

Circulating tumour DNA (ctDNA) can provide critical insights into tumour genomic landscapes [381], or predict recurrences after surgery [209, 382]. The minimally invasive nature of ctDNA analysis also enables longitudinal tracking of cancer genomic changes without the discomfort, costs and risk associated with invasive tumor biopsies. Through the Royal Marsden's Biomedical Research Centre core facility, I have coordinated longitudinal blood collection for each patient recruited into the ICONIC trial. 30 ml peripheral blood is collected from each patient on a 2-weekly basis during perioperative FLOT-A treatment and regularly at each follow-up visit.

In a batch-wise approach ctDNA will be extracted from stored plasma for mutation analysis. We have chosen to utilise ddPCR rather than cfDNA seq assay developed in Chapters 3 and 4 as mutations which are usually truncal in GOA e.g. *TP53* [383] will be targeted rather than multiple whole genes, avoiding the need for a predesigned target sequencing panel. Thus in the interest of time and cost effectiveness ddPCR will be performed on known *TP53* variants to:

- track tumour mutation VAF as a surrogate marker of response to treatment

- monitor ctDNA clearance or persistence after surgery and following completion of curative intent treatment to detect tumour recurrence early during follow-up. ctDNA mutation tracking can detect minimal residual disease (MRD) and identify earlier which patients are at risk of cancer recurrence. This has been effectively demonstrated in breast cancer [382], where information on driver, likely clonal mutations from baseline primary tumour are used to develop personalised ddPCR assays to track the presence of ctDNA at different time points. For example, post-surgery to identify patients with detectable MRD inferring that FLOT-A treatment has failed to eradicate the disease and therefore these patients would require postoperative systemic treatment to check whether residual ctDNA gets cleared through more treatment, or if more of the same treatment does never achieve this which may suggest a switch in therapy is required. Switch therapy is not a feature of the ICONIC trial but could be considered in future trials if serial ctDNA sampling is informative. Following completion of treatment, serial ctDNA monitoring should demonstrate disease recurrence in the plasma, months before it is clinically or radiologically detected allowing for much earlier intervention while the disease burden remains low.

6.14.4 Faecal microbiome

Intestinal microbiota can affect tumour response to immunotherapies and chemotherapy agents [366]. The ICONIC trial collects stool samples at two time points: before treatment initiation and immediately prior to tumour resection. Samples are collected at two timepoints to assess the dynamic changes in the microbiota between those who achieve pCR vs. suboptimal

response vs. no response and investigate the role of the intestinal microbiome as a candidate biomarker of immunotherapy response. This will establish a biobank resource to investigate the microbiome in the future.

6.14.5 Recurrence biopsies

Patients with a recurrence after surgery will have a repeat biopsy taken to assess the genomic and immunologic landscape of tumour recurrences. These will be analysed by exome sequencing to reassess the tumour mutational burden [358] and identify potential genetic drivers of chemo-immunotherapy resistance present at the time of recurrence but absent in baseline biopsies. RNA sequencing will also be performed to define transcriptomic changes that may contribute to recurrences [384, 385]. Furthermore, immune cell infiltrates will be quantified by computational deconvolution from RNA sequencing and by multiparametric IHC which will reveal potential changes in the immune landscape. Overall, this will show whether and how recurrent tumours change their molecular and immunological characteristics after FLOT-A and surgery which may provide insights into mechanisms of resistance and targetable characteristics in recurrent GOAs.

6.15 Results of the safety run-in phase

The ICONIC study first opened for recruitment in July 2017 however underwent amendment shortly after, following a change in the standard of care of perioperative treatment for GOAs. The trial was halted while the amendment was prepared and approved to avoid treating some patients with FOLFOX-A combination and some with the amended FLOT-A combination.

The amended ICONIC trial was approved and opened to recruitment in February 2018 (Table 6.4). At the time of writing, six patients have been enrolled in the study and completed preoperative treatment within the safety run in phase.

Milestone	Timeline
Final protocol version 1	20.04.2017
REC approval (protocol version 1)	20.06.2017
HRA approval (protocol version 1)	21.06.2017
Trial opened (protocol version 1)	13.07.2017
Final protocol version 2	23.11.2017
REC approval (amended protocol version 2)	01.01.2018
HRA approval (amended protocol version 2)	17.01.2018
Accrual opened (amended protocol version 2)	09.02.2018
1st patient consented (amended protocol version 2)	12.02.2018
Completion of safety run-in phase (n=6)	19.03.2019

Table 6.4: ICONIC study timelines and key milestones

Eligible patients were enrolled into the 3+3 design dose finding stage where standard dose FLOT was administered with 10mg/kg i.v. avelumab every two weeks (at dose level 0). DLTs were assessed for a 28 day period thereafter, the safety review committee (SRC) was convened to review adverse events and DLT data of the first 3+3 patient cohort, in order to make a formal dosing recommendation for the main efficacy phase.

In the first 3 patient cohort, only one patient experienced a DLT of cardiac chest pain during the 5FU infusional component of cycle 2. Although this is an uncommon but recognised side effect from the standard of care chemotherapy agent and reported in the avelumab investigator's brochure, it required discontinuation from FLOT-A trial treatment. However no dose

changes were required based on the DLT criteria set out in the trial protocol and a further 3 patients were recruited into the second cohort of the safety run-in phase. A total of one out of six patients experienced a DLT during the 28 day DLT monitoring period, thus avelumab dose level 0 (10 mg/kg) was established as the maximum administered dose in combination with standard dose FLOT, for the efficacy phase.

During pre-operative FLOT-A all patients experienced at least one grade 1-2 adverse event (AE) as per CTCAE toxicity grading criteria, most commonly diarrhoea occurring in 5/6 patients, fatigue, nausea, peripheral neuropathy and hypokalaemia (all occurring in 4/6 patients) and 3/6 patients experienced at least one grade 3-4 AE including neutropenia and elevated liver enzymes (1 patient), thrombotic event (1 patient) and cardiac chest pain (1 patient). 4/6 patients reported any grade chest pain and underwent cardiac work up: one episode was due to a pulmonary embolus, one of likely gastrointestinal origin and two cardiac in nature. One event of cardiac chest pain occurred during the DLT period (described above) and one other patient experienced chest pain which was not clearly cardiac during the first presentation in the DLT monitoring period and subsequent treatment with FLOT alone, without avelumab, caused recurrence of the symptoms demonstrating that this was not due to the combination FLOT-A but highly likely to be 5FU induced. Overall, no unexpected immune- related adverse events occurred and the overall side effect profile was consistent with what is expected for FLOT or avelumab.

During the entire pre-operative treatment period, three of six patients completed four cycles of FLOT-A. One patient discontinued avelumab due to diarrhoea, two patients discontinued FLOT-A due to cardiac chest pain and switched to a regimen without 5FU. All six patients have undergone surgery at data cut-off, without unexpected complications (Table 6.5).

Patient no. (age)	Pre-operative staging and location	Pre-operative systemic treatment	Surgical procedure	Post-op days in ITU or HDU / surgical ward (days)	90 day post-op complications (Clavien-Dindo grade)
1. (64 y)	T3N2 GOJ	FLOT-A x3 (avelumab discontinued due to diarrhoea) FLOT x1	Oesophago-gastrectomy	4 /12	Chest infection (II)
2. (70 y)	T3N2 GOJ	FLOT-A x2 (discontinued due to cardiac toxicity) raltitrexed/ oxaliplatin x2	Oesophago-gastrectomy	7 / 18	SVT (II) Pulmonary collapse/ effusion (IIIa)
3. (74 y)	T2N0 Oesophagus	FLOT-A x4	Oesophago-gastrectomy	29 / 48	Anastomotic leak (IIIa)
4. (26 y)	T2N0 Gastric	FLOT-A x2 FLOT x1 (discontinued due to cardiac toxicity) raltitrexed/ oxaliplatin x1	Gastrectomy	1 / 16	Infection, unknown source (II)
5. (51 y)	T2N0 Gastric	FLOT-A x4	Subtotal gastrectomy	0 / 9	Subacute bowel obstruction (II)
6. (64 y)	T3N0 Oesophagus	FLOT-A x4	Oesophago-gastrectomy	4 / 15	Chest infection (II)

Table 6.5: Summary of six patients enrolled in the safety run-in phase, detailing the treatment regimen and the peri-operative safety.

Overall, this data demonstrates that FLOT can be safely combined with the PD-L1-inhibitor avelumab and the combination has a manageable safety profile at dose level 0 (standard dose FLOT + 10 mg/kg avelumab).

6.15.1 Multiparametric Immunohistochemistry of matched pre- and on-treatment biopsies.

Multiparametric analysis has been performed as a pilot study on three matched baseline and early on-treatment biopsies, by Dr Katharina von Loga.

The Perkin Elmer OPAL assay has been used to stain for CD8 (cytotoxic T-cell), CD4 (helper T-cell), CD45RO (memory T-cell), FOXP3 (regulatory T-cell) and cytokeratin-positive cancer cells and PD-L1 (Figure 6.4: IHC preformed by Dr K von Loga, Consultant Molecular Pathologist, RMH).

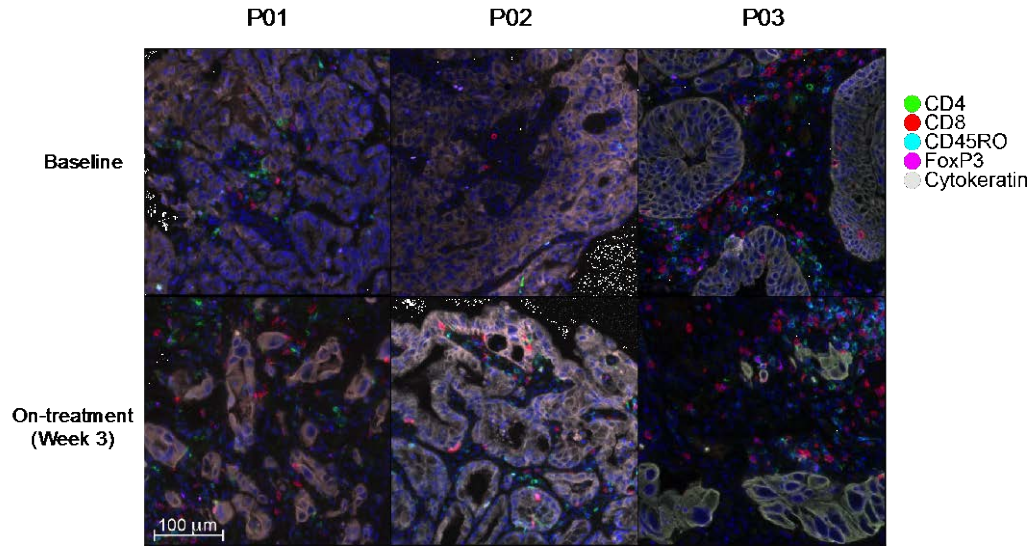


Figure 6.4: Comparison of pre-treatment and week three biopsies obtained via endoscopy for ICONIC patients P01, P02 and P03.
Each panel multispectral representative images at 20x magnification. IHC performed with Opal protocol for simultaneous detection of 5 tissue biomarkers plus nuclear counterstain to visualise spatial distribution of immune and cancer cells. Colour code: white-grey epithelium (normal and malignant); blue- DAPI counterstain; green- CD4 stain for T-helper cells; red- CD8 stain for cytotoxic T-cells; cyan- CD45RO stain for T-memory cells pink- nuclear FOXP3 stain for regulatory T-cell.

The on-treatment biopsies demonstrated an increase in CD45RO+ (cyan) T-cell numbers in all three patients and an increase in CD8+ (red) T-cells in two out of three patients (Figure 6.5). Two out of three patients showed a decrease in CD4+ and Treg cel numbers.

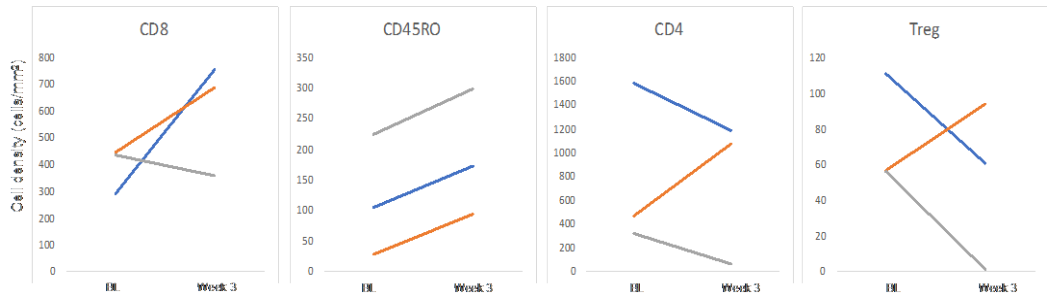


Figure 6.5: Quantification of first three evaluable biopsy pairs demonstrating dynamic changes in T-cell populations between baseline (BL) and week three biopsies.

This preliminary data on three paired samples demonstrates that the multiparametric IHC works for simultaneous detection of tissue biomarkers, providing the ability to visualise and quantify changes in the intratumoural immune infiltrate over the course of therapy.

6.15.2 Patient derived organoids

PDOs have emerged as robust pre-clinical models, showing a high degree of similarity to original patient tumour and can therefore be used to further predict response to treatment and understand mechanisms of acquired resistance. PDOs are being established from pre-treatment and on-treatment (week 3) endoscopic guided biopsies (Figure 6.6, PDOs established and maintained by Beatrice Griffiths, Scientific Officer, Gerlinger Lab). Based on ongoing experience within the lab, the expected success rate of PDO culture is ~30%.

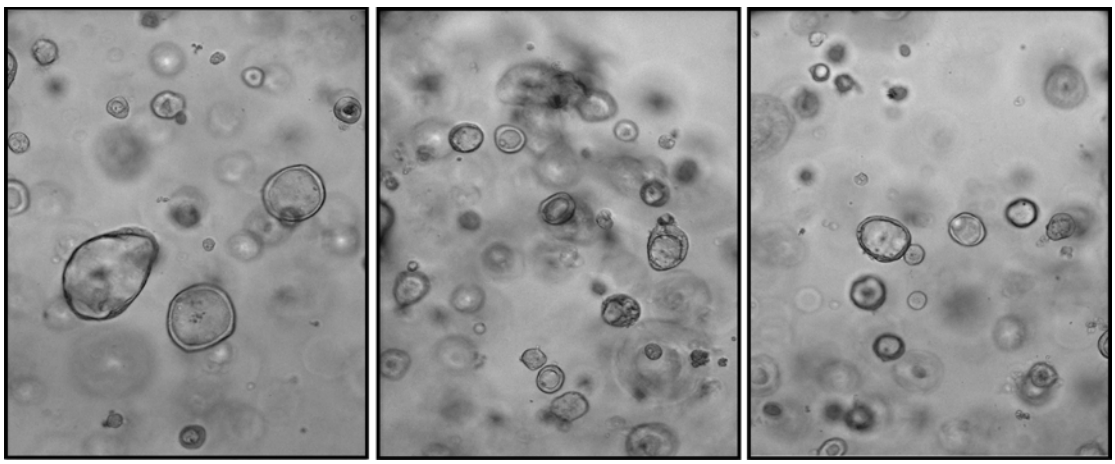


Figure 6.6: ICONIC patient P03 derived organoid established from fresh endoscopic guided biopsy of untreated early stage GOA and cultured in Matrigel *in vitro*.
Cells imaged at passage 10 at x10 magnification.

As the PDO culture recapitulates cellular architecture together with stromal cells, the PDOs will be treated with FLOT alone and in combination with other agents to assess induction of immunogenic cell death.

6.16 Summary and Discussion

The prognosis of gastric and oesophageal cancers is poor and many trials have failed to improve survival through multimodality strategies. Novel checkpoint inhibitor immunotherapy agents have shown some improvements in response rate however, clinical benefit is limited to a small proportion of patients. Therefore we developed the ICONIC clinical trial in early stage GOA by combining FLOT chemotherapy with avelumab to generate synergies and hence sufficient immune stimulation through the cancer-immunity cycle.

ICONIC was the first clinical trial opened in early stage GOAs assessing the combination of peri-operative chemo-immunotherapy. We promptly updated the chemotherapy backbone from FOLFOX to FLOT in light of the landmark FLOT-4 trial data. Thereafter, 3 further trials have opened using FLOT plus checkpoint inhibitors: i) perioperative FLOT plus atezolizumab in the phase II DANTE trial opened in September 2018 (NCT03421288), ii) perioperative FLOT plus avelumab followed by adjuvant avelumab for up to 1 year in the phase II MONEO trial opened to recruitment in June 2019 (NCT03979131) and iii) pembrolizumab plus FLOT chemotherapy (amendment of initial study using XP or FP chemotherapy) as neoadjuvant/adjuvant treatment in the KEYNOTE-585 expected to open to recruitment in October 2019 following protocol amendment (NCT03221426). Thus, the field recognises this as a clinically promising strategy.

A key feature of the ICONIC trial is the translational research sub-study, with extensive sample collection to provide important insights into which patients are ideal candidates for combination therapy given the additional toxicities

that are likely to occur, identify predictive biomarkers for treatment response and immunotherapy toxicities and provide understanding into the mechanisms of action of checkpoint inhibitors on the tumour immune microenvironment.

Following completion of the 3+3 safety run-in phase of the trial, recruitment was halted to review safety data emerging during the 28-day DLT period. Of the first 6 patients, adverse events were consistent with those seen with FLOT chemotherapy and/or recognised with immune checkpoint inhibitors including diarrhoea, fatigue, nausea, neutropenia and elevated liver enzymes. However chest pain occurred in 4/6 (66%) patients with variable aetiologies, only one of which was deemed a DLT by the SRC. This will continue to be closely monitored and assessed during the efficacy phase. On this basis, the SRC concluded that the combination of FLOT with avelumab was safe to proceed to the efficacy phase of the trial at standard dose FLOT + 10mg/kg avelumab, with ongoing monitoring of chest pain to ascertain whether we are observing the emergence of a new safety signal with this novel combination treatment.

Through the translational sub-study we have performed multiparametric immunohistochemistry analysis of 3 paired biopsies (pre-treatment and on-treatment). Preliminary biomarkers have supported an increase in CD8+ cytotoxic T-cells (in 2 out of 3 cases) and a decrease in T-regulatory cells (in 2 out of 3 cases). Although these results are preliminary, this activation of T-cell responses is recognised in transforming tumours into immunologically hot tumours, which is consistent with what we hypothesised, supporting the

overall trial strategy. Although, this will require reassessment once more cases are available in the efficacy phase. We have demonstrated proof of concept for the culture of PDOs as a preclinical model of early stage GOA. The development of these organoid cultures represents the first step that is required to evaluate immunogenic cell death. Although checkpoint inhibitor antibodies have been used in trials for advanced GOA for some years, there are no preclinical models that allow us to investigate ICD induction in generating a treatment response in early stage disease [386]. To date, one out of 11 biopsies have successfully generated PDOs. PDO take rate can be influenced by many factors including tumour content, tumour/stroma ratio and *in vitro* conditions. Further refinement of the organoid culture system is currently underway, involving alterations to sample processing and media composition.

Future correlation of genetic information, immune infiltrates and response data will reveal how the genetic profile of individual cancers influences immune infiltrates and whether the combined analysis of genetic alterations and immune cell infiltrates allows more accurate predictions of immunotherapy outcome. A limitation of our single arm design is that it will be difficult to define whether any of the biomarkers we will assess in pre-treatment biopsies are prognostic or predictive. To address this limitation, we will analyse any of the biomarkers that are associated with a pCR in ICONIC with pre-treatment biopsies in 40 patients who received perioperative FLOT alone and surgery at RMH, to assess which of the markers are specifically associated with FLOT-A response but not FLOT.

CHAPTER 7: FINAL DISCUSSION

The work included in this thesis focuses on the study of cfDNA to define the cancer genome in mCRCs and inform clinical application. In addition, I have developed a clinical trial investigating a novel combination therapy for resectable GOAs, which utilises chemotherapy agents whose resistance mechanisms have been studied in the previous chapters, combined with an anti-PDL1-antibody.

7.1 Establishing a high sensitivity cfDNA sequencing assay for mutation detection and copy number analysis

I have developed an ultra-sensitive targeted sequencing assay using molecular barcode and duplexCaller error correction for the analysis of cfDNA from only 25 ng of starting material. This amount was obtained from 9 ml of EDTA blood in the majority of patients with mCRC (Figure 3.6) assuring that our technology can be widely applied.

I applied the assay to cfDNA samples from 28 patients with mCRCs whose tumours had previously undergone genetic profiling. All 91 known mutations were re-identified in the cfDNA of 89% patients (Figure 4.1). The false negative cases were due to low tumour content within the cfDNA, confirmed by orthogonal validation with ddPCR (Figure 4.2).

With a good concordance with matched solid tumour biopsies, cfDNA-seq can provide insights into the mutational landscape as well as identify acquired drug resistance mechanisms which can guide clinical management. Until now, cfDNA analysis has been evaluated in small cohorts in the setting of clinical trials, however further validation is required to define the VAF

thresholds to infer response to treatment and MRD estimation before cfDNA can be used in routine clinical practice. An ongoing study in stage II and III CRC (TRACC trial NCT04050345) is investigating whether cfDNA analysis can help identify MRD and relapse earlier than existing radiological/CEA biomarker methods. In metastatic disease, cfDNA has been most widely studied in metastatic breast cancer, where studies have shown the advantages of cfDNA analysis for rapid, rare mutation testing including in patients with non-biopsiable disease [387] and in predicting response to treatment and OS [388]. Similarly in lung cancer, cfDNA analysis has been successfully used to detect actionable mutations and to track acquired resistance [389], however cfDNA is currently only tested in clinical practice where tumour tissue is insufficient for genomic profiling.

7.2 Current applications of cfDNA-seq assay

The cfDNA-seq assay developed in Chapters 3 and 4 is currently being used in the Royal Marsden's Centre for Molecular Pathology, a clinically accredited laboratory, to develop a routine cfDNA sequencing assay and has a role specifically where tumour tissue sequencing fails on *RAS*/*RAF* pathway genes, that can affect first line mCRC treatment options.

The developed ultra-sensitive assay can been used for tumour genotyping and detection of subclonal mutations that evolve during treatment or tumour progression, for example this assay was used in a study that investigated the clonality of mutations driving acquired cetuximab resistance in mCRC [215]. This study confirmed that acquired cetuximab resistance is usually polyclonal as previously shown [96], and the study demonstrated for the first time that all resistance driver mutations taken together in each individual patient were

confined to a median of 21% of all cancer cells in that patient. cfDNA-seq hence demonstrated that genetic drug resistance drivers were undetectable in a large proportion of cancer cells in the population. This supported the existence of a new mechanism of non-genetic cetuximab resistance that was mediated by microenvironmental remodelling which my host lab has identified through the analysis of biopsies from the same patients and in vitro studies. Overall, the cfDNA-seq assay can not only foster translational research to improve personalised medicine approaches but also provide insights into cancer clonal dynamics and cancer evolution which were not previously possible using tumour tissue biopsies alone.

Since the start of this PhD other groups have also developed NGS error-corrected cfDNA sequencing methods such as Guardant 360 [116] and Capp-Seq [105, 194], however these methods use proprietary technologies rather than off-the-shelf reagents, which limits their application to answer research questions requiring the design of customised cfDNA sequencing panels. Furthermore these assays require higher input cfDNA amount (Guardant 360 10-50 ng and CAPP-seq 32 ng) and CAPP-seq incorporates a background error correction model that requires up to 30 healthy donor samples to be sequenced for digital error suppression, compared to the six required for our assay. This can make it difficult for academic research where re-sequencing large numbers of HD samples is challenging due to sample availability and cost. Another advantage of the cfDNA-seq assay is the limit of detection of <0.15% whereas the detection limit of CAPP-seq is 1 in 5000 mutant:wild-type molecules i.e. 0.02% VAF as described in the initial publication of this technique [105], however the commercially marketed AVENIO technology is described to achieve 0.5% sensitivity for point

mutations and 1% for indels [390]. Finally, cfDNA-seq also has the ability to generate simultaneous genome wide CNA profiles from off-target reads.

7.3 Development of a clinical trial using novel chemo-immunotherapy combination for GOA

I developed the investigator led ICONIC trial of perioperative FLOT chemotherapy in combination with the anti-PD-L1 antibody avelumab in resectable GOAs. This trial aims to assess the safety and feasibility of chemo-immunotherapy combination in the peri-operative setting and assess the pathological complete response rate as a surrogate marker of efficacy.

Since the conception of this study, combination chemo-immunotherapy has been recognised as a clinically promising therapeutic strategy within the field of GOAs with three other large trials opening with FLOT combined with checkpoint inhibitors (NCT03421288, NCT03979131 and NCT03221426).

A key aspect of the ICONIC trial, and the relevance to this thesis, is in the field of biomarker development as efforts are continuously being made to identify and stratify patients for whom such treatments will be effective. A vital question that will be answered is why such treatments do not work in some patients and how resistance develops. One method of doing this is through analysing pre-treatment, early on-treatment and recurrence biopsies to assess immunogenic cell death and understand how recurrent tumours change their genomic and immunological characteristics after FLOT-A. cfDNA will also be used for recurrence prediction as radiological analysis can be challenging in localised or locally recurrent GOAs. Sample analysis will form part of future work.

7.4 Conclusions and Future Directions

The work presented in this thesis represents contributions to current knowledge and aspects that have the potential for translation into clinical practice. The earlier chapters incorporate a review of the current landscape of technologies used to analyse cfDNA as a biomarker, and the results of experiments carried out to develop a novel cfDNA-seq assay. The optimisations performed in Chapters 3 and 4 resolve the unmet need of designing an ultra-deep error-corrected cfDNA sequencing technology for ultrasensitive mutation detection and copy number analysis in mCRC with a single assay using low cfDNA inputs. Proof of concept data was acquired from a real-world population of a cohort of mCRC patients.

An alternative method of biomarker investigation is through whole exome sequencing approaches. Chapter 5 describes the application of the cfDNA-seq assay to deep WES which has not been performed in this tumour type to date. Results presented in this chapter are an important first evaluation of cfDNA-WES-seq in understanding this technique and its associated limitations, which forms part of ongoing work within a European Research Council grant.

Although we [215] and others [188, 189, 201] have identified potential resistance drivers to targeted therapies in mCRC, a complete understanding of events remains elusive and requires further investigation through exome wide sequencing, functional studies and assessment beyond genetics. The first step in investigating this will be in advancing longitudinal deep exome cfDNA sequencing technology with simultaneous development of a data analysis and interpretation framework to assess the evolutionary nature of

cancer. A key advantage of surveying multiple timepoints is that it will allow us to use the emerging changes to infer phylogenies with much higher accuracy than possible from single timepoint data. This could also enable mapping of the drug resistance driver landscapes of chemotherapies, which are currently poorly understood.

REFERENCES

1. CRUK. *Bowel cancer statistics*. 2016 [cited 2018 September]; Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer.
2. DeSantis, C.E., et al., *Cancer treatment and survivorship statistics, 2014.* CA Cancer J Clin, 2014. **64**(4): p. 252-71.
3. Kinzler, K.W. and B. Vogelstein, *Lessons from hereditary colorectal cancer.* Cell, 1996. **87**(2): p. 159-70.
4. Campbell, P.J., et al., *Pan-cancer analysis of whole genomes.* Nature, 2020. **578**(7793): p. 82-93.
5. Cancer Genome Atlas, N., *Comprehensive molecular characterization of human colon and rectal cancer.* Nature, 2012. **487**(7407): p. 330-7.
6. Cunningham, J.M., et al., *Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability.* Cancer Res, 1998. **58**(15): p. 3455-60.
7. Loeb, L.A., *Cancer cells exhibit a mutator phenotype.* Adv Cancer Res, 1998. **72**: p. 25-56.
8. Kondo, Y. and J.P. Issa, *Epigenetic changes in colorectal cancer.* Cancer Metastasis Rev, 2004. **23**(1-2): p. 29-39.
9. Lao, V.V. and W.M. Grady, *Epigenetics and colorectal cancer.* Nat Rev Gastroenterol Hepatol, 2011. **8**(12): p. 686-700.
10. Forbes, S.A., et al., *COSMIC: High-Resolution Cancer Genetics Using the Catalogue of Somatic Mutations in Cancer.* Curr Protoc Hum Genet, 2016. **91**: p. 10 11 1-10 11 37.
11. Gonzalez-Perez, A., et al., *IntOGen-mutations identifies cancer drivers across tumor types.* Nat Methods, 2013. **10**(11): p. 1081-2.
12. Cunningham, D., et al., *Colorectal cancer.* Lancet, 2010. **375**(9719): p. 1030-47.
13. Van Cutsem, E., et al., *ESMO consensus guidelines for the management of patients with metastatic colorectal cancer.* Ann Oncol, 2016. **27**(8): p. 1386-422.
14. Amin, M.B., et al., *The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging.* CA Cancer J Clin, 2017. **67**(2): p. 93-99.
15. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
16. Fidler, I.J., *The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited.* Nat Rev Cancer, 2003. **3**(6): p. 453-8.
17. Valastyan, S. and R.A. Weinberg, *Tumor metastasis: molecular insights and evolving paradigms.* Cell, 2011. **147**(2): p. 275-92.
18. Stratton, M.R., P.J. Campbell, and P.A. Futreal, *The cancer genome.* Nature, 2009. **458**(7239): p. 719-24.
19. The Cancer Genome Atlas, N., et al., *Comprehensive molecular characterization of human colon and rectal cancer.* Nature, 2012. **487**: p. 330.
20. Verdaguer, H., T. Sauri, and T. Macarulla, *Predictive and prognostic biomarkers in personalized gastrointestinal cancer treatment.* J Gastrointest Oncol, 2017. **8**(3): p. 405-417.

21. Harvard, B.I.o.M.a. *SNP6 copy number analysis (GISTIC2)*. [cited 2016; Available from: www.http://firebrowse.org/#.
22. Bettstetter, M., et al., *Distinction of hereditary nonpolyposis colorectal cancer and sporadic microsatellite-unstable colorectal cancer through quantification of MLH1 methylation by real-time PCR.* Clin Cancer Res, 2007. **13**(11): p. 3221-8.
23. Peeters, M., et al., *Mutant KRAS codon 12 and 13 alleles in patients with metastatic colorectal cancer: assessment as prognostic and predictive biomarkers of response to panitumumab.* J Clin Oncol, 2013. **31**(6): p. 759-65.
24. Douillard, J.Y., et al., *Final results from PRIME: randomized phase III study of panitumumab with FOLFOX4 for first-line treatment of metastatic colorectal cancer.* Ann Oncol, 2014. **25**(7): p. 1346-55.
25. Douillard, J.Y., et al., *Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer.* N Engl J Med, 2013. **369**(11): p. 1023-34.
26. Van Cutsem, E., et al., *Fluorouracil, leucovorin, and irinotecan plus cetuximab treatment and RAS mutations in colorectal cancer.* J Clin Oncol, 2015. **33**(7): p. 692-700.
27. Peeters, M., et al., *Analysis of KRAS/NRAS Mutations in a Phase III Study of Panitumumab with FOLFIRI Compared with FOLFIRI Alone as Second-line Treatment for Metastatic Colorectal Cancer.* Clin Cancer Res, 2015. **21**(24): p. 5469-79.
28. Aapro, M., et al., *A randomized phase III study evaluating the efficacy and safety of NEPA, a fixed-dose combination of netupitant and palonosetron, for prevention of chemotherapy-induced nausea and vomiting following moderately emetogenic chemotherapy.* Ann Oncol, 2014. **25**(7): p. 1328-33.
29. Schwartzberg, L.S., et al., *PEAK: a randomized, multicenter phase II study of panitumumab plus modified fluorouracil, leucovorin, and oxaliplatin (mFOLFOX6) or bevacizumab plus mFOLFOX6 in patients with previously untreated, unresectable, wild-type KRAS exon 2 metastatic colorectal cancer.* J Clin Oncol, 2014. **32**(21): p. 2240-7.
30. Tran, B., et al., *Impact of BRAF mutation and microsatellite instability on the pattern of metastatic spread and prognosis in metastatic colorectal cancer.* Cancer, 2011. **117**(20): p. 4623-32.
31. Cremolini, C., et al., *BRAF codons 594 and 596 mutations identify a new molecular subtype of metastatic colorectal cancer at favorable prognosis.* Ann Oncol, 2015. **26**(10): p. 2092-7.
32. Seymour, M.T., et al., *Panitumumab and irinotecan versus irinotecan alone for patients with KRAS wild-type, fluorouracil-resistant advanced colorectal cancer (PICCOLO): a prospectively stratified randomised trial.* Lancet Oncol, 2013. **14**(8): p. 749-59.
33. Pietrantonio, F., et al., *Predictive role of BRAF mutations in patients with advanced colorectal cancer receiving cetuximab and panitumumab: a meta-analysis.* Eur J Cancer, 2015. **51**(5): p. 587-94.
34. Boland, C.R. and A. Goel, *Microsatellite instability in colorectal cancer.* Gastroenterology, 2010. **138**(6): p. 2073-2087 e3.
35. Le, D.T., et al., *PD-1 Blockade in Tumors with Mismatch-Repair Deficiency.* New England Journal of Medicine, 2015. **372**(26): p. 2509-2520.

36. Hewish, M., et al., *Mismatch repair deficient colorectal cancer in the era of personalized treatment.* Nat Rev Clin Oncol, 2010. **7**(4): p. 197-208.
37. Goossens, N., et al., *Cancer biomarker discovery and validation.* Transl Cancer Res, 2015. **4**(3): p. 256-269.
38. Duffy, M.J., et al., *Tumor markers in colorectal cancer, gastric cancer and gastrointestinal stromal cancers: European group on tumor markers 2014 guidelines update.* Int J Cancer, 2014. **134**(11): p. 2513-22.
39. Andre, T., et al., *Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer.* N Engl J Med, 2004. **350**(23): p. 2343-51.
40. Weiser, M.R., W.R. Jarnagin, and L.B. Saltz, *Colorectal cancer patients with oligometastatic liver disease: what is the optimal approach?* Oncology (Williston Park), 2013. **27**(11): p. 1074-8.
41. de Gramont, A., et al., *Leucovorin and fluorouracil with or without oxaliplatin as first-line treatment in advanced colorectal cancer.* J Clin Oncol, 2000. **18**(16): p. 2938-47.
42. Douillard, J.Y., et al., *Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial.* Lancet, 2000. **355**(9209): p. 1041-7.
43. Saltz, L.B., et al., *Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. Irinotecan Study Group.* N Engl J Med, 2000. **343**(13): p. 905-14.
44. Colucci, G., et al., *Phase III randomized trial of FOLFIRI versus FOLFOX4 in the treatment of advanced colorectal cancer: a multicenter study of the Gruppo Oncologico Dell'Italia Meridionale.* J Clin Oncol, 2005. **23**(22): p. 4866-75.
45. Adams, R.A., et al., *Intermittent versus continuous oxaliplatin and fluoropyrimidine combination chemotherapy for first-line treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial.* Lancet Oncol, 2011. **12**(7): p. 642-53.
46. Cremolini, C., et al., *FOLFOXIRI plus bevacizumab versus FOLFIRI plus bevacizumab as first-line treatment of patients with metastatic colorectal cancer: updated overall survival and molecular subgroup analyses of the open-label, phase 3 TRIBE study.* Lancet Oncol, 2015. **16**(13): p. 1306-15.
47. Douillard, J.Y., et al., *Final results from PRIME: randomized phase III study of panitumumab with FOLFOX4 for first-line treatment of metastatic colorectal cancer.* Ann Oncol, 2014. **25**(7): p. 1346-1355.
48. Giantonio, B.J., et al., *Bevacizumab in combination with oxaliplatin, fluorouracil, and leucovorin (FOLFOX4) for previously treated metastatic colorectal cancer: results from the Eastern Cooperative Oncology Group Study E3200.* J Clin Oncol, 2007. **25**(12): p. 1539-44.
49. Hurwitz, H.I., et al., *Bevacizumab in combination with fluorouracil and leucovorin: an active regimen for first-line metastatic colorectal cancer.* J Clin Oncol, 2005. **23**(15): p. 3502-8.
50. Saltz, L.B., et al., *Bevacizumab in combination with oxaliplatin-based chemotherapy as first-line therapy in metastatic colorectal cancer: a randomized phase III study.* J Clin Oncol, 2008. **26**(12): p. 2013-9.
51. Van Cutsem, E., et al., *Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer.* N Engl J Med, 2009. **360**(14): p. 1408-17.
52. Van Cutsem, E., et al., *Addition of aflibercept to fluorouracil, leucovorin, and irinotecan improves survival in a phase III randomized trial in patients with*

metastatic colorectal cancer previously treated with an oxaliplatin-based regimen. J Clin Oncol, 2012. **30**(28): p. 3499-506.
53. Ciardiello, F., *Maintenance therapy for metastatic colorectal cancer.* Lancet Oncol, 2015. **16**(15): p. 1444-1445.
54. Esin, E. and S. Yalcin, *Maintenance strategy in metastatic colorectal cancer: A systematic review.* Cancer Treat Rev, 2016. **42**: p. 82-90.
55. Goey, K.K.H., et al., *Maintenance treatment with capecitabine and bevacizumab versus observation in metastatic colorectal cancer: updated results and molecular subgroup analyses of the phase 3 CAIRO3 study.* Ann Oncol, 2017. **28**(9): p. 2128-2134.
56. Sonbol, M.B., et al., *The Role of Maintenance Strategies in Metastatic Colorectal Cancer: A Systematic Review and Network Meta-analysis of Randomized Clinical Trials.* JAMA Oncol, 2020. **6**(3): p. e194489.
57. Bokemeyer, C., et al., *FOLFOX4 plus cetuximab treatment and RAS mutations in colorectal cancer.* Eur J Cancer, 2015. **51**(10): p. 1243-52.
58. Bokemeyer, C., et al., *Addition of cetuximab to chemotherapy as first-line treatment for KRAS wild-type metastatic colorectal cancer: pooled analysis of the CRYSTAL and OPUS randomised clinical trials.* Eur J Cancer, 2012. **48**(10): p. 1466-75.
59. Bokemeyer, C., et al., *Fluorouracil, leucovorin, and oxaliplatin with and without cetuximab in the first-line treatment of metastatic colorectal cancer.* J Clin Oncol, 2009. **27**(5): p. 663-71.
60. Bokemeyer, C., et al., *Efficacy according to biomarker status of cetuximab plus FOLFOX-4 as first-line treatment for metastatic colorectal cancer: the OPUS study.* Ann Oncol, 2011. **22**(7): p. 1535-46.
61. Maughan, T.S., et al., *Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial.* Lancet, 2011. **377**(9783): p. 2103-14.
62. Tveit, K.M., et al., *Phase III trial of cetuximab with continuous or intermittent fluorouracil, leucovorin, and oxaliplatin (Nordic FLOX) versus FLOX alone in first-line treatment of metastatic colorectal cancer: the NORDIC-VII study.* J Clin Oncol, 2012. **30**(15): p. 1755-62.
63. Hurwitz, H., et al., *Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer.* N Engl J Med, 2004. **350**(23): p. 2335-42.
64. Cunningham, D., et al., *Bevacizumab plus capecitabine versus capecitabine alone in elderly patients with previously untreated metastatic colorectal cancer (AVEX): an open-label, randomised phase 3 trial.* Lancet Oncol, 2013. **14**(11): p. 1077-1085.
65. Venook A, N.D., Lenz HJ, Innocenti F, Mahoney M, O'Neil B, Shaw J, Polite B, Hochster H, Atkins J, Goldberg R, Mayer R, Schilsky R, Bertagnolli M, Blanke C., *CALGB/SWOG 80405: Phase III trial of irinotecan/5-FU/leucovorin (FOLFIRI) or oxaliplatin/5-FU/leucovorin (mFOLFOX6) with bevacizumab (BV) or cetuximab (CET) for patients (pts) with KRAS wild-type (wt) untreated metastatic adenocarcinoma of the colon or rectum (MCRC).* 2014. **32(15 Suppl)**.
66. Van Cutsem, E., et al., *Addition of aflibercept to fluorouracil, leucovorin, and irinotecan improves survival in a phase III randomized trial in patients with*

metastatic colorectal cancer previously treated with an oxaliplatin-based regimen. (1527-7755 (Electronic)).
67. Tabernero, J., et al., *Aflibercept versus placebo in combination with fluorouracil, leucovorin and irinotecan in the treatment of previously treated metastatic colorectal cancer: prespecified subgroup analyses from the VELOUR trial.* (1879-0852 (Electronic)).
68. Lievre, A., et al., *Bevacizumab plus FOLFIRI or FOLFOX in chemotherapy-refractory patients with metastatic colorectal cancer: a retrospective study.* BMC Cancer, 2009. **9**: p. 347.
69. Beretta, G.D., et al., *FOLFIRI + bevacizumab as second-line therapy for metastatic colorectal cancer pretreated with oxaliplatin: a pooled analysis of published trials.* Med Oncol, 2013. **30**(1): p. 486.
70. Peeters, M., et al., *Randomized phase III study of panitumumab with fluorouracil, leucovorin, and irinotecan (FOLFIRI) compared with FOLFIRI alone as second-line treatment in patients with metastatic colorectal cancer.* J Clin Oncol, 2010. **28**(31): p. 4706-13.
71. Sobrero, A.F., et al., *EPIC: phase III trial of cetuximab plus irinotecan after fluoropyrimidine and oxaliplatin failure in patients with metastatic colorectal cancer.* J Clin Oncol, 2008. **26**(14): p. 2311-9.
72. Cunningham, D., et al., *Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer.* N Engl J Med, 2004. **351**(4): p. 337-45.
73. Hecht, J.R., et al., *Panitumumab monotherapy in patients with previously treated metastatic colorectal cancer.* Cancer, 2007. **110**(5): p. 980-8.
74. Mayer, R.J., et al., *Randomized trial of TAS-102 for refractory metastatic colorectal cancer.* N Engl J Med, 2015. **372**(20): p. 1909-19.
75. Grothey, A., et al., *Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international, multicentre, randomised, placebo-controlled, phase 3 trial.* Lancet, 2013. **381**(9863): p. 303-12.
76. Li, J., et al., *Regorafenib plus best supportive care versus placebo plus best supportive care in Asian patients with previously treated metastatic colorectal cancer (CONCUR): a randomised, double-blind, placebo-controlled, phase 3 trial.* Lancet Oncol, 2015. **16**(6): p. 619-29.
77. Falcone, A., et al., *Phase III trial of infusional fluorouracil, leucovorin, oxaliplatin, and irinotecan (FOLFOXIRI) compared with infusional fluorouracil, leucovorin, and irinotecan (FOLFIRI) as first-line treatment for metastatic colorectal cancer: the Gruppo Oncologico Nord Ovest.* J Clin Oncol, 2007. **25**(13): p. 1670-6.
78. Tabernero, J., et al., *Ramucirumab versus placebo in combination with second-line FOLFIRI in patients with metastatic colorectal carcinoma that progressed during or after first-line therapy with bevacizumab, oxaliplatin, and a fluoropyrimidine (RAISE): a randomised, double-blind, multicentre, phase 3 study.* Lancet Oncol, 2015. **16**(5): p. 499-508.
79. Souglakos, J., et al., *FOLFOXIRI (folinic acid, 5-fluorouracil, oxaliplatin and irinotecan) vs FOLFIRI (folinic acid, 5-fluorouracil and irinotecan) as first-line treatment in metastatic colorectal cancer (MCC): a multicentre randomised phase III trial from the Hellenic Oncology Research Group (HORG).* Br J Cancer, 2006. **94**(6): p. 798-805.

80. Khan, K.H., et al., *Longitudinal Liquid Biopsy and Mathematical Modeling of Clonal Evolution Forecast Time to Treatment Failure in the PROSPECT-C Phase II Colorectal Cancer Clinical Trial.* Cancer Discov, 2018. **8**(10): p. 1270-1285.
81. Loupakis, F., et al., *KRAS codon 61, 146 and BRAF mutations predict resistance to cetuximab plus irinotecan in KRAS codon 12 and 13 wild-type metastatic colorectal cancer.* Br J Cancer, 2009. **101**(4): p. 715-21.
82. Bertotti, A., et al., *The genomic landscape of response to EGFR blockade in colorectal cancer.* Nature, 2015. **526**(7572): p. 263-7.
83. Sartore-Bianchi, A., et al., *PIK3CA mutations in colorectal cancer are associated with clinical resistance to EGFR-targeted monoclonal antibodies.* Cancer Res, 2009. **69**(5): p. 1851-7.
84. Valtorta, E., et al., *KRAS gene amplification in colorectal cancer and impact on response to EGFR-targeted therapy.* Int J Cancer, 2013. **133**(5): p. 1259-65.
85. Montero, A.J., et al., *Epigenetic inactivation of EGFR by CpG island hypermethylation in cancer.* Cancer Biol Ther, 2006. **5**(11): p. 1494-501.
86. Scartozzi, M., et al., *Epidermal growth factor receptor gene promoter methylation in primary colorectal tumors and corresponding metastatic sites: a new perspective for an "old" therapeutic target.* Anal Quant Cytol Histol, 2009. **31**(6): p. 417-23.
87. Scartozzi, M., et al., *Epidermal growth factor receptor (EGFR) gene promoter methylation and cetuximab treatment in colorectal cancer patients.* Br J Cancer, 2011. **104**(11): p. 1786-90.
88. Fisher, R., J. Larkin, and C. Swanton, *Inter and intratumour heterogeneity: a barrier to individualized medical therapy in renal cell carcinoma?* Front Oncol, 2012. **2**: p. 49.
89. Gerlinger, M., et al., *Intratumor heterogeneity and branched evolution revealed by multiregion sequencing.* N Engl J Med, 2012. **366**(10): p. 883-892.
90. Lipinski, K.A., et al., *Cancer Evolution and the Limits of Predictability in Precision Cancer Medicine.* Trends Cancer, 2016. **2**(1): p. 49-63.
91. Huxley, J., *The Biological Aspects of Cancer.* 1958: Harcourt, Brace, New York.
92. Swanton, C., *Intratumor heterogeneity: evolution through space and time.* (1538-7445 (Electronic)).
93. Mandel P Fau - Metais, P. and P. Metais, *[Not Available].* (0037-9026 (Print)).
94. Cai, X., et al., *Accessing Genetic Information with Liquid Biopsies.* Trends Genet, 2015. **31**(10): p. 564-75.
95. Wang, R., et al. *Cell-free circulating tumor DNA analysis for breast cancer and its clinical utilization as a biomarker*. Oncotarget, 2017. **8**, 75742-75755 DOI: 10.18632/oncotarget.20608.
96. Bettegowda, C., et al., *Detection of circulating tumor DNA in early- and late-stage human malignancies.* Sci Transl Med, 2014. **6**(224): p. 224ra24.
97. Diaz, L.A., Jr. and A. Bardelli, *Liquid biopsies: genotyping circulating tumor DNA.* J Clin Oncol, 2014. **32**(6): p. 579-86.
98. Diehl, F., et al., *Circulating mutant DNA to assess tumor dynamics.* Nat Med, 2008. **14**(9): p. 985-90.

99. Lo, Y.M., et al., *Rapid clearance of fetal DNA from maternal plasma.* Am J Hum Genet, 1999. **64**(1): p. 218-24.
100. Forshew, T., et al., *Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA.* Sci Transl Med, 2012. **4**(136): p. 136ra68.
101. Heitzer, E., et al., *Circulating tumor cells and DNA as liquid biopsies.* (1756-994X (Print)).
102. Abbosh, C., et al., *Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution.* Nature, 2017. **545**(7655): p. 446-451.
103. Murtaza, M.A.-O.h.o.o., et al., *Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer.* (2041-1723 (Electronic)).
104. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
105. 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.
106. Maier, J., et al., *Detection of mutant free circulating tumor DNA in the plasma of patients with gastrointestinal stromal tumor harboring activating mutations of CKIT or PDGFRA.* Clin Cancer Res, 2013. **19**(17): p. 4854-67.
107. Koeppel, F.A.-O.h.o.o., et al., *Whole exome sequencing for determination of tumor mutation load in liquid biopsy from advanced cancer patients.* (1932-6203 (Electronic)).
108. Misale, S., et al., *Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer.* Nature, 2012. **486**(7404): p. 532-6.
109. Bardelli, A. and K. Pantel, *Liquid Biopsies, What We Do Not Know (Yet).* (1878-3686 (Electronic)).
110. Siravegna, G. and A. Bardelli, *Genotyping cell-free tumor DNA in the blood to detect residual disease and drug resistance.* Genome Biol, 2014. **15**(8): p. 449.
111. Aung, K.L., et al., *Current status and future potential of somatic mutation testing from circulating free DNA in patients with solid tumours.* Hugo J, 2010. **4**(1-4): p. 11-21.
112. Dawson, S.J., et al., *Analysis of circulating tumor DNA to monitor metastatic breast cancer.* N Engl J Med, 2013. **368**(13): p. 1199-209.
113. van Ginkel, J.H., et al., *Preanalytical blood sample workup for cell-free DNA analysis using Droplet Digital PCR for future molecular cancer diagnostics.* Cancer Med, 2017. **6**(10): p. 2297-2307.
114. Diehl, F., et al., *BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions.* Nat Methods, 2006. **3**(7): p. 551-9.
115. Perakis, S. and M.R. Speicher, *Emerging concepts in liquid biopsies.* BMC Med, 2017. **15**(1): p. 75.
116. Lanman, R.B., et al., *Analytical and Clinical Validation of a Digital Sequencing Panel for Quantitative, Highly Accurate Evaluation of Cell-Free Circulating Tumor DNA.* PLoS One, 2015. **10**(10): p. e0140712.
117. Newman, A.M., et al., *Integrated digital error suppression for improved detection of circulating tumor DNA.* Nat Biotechnol, 2016. **34**(5): p. 547-555.
118. Elazezy, M. and S.A. Joosse, *Techniques of using circulating tumor DNA as a liquid biopsy component in cancer management.* (2001-0370).

119. Gevensleben, H., et al., *Noninvasive detection of HER2 amplification with plasma DNA digital PCR.* Clin Cancer Res, 2013. **19**(12): p. 3276-84.
120. Sorber, L., et al., *Circulating cell-free nucleic acids and platelets as a liquid biopsy in the provision of personalized therapy for lung cancer patients.* (1872-8332).
121. Chang, F. and M.M. Li, *Clinical application of amplicon-based next-generation sequencing in cancer.* Cancer Genet, 2013. **206**(12): p. 413-9.
122. Dapprich, J., et al., *The next generation of target capture technologies - large DNA fragment enrichment and sequencing determines regional genomic variation of high complexity.* BMC Genomics, 2016. **17**(1): p. 486.
123. Kennedy, S.R., et al., *Detecting ultralow-frequency mutations by Duplex Sequencing.* Nat Protoc, 2014. **9**(11): p. 2586-606.
124. Bos, M.K., et al., *Whole exome sequencing of cell-free DNA – A systematic review and Bayesian individual patient data meta-analysis.* Cancer Treatment Reviews, 2020. **83**: p. 101951.
125. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* (1097-0215).
126. Torre, L.A., et al., *Global Cancer Incidence and Mortality Rates and Trends--An Update.* Cancer Epidemiol Biomarkers Prev, 2016. **25**(1): p. 16-27.
127. Machlowska, J., et al., *Gastric Cancer: Epidemiology, Risk Factors, Classification, Genomic Characteristics and Treatment Strategies.* Int J Mol Sci, 2020. **21**(11).
128. Pabla, B.S., et al., *Increased Incidence and Mortality of Gastric Cancer in Immigrant Populations from High to Low Regions of Incidence: A Systematic Review and Meta-Analysis.* Clin Gastroenterol Hepatol, 2020. **18**(2): p. 347-359 e5.
129. Rice, T.W., D.T. Patil, and E.H. Blackstone, *8th edition AJCC/UICC staging of cancers of the esophagus and esophagogastric junction: application to clinical practice.* Ann Cardiothorac Surg, 2017. **6**(2): p. 119-130.
130. Cunningham, D., et al., *Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer.* N Engl J Med, 2006. **355**(1): p. 11-20.
131. Ychou, M., et al., *Perioperative chemotherapy compared with surgery alone for resectable gastroesophageal adenocarcinoma: an FNCLCC and FFCD multicenter phase III trial.* J Clin Oncol, 2011. **29**(13): p. 1715-21.
132. Al-Batran, S.-E., et al., *Perioperative chemotherapy with docetaxel, oxaliplatin, and fluorouracil/leucovorin (FLOT) versus epirubicin, cisplatin, and fluorouracil or capecitabine (ECF/ECX) for resectable gastric or gastroesophageal junction (GEJ) adenocarcinoma (FLOT4-AIO): A multicenter, randomized phase 3 trial.* Journal of Clinical Oncology, 2017. **35**(15_suppl): p. 4004-4004.
133. Salah-Eddin Al-Batran, N.H., Harald Schmalenberg, Hans-Georg Kopp, Georg Martin Haag, Kim Barbara Luley, Wolff H. Schmiegel, Gunnar Folprecht, Stephan Probst, Nicole Prasnikar, Peter C. Thuss-Patience, Wolfgang Fischbach, Jorg Trojan, Michael Koenigsmann, Claudia Pauligk, Thorsten Oliver Goetze, Elke Jaeger, Johannes Meiler, Martin H. Schuler, Ralf Hofheinz, *Perioperative chemotherapy with docetaxel, oxaliplatin, and fluorouracil/leucovorin (FLOT) versus epirubicin, cisplatin, and fluorouracil or capecitabine (ECF/ECX) for resectable gastric or gastroesophageal junction*

(GEJ) adenocarcinoma (FLOT4-AIO): A multicenter, randomized phase 3 trial. J Clin Oncol, 2017. **35**.
134. Al-Batran, S.E., et al., *Perioperative chemotherapy with fluorouracil plus leucovorin, oxaliplatin, and docetaxel versus fluorouracil or capecitabine plus cisplatin and epirubicin for locally advanced, resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4): a randomised, phase 2/3 trial.* (1474-547X).
135. Al-Batran, S.E., et al., *Perioperative chemotherapy with fluorouracil plus leucovorin, oxaliplatin, and docetaxel versus fluorouracil or capecitabine plus cisplatin and epirubicin for locally advanced, resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4): a randomised, phase 2/3 trial.* Lancet, 2019. **393**(10184): p. 1948-1957.
136. Bose, K., et al., *Perioperative Therapy of Oesophagogastric Adenocarcinoma: Mainstay and Future Directions.* (1687-6121).
137. Waldman, A.D., J.M. Fritz, and M.J. Lenardo, *A guide to cancer immunotherapy: from T cell basic science to clinical practice.* Nat Rev Immunol, 2020. **20**(11): p. 651-668.
138. Brignone, C., et al., *First-line chemoimmunotherapy in metastatic breast carcinoma: combination of paclitaxel and IMP321 (LAG-3Ig) enhances immune responses and antitumor activity.* J Transl Med, 2010. **8**: p. 71.
139. Fridlender, Z.G., et al., *Chemotherapy delivered after viral immunogene therapy augments antitumor efficacy via multiple immune-mediated mechanisms.* Mol Ther, 2010. **18**(11): p. 1947-59.
140. Ghansah, T., et al., *Dendritic cell immunotherapy combined with gemcitabine chemotherapy enhances survival in a murine model of pancreatic carcinoma.* Cancer Immunol Immunother, 2013. **62**(6): p. 1083-91.
141. Kodumudi, K.N., et al., *A novel chemoimmunomodulating property of docetaxel: suppression of myeloid-derived suppressor cells in tumor bearers.* Clin Cancer Res, 2010. **16**(18): p. 4583-94.
142. Alsina, M., I. Gullo, and F. Carneiro, *Intratumoral heterogeneity in gastric cancer: a new challenge to face.* Ann Oncol, 2017. **28**(5): p. 912-913.
143. Murugaesu, N., et al., *Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy.* Cancer Discov, 2015. **5**(8): p. 821-831.
144. Gao, J., et al., *Circulating tumor DNA functions as an alternative for tissue to overcome tumor heterogeneity in advanced gastric cancer.* Cancer Sci, 2017. **108**(9): p. 1881-1887.
145. Pectasides, E., et al., *Genomic Heterogeneity as a Barrier to Precision Medicine in Gastroesophageal Adenocarcinoma.* Cancer Discov, 2018. **8**(1): p. 37-48.
146. Jin, Y., et al., *The predicting role of circulating tumor DNA landscape in gastric cancer patients treated with immune checkpoint inhibitors.* Mol Cancer, 2020. **19**(1): p. 154.
147. Maron, S.B., et al., *Circulating Tumor DNA Sequencing Analysis of Gastroesophageal Adenocarcinoma.* Clin Cancer Res, 2019. **25**(23): p. 7098-7112.
148. 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.

149. 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.
150. Tie, J., et al., *Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer.* Ann Oncol, 2015. **26**(8): p. 1715-22.
151. Moorcraft, S.Y., et al., *Investigating the feasibility of tumour molecular profiling in gastrointestinal malignancies in routine clinical practice.* Ann Oncol, 2018. **29**(1): p. 230-236.
152. James T. Robinson, H.T., Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov, *Integrative Genomics Viewer.* Nat Biotechnol, 2011. **29**: p. 24-26.
153. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
154. Talevich, E., et al., *CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing.* PLoS Comput Biol, 2016. **12**(4): p. e1004873.
155. Strickler, J.H., et al., *Genomic Landscape of Cell-Free DNA in Patients with Colorectal Cancer.* (2159-8290).
156. Siravegna, G., et al., *Integrating liquid biopsies into the management of cancer.* Nat Rev Clin Oncol, 2017. **14**(9): p. 531-548.
157. Heitzer, E., et al., *Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing.* Genome Med, 2013. **5**(4): p. 30.
158. Haber, D.A. and V.E. Velculescu, *Blood-Based Analyses of Cancer: Circulating Tumor Cells and Circulating Tumor DNA.* Cancer Discov, 2014. **4**(6): p. 650-661.
159. Leary, R.J., et al., *Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing.* Sci Transl Med, 2012. **4**(162): p. 162ra154.
160. Gerlinger, M., et al., *Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing.* Nat Genet, 2014. **46**(3): p. 225-233.
161. Landau, D.A., et al., *Evolution and impact of subclonal mutations in chronic lymphocytic leukemia.* Cell, 2013. **152**(4): p. 714-726.
162. Bolli, N., et al., *Heterogeneity of genomic evolution and mutational profiles in multiple myeloma.* (2041-1723).
163. Samorodnitsky, E., et al., *Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing.* Human mutation, 2015. **36**(9): p. 903-914.
164. Technologies, A., *SureSelectXT HS Target Enrichment System for Illumina Multiplexed Sequencing*. 2017.
165. Schmitt, M.W., et al., *Detection of ultra-rare mutations by next-generation sequencing.* Proc Natl Acad Sci U S A, 2012. **109**(36): p. 14508-13.
166. Ståhlberg, A., et al., *Simple, multiplexed, PCR-based barcoding of DNA enables sensitive mutation detection in liquid biopsies using sequencing.* Nucleic acids research, 2016. **44**(11): p. e105-e105.

167. Arbeithuber, B., K.D. Makova, and I. Tiemann-Boege, *Artifactual mutations resulting from DNA lesions limit detection levels in ultrasensitive sequencing applications.* DNA Research, 2016. **23**(6): p. 547-559.
168. Moorcraft SY, Gonzalez D, Cunningham D, Jones T, Walker B. A., Peckitt C, Yuan L. C., Frampton M, Begum R, Eltahir Z, Wotherspoon A, Teixeira Mendes L. S., Hulkki Wilson S, Gillbanks A, Baratelli C, Fotiadis N, Patel A, Braconi C, Valeri N, Gerlinger M, Rao S, Watkins D, Chau I, Starling N, *Investigating the feasibility of tumour molecular profiling in gastrointestinal malignancies in routine clinical practice.* Annals of Oncology, 2017.
169. Merker, J.D., et al., *Circulating Tumor DNA Analysis in Patients With Cancer: American Society of Clinical Oncology and College of American Pathologists Joint Review.* Arch Pathol Lab Med, 2018.
170. Luke, J.J., et al., *Realizing the potential of plasma genotyping in an age of genotype-directed therapies. LID - 10.1093/jnci/dju214 [doi] LID - dju214 [pii].* (1460-2105).
171. Jiang, P., et al., *Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients.* Proc Natl Acad Sci U S A, 2015. **112**(11): p. E1317-25.
172. Underhill, H.R., et al., *Fragment Length of Circulating Tumor DNA.* PLoS Genet, 2016. **12**(7): p. e1006162.
173. Raymond, C.K., et al., *Collection of cell-free DNA for genomic analysis of solid tumors in a clinical laboratory setting.* PLoS One, 2017. **12**(4): p. e0176241.
174. Grasso, C.S., et al., *Genetic Mechanisms of Immune Evasion in Colorectal Cancer.* (2159-8290).
175. Abdallah, E.A., et al., *Thymidylate synthase expression in circulating tumor cells: a new tool to predict 5-fluorouracil resistance in metastatic colorectal cancer patients.* (1097-0215).
176. Diaz, L.A., Jr., et al., *The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers.* Nature, 2012. **486**(7404): p. 537-40.
177. Atlas, T.C.G., *Comprehensive molecular characterization of human colon and rectal cancer.* Nature, 2012. **487**(7407): p. 330-7.
178. Cerami, E., et al., *The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.* Cancer Discov, 2012. **2**(5): p. 401-4.
179. Gao, J., et al., *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.* Sci Signal, 2013. **6**(269): p. pl1.
180. Schwaederle, M., et al., *Genomic Alterations in Circulating Tumor DNA from Diverse Cancer Patients Identified by Next-Generation Sequencing.* Cancer Res, 2017. **77**(19): p. 5419-5427.
181. Chung, J., et al., *The minimal amount of starting DNA for Agilent's hybrid capture-based targeted massively parallel sequencing.* Sci Rep, 2016. **6**: p. 26732.
182. MacConaill, L.E., et al., *Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing.* BMC Genomics, 2018. **19**(1): p. 30.
183. Sinha R, S.G., Gulati GS, Ezran C, Travaglini KJ, Wei E, Chan CKF, Nabhan AN, Su T, Morganti RM, Conley SD, Chaib H, Red-Horse K, Longaker MT,

Snyder MP, Krasnow MA, Weissman ML *Index switching causes "spreading-of-signal" among multiplexed samples in Illumina HiSeq 4000 DNA sequencing.* BioRxiv Preprint, 2018.
184. Vodak D, L.S., Nakken S, Aasheim LB, Holte H, Bai B, Myklebost O, Meza-Zepeda LA, Hovig E, *Sample-Index Misassignment Impacts Tumour Exome Sequencing.* Scientific Reports, 2018. **8**(5307).
185. Illumina *Effects of Index Misassignment on Multiplexing and Downstream Analysis*. 2017.
186. Illumina *cBot specification Sheet*. 2011.
187. Illumina *HiSeq 2500 Sequencing System*. Specification Sheet, 2015.
188. Bardelli, A., et al., *Amplification of the MET receptor drives resistance to anti-EGFR therapies in colorectal cancer.* Cancer Discov, 2013. **3**(6): p. 658-73.
189. Bertotti, A., et al., *The genomic landscape of response to EGFR blockade in colorectal cancer.* (1476-4687).
190. Nguyen, P., et al., *Identification of errors introduced during high throughput sequencing of the T cell receptor repertoire.* BMC Genomics, 2011. **12**(1): p. 106.
191. Jiang, T., et al., *The diagnostic value of circulating cell free DNA quantification in non-small cell lung cancer: A systematic review with meta-analysis.* Lung Cancer, 2016. **100**: p. 63-70.
192. Mouliere, F., et al., *High fragmentation characterizes tumour-derived circulating DNA.* PLoS One, 2011. **6**(9): p. e23418.
193. Newman, A.M., et al., *Integrated digital error suppression for improved detection of circulating tumor DNA.* Nature biotechnology, 2016. **34**(5): p. 547-555.
194. Vessies, D.C., et al., *Abstract 1381: Technical validation of Roche AVENIO sequencing platform for liquid biopsies.* Cancer Research, 2019. **79**(13 Supplement): p. 1381.
195. Roche, *AVENIO ctDNA Expanded Kit*, in *Roche Sequencing*. 2018.
196. Haan, J.C., et al., *Genomic landscape of metastatic colorectal cancer.* (2041-1723 (Electronic)).
197. Kopetz, S., et al., *Encorafenib, Binimetinib, and Cetuximab in BRAF V600E-Mutated Colorectal Cancer.* (1533-4406 (Electronic)).
198. Lenz, H.-J., et al., *Nivolumab (NIVO) + low-dose ipilimumab (IPI) as first-line (1L) therapy in microsatellite instability-high/DNA mismatch repair deficient (MSI-H/dMMR) metastatic colorectal cancer (mCRC): Clinical update.* Journal of Clinical Oncology, 2019. **37**(15_suppl): p. 3521-3521.
199. Pereira, A.A.L.A.-O.h.o.o., et al., *Clinical utility of circulating cell-free DNA in advanced colorectal cancer.* (1932-6203).
200. Gerlinger, M., et al., *Cancer: evolution within a lifetime.* (1545-2948).
201. Misale, S., et al., *Emergence of KRAS mutations and acquired resistance to anti EGFR therapy in colorectal cancer.* Nature, 2012. **486**(7404): p. 532-536.
202. Fleming, N.I., et al., *SMAD2, SMAD3 and SMAD4 mutations in colorectal cancer.* Cancer Res, 2013. **73**(2): p. 725-35.
203. Liao, X., et al., *Clinicopathological characterization of SMAD4-mutated intestinal adenocarcinomas: A case-control study.* PLOS ONE, 2019. **14**(2): p. e0212142.

204. Zhou, J.Y., et al., *The association between two polymorphisms in the TS gene and risk of cancer: a systematic review and pooled analysis.* (1097-0215 (Electronic)).
205. Brannon, A.R., et al., *Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions.* Genome Biol, 2014. **15**(8): p. 454.
206. Diehl, F. and L.A. Diaz, Jr., *Digital quantification of mutant DNA in cancer patients.* (1040-8746 (Print)).
207. Dogruluk, T., et al., *Identification of Variant-Specific Functions of PIK3CA by Rapid Phenotyping of Rare Mutations.* Cancer research, 2015. **75**(24): p. 5341-5354.
208. Wallin, J.J., et al., *Active PI3K pathway causes an invasive phenotype which can be reversed or promoted by blocking the pathway at divergent nodes.* (1932-6203 (Electronic)).
209. Phallen, J., et al., *Direct detection of early-stage cancers using circulating tumor DNA.* Sci Transl Med, 2017. **9**(403).
210. Xie, M., et al., *Age-related mutations associated with clonal hematopoietic expansion and malignancies.* Nat Med, 2014. **20**(12): p. 1472-8.
211. Coombs, C.C., et al., *Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes.* Cell Stem Cell, 2017. **21**(3): p. 374-382.e4.
212. Genovese, G., et al., *Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence.* New England Journal of Medicine, 2014. **371**(26): p. 2477-2487.
213. Jaiswal, S., et al., *Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes.* New England Journal of Medicine, 2014. **371**(26): p. 2488-2498.
214. Fearon, E.R. and B. Vogelstein, *A genetic model for colorectal tumorigenesis.* (0092-8674 (Print)).
215. Woolston, A., et al., *Genomic and Transcriptomic Determinants of Therapy Resistance and Immune Landscape Evolution during Anti-EGFR Treatment in Colorectal Cancer.* Cancer cell, 2019. **36**(1): p. 35-50.e9.
216. Pinkel, D., et al., *High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays.* Nature Genetics, 1998. **20**(2): p. 207-211.
217. Alonso, M.H., et al., *Comprehensive analysis of copy number aberrations in microsatellite stable colon cancer in view of stromal component.* Br J Cancer, 2017. **117**(3): p. 421-431.
218. Illumina. *Whole Genome Sequencing*. 2019 [cited 2019 July]; Available from: https://emea.illumina.com/techniques/sequencing/dna-sequencing/whole-genome-sequencing.html?langsel=/gb/.
219. Hovelson, D.H., et al., *Rapid, ultra low coverage copy number profiling of cell-free DNA as a precision oncology screening strategy.* Oncotarget, 2017. **8**(52): p. 89848-89866.
220. Bergheim, J., et al., *Potential of quantitative SEPT9 and SHOX2 methylation in plasmatic circulating cell-free DNA as auxiliary staging parameter in colorectal cancer: a prospective observational cohort study.* Br J Cancer, 2018. **118**(9): p. 1217-1228.
221. Ghatalia, P., et al., *Clinical Utilization Pattern of Liquid Biopsies (LB) to Detect Actionable Driver Mutations, Guide Treatment Decisions and Monitor*

Disease Burden During Treatment of 33 Metastatic Colorectal Cancer (mCRC) Patients (pts) at a Fox Chase Cancer Center GI Oncology Subspecialty Clinic. Front Oncol, 2018. **8**: p. 652.

222. Osumi, H., et al., *Clinical relevance of circulating tumor DNA assessed through deep sequencing in patients with metastatic colorectal cancer.* Cancer Med, 2019. **8**(1): p. 408-417.
223. Takayama, Y., et al., *Monitoring circulating tumor DNA revealed dynamic changes in KRAS status in patients with metastatic colorectal cancer.* Oncotarget, 2018. **9**(36): p. 24398-24413.
224. Tabernero, J., et al., *Analysis of circulating DNA and protein biomarkers to predict the clinical activity of regorafenib and assess prognosis in patients with metastatic colorectal cancer: a retrospective, exploratory analysis of the CORRECT trial.* Lancet Oncol, 2015. **16**(8): p. 937-48.
225. Kim, S.T., et al., *Prospective blinded study of somatic mutation detection in cell-free DNA utilizing a targeted 54-gene next generation sequencing panel in metastatic solid tumor patients.* Oncotarget, 2015. **6**(37): p. 40360-40369.
226. McGranahan, N. and C. Swanton, *Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future.* (1097-4172 (Electronic)).
227. Wan, R., et al., *Comprehensive Analysis of the Discordance of EGFR Mutation Status between Tumor Tissues and Matched Circulating Tumor DNA in Advanced Non-Small Cell Lung Cancer.* (1556-1380 (Electronic)).
228. Han, C.B., et al., *Concordant KRAS mutations in primary and metastatic colorectal cancer tissue specimens: a meta-analysis and systematic review.* (1532-4192 (Electronic)).
229. Khakoo, S., et al., *Circulating tumour DNA, a promising biomarker for the management of colorectal cancer.* Crit Rev Oncol Hematol 2018. **122**: p. 72-82.
230. McEvoy, A.C., et al., *Correlation between circulating tumour DNA and metabolic tumour burden in metastatic melanoma patients.* (1471-2407 (Electronic)).
231. Saluja, H., et al., *The Use of Circulating Tumor DNA for Prognosis of Gastrointestinal Cancers.* Frontiers in Oncology, 2018. **8**: p. 275.
232. Ueda, M., et al., *Somatic mutations in plasma cell-free DNA are diagnostic markers for esophageal squamous cell carcinoma recurrence.* (1949-2553 (Electronic)).
233. Helleday, T., et al., *DNA repair pathways as targets for cancer therapy.* Nature Reviews Cancer, 2008. **8**(3): p. 193-204.
234. Broustas, C.G. and H.B. Lieberman, *DNA Damage Response Genes and the Development of Cancer Metastasis.* Radiation Research, 2014. **181**(2): p. 111-130.
235. Zhao, Z., G. Zhang, and W. Li, *Elevated Expression of ERCC6 Confers Resistance to 5-Fluorouracil and Is Associated with Poor Patient Survival in Colorectal Cancer.* (1557-7430 (Electronic)).
236. Seetharam, R.N., et al., *Oxaliplatin Resistance Induced by ERCC1 Up-regulation Is Abrogated by siRNA-mediated Gene Silencing in Human Colorectal Cancer Cells.* Anticancer Research, 2010. **30**(7): p. 2531-2538.
237. Lenz, H.J., et al., *ERCC-1 gene expression levels and outcome to FOLFOX chemotherapy in patients enrolled in CONFIRM1 and CONFIRM2.* Journal of Clinical Oncology, 2008. **26**(15_suppl): p. 4131-4131.

238. Reed, E., *ERCC1 and Clinical Resistance to Platinum-Based Therapy.* Clinical Cancer Research, 2005. **11**(17): p. 6100.
239. Kheirelseid, E.A.H., et al., *Mismatch repair protein expression in colorectal cancer.* Journal of gastrointestinal oncology, 2013. **4**(4): p. 397-408.
240. Li, Q., et al., *Association between the level of ERCC-1 expression and the repair of cisplatin-induced DNA damage in human ovarian cancer cells.* (0250-7005 (Print)).
241. Nadal, R., et al., *Microtubule-regulatory phosphoproteins and NER system are involved in platinum and paclitaxel-based chemotherapy resistance in ovarian cancer.* Journal of Clinical Oncology, 2007. **25**(18_suppl): p. 5567-5567.
242. Rasio, D., et al., *Characterization of the human homologue of RAD54: a gene located on chromosome 1p32 at a region of high loss of heterozygosity in breast tumors.* (0008-5472 (Print)).
243. Matsuda, M., et al., *Mutations in the RAD54 recombination gene in primary cancers.* Oncogene, 1999. **18**: p. 3427.
244. AlDubayan, S.H., et al., *Inherited DNA-Repair Defects in Colorectal Cancer.* (1537-6605 (Electronic)).
245. Singh, B., et al., *Human REV3 DNA Polymerase Zeta Localizes to Mitochondria and Protects the Mitochondrial Genome.* PLOS ONE, 2015. **10**(10): p. e0140409.
246. Martin, S.K. and R.D. Wood, *DNA polymerase ζ in DNA replication and repair.* Nucleic Acids Research, 2019. **47**(16): p. 8348-8361.
247. Wang, H., et al., *REV3L confers chemoresistance to cisplatin in human gliomas: the potential of its RNAi for synergistic therapy.* Neuro-oncology, 2009. **11**(6): p. 790-802.
248. Yang, L., et al., *REV3L, a Promising Target in Regulating the Chemosensitivity of Cervical Cancer Cells.* PLOS ONE, 2015. **10**(3): p. e0120334.
249. Zhu, X., et al., *REV3L, the catalytic subunit of DNA polymerase zeta, is involved in the progression and chemoresistance of esophageal squamous cell carcinoma.* Oncol Rep, 2016. **35**(3): p. 1664-70.
250. Brondello, J.M., et al., *Novel evidences for a tumor suppressor role of Rev3, the catalytic subunit of Pol zeta.* (1476-5594 (Electronic)).
251. Choi, M., T. Kipps, and R. Kurzrock, *ATM Mutations in Cancer: Therapeutic Implications.* Molecular Cancer Therapeutics, 2016. **15**(8): p. 1781.
252. Ahn, J.-Y., et al., *Acquired resistance to 5-fluorouracil via HSP90/Src-mediated increase in thymidylate synthase expression in colon cancer.* Oncotarget, 2015. **6**(32): p. 32622-32633.
253. Zill, O.A., et al., *The Landscape of Actionable Genomic Alterations in Cell-Free Circulating Tumor DNA from 21,807 Advanced Cancer Patients.* Clin Cancer Res, 2018. **24**(15): p. 3528-3538.
254. Lange, S.S., et al., *The Polymerase Activity of Mammalian DNA Pol ζ Is Specifically Required for Cell and Embryonic Viability.* PLOS Genetics, 2016. **12**(1): p. e1005759.
255. Decker, B., et al., *Rare, protein-truncating variants in ATM, CHEK2 and PALB2, but not XRCC2, are associated with increased breast cancer risks.* (1468-6244 (Electronic)).

256. Jiang, X., et al., *The FATC domains of PIKK proteins are functionally equivalent and participate in the Tip60-dependent activation of DNA-PKcs and ATM.* (0021-9258 (Print)).
257. Seidel, J.J., C.M. Anderson, and E.H. Blackburn, *A novel Tel1/ATM N-terminal motif, TAN, is essential for telomere length maintenance and a DNA damage response.* Molecular and cellular biology, 2008. **28**(18): p. 5736-5746.
258. Blasius, M. and J. Bartek, *ATM targets hnRNPK to control p53.* (1551-4005 (Electronic)).
259. Weber, A.M., et al., *Phenotypic consequences of somatic mutations in the ataxia-telangiectasia mutated gene in non-small cell lung cancer.* Oncotarget; Vol 7, No 38, 2016.
260. Liu, J., et al., *The Differential Expression of Core Genes in Nucleotide Excision Repair Pathway Indicates Colorectal Carcinogenesis and Prognosis.* BioMed Research International, 2018. **2018**: p. 10.
261. Van der Jeught, K., et al., *Drug resistance and new therapies in colorectal cancer.* World journal of gastroenterology, 2018. **24**(34): p. 3834-3848.
262. Head, S.R., et al., *Library construction for next-generation sequencing: overviews and challenges.* BioTechniques, 2014. **56**(2): p. 61-passim.
263. Aigrain, L., Y. Gu, and M.A. Quail, *Quantitation of next generation sequencing library preparation protocol efficiencies using droplet digital PCR assays - a systematic comparison of DNA library preparation kits for Illumina sequencing.* BMC genomics, 2016. **17**: p. 458-458.
264. Butler, T.M., et al., *Exome Sequencing of Cell-Free DNA from Metastatic Cancer Patients Identifies Clinically Actionable Mutations Distinct from Primary Disease.* (1932-6203 (Electronic)).
265. Dietz, S., et al., *Low Input Whole-Exome Sequencing to Determine the Representation of the Tumor Exome in Circulating DNA of Non-Small Cell Lung Cancer Patients.* (1932-6203 (Electronic)).
266. Choi, M., T. Kipps, and R. Kurzrock, *ATM Mutations in Cancer: Therapeutic Implications.* Mol Cancer Ther, 2016. **15**(8): p. 1781-91.
267. Mateo, J., et al., *DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer.* (1533-4406 (Electronic)).
268. Kwok, M., et al., *ATR inhibition induces synthetic lethality and overcomes chemoresistance in TP53- or ATM-defective chronic lymphocytic leukemia cells.* Blood, 2016. **127**(5): p. 582-95.
269. Min, A., et al., *AZD6738, A Novel Oral Inhibitor of ATR, Induces Synthetic Lethality with ATM Deficiency in Gastric Cancer Cells.* Mol Cancer Ther, 2017. **16**(4): p. 566-577.
270. Mateo, J., et al., *DNA Repair in Prostate Cancer: Biology and Clinical Implications.* (1873-7560 (Electronic)).
271. Rafiei, S., et al., *ATM Loss Confers Greater Sensitivity to ATR Inhibition Than PARP Inhibition in Prostate Cancer.* Cancer Res, 2020. **80**(11): p. 2094-2100.
272. De Bono JS, T.D., Caldwell R, Terbuch A, Goh BC, Heong V, et al., *First-in-human trial of the oral ataxia telangiectasia and Rad3-related (ATR) inhibitor BAY 1895344 in patients (pts) with advanced solid tumors.* J Clin Oncol., 2019. **37**.
273. Lloyd, R.L., et al., *Combined PARP and ATR inhibition potentiates genome instability and cell death in ATM-deficient cancer cells.* Oncogene, 2020. **39**(25): p. 4869-4883.

274. Christenson, E.S. and E.S. Antonarakis, *PARP inhibitors for homologous recombination-deficient prostate cancer.* (1744-7623 (Electronic)).
275. Coupland, V.H., et al., *Incidence and survival of oesophageal and gastric cancer in England between 1998 and 2007, a population-based study.* BMC cancer, 2012. **12**: p. 11-11.
276. CRUK. *Gastric and oesophageal cancer statistics*. 2013 [cited 2016 May]; Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer.
277. Torre, L.A., et al., *Global cancer statistics, 2012.* (1542-4863 (Electronic)).
278. Smyth, E.C., et al., *Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†.* Annals of Oncology, 2016. **27**(suppl_5): p. v38-v49.
279. Ajani Ja Fau - D'Amico, T.A., et al., *Gastric Cancer, Version 3.2016, NCCN Clinical Practice Guidelines in Oncology.* (1540-1413 (Electronic)).
280. Cunningham, D., et al., *Peri-operative chemotherapy with or without bevacizumab in operable oesophagogastric adenocarcinoma (UK Medical Research Council ST03): primary analysis results of a multicentre, open-label, randomised phase 2-3 trial.* Lancet Oncol, 2017. **18**(3): p. 357-370.
281. Stahl, M., et al., *Perioperative chemotherapy with or without epidermal growth factor receptor blockade in unselected patients with locally advanced oesophagogastric adenocarcinoma: Randomized phase II study with advanced biomarker program of the German Cancer Society (AIO/CAO STO-0801).* Eur J Cancer, 2018. **93**: p. 119-126.
282. Lordick, F., et al., *Capecitabine and cisplatin with or without cetuximab for patients with previously untreated advanced gastric cancer (EXPAND): a randomised, open-label phase 3 trial.* Lancet Oncol, 2013. **14**(6): p. 490-9.
283. Waddell, T., et al., *Epirubicin, oxaliplatin, and capecitabine with or without panitumumab for patients with previously untreated advanced oesophagogastric cancer (REAL3): a randomised, open-label phase 3 trial.* Lancet Oncol, 2013. **14**(6): p. 481-9.
284. Catenacci, D.V.T., et al., *Rilotumumab plus epirubicin, cisplatin, and capecitabine as first-line therapy in advanced MET-positive gastric or gastro-oesophageal junction cancer (RILOMET-1): a randomised, double-blind, placebo-controlled, phase 3 trial.* The Lancet. Oncology, 2017. **18**(11): p. 1467-1482.
285. Shah, M.A., et al., *METGastric: A phase III study of onartuzumab plus mFOLFOX6 in patients with metastatic HER2-negative (HER2-) and MET-positive (MET+) adenocarcinoma of the stomach or gastroesophageal junction (GEC).* Journal of Clinical Oncology, 2015. **33**(15_suppl): p. 4012-4012.
286. Fuchs, C.S., et al., *Ramucirumab monotherapy for previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (REGARD): an international, randomised, multicentre, placebo-controlled, phase 3 trial.* Lancet, 2014. **383**(9911): p. 31-39.
287. Wilke, H., et al., *Ramucirumab plus paclitaxel versus placebo plus paclitaxel in patients with previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (RAINBOW): a double-blind, randomised phase 3 trial.* Lancet Oncol, 2014. **15**(11): p. 1224-35.

288. Bang, Y.J., et al., *Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial.* Lancet, 2010. **376**(9742): p. 687-97.
289. Smyth, E.C., et al., *Safety and Efficacy of the Addition of Lapatinib to Perioperative Chemotherapy for Resectable HER2-Positive Gastroesophageal Adenocarcinoma: A Randomized Phase 2 Clinical Trial.* JAMA Oncology, 2019. **5**(8): p. 1181-1187.
290. Fernando Rivera, P.J.-F., Pilar Garcia Alfonso, Javier Gallego, M. Luisa Limon, Maria Alsina, Luis Lopez-Gomez, MaCarmen Galan, Esther Falco, Jose Luis Manzano, Encarnación González, Enrique Aranda, Eva Fernandez, Monica Jorge, *NEOHX study: Perioperative treatment with trastuzumab in combination with capecitabine and oxaliplatin (XELOX-T) in patients with HER-2 resectable stomach or esophagogastric junction (EGJ) adenocarcinoma—18 m DFS analysis.* J Clin Oncol 33 (suppl 3; abstr 107), 2015.
291. ClinicalTrials.gov. *Neoadjuvant study using trastuzumab or trastuzumab with pertuzumab in gastric or gastroesophageal junction adenocarcinoma (INNOVATION)*. 2016 March 2016 [cited 2016 May]; Available from: https://clinicaltrials.gov/ct2/show/NCT02205047.
292. Bang, Y.J., et al., *Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial.* Lancet, 2010. **376**(9742): p. 687-97.
293. Catenacci, D.V.T., et al., *Rilotumumab plus epirubicin, cisplatin, and capecitabine as first-line therapy in advanced MET-positive gastric or gastro-oesophageal junction cancer (RILOMET-1): a randomised, double-blind, placebo-controlled, phase 3 trial.* Lancet Oncol, 2017. **18**(11): p. 1467-1482.
294. Fuchs, C.S., et al., *Ramucirumab monotherapy for previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (REGARD): an international, randomised, multicentre, placebo-controlled, phase 3 trial.* Lancet, 2014. **383**(9911): p. 31-39.
295. Rivera, F., et al., *Perioperative trastuzumab, capecitabine and oxaliplatin in patients with HER2-positive resectable gastric or gastro-oesophageal junction adenocarcinoma: NEOHX phase II trial.* Eur J Cancer, 2021. **145**: p. 158-167.
296. Shah, M.A., et al., *Effect of Fluorouracil, Leucovorin, and Oxaliplatin With or Without Onartuzumab in HER2-Negative, MET-Positive Gastroesophageal Adenocarcinoma: The METGastric Randomized Clinical Trial.* JAMA Oncol, 2017. **3**(5): p. 620-627.
297. Wagner, A.D., et al., *EORTC-1203-GITCG - the "INNOVATION"-trial: Effect of chemotherapy alone versus chemotherapy plus trastuzumab, versus chemotherapy plus trastuzumab plus pertuzumab, in the perioperative treatment of HER2 positive, gastric and gastroesophageal junction adenocarcinoma on pathologic response rate: a randomized phase II-intergroup trial of the EORTC-Gastrointestinal Tract Cancer Group, Korean Cancer Study Group and Dutch Upper GI-Cancer group.* BMC Cancer, 2019. **19**(1): p. 494.

298. Wilke, H., et al., *Ramucirumab plus paclitaxel versus placebo plus paclitaxel in patients with previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (RAINBOW): a double-blind, randomised phase 3 trial.* Lancet Oncol, 2014. **15**(11): p. 1224-35.
299. Chau, I., *Clinical Development of PD-1/PD-L1 Immunotherapy for Gastrointestinal Cancers: Facts and Hopes.* (1078-0432 (Print)).
300. Topalian, S.L., et al., *Safety, activity, and immune correlates of anti-PD-1 antibody in cancer.* N Engl J Med, 2012. **366**(26): p. 2443-54.
301. ClincalTrials.gov. *Neoadjuvant anti-PD-1, nivolumab in resectable NSCLC (NA_00092076)*. 2016 April 2016 [cited 2016 May]; Available from: https://clinicaltrials.gov/ct2/show/NCT02259621?term=NCT02259621.
302. ClinicalTrials.gov. *A pilot window-of-opportunity study of the anti-PD-1 antibody pembrolizumab in patients with resectable malignant pleural mesothelioma*. 2016 March 2016 [cited 2016 May]; Available from: https://clinicaltrials.gov/ct2/show/NCT02707666?term=NCT02707666.
303. ClinalTrials.gov. *Pembrolizumab, combination chemotherapy, and radiation therapy before surgery in treating adult patients with locally advanced gastroesophageal junction or gastric cardia cancer that can be removed by surgery*. 2016 April 2016 [cited 2016 May]; Available from: https://clinicaltrials.gov/ct2/show/NCT02730546?term=NCT02730546.
304. ClinalTrials.gov. *Study of neoadjuvant nivolumab in patients with non-metastatic stage II-IV clear cell renal cell carcinoma*. 2016 February 2016 [cited 2016 May]; Available from: https://clinicaltrials.gov/ct2/show/NCT02575222?term=NCT02575222.
305. Ohigashi, Y., et al., *Clinical significance of programmed death-1 ligand-1 and programmed death-1 ligand-2 expression in human esophageal cancer.* Clin Cancer Res, 2005. **11**(8): p. 2947-53.
306. Sun, J., et al., *PD-L1 expression analysis in gastric carcinoma tissue and blocking of tumor-associated PD-L1 signaling by two functional monoclonal antibodies.* Tissue Antigens, 2007. **69**(1): p. 19-27.
307. Manish A. Shah, J.B., Lin Shen, Peter C. Enzinger, Qiao Li, Ildiko Csiki, Minori Koshiji, Toshihiko Doi, *Pembrolizumab (MK-3475) for previously treated metastatic adenocarcinoma or squamous cell carcinoma of the esophagus: Phase II KEYNOTE-180 study.* J Clin Oncol, 2016. **34 (suppl 4S; abstr TPS189)**.
308. Chung, H.C., et al., *Avelumab (anti-PD-L1) as first-line switch-maintenance or second-line therapy in patients with advanced gastric or gastroesophageal junction cancer: phase 1b results from the JAVELIN Solid Tumor trial.* J Immunother Cancer, 2019. **7**(1): p. 30.
309. Doi, T., et al., *Safety and Antitumor Activity of the Anti-Programmed Death-1 Antibody Pembrolizumab in Patients With Advanced Esophageal Carcinoma.* J Clin Oncol, 2018. **36**(1): p. 61-67.
310. Janjigian, Y.Y., et al., *CheckMate-032 Study: Efficacy and Safety of Nivolumab and Nivolumab Plus Ipilimumab in Patients With Metastatic Esophagogastric Cancer.* J Clin Oncol, 2018. **36**(28): p. 2836-2844.
311. Kang, Y.K., et al., *Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial.* Lancet, 2017. **390**(10111): p. 2461-2471.

312. Muro, K., et al., *Pembrolizumab for patients with PD-L1-positive advanced gastric cancer (KEYNOTE-012): a multicentre, open-label, phase 1b trial.* Lancet Oncol, 2016. **17**(6): p. 717-726.
313. Segal, N.H., et al., *Preliminary data from a multi-arm expansion study of MEDI4736, an anti-PD-L1 antibody.* Journal of Clinical Oncology, 2014. **32**(15_suppl): p. 3002-3002.
314. Shitara, K., et al., *Pembrolizumab versus paclitaxel for previously treated, advanced gastric or gastro-oesophageal junction cancer (KEYNOTE-061): a randomised, open-label, controlled, phase 3 trial.* Lancet, 2018. **392**(10142): p. 123-133.
315. Poklepovic As Fau - Carvajal, R.D. and R.D. Carvajal, *Prognostic Value of Low Tumor Burden in Patients With Melanoma.* (0890-9091 (Print)).
316. Martincorena, I. and P.J. Campbell, *Somatic mutation in cancer and normal cells.* (1095-9203 (Electronic)).
317. Hegde, P.S., V. Karanikas, and S. Evers, *The Where, the When, and the How of Immune Monitoring for Cancer Immunotherapies in the Era of Checkpoint Inhibition.* Clinical Cancer Research, 2016. **22**(8): p. 1865.
318. Bonaventura, P., et al., *Cold Tumors: A Therapeutic Challenge for Immunotherapy.* Frontiers in immunology, 2019. **10**: p. 168-168.
319. Galon, J. and D. Bruni, *Approaches to treat immune hot, altered and cold tumours with combination immunotherapies.* Nat Rev Drug Discov, 2019. **18**(3): p. 197-218.
320. Chen, D.S. and I. Mellman, *Oncology meets immunology: the cancer-immunity cycle.* Immunity, 2013. **39**(1): p. 1-10.
321. Kroemer, G., et al., *Immunogenic cell death in cancer therapy.* (1545-3278 (Electronic)).
322. Zhou, J., et al., *Immunogenic cell death in cancer therapy: Present and emerging inducers.* J Cell Mol Med, 2019. **23**(8): p. 4854-4865.
323. Fucikova, J., et al., *Detection of immunogenic cell death and its relevance for cancer therapy.* Cell Death Dis, 2020. **11**(11): p. 1013.
324. Bezu, L., et al., *Combinatorial strategies for the induction of immunogenic cell death.* (1664-3224 (Print)).
325. Kepp, O., et al., *Consensus guidelines for the detection of immunogenic cell death.* Oncoimmunology, 2014. **3**(9): p. e955691-e955691.
326. Zitvogel, L., et al., *Mechanism of action of conventional and targeted anticancer therapies: reinstating immunosurveillance.* Immunity, 2013. **39**(1): p. 74-88.
327. Pfirschke, C., et al., *Immunogenic Chemotherapy Sensitizes Tumors to Checkpoint Blockade Therapy.* Immunity, 2016. **44**(2): p. 343-54.
328. !!! INVALID CITATION !!! {Boyerinas, 2015 #486;Juliá, 2018 #583}.
329. Juliá, E.P., et al., *Avelumab, an IgG1 anti-PD-L1 Immune Checkpoint Inhibitor, Triggers NK Cell-Mediated Cytotoxicity and Cytokine Production Against Triple Negative Breast Cancer Cells.* Frontiers in immunology, 2018. **9**: p. 2140-2140.
330. Hato, S.V., et al., *Molecular pathways: the immunogenic effects of platinum-based chemotherapeutics.* (1078-0432 (Print)).
331. Vincent, J., et al., *5-Fluorouracil selectively kills tumor-associated myeloid-derived suppressor cells resulting in enhanced T cell-dependent antitumor immunity.* Cancer Res, 2010. **70**(8): p. 3052-61.

332. Wang, W., S. Qin, and L. Zhao, *Docetaxel enhances CD3+ CD56+ cytokine-induced killer cells-mediated killing through inducing tumor cells phenotype modulation.* (1950-6007 (Electronic)).
333. Michels, T., et al., *Paclitaxel promotes differentiation of myeloid-derived suppressor cells into dendritic cells in vitro in a TLR4-independent manner.* (1547-6901 (Electronic)).
334. Zhang, P., et al., *Chemopreventive agents induce programmed death-1-ligand 1 (PD-L1) surface expression in breast cancer cells and promote PD-L1-mediated T cell apoptosis.* (0161-5890 (Print)).
335. Sevko, A., et al., *Antitumor effect of paclitaxel is mediated by inhibition of myeloid-derived suppressor cells and chronic inflammation in the spontaneous melanoma model.* (1550-6606 (Electronic)).
336. Tesniere, A., et al., *Immunogenic death of colon cancer cells treated with oxaliplatin.* Oncogene, 2010. **29**(4): p. 482-91.
337. Wang, Y.-J., et al., *Immunogenic effects of chemotherapy-induced tumor cell death.* Genes & Diseases, 2018.
338. Alderson, D., et al., *Neoadjuvant cisplatin and fluorouracil versus epirubicin, cisplatin, and capecitabine followed by resection in patients with oesophageal adenocarcinoma (UK MRC OE05): an open-label, randomised phase 3 trial.* (1474-5488 (Electronic)).
339. Cunningham, D., et al., *Capecitabine and oxaliplatin for advanced esophagogastric cancer.* N Engl J Med, 2008. **358**(1): p. 36-46.
340. S-E. Al-Batran, C.P., N. Homann, H. Schmalenberg, H.-G. Kopp, G.M. Haag, K. Luley, G. Folprecht, S. Probst, P. Thuss-Patience, J. Trojan, M. Koenigsmann, U. Lindig, M. Pohl, S. Kasper, M. Möhler, T. Goetze, M. Schuler, E. Jaeger, R.D. Hofheinz, *New Data Confirms Superiority of Docetaxel Based Triplet Therapy in Gastric Cancer*, in *Docetaxel, oxaliplatin, and fluorouracil/leucovorin (FLOT) for resectable esophagogastric cancer: updated results from multicenter, randomized phase 3 FLOT4-AIO trial (German Gastric Group at AIO)*, E. Congress, Editor. 2017.
341. Kepp O Fau - Senovilla, L., G. Senovilla L Fau - Kroemer, and G. Kroemer, *Immunogenic cell death inducers as anticancer agents.* (1949-2553 (Electronic)).
342. Al-Batran, S.E., et al., *Histopathological regression after neoadjuvant docetaxel, oxaliplatin, fluorouracil, and leucovorin versus epirubicin, cisplatin, and fluorouracil or capecitabine in patients with resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4-AIO): results from the phase 2 part of a multicentre, open-label, randomised phase 2/3 trial.* Lancet Oncol, 2016. **17**(12): p. 1697-1708.
343. Achilli, P., et al., *Tumor response evaluation after neoadjuvant chemotherapy in locally advanced gastric adenocarcinoma: a prospective, multi-center cohort study.* (2078-6891 (Print)).
344. Fareed, K.R., et al., *Tumour regression grade (TRG) analyses in patients with resectable gastro-oesophageal adenocarcinomas treated with platinum-based neoadjuvant chemotherapy.* (1365-2559 (Electronic)).
345. Schmidt, T., et al., *Prognostic value of histopathological regression in 850 neoadjuvantly treated oesophagogastric adenocarcinomas.* Br J Cancer, 2014. **110**(7): p. 1712-20.

346. Noble, F., et al., *Refining pathological evaluation of neoadjuvant therapy for adenocarcinoma of the esophagus.* (2219-2840 (Electronic)).
347. Mandard, A.M., et al., *Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations.* Cancer, 1994. **73**(11): p. 2680-6.
348. D. Cunningham, E.S., S. Stenning, L. Stevenson, C. Robb, W. Allum, H. Grabsch, D. Alderson, A. Riddell, S. Chua, T. Crosby, R. Mason, M. Griffin, W. Mansoor, F. Coxon, S. Falk, S. Rowley, K. Sumpter, J. Blazeby, R. Langley, *2201 Peri-operative chemotherapy ± bevacizumab for resectable gastro-oesophageal adenocarcinoma: Results from the UK Medical Research Council randomised ST03 trial* European Journal of Cancer, 2015. **51**(S400).
349. Elizabeth Catherine Smyth, R.E.L., Sally P. Stenning, Laura Stevenson, Claire Robb, William H. Allum, Heike Grabsch, Derek Alderson, Angela M. Riddell, Thomas Crosby, Robert Mason, Michael Griffin, Wasat Mansoor, Fareeda Y. Coxon, Stephen Falk, Debra Furniss, Gary William Middleton, Katherine Anne Sumpter, Jane M. Blazeby, David Cunningham, *ST03: A randomized trial of perioperative epirubicin, cisplatin plus capecitabine (ECX) with or without bevacizumab (B) in patients (pts) with operable gastric, esophagogastric junction (OGJ), or lower esophageal adenocarcinoma.* J Clin Oncol., 2013. **0520;31(15)**.
350. Gibney, G.T., L.M. Weiner, and M.B. Atkins, *Predictive biomarkers for checkpoint inhibitor-based immunotherapy.* The Lancet. Oncology, 2016. **17**(12): p. e542-e551.
351. Havel, J.J., D. Chowell, and T.A. Chan, *The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy.* Nature reviews. Cancer, 2019. **19**(3): p. 133-150.
352. Borghaei, H., et al., *Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer.* (1533-4406 (Electronic)).
353. Garon, E.B., et al., *Pembrolizumab for the treatment of non-small-cell lung cancer.* (1533-4406 (Electronic)).
354. Larkin, J., et al., *Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma.* (1533-4406 (Electronic)).
355. Wolchok, J.D., et al., *Updated results from a phase III trial of nivolumab (NIVO) combined with ipilimumab (IPI) in treatment-naive patients (pts) with advanced melanoma (MEL) (CheckMate 067).* Journal of Clinical Oncology, 2016. **34**(15_suppl): p. 9505-9505.
356. Hamid, O., et al., *A prospective phase II trial exploring the association between tumor microenvironment biomarkers and clinical activity of ipilimumab in advanced melanoma.* (1479-5876 (Electronic)).
357. Zeng, D.Q., et al., *Prognostic and predictive value of tumor-infiltrating lymphocytes for clinical therapeutic research in patients with non-small cell lung cancer.* (1949-2553 (Electronic)).
358. Snyder, A., et al., *Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma.* New England Journal of Medicine, 2014. **371**(23): p. 2189-2199.
359. Van Allen, E.M., et al., *Genomic correlates of response to CTLA-4 blockade in metastatic melanoma.* (1095-9203 (Electronic)).
360. Rizvi, N.A., et al., *Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer.* Science, 2015. **348**(6230): p. 124-8.

361. Rosenberg, J.E., et al., *Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial.* (1474-547X (Electronic)).
362. Ji, R.-R., et al., *An immune-active tumor microenvironment favors clinical response to ipilimumab.* Cancer Immunology, Immunotherapy, 2012. **61**(7): p. 1019-1031.
363. Ribas, A., et al., *Association of response to programmed death receptor 1 (PD-1) blockade with pembrolizumab (MK-3475) with an interferon-inflammatory immune gene signature.* Journal of Clinical Oncology, 2015. **33**(15_suppl): p. 3001-3001.
364. Higgs, B.W., et al., *Relationship of baseline tumoral IFNγ mRNA and PD-L1 protein expression to overall survival in durvalumab-treated NSCLC patients.* Journal of Clinical Oncology, 2016. **34**(15_suppl): p. 3036-3036.
365. Tumeh, P.C., et al., *PD-1 blockade induces responses by inhibiting adaptive immune resistance.* (1476-4687 (Electronic)).
366. Zitvogel, L., et al., *Microbiome and Anticancer Immunosurveillance.* Cell, 2016. **165**(2): p. 276-287.
367. Iida, N., et al., *Commensal Bacteria Control Cancer Response to Therapy by Modulating the Tumor Microenvironment.* Science, 2013. **342**(6161): p. 967.
368. Viaud, S., et al., *The Intestinal Microbiota Modulates the Anticancer Immune Effects of Cyclophosphamide.* Science, 2013. **342**(6161): p. 971.
369. Matson, V., et al., *The commensal microbiome is associated with anti–PD-1 efficacy in metastatic melanoma patients.* Science, 2018. **359**(6371): p. 104.
370. Chaput, N., et al., *Baseline gut microbiota predicts clinical response and colitis in metastatic melanoma patients treated with ipilimumab.* Annals of oncology : official journal of the European Society for Medical Oncology, 2017. **28**(6): p. 1368-1379.
371. Gopalakrishnan, V., et al., *Gut microbiome modulates response to anti–PD-1 immunotherapy in melanoma patients.* Science, 2018. **359**(6371): p. 97.
372. Routy, B., et al., *Gut microbiome influences efficacy of PD-1–based immunotherapy against epithelial tumors.* Science, 2018. **359**(6371): p. 91.
373. Kelly, R.J., *Immunotherapy for Esophageal and Gastric Cancer.* American Society of Clinical Oncology Educational Book, 2017(37): p. 292-300.
374. Chen, Z., et al., *Chromosomal instability of circulating tumor DNA reflect therapeutic responses in advanced gastric cancer.* Cell death & disease, 2019. **10**(10): p. 697-697.
375. Castro, A., et al., *Elevated neoantigen levels in tumors with somatic mutations in the HLA-A, HLA-B, HLA-C and B2M genes.* BMC medical genomics, 2019. **12**(Suppl 6): p. 107-107.
376. Shin, D.S., et al., *Primary Resistance to PD-1 Blockade Mediated by JAK1/2 Mutations.* Cancer Discov, 2017. **7**(2): p. 188-201.
377. Institute, U.o.C.B. *Single Sample Gene Set Enrichment Analysis*. 2018 [cited 2019 May 2019]; Available from: http://software.broadinstitute.org/cancer/software/genepattern/modules/docs/ssGSEAProjection/4.
378. Turan, T., et al., *Immune oncology, immune responsiveness and the theory of everything.* Journal for ImmunoTherapy of Cancer, 2018. **6**(1): p. 50.

379. Kulangara, K., et al., *Clinical Utility of the Combined Positive Score for Programmed Death Ligand-1 Expression and the Approval of Pembrolizumab for Treatment of Gastric Cancer.* Arch Pathol Lab Med, 2019. **143**(3): p. 330-337.
380. Galon, J.A.-O.h.o.o. and D. Bruni, *Approaches to treat immune hot, altered and cold tumours with combination immunotherapies.* (1474-1784 (Electronic)).
381. Mok, T., et al., *Detection and Dynamic Changes of EGFR Mutations from Circulating Tumor DNA as a Predictor of Survival Outcomes in NSCLC Patients Treated with First-line Intercalated Erlotinib and Chemotherapy.* Clin Cancer Res, 2015. **21**(14): p. 3196-203.
382. Garcia-Murillas, I., et al., *Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer.* Sci Transl Med, 2015. **7**(302): p. 302ra133.
383. The Cancer Genome Atlas Research, N., et al., *Comprehensive molecular characterization of gastric adenocarcinoma.* Nature, 2014. **513**: p. 202.
384. Nowicki Ts Fau - Hu-Lieskovan, S., A. Hu-Lieskovan S Fau - Ribas, and A. Ribas, *Mechanisms of Resistance to PD-1 and PD-L1 Blockade.* (1540-336X (Electronic)).
385. Hugo, W., et al., *Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma.* Cell, 2016. **165**(1): p. 35-44.
386. Steele, N.G., et al., *An Organoid-Based Preclinical Model of Human Gastric Cancer.* Cellular and Molecular Gastroenterology and Hepatology, 2019. **7**(1): p. 161-184.
387. Turner, N.C., et al., *Circulating tumour DNA analysis to direct therapy in advanced breast cancer (plasmaMATCH): a multicentre, multicohort, phase 2a, platform trial.* Lancet Oncol, 2020. **21**(10): p. 1296-1308.
388. Fernandez-Garcia, D., et al., *Plasma cell-free DNA (cfDNA) as a predictive and prognostic marker in patients with metastatic breast cancer.* Breast Cancer Res, 2019. **21**(1): p. 149.
389. Esposito Abate, R., et al., *Next Generation Sequencing-Based Profiling of Cell Free DNA in Patients with Advanced Non-Small Cell Lung Cancer: Advantages and Pitfalls.* Cancers (Basel), 2020. **12**(12).
390. Choi, J., et al., *Abstract 3648: Performance of the AVENIO ctDNA assays across multiple high-throughput next-generation sequencing platforms.* Cancer Research, 2018. **78**(13 Supplement): p. 3648.