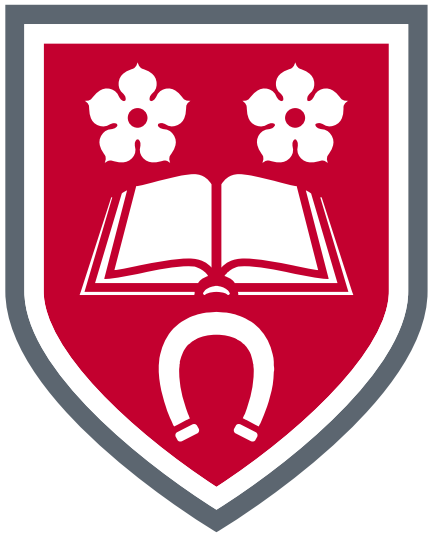
Using next-generation sequencing to interrogate circulating tumour DNA in thoracic malignancies and as a diagnostic tool

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Luke Joseph Martinson

BSc Honours (Newcastle University)
MRes (Newcastle University)



Leicester Cancer Research Centre
University of Leicester
October 2018

Abstract

Non-small cell lung cancer (NSCLC) and malignant pleural mesothelioma (MPM) represent two of the most aggressive solid tumours, where treatment is frequently ineffective, likely due to intratumour heterogeneity and associated resistance to therapy. This thesis primarily assessed circulating tumour DNA (ctDNA) in the characterisation of these diseases within a phylogenetic framework, specifically through the targeting of variants (SNVs) acquired early during tumour evolution (clonal variants). Another project evaluated whether targeted next-generation sequencing (NGS) could resolve diagnostic ambiguity by differentiating between metastases and independent second primary malignancies.

As part of Lung TRACERx, clonal tumour variants were detected within pre-surgical cfDNA of 13/41 (32%) early-stage NSCLC patients using a deep NGS approach targeting eleven genes (>150 hotspots). Of 14 stage I adenocarcinomas, none were detected with this technology. Longitudinal profiling provided evidence of minimal residual disease following surgery, with ctDNA detection antedating clinical relapse by up to five months. Despite being a generic approach (not patient-specific), this assay targeted ≥ 1 clonal tumour variants in 41/43 (95%) patients, indicating that the majority of “hotspot” SNVs are acquired early during NSCLC evolution and hence represent ideal targets for intervention.

In 11 patients with MPM, clonal variants selected from multiregional exome data, were analysed in pre-surgical cfDNA using patient-specific ddPCR assays. Four of 11 patients (36.3%) were detected, and interestingly these preliminary findings suggest ctDNA detection may have prognostic value in MPM and potentially could be used to preoperatively stratify patients with regards to invasive surgery.

Lastly, in an unambiguous diagnostic cohort, 19/20 (95%) cases were accurately diagnosed as metastatic or independent primary by inferring clonal origin through NGS. Subsequently, ambiguous case studies were analysed, where results suggest NGS could aid clinicians in providing precise diagnosis for these patients. Together, the results of this thesis highlight how NGS technologies are fast approaching the clinic, but magnify the challenges we face in bringing about the genomic revolution.

Acknowledgements

Firstly, I wish to extend my greatest thanks to my supervisors, Professors Jacqui Shaw and Dean Fennell, for their expert insight, kindness and support. Their passion for science is truly infectious, and I aspire to follow in their footsteps.

For my work involved in the Lung TRACERx study, I would like to particularly thank Mrs Joan Riley and Dr Christopher Abbosh. I extend my gratitude to Professor Charles Swanton for allowing me to be involved in this exciting study.

To Miss Annabel Sharkey and Mr Alan Dawson, I am grateful for all of your specialist surgical and clinical input on the MEDUSA study. I would also like to thank the CBS Histology Facility for their continued support of MEDUSA.

To Professor John Le Quesne, Dr David Moore and Dr Caroline Cowley, it has been fantastic to work with you on the PRIMINGS project and I am excited about the future developments with this work.

I am indebted to Drs Robert Hastings and Garth Wilson for all of their bioinformatics support for various projects. I send my regards to the UoL High Performance Computing services for the use of ALICE and SPECTRE systems.

To all the members of the Circulating Nucleic Acids group past and present (there are too many to name), it has been brilliant being part of such a close-knit bunch of people; you fill every day with laugher and joy.

Finally, I would like to all my family and friends for their support, understanding and most importantly, the occasional sanity check! A special mention to Neal Rimmer, Barbara Ottolini and Mark Openshaw for both work-related endeavours and those ‘prescribed’ trips to 45 West.

I wish to acknowledge Cancer Research UK for their generous funding enabling me to pursue this PhD.

Related Publications

1. Abbosh, C., Birkbak, N. J., Wilson, G. A., Jamal-Hanjani, M., Constantin, T., Salari, R., Le Quesne, J., Moore, D. A., Veeriah, S., Rosenthal, R., Marafioti, T., Kirkizlar, E., Watkins, T. B. K., McGranahan, N., Ward, S., **Martinson, L**., Riley, J., Fraioli, F., Al Bakir, M., Gronroos, E., Zambrana, F., Endozo, R., Bi, W. L., Fennessy, F. M., Sponer, N., Johnson, D., Laycock, J., Shafi, S., Czyzewska-Khan, J., Rowan, A., Chambers, T., Matthews, N., Turajlic, S., Hiley, C., Lee, S. M., Forster, M. D., Ahmad, T., Falzon, M., Borg, E., Lawrence, D., Hayward, M., Kolvekar, S., Panagiotopoulos, N., Janes, S. M., Thakrar, R., Ahmed, A., Blackhall, F., Summers, Y., Hafez, D., Naik, A., Ganguly, A., Kareht, S., Shah, R., Joseph, L., Marie Quinn, A., Crosbie, P. A., Naidu, B., Middleton, G., Langman, G., Trotter, S., Nicolson, M., Remmen, H., Kerr, K., Chetty, M., Gomersall, L., Fennell, D. A., Nakas, A., Rathinam, S., Anand, G., Khan, S., Russell, P., Ezhil, V., Ismail, B., Irvin-Sellers, M., Prakash, V., Lester, J. F., Kornaszewska, M., Attanoos, R., Adams, H., Davies, H., Oukrif, D., Akarca, A. U., Hartley, J. A., Lowe, H. L., Lock, S., Iles, N., Bell, H., Ngai, Y., Elgar, G., Szallasi, Z., Schwarz, R. F., Herrero, J., Stewart, A., Quezada, S. A., Peggs, K. S., Van Loo, P., Dive, C., Lin, C. J., Rabinowitz, M., Aerts, H., Hackshaw, A., Shaw, J. A., Zimmermann, B. G. and Swanton, C. (2017) Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature*, 545(7655), pp. 446-451.

2. Trigg, R. M., **Martinson, L. J**., Parpart-Li, S. and Shaw, J. A. (2018) Factors that influence quality and yield of circulating-free DNA: A systematic review of the methodology literature. *Heliyon*, 4(7), pp. e00699.

*Other manuscripts are currently in draft.

Table of Contents

List of Tables

List of Figures

Abbreviations

Acronym	Definition
µg	microgram
µL	microlitre
µM	micromolar
6-FAM™	6-carboxyfluorescein
AIS	adenocarcinoma in situ (lung)
ALK	anaplastic lymphoma kinase
APC	Adenomatous polyposis coli
BAP1	BRCA1 associated protein-1 (ubiquitin carboxy-terminal hydrolase)
BHQ-1	black hole quencher-1
BLAST	Basic Local Alignment Search Tool
bp	base pair
BRAF	v-raf murine sarcoma viral oncogene homologue B1
CDKN2A	cyclin-dependent kinase Inhibitor 2A
cfDNA	circulating free DNA
CHPv2	Ion AmpliSeq Cancer Hotspot Panel v2
CI	confidence interval
CNV	copy number variation
c-SCLC	combined small cell lung cancer
CT	computed tomography
C_T	cycle threshold
CTC	circulating tumour cell
ctDNA	circulating tumour DNA
ddPCR	Droplet Digital PCR
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dsDNA	double stranded DNA
EGFR	epidermal growth factor receptor
EML4	echinoderm microtubule-associated protein-like 4
EPD	extended pleurectomy decortication
EPP	extrapleural pnuemonectomy
FFPE	formalin-fixed, paraffin-embedded
FGFR	fibroblast growth factor receptor 1
g	gram
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
gDNA	genomic DNA
H&E	haematoxylin and eosin
HEX	Hexachlorofluorescein
hgDNA	human genomic DNA
HR	hazard ratio

iMig	International Mesothelioma Interest Group
KRAS	Kirsten rat sarcoma viral oncogene homolog
LOD	limit of detection
LOH	loss of heterozygosity
LUAD	lung adenocarcinoma
LUSC	lung squamous cell carcinoma
M	molar
MEDUSA	Mesothelioma Evolution: DrUgabble Somatic Alterations
MET	hepatocyte growth factor receptor
mg	milligram
MGB/NFQ	minor groove binder / non-fluorescent quencher
MIA	minimally invasive adenocarcinoma (lung)
mL	millilitre
mM	millimolar
mol	mole
MPM	malignant pleural mesothelioma
MRD	minimal residual disease
NCI	National Cancer Institute
NF2	neurofibromin 2
ng	nanogram
NGS	next-generation sequencing
nM	nanomolar
NRAS	neuroblastoma rat sarcoma viral oncogene homolog
NSCLC	non-small cell lung carcinoma
nt	nucleotide
NTC	no template control
ºC	degrees Celsius
OS	overall survival
OST	Oncomine Solid Tumour DNA panel
P	P-value
p	short chromosome arm
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFS	progression-free survival
pg	picogram
PIK3CA	phosphoinositide-3-kinase, catalytic, alpha polypeptide
pM	picomolar
PRIMINGS	PRImary vs MetastasIs using Next-Generation Sequencing
PTEN	phosphatase and tensin homolog
q	long chromosome arm
qPCR	quantitative PCR
R^2	regression coefficient
RB1	retinoblastoma protein
RET	rearranged during transfection

RNA	ribonucleic acid
ROS1	c-rose oncogene 1, receptor tyrosine kinase
RPPH1	ribonuclease P RNA component H1
RT	room temperature
SCLC	small cell lung carcinoma
SCNA	somatic copy number alteration
SNV	single nucleotide variant
T_a	annealing temperature
TE	Tris-EDTA
TKI	tyrosine kinase inhibitor
T_m	melting temperature
tNGS	targeted next-generation sequencing
TNM	tumour-node-metastasis
TP53	tumour protein p53
TRACERx	Tracking Non–Small-Cell Lung Cancer Evolution through Therapy
VIC	2'-chloro-7'-phenyl-1,4-dichloro-6-carboxyfluorescein
WES	whole exome sequencing
WGS	whole genome sequencing
WT	wild-type

Chapter 1 General Introduction

1.1 Intratumour heterogeneity in cancer

1.1.1 Origins of intratumour heterogeneity

The understanding that solid tumours are comprised of different subsets of cells is not a novel observation, particularly at a histological level, where differing cellular morphology has formed the basis of various tumour grading and prognostic classification systems for decades (Bloom and Richardson 1957, Heppner and Miller 1983). However, following the development of massively parallel sequencing technologies, our ability to dissect the genomic heterogeneity present within individual solid tumours has been revolutionised in regards to both the scale of this complexity and importantly how this could impact patient outcomes (Bedard *et al.* 2013). This complexity, termed intratumour heterogeneity now represents a major area of research, with analyses of most major solid tumour types being conducted, hoping to identify key targets for efficacious treatment and to derive new stratification methods for patients.

It is clear that phenotypic tumour heterogeneity is caused by a combination of genetic and non-genetic causes. For the latter, these include epigenetic alterations, heterogeneous tumour microenvironments, stochastic gene expression patterns and hierarchical differentiation of cancer stem cells (Brock *et al.* 2009, Huh and Paulsson 2011, Marusyk *et al.* 2012). When combined with diverse genetic alterations exhibited within tumours, our ability to comprehend this complexity is still relatively in its infancy. Nonetheless, ever since it became known that tumours can arise from oncogenic mutations within single cells, the temporal and spatial heterogeneity observed within tumours is as a consequence of the same principles that brought organisms to existence.

1.1.2 The Clonal Evolution of tumours

Natural selection is a process occurring on a cellular level, as well as at the level of an individual. The intratumour heterogeneity observed in tumours can be attributed to a form of *in situ* Darwinian evolution (Greaves and Maley 2012, McGranahan and Swanton 2017). First proposed by Peter Nowell in 1976, the clonal evolution model of cancer relies upon the accumulation of genomic aberrations (e.g. point mutations and copy number alterations) which confer a selective advantage over other cells and hence are positively selected. The genetic instability hallmark of tumour cells leads to the further mutation accumulation, creating a highly heterogeneous, highly permissive tumour mass (Nowell 1976). Although linear evolution has been reported, whereby ancestral cancer clones are outcompeted by their "fitter" progeny, tumour heterogeneity is better attributed to a branched evolution model (Anderson *et al.* 2011).

Branched evolution exhibits subclonal expansion whereby tumours become comprised of a "tree of cells", with each individual branch possessing private mutations and importantly individual phenotypic traits. For the latter, this is believed to be a key mechanism of therapeutic resistance as certain mutations can facilitate escape mechanisms and hence persist and seed metastases (Caldas 2012). Despite this diversity, clonal evolution stipulates a common ancestor, meaning that certain mutations that were acquired early during the tumours evolutionary history are present in every tumour cell (Yates and Campbell 2012). These mutations, termed clonal mutations, form the "trunk" of the tree and are likely to represent priorities for therapeutic targeting for optimal efficacy (Figure 1.1).

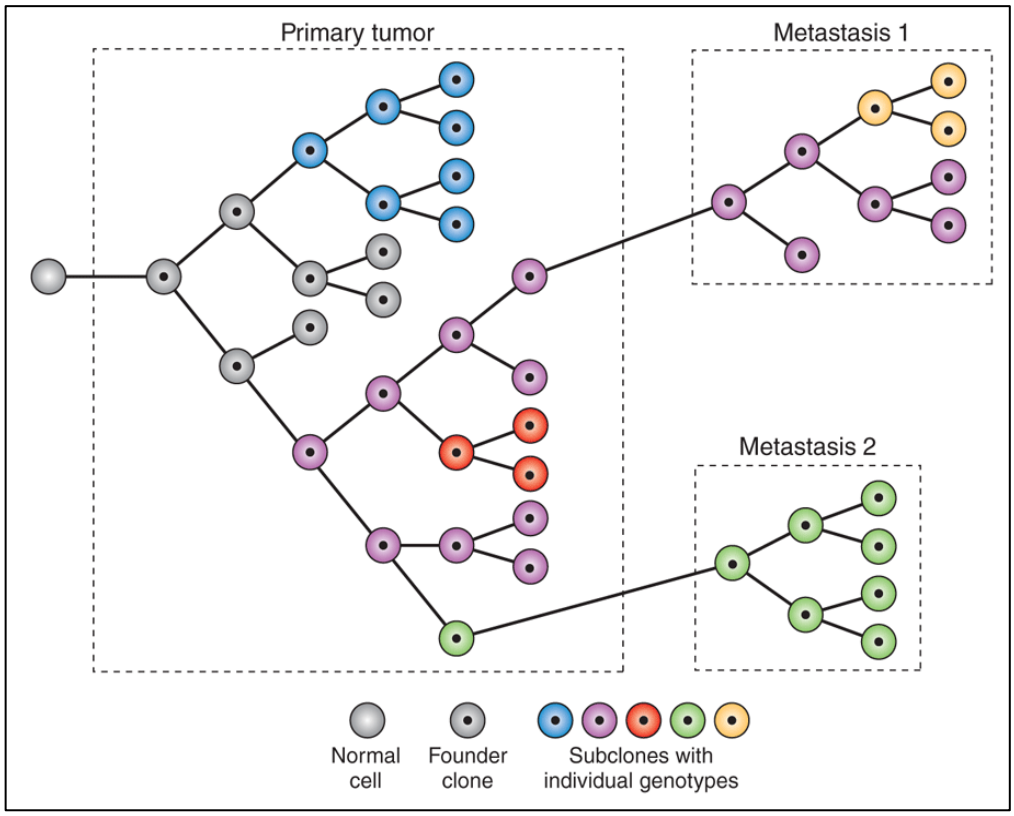


Figure 1.1 – The branched evolution of tumours. The central dot indicates the initial (clonal) somatic alteration acquired by a tumour, and the different colours represent the presence of additional mutations leading to subclonal diversification. Taken from Caldas, 2012.

Our understanding of branched evolution has been elucidated following the establishment of next-generation sequencing (NGS) technology. For example, in the landmark study by Gerlinger and colleagues, they conducted multiregional whole-exome sequencing (WES) on four cases of clear cell renal cell carcinoma. It was revealed that between 63% and 69% of all somatic mutations detected were not detected across every tumour region. However, through phylogenetic reconstruction, the group were capable of differentiating between early, clonal events and later mutations in genes such as *VHL, SETD2 and BAP1* (Gerlinger *et al.* 2012). In this study, *VHL* inactivation was characterised as a ubiquitous (clonal) event, whereas mutations to *SETD2* and *BAP1* were often subclonal. It is therefore logical to assume that the successful identification and exploitation of these early mutations is crucial in providing truly curative effects for patients.

The large scale genomic analyses the group performed were highly reliant on the acquisition of multiregional tumour material; an aspect that is not applicable in a

clinical setting (Bedard *et al.* 2013). Currently, the gold standard method for accurately assessing tumour status both prior and during cancer treatment is *via* the resection of tumour tissue (biopsy). However, it is likely that a single tumour biopsy will underestimate the mutational burden of heterogeneous tumours, meaning studies such as these are vital for building a framework for the mutational disease process (Gerlinger *et al.* 2012). Other studies have been conducted across a multitude of different malignancies including breast (Yates *et al.* 2015) , lung (de Bruin *et al.* 2014, Jamal-Hanjani *et al.* 2017) and colorectal adenocarcinoma (Sottoriva *et al.* 2015).

In addition to their potential deficiencies in assessing intratumour heterogeneity, tissue biopsies are temporal 'snapshots' that cannot accurately assess the dynamics of tumour evolution. The acquisition of tissue biopsies at multiple time points is both impractical and highly morbid for the patient. Therefore, much effort is now being put into the development of minimally-invasive blood based procedures which can accurately characterise and monitor cancer evolution in real-time (Diaz and Bardelli 2014).

1.2 Circulating-free DNA (cfDNA)

1.2.1 Background

Published in 1948, Mandel and Metais first discovered the presence of circulating nucleic acids (later confirmed as DNA) in the blood (Mandel and Metais 1948). Yet, the application of this knowledge remained dormant for three decades, receiving little clinical interest. Awareness reignited in 1977 following the discovery that circulating-free DNA (cfDNA) titre was significantly higher in the blood of pancreatic cancer patients. Furthermore, the group showed cfDNA levels reduced following the administration of chemotherapy (Leon *et al.* 1977). As the understanding of genetics has advanced, the potential of cfDNA as a diagnostic, monitoring and prognostic tool in cancer research has rapidly increased.

Both healthy and abnormal cells are known to release cfDNA into the blood. Yet, the precise mechanisms underlying this release are still unclear; the likely routes being *via* apoptosis, necrosis and active release. Gel electrophoresis of DNA derived from apoptotic cells generally exhibits a characteristic 'ladder' of fragment sizes, due to the cleavage of chromatin DNA into internucleosomal fragments (~180bp). cfDNA fragmentation is often consistent with this electrophoretic pattern, indicating apoptotic origin (Jahr *et al.* 2001).

Studies have proposed a role for necrosis in cfDNA release, as not uncommonly, cancer patients exhibit much larger cfDNA fragments (up to 10,000bp) (Jahr *et al.* 2001). Nevertheless, the observation that cfDNA levels often decrease following radiation therapy, indicate that necrosis is not likely to be a major cfDNA release mechanism (Leon *et al.* 1977). Taken together, the apoptosis/necrosis theory is consistent with the higher cell turnover exhibited by cancer cells.

Active release from viable cells is considered as another potential route of cfDNA release. Cell culture experiments have shown that cells can actively release cfDNA *via* exosomes and even alluded to a potential biological role of cfDNA in the stimulation of tumour cells (Fernando *et al.* 2017, W. Wang *et al.* 2017). It seems likely that several mechanisms collectively contribute to cfDNA release into circulation.

1.2.2 cfDNA as a biomarker

Along with circulating tumour cells (CTCs), cfDNA analysis has become synonymous with the term 'liquid biopsy', first proposed by the Pantel group (Diaz and Bardelli 2014). Although cfDNA levels are generally increased in cancer patients, the sole use of cfDNA titre as a cancer biomarker is considered weak, as many other conditions including diabetes and myocardial infarction also exhibit elevated cfDNA levels (Leon *et al.* 1977). The potential of cfDNA as an accurate biomarker is predominantly rooted in its ability to detect genetic aberrations symptomatic of tumour cells. Since 1994, specific point mutations have been successfully detected in cfDNA. Sorenson and colleagues were first to demonstrate the presence of point mutations of *KRAS* in plasma of pancreatic cancer patients, emphasising a tumour origin of cfDNA (Sorenson *et al.* 1994).

It has been shown that tumour-derived component of cfDNA, or circulating-tumour DNA (ctDNA), can be detected in the blood, amongst a largely healthy cfDNA background. As cancer is quintessentially a genetic disease, the information obtained *via* cfDNA analysis could provide a very detailed mutational profile of tumours, even in the absence of tumour tissue DNA. ctDNA analysis has since been shown capable of detecting a wide range of genetic and epigenetic alterations, including point mutations, loss of heterozygosity, somatic copy number alterations (SCNA) and hypermethylation (Jung *et al.* 2010).

The establishment of ctDNA analysis as a biomarker is not without challenges, particularly in relation to its short fragment size and its low fractional abundance in many patients, especially in earlier stages of disease. Some studies have also shown that ctDNA is generally shorter in fragment length than that derived from healthy cells, meaning detection methods need to account for this in NGS panel design and other PCR-based methods (Mouliere and Rosenfeld 2015). Likewise, the highly abundant wild-type cfDNA background means that highly sensitive and specific methods are required for ctDNA study; a problem that is now becoming circumvented with methods such as Droplet Digital PCR (ddPCR) and deep sequencing technologies.

A key development in recent NGS technology is centred on improved limit of detection (LOD) when analysing cfDNA. Initially ctDNA detection was limited to samples from patients with significant tumour burden, where variant allele fractions (VAFs) were high (> 2 %), meaning that cfDNA analysis would typically be limited to more advanced/metastatic stages. The reduction of LOD in recent years however, has extended the potential of cfDNA analysis to all tumour stages. Modern NGS technologies now offer interrogation of cfDNA below 0.1% VAF (Newman *et al.* 2014). This improvement in sensitivity is likely to have a radical impact on the potential of cfDNA for the molecular analysis of cancer, namely as a biomarker for early detection, the identification of minimal residual disease (MRD), the assessment of treatment response and tumour evolution.

The utility of cfDNA analysis with respect to the early cancer diagnosis is unsurprisingly an area of great interest, but is also the most difficult to execute. Studies have shown that ctDNA correlates with tumour burden, meaning that a small tumour size often constrains the ability to detect ctDNA, at least when using current technologies and "typical" DNA input quantities (~5-20 ng cfDNA) (Dawson *et al.* 2013, Schreuer *et al.* 2016). One large study found that across multiple malignancies 23/49 (47%) of stage I cancers had ctDNA detected, compared to 113/136 (83%) of those with stage IV disease (Bettegowda *et al.* 2014). Detection of stage I malignancies in plasma is encouraging however, there has been shown to be great diversity between different solid tumour types and subtypes. Another problem with cfDNA for early detection, is that in the absence of tumour tissue as a reference point, the differentiation between individuals with early-stage malignancy and healthy persons will be difficult; it is widely known that cancer-associated mutations accumulate in a healthy ageing population, meaning "over-diagnosis" is a potential confounder (Genovese *et al.* 2015, Zink *et al.* 2017).

The detection of ctDNA in patients has been shown to be of prognostic value. Hadano and colleagues demonstrated ctDNA-positive patients following surgery for pancreatic ductal adenocarcinoma had a significantly poorer overall survival than ctDNA-negative patients (13.6 vs 27.6 months) (Hadano *et al.* 2016). Another study focused on metastatic breast cancer showed that ctDNA copies

per millimetre of plasma correlated with poorer survival; a trend that was not observed with the levels of breast tumour marker, CA 15-3 (Dawson *et al.* 2013).

In patients receiving radical surgery, ctDNA analysis is an attractive option for the detection of MRD, as longitudinal blood sampling is a minimally-invasive and efficient process (Wan *et al.* 2017). There are currently no effective means of clinically distinguishing between patients that are disease-free and those with MRD following curative surgery, as imaging and clinical examination are not yet capable of such sensitivity (Tie *et al.* 2016). However, in the research setting cfDNA analysis has been shown capable of this from various studies. One study provided evidence of breast cancer dormancy (MRD) by the analysis of cfDNA from patients otherwise considered cured of their disease. They observed high-level amplification of genes with oncogenic potential (e.g. *USP17L2* (*DUB3*), *BRF1*, *MTA1*, and *JAG2*) in cfDNA, concordant with that observed of matched primary tumour DNA. It was also clear that SCNA analysis of cfDNA was capable of distinguishing between breast cancer patients and healthy controls in the cohort (Shaw *et al.* 2012). Likewise, another study focused on stage II colorectal cancer, post-surgically found ctDNA detectable in 11/14 (79%) patients which subsequently relapsed within a median 27 month follow-up period (Tie *et al.* 2016). By comparison, only 16/164 (9.8%) ctDNA-negative patients experienced relapse in this study. It is therefore hoped cfDNA analysis could be used in the clinical setting to inform upon the success of radical treatment and accordingly affect clinical decision-making.

It has become clear that despite slow beginnings, cfDNA analysis now has applications throughout the cancer patient journey. One area now subject to much research is whether cfDNA has significant utility in the representation of tumour evolution and tumour heterogeneity.

1.2.3 Using cfDNA analysis to study intratumour heterogeneity

A key challenge of understanding intratumour heterogeneity is rooted in a lack of efficient methods for tracking the evolution of malignancies throughout the course of treatment. Due to tissue biopsy being an invasive surgical procedure, the genomic assessment of tumours cannot be applied at several time points, without significant patient morbidity, high costs and burden to the healthcare system. Several studies have shown that cfDNA analysis can assist in this setting.

As a method of tumour monitoring, cfDNA analysis could eclipse the solid biopsy. As the acquisition of blood is minimally-invasive, cfDNA analysis could potentially monitor genomic changes over time, through follow-up blood sampling; an aspect conventional tissue analysis would be unable to achieve. Therefore, assessing the temporal evolution of tumours *via* cfDNA analysis could provide novel insights into intratumour heterogeneity. As mentioned previously, it is likely that solid biopsies may underrepresent the heterogeneity of tumours. By contrast, cfDNA is not subject to selection bias, and hence could theoretically detect 'fingerprints' of all sub-clones present in the disease (Diaz and Bardelli 2014) (Figure 1.2).

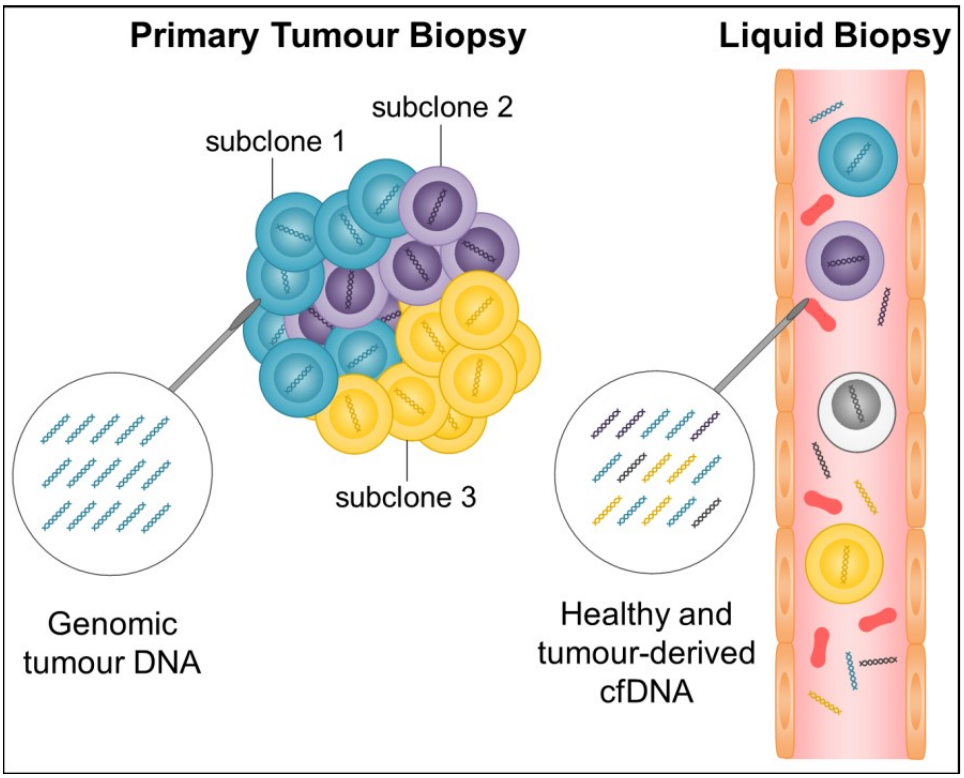


Figure 1.2 – Tissue vs. liquid biopsy. Solid tissue biopsies can be subject to spatial selection, while cfDNA may potentially all subclones present in the primary tumour. Kindly provided by Dr Ricky Trigg.

Concordance between SNVs detected by cfDNA analysis and that of their primary tumour, as well as cfDNA correlating with tumour burden has been well studied. Rothé and colleagues showed that 76% (13/17) of the mutations identified in cfDNA were concordant with that of their primary and metastatic tumours in breast cancer (Rothe *et al.* 2014). Another study assessed the concordance between 'RAS pathway mutations' (*KRAS*, *BRAF* and *NRAS*) in colorectal adenocarcinoma between cfDNA and tumour using droplet digital PCR (ddPCR). 97/100 (97%) cases showed concordance, with interestingly 8/100 (8%) of these cases finding additional variants in the cfDNA; suggestion that cfDNA analysis can provide a more comprehensive assessment of tumour heterogeneity versus the conventional solid tissue biopsy (Siravegna *et al.* 2015). Furthermore, this study provided insight into how clonal evolution and selective pressures can be inferred through the longitudinal monitoring of cfDNA. In the study, patients with resistance to EGFR blockade using cetuximab and panitumumab exhibited the emergence of *KRAS* mutations in plasma. Yet following treatment withdrawal, these *KRAS*-mutant clones declined and sensitivity to treatment was regained.

Even if cfDNA cannot comprehensively account for intratumour heterogeneity, there is evidence that it can detect the dominant subclones which possess increased invasive and/or metastatic potential. Those subclones responsible for metastasis may be increasingly represented in the blood. Such a belief is true of CTCs, attributed to their mesenchymal-like phenotype (Krebs *et al.* 2014, Wu *et al.* 2015). Therefore, matched analysis with CTCs could provide a framework for assessing the accuracy of cfDNA analysis in identifying these metastatic sub-clones (Kidess and Jeffrey 2013). Indeed, a recent study in breast cancer showed that the mutational profile of cfDNA showed complete concordance between individual CTCs in the metastatic setting, as well as certain mutations that were not present in the primary tumour biopsy; indicative of either a minor sub-clone or a later aberration that developed during disease progression (Shaw *et al.* 2017)

The monitoring of emerging resistance to treatment using cfDNA is quickly approaching the clinic. Murtaza and colleagues conducted serial sampling and whole-exome sequencing on cfDNA from six patients whilst receiving therapy

(Murtaza *et al.* 2013). By detecting mutations in cfDNA, the group were capable of showing that the emergence of therapeutic resistance correlated with the increasing VAFs of certain mutations. This included a truncating mutation in *RB1* following treatment with cisplatin and an activating *PIK3CA* p.E545K mutation while on Paclitaxel treatment. In breast cancer, the emergence of *ESR1* mutations with patients receiving endocrine therapy has been shown in recent studies (Fribbens *et al.* 2016, Page *et al.* 2017). Tumour dynamics and evolution of cancer is fast becoming uncovered through cfDNA analysis, but certain considerations and refinements are critical for future success.

1.3 Clonal haematopoiesis of indiscriminate potential (CHIP)

One confounder to cfDNA analysis is the presence of low frequency variants that are not tumour-derived, but rather aging-related somatic mutations acquired by haematopoietic cells (Zink *et al.* 2017). This phenomenon, termed clonal haematopoiesis of indiscriminate potential (CHIP) has been shown to occur more frequently in patients with cancer (solid and haematological malignancies) and with increasing age; a critical consideration as the vast majority of patients with solid tumours are > 50 years (Genovese *et al.* 2015). As white blood cells (WBC) contribute to cfDNA fraction, they have the potential to confound cfDNA-based assays, and a number of cancer consensus genes like *JAK2*, *TP53* and *KRAS* have all been shown to have CHIP-associated variants (Hu *et al.* 2018). With the development of highly sensitive technologies, the detection of CHIP variants is increasing and without being accounted for, there will be a tendency to incorrectly attribute them to intratumour heterogeneity.

One common strategy that can account for CHIP in addition to germline variants, is to sequence DNA derived from leukocytes (buffy layer) as a control, however costs of NGS have meant that these samples are often neglected, or are only sequenced by WES along with the primary tumour tissue sample. The weakness of the latter is that CHIP variants are frequently below 1% VAF and therefore WES depth is often not deep enough to detect these variants. The detection of CHIP variants may be an independent prognostic factor in patients with solid tumours, as those observed in patients with cardiovascular disease and haematological malignancies (Jaiswal *et al.* 2014, Jaiswal *et al.* 2017).

1.4 Thesis aims and hypotheses

The aim of this thesis is to use high-throughput next generation sequencing and related technologies including ddPCR to investigate the utility of ctDNA detection in patients with early-stage NSCLC and MPM, as well as for resolving an important diagnostic challenge; the differentiation between metastatic and independent second primary tumours.

This thesis is comprised of three results chapters which address specific objectives as follows:

i. Pre-surgical and longitudinal detection of ctDNA in patients that have received radical treatment for early-stage non-small cell lung cancer (NSCLC), as part of the Lung TRACERx study (Chapter 3):
ii. Preliminary analyses that determine whether ctDNA is detectable in patients with malignant pleural mesothelioma (MPM) that have received surgery as a primary treatment, using a patient-specific, phylogenetic approach (Chapter 4):
iii. An assessment of targeted NGS (tNGS) as a tool to differentiate between metastases and independent primary malignancies using FFPE tumour tissues (Chapter 5).

This thesis will investigate the hypotheses that:

i. ctDNA is detectable in the cfDNA of patients with early stage NSCLC and with MPM, and ctDNA detection will be optimal due to the targeting of 'clonal' single nucleotide variants, as confirmed from multiregional tumour WES data (Chapters 3 and 4); and
ii. tNGS can accurately diagnose tumour pairs as primary - metastatic cases or independent primary pairs without pathological diagnosis input (Chapter 5).

Chapter 2 Materials and Methods

2.1 Statement of regulatory approval

All work using human tissue samples, including whole blood, fresh-frozen and FFPE tissue samples was conducted within a Good Clinical Laboratory Practice (GCLP) compliant laboratory.

2.2 Patient tissue

Patient tissue for this thesis was collected in accordance with the World Medical Association Declaration of Helsinki entitled 'Ethical Principles for Medical Research Involving Human Subjects' and follows the terms and conditions of the ethical approval given to the study. The TRACERx study (https://clinicaltrials.gov/ct2/show/NCT01888601) is approved by an independent Research Ethics Committee (REC) (Reference, 13/LO/1546). The MEDUSA study has been approved by University Hospitals of Leicester (Ethics number, UHL 131283) and the PRIMINGS study was approved by the NRES Committee East Midlands - Nottingham 2 (reference 14/EM/1265). Metastatic NSCLC samples were obtained from the Leicester Cancer Research Biobank (REC 13/EM/0196, Study number UHL 0429, Biobank UHL number, UHL11274).

2.3 Tissue and DNA Isolation

2.3.1 Isolation of plasma and leukocytes from whole blood

All Blood samples were collected in EDTA tubes and blood components (leukocytes and plasma) were isolated ≤ 2 hours post-venepuncture. Freshly collected whole blood was centrifuged for 10 minutes at 1,000 x g at 4°C to separate blood components. Plasma was carefully transferred to clean plastic ware, avoiding contamination from the buffy coat and packed erythrocytes, and centrifuged for a further 10 minutes at 2,000 x g at 4°C, before being aliquoted into 1.5mL tubes. Leukocytes; a source of germline control DNA, were isolated by carefully pipetting the buffy coat layer and placed in 1.5mL tubes. Both plasma and buffy coat aliquots were placed at -80°C for later use. All contaminated material was disposed of in a clinical waste container or in an autoclave bag following overnight soaking in Distel laboratory disinfectant.

2.3.2 cfDNA extraction from plasma

cfDNA was isolated from 3mL of plasma using the QIAamp® Circulating nucleic acid kit (Qiagen), according to the manufacturer's protocol. This kit uses a silica membrane-based DNA purification technology, where using several washing steps, contaminants (e.g. proteins, RNA) are effectively removed. The protocol was as follows:

3mL of plasma was placed into a 15mL tube and spun at 1000 x g in a bench centrifuge for 5 minutes. 3mL Plasma was separated from any potential pellet formed, before being added to 300µL Proteinase K and 2.4mL Buffer ACL in a 50mL Falcon tube and vortexed for 30 seconds. Samples were placed in a water bath at 60°C for 30 minutes. Following this, 5.4mL of Buffer ACB was added to the tube and vortexed for 30 seconds and placed on ice for 5 minutes. Subsequently the samples were transferred into a 20mL extender affixed to QIAamp Mini columns mounted on a QIAvac 24 Plus vacuum manifold with VacConnectors. A vacuum was subsequently applied, resulting in the passing of the lysate through the column membrane, whilst retaining DNA. The vacuum was switched off, the 20mL extender was removed and the column was washed using 600µL, 700µL and 700µL of Buffer ACW1, ACW2 and 100% ethanol respectively (with vacuum reapplied with each wash reagent). Following the ethanol wash, the column was removed from the QIAvac 24 Plus and placed into a clean 2mL collection tube and centrifuged at 13,100 x g for 3 minutes. In order to ensure the membrane was free of ethanol, the column was placed into a clean 2mL collection tube and placed in a 56°C incubator for 10 minutes with the lid open to dry the membrane. Finally, the column was placed in a 1.5mL Lo-Bind Eppendorf tube, 150µL of elution Buffer AVE was added to the column, incubated for 3 minutes at RT, before being centrifuged at 13,100 x g for 1 minute to elute the nucleic acids. DNA was quantified by the 4200 TapeStation System (Agilent Technologies) (see section 2.4.2). Eluted DNA was stored at 4°C for immediate use (up to 6 weeks) or at -20°C for later use.

2.3.3 cfTNA extraction from plasma

cfDNA and cfRNA; collectively termed Cell-Free Total Nucleic Acid (cfTNA), was isolated manually using the MagMAX™ cfTNA kit (Thermo Fisher Scientific), according to the manufacturer's instructions. This uses a magnetic bead-based purification technology which removes centrifugation steps required in silica membrane-based methods. The protocol was as follows:

4mL plasma was combined with 60µL MagMAX™ cfTNA Proteinase K and 2mL MagMAX™ cfTNA Lysis/Binding Solution and incubated for 30 minutes at 65°C with shaking at 1,000 rpm. Samples were placed on ice for 5 minutes, before a further 3mL of Lysis/Binding Solution was added. 120µL of well-vortexed MagMAX™ cfTNA Magnetic Beads were then added and the tube was mixed for 10 minutes at 1,000 rpm. Subsequently, beads were pelleted by placing the 50mL tube on a DynaMag™-50 Magnet stand, the supernatant was discarded prior to bead re-suspension using 1mL of Wash Solution 1. The bead slurry was transferred to a 1.5mL tube and placed on a DynaMag™-2 Magnet to pellet the beads, before discarding the supernatant. Following this, two 1 mL 80% ethanol washes occurred, where the pellet was re-suspended and then returned to the DynaMag™-2 Magnet. The beads were subsequently air-dried for 5 minutes, before being re-suspended with 400µL MagMAX™ cfTNA Elution Solution and vortexed at high speed for 5 minutes. The tube was returned to the magnet to capture beads and then the supernatant was transferred to a clean 1.5mL tube. 500µL of Lysis/Binding Solution and 10µL Magnetic Beads were added and the tube was vortexed for 5 minutes, in order to concentrate cfTNA. The tube was then returned to the magnet, beads were pelleted before the supernatant was removed, and 1mL of Wash Solution 1 was used to wash the beads by mixing and returned to the magnet, with the supernatant being removed. Two 1mL 80% ethanol washes (as previously described) then occurred, before the pelleted beads were air-dried for 3 minutes. Finally, 50µL of MagMAX™ cfTNA Elution Solution was used to re-suspend the beads by mixing by vortex for 5 minutes, then was returned to the magnet. The supernatant was transferred to a fresh Lo-Bind Eppendorf and cfTNA was stored at -80°C to preserve cfRNA integrity.

2.3.4 Lyophilisation and reconstitution of cfDNA

For cfDNA samples with limiting concentrations, samples were lyophilised (freeze-dried/sublimated) and reconstituted in a smaller volume to allow subsequent analysis (e.g. ddPCR, NGS library preparation). The lids of LoBind® microcentifuge tubes (Eppendorf) tubes containing cfDNA were pierced with a hypodermic needle, before being frozen at -80°C for ≥2 hours. Frozen samples were subsequently lyophilised using the VirTis BenchTop Pro Freeze Dryer (SP Scientific) at a temperature and vacuum of -60°C and 100mTorr for 3 hours. Following lyophilisation, cfDNA was reconstituted with an appropriate volume of TE Buffer (Invitrogen) and stored at 4°C for immediate use, or -20°C for later use.

2.3.5 Extraction of genomic DNA from Leukocytes

All germline DNA was isolated from 200µL buffy coat sample using the QIAamp® DNA Blood Mini Kit (Qiagen), according to the manufacturer's protocol. The protocol was as follows:

200µL of buffy coat sample was combined with 20µL protease in a 1.5mL Eppendorf and vortexed for 15 seconds. 200µL Buffer AL was added to the sample vortexed for a further 15 seconds before being incubated in a 56°C water bath for 10 minutes. 200µL 100% ethanol was then added, vortexed for 15 seconds and then the mixture was added to a QIAamp® Mini Spin column placed in a 2mL collection tube. The spin column was subsequently centrifuged at 4,300 x g for 1 minute or until all of the mixture passed through the column. The column was then transferred to a fresh 2mL collection tube and was washed by 500µL of Buffer AW1 and centrifuged at 4,300 x g for 1 minute. Again, the column was transferred to a new 2mL collection tube, before 500µL Buffer AW2 was added and was then centrifuged at 13,100 x g for 3 minutes. Flow-through was then discarded and the column was returned to the same collection tube and subsequently centrifuged for a further 1 minute at 13,100 x g to completely dry the membrane. Finally, the column was transferred to a clean 1.5mL Lo-Bind Eppendorf tube, 200µL Buffer AE was added to the column, incubated for 5 minutes at RT, before being centrifuged for 1 minute at 4,300 x g to elute the DNA. Germline DNA quantified on the Qubit 2.0 fluorometer (Invitrogen) using

the Qubit™ dsDNA BR Assay Kit reagents (Invitrogen), with a detection range of 100 pg/μL – 1000 ng/μL (see section 2.4.3). DNA was stored at 4°C.

2.3.6 Extraction of genomic DNA from fresh tumour tissue

Fresh-Frozen tumour tissue (stored at -80°C) was sectioned by the Histology Facility, Core Biotechnology Services, University of Leicester. Sectioned tissue (approximately 5 x 5μm sections per tissue region) were immediately placed in 180μL Tissue Lysis Buffer ATL (Qiagen), before immediate DNA extraction using the QIAamp® DNA Mini Kit (Qiagen). H&E staining was conducted for each tissue, in order for tumour content to be assessed by a qualified histopathologist (David Moore), before being selected for whole exome sequencing (WES). The DNA was isolated using the following protocol:

Cut sections in 180μL Buffer ATL were incubated with 20μL Proteinase K (Qiagen) in a 1.5mL Eppendorf, vortexed well and incubated at 56°C for 24 hours, or until the tissue was completely lysed. Following this, 200μL Buffer AL was added, vortexed for 15 seconds, before being placed in a hot block for 10 minutes at 70°C. 200μL 100% ethanol was added, well mixed and then briefly centrifuged. The mixture was subsequently applied to a QIAamp® Mini Spin Column and centrifuged at 6,000 x g for 1 minute. The column was then transferred to a clean 2mL collection tube, where 500μL Buffer AW1 was added to the column, spun at 6,000 x g for 1 minute. Likewise, 500μL Buffer AW2 was added, then centrifuged at 13,100 x g for 3 minutes. The collection tube was emptied of its contents and then was placed back under the column and then centrifuged for a further 1 minute at 13,100 x g to ensure the column membrane is well-dried. Finally, the column was placed in a 1.5mL Lo-Bind Eppendorf tube, 100μL elution Buffer AE was added and incubated at RT for 5 minute, then centrifuged at 6,000 x g for 1 minute to elute DNA. All DNA was consequently quantified using the Qubit® Fluorimeter 2.0 (Invitrogen) using the DNA high sensitivity kit. DNA was stored at 4^0C.

2.3.7 DNA extraction from FFPE tissue

2.3.7.1 PRIMINGS Unambiguous cohort (Phase A)

Prior to DNA extraction, regions of tumour tissue containing the highest tumour content were identified by a consultant pathologist (Dr David Moore) *via* haematoxylin and eosin (H&E) staining analysis. Subsequently, the region with the highest tumour content was cored from the corresponding FFPE blocks using a 1mm diameter corer and sent for DNA extraction. All FFPE DNA was extracted using the GeneRead^{TM} DNA FFPE Kit (Qiagen) in accordance with the manufacturer's instructions. This kit includes the enzyme, Uracil DNA glycosylase (UNG), which allows the enzymatic removal of cytosine deamination artefacts (caused by formalin fixation process), which otherwise would lead to artificial C>T mutations being called by next-generation sequencing, reducing the quality of results. The protocol for the GeneRead^{TM} DNA FFPE extraction was as follows:

Tissue cores were deparaffinised using 155µL deparrafinisation solution and incubated at 56°C for 3 minutes in a 1.5mL Safelock tube. 55µL RNase-free water, 25µL Buffer FTB and 20µL Proteinase K were added and well vortexed, then placed in an incubator at 56°C for a minimum of 24 hours to ensure adequate protein digestion. By this stage clear separation between deparaffinisation solution (blue) and the remaining reagents and tissue (clear) formed; the clear solution was isolated and placed into a fresh 1.5mL Safelock tube. The tube containing digested tissue was then placed in a heat block at 90°C for 1 hour. Following this, the clear phase was transferred into a fresh safelock tube, leaving behind any potential deparaffinisation solution/denatured proteins (white lumps) present. 115µL RNase-free water and 35µL UNG were added and mixed well then incubated at 50°C for 1 hour. Subsequently, 2µL RNase A was added, incubated for 2 minutes at room temperature (RT), before the addition of 250µL of Buffer AL (lysis buffer) and 250µL 100% ethanol respectively. The lysate was then transferred into a MinElute column inside a 2mL collection tube and centrifuged for 1 minute at 13,100 x g. The flow-through was discarded before two washes of the column with equal 500µL volumes of Buffer AW1 and AW2, centrifuging at 13,100 x g for 1 minute for each respectively and discarding the flow-through. 250µL of 100% ethanol was then added to the column and

centrifuged for 13,100 x g for 1 minute. The column was subsequently transferred to a clean collection tube and was centrifuged for another 1 minute at 13,100 x g to ensure the column membrane was free of contaminants. Finally, the column was transferred into a 1.5mL Lo-Bind Eppendorf tube and DNA was eluted using 60µL Buffer ATE, then quantified using firstly Qubit™, then by RT-qPCR (see sections X and X). All DNA was stored at 4°C.

2.3.7.2 PRIMINGS Ambiguous cohort (Phase B)

DNA was isolated from FFPE tissue cores for the ambiguous PRIMINGS cases (n =15 samples) using a largely automated procedure on the KingFisher™ Flex Purification System (Thermo Fisher Scientific) with the MagMAX™ FFPE DNA/RNA Ultra Kit (Thermo Fisher Scientific).

Tissue cores in Safelock 1.5mL tubes (Eppendorf) were firstly deparaffinised using 1mL xylene and vortexed briefly and placed on a heat block at 50°C for 3 minutes to melt the paraffin. Following this, each tube was centrifuged at 13,100 x g for 2 minutes to pellet the tissue. The xylene was subsequently removed, before the tissue was washed twice with 1mL of 100% ethanol, vortexed, and centrifuged at 13,100 x g for 2 minutes. All ethanol was removed from the tubes and the pellets were air-dried at RT for 30 mins. 210µL Protease Solution was added and gently mixed before being incubated at 55°C for 1 hour. Following this, the tubes were transferred to another heat block at 90°C for a further 1 hour, then centrifuged briefly to collect any condensation droplets.

DNA processing plates were then set up in preparation for the isolation of DNA using the KingFisher™ Flex Purification System using the "A31881_FLEX_std_DNA" program, according to the manufacturer's instructions. Eluted DNA was placed at 4°C following quantitation using the TapeStation instrument (see section 2.4.2).

2.4 Analytical methods

2.4.1 Polymerase chain reaction (PCR)

2.4.1.1 Oligonucleotide sequences

Some of the oligonucleotides used in this thesis are found within the results section of Chapter 4 (see section 4.3.2.2.). The remaining oligonucleotide sequences are found in Table 2.1.

Assay	Component	Oligonucleotide sequence (5' to 3')
CDKN2A (83bp)	**F (common)**	AAGTCGTGGCCTTTCAACAC
	R	AGTGCGCTTCTGAGACTCTT
	P (common)	TCTCCATCTGGCTTGGAA (FAM-MGB)
CDKN2A (61bp)	**F (common)**	AAGTCGTGGCCTTTCAACAC
	R	CAGCCATTTCGTTCCCTTCC
	P (common)	TCTCCATCTGGCTTGGAA (FAM-MGB)
RPPH1 (62bp)	**F**	CGGAGGGAAGCTCATCAGTG
	R	GACATGGGAGTGGAGTGACA
	P	CACGAGCTGAGTGCGT (VIC-MGB)
GAPDH (95bp)	**F (common)**	GGCTAGCTGGCCCGATTT
	R	GGACACAAGAGGACCTCCATAAA
	P (common)	ATGCTTTTCCTAGATTATTC (FAM-MGB)
GAPDH (186bp)	**F (common)**	GGCTAGCTGGCCCGATTT
	R	TTTGCGGTGGAAATGTCCTT
	P (common)	ATGCTTTTCCTAGATTATTC(FAM-MGB)

Table 2.1 – Oligonucleotide sequences for assays used in Chapters 4 and 5. Forward primer (F), reverse primer (R) and probe (P) sequences are shown. Included are two *CDKN2A* copy number assays of differing amplicon lengths, along with the reference assay targeting *RPPH1*. The DNA fragmentation assays used in Chapter 5 are also shown.

2.4.1.2 Assay design, validation and optimisation

All in-house oligonucleotides were designed using the Primer3 primer design tool (Whitehead Institute, MIT). Sequences of interest were aligned to the human genome using the NCBI BLASTN tool, to confirm specificity. Following this, sequences of primer/probe candidates were entered on the OligoAnalyzer 3.1 (Integrated DNA Technologies) software, to predict their tendency to form hairpins and homodimers/heterodimers. As a general rule, oligonucleotides with a secondary structure ΔG value more negative than -6 kcal/mol were excluded. Once satisfactory, the amplicon length produced by the primers was verified using the UCSC *in silico* PCR tool using the GRCh37/hg19 reference genome. Where possible, the following parameters were followed: GC content of 40-60%,

length of 15-25 bp, primer T_m of 55-62°C, probe T_m of 65-72°C, ≤ 3 consecutive G/C bases, a GC-clamp of ≤ 3 G/C bases, and an amplicon length of <100bp (particularly for cfDNA analysis). Desalted primers were delivered lyophilised and were subsequently reconstituted using sterile TE Buffer (Invitrogen, 10 mM Tris-HCl (pH 8.0) 0.1 mM EDTA) to a stock concentration of 200 µM.

For probe-based assays, designed primers were wet-lab validated using Fast SYBR® Green melt curve analysis, prior to the introduction of the probe, in order to assess primer specificity, efficiency and potential secondary structure formation. Each primer was diluted to a 10 µM working stock and a standard curve was constructed using serially diluted (1:2 dilution; 7 points) human genomic DNA (Roche) ranging from 20 – 0.3125ng per well. Each dilution was ran in triplicate, with a no template control (NTC). A total reaction volume of 10µL was comprised of the following; 5µL Fast SYBR® Green Master Mix (2X) (Applied Biosystems), 0.6µL forward primer (600nM final concentration), 0.6µL reverse primer (600nM final concentration), 0.2µL sterile TE Buffer (Invitrogen) and 3.6µL template DNA. Reactions were set up in MicroAmp® Fast Optical 96-well reaction plates (Applied Biosystems), sealed with MicroAmp® adhesive film (Applied Biosystems), centrifuged and run on the StepOnePlus™ Real-Time PCR System (Applied Biosystems) under the following conditions (Table 2.2):

Stage	Temp (°C)	Time	
Enzyme activation	95	20 secs	
Denaturation	95	3 secs	X 50
Annealing/extension	60	30 secs	
Melt curve	60 to 95*	n/a	

*fluorescence readings every increase of 0.3°C.

Table 2.2 – Cycling conditions for SYBR-Green qPCR.

The standard curve efficiency was calculated as:

$$\text{Efficiency } (\%) = \left(10^{\frac{-1}{\text{slope}}} - 1\right) \times 100$$

An efficiency between 90% and 110%, and an R^2 > 0.98 was considered acceptable. Secondary products, where applicable were identified through the inspection of the melt curve (Figure 2.1).

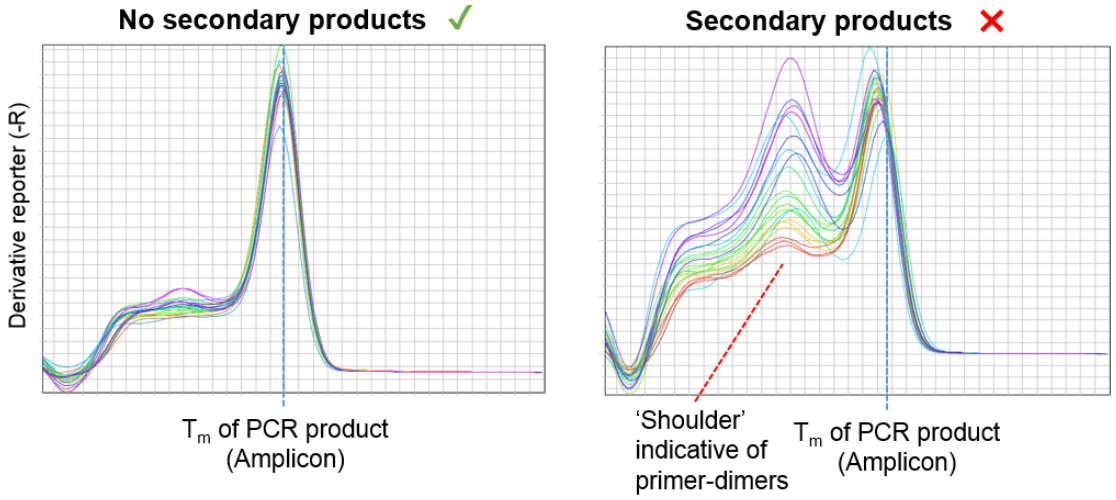


Figure 2.1 – Melt curve examples to assess primer quality. Primers with little/no secondary product formation will have a single strong peak (left), whereas those with secondary product formation will possess a 'shoulder' which will become more dominant as the concentration of DNA decreases (right).

Once primers were satisfactory, probe-based optimisation then occurred. For qPCR, SYBR® Green was replaced by a TaqMan® probe and assay efficiency was assessed in the same way, using a standard curve of serially diluted human genomic DNA. As Droplet Digital PCR (ddPCR) does not require near-perfect efficiency, a standard curve is not required. ddPCR assays were optimised by using a temperature gradient (typically 53–63°C) in order to identify the optimum annealing temperature (T_a) of each assay. Optimal T_a was defined by the temperature which allowed the largest discrimination between positive and negative droplets, whilst avoiding non-specific amplification.

2.4.1.3 Quantitative real-time PCR (qPCR)

Real-time qPCR provides relative quantitation of DNA, through the generation of a standard curve of known DNA concentrations; in this case, by using human genomic DNA (HGD) (Roche). Depending on the DNA to be quantified (FFPE-DNA or cfDNA), different standard curves were created. For the usually more concentrated FFPE-DNA, a larger standard curve ranging from 20ng – 0.078 ng / well (serial 1:2 dilution; 9 points) was used, whereas for cfDNA this ranged from 5ng – 0.078 ng / well (serial 1:2 dilution; 7 points). Each TaqMan® assay comprised unlabelled forward and reverse primers (reaction concentration, 600nM) and a 5' 6-FAM™, 3' MGB/NFQ-labelled hydrolysis probe (reaction concentration, 200nM). TaqMan® probes utilise a 3' Minor groove binder (MGB) and a 3' nonfluorescent quencher (NFQ), in order to enhance PCR-based applications. A 3' MGB moiety is included to increase the melting temperature (T_m) of the probe (~10°C) and to stabilise probe-target hybrids. A 3' NFQ is incorporated to absorb the signal from the 5' fluorescent dye (e.g. FAM), which decreases background signal and thereby increases assay sensitivity and precision. Typically the assay used for qPCR was an in-house designed assay targeting the *GAPDH* gene (95bp) (for oligonucleotide sequences, see section 2.4.1.1).

Each 10µL PCR reaction was comprised of the following; 5µL TaqMan® Fast Universal PCR Master Mix (2X) (Applied Biosystems), 1.4µL Forward primer/Reverse primer/Probe mix (F/R/P) and 3.6µL DNA (sample or standard DNA). qPCR was conducted using StepOnePlus™ Real-Time PCR System using the following conditions (Table 2.3):

Stage	Temp (°C)	Time	
Enzyme activation	95	20 secs	
Denaturation	95	3 secs	X 50
Annealing/extension	60	30 secs	

Table 2.3 – Cycling conditions for real-time qPCR.

2.4.1.4 Droplet Digital PCR (ddPCR)

Droplet Digital PCR (ddPCR) allows the absolute quantitation of DNA, without the need for standard curve generation. This is possible, as it utilises microfluidic and

emulsion-PCR based technology to partition single DNA samples into thousands (>10,000) of individual reactions, in the form of oil-emulsion nanolitre-sized droplets. For probe-based assays, ddPCR was conducted using two assays; a 'test' assay and a 'reference' assay. A 'test' assay refers to a primer/probe mix that targets either a specific mutation (mutant probe), or a region of a gene subject to copy number assessment. A 'reference' assay refers to a primer/probe mix that either targets the wild-type sequence for the mutation assay of interest (WT probe), or an assay targeting a gene that is not usually subject to copy number alteration (for copy number experiments), thereby being used an internal control. In regards to this thesis, 'test' assay probes were FAM-labelled, whereas 'reference' assays were either HEX or VIC-labelled. All experiments included a positive and negative control, where possible, in addition to a no template control (NTC) comprised of sterile dH_2O to ensure reactions were free of contamination. The ddPCR protocol was as follows:

22µL reaction mixes were produced in 8-strip 0.2mL PCR tubes containing the following reagents: 11µL ddPCR Supermix for Probes (No dUTP) (Bio-Rad), 1.1µL 'test' assay, 1.1µL 'reference' assay and 8.8µL of DNA/dH_2O. 20µL of each reaction mix was subsequently transferred to a single well of a DG8™ Cartridge (Bio-Rad) along with 70µL Droplet Generation Oil for Probes (Bio-Rad), then sealed using a DG8™ gasket (Bio-Rad). In the event that some cartridge wells were empty, 10µL of ddPCR 2 x Buffer Control (Bio-Rad) and 10µL sterile dH_2O was mixed and added to each of the empty cartridge wells. Sealed cartridges were then placed into the QX200™ Droplet Generator (Bio-Rad), to allow the generation of >10,000 nanolitre sized oil droplets, according to the manufacturer's instructions. Following this, generated droplets were carefully transferred to Twin-Tec 96-Well x 250µL Semi-Skirted PCR plates (Eppendorf) and sealed with a Pierceable Foil Heat Seal (Bio-Rad) at 180°C using the PX1™ PCR Plate Sealer (Bio-Rad). The sealed plated then underwent PCR in the C1000 Touch™ Thermal Cycler (Bio-Rad) with the following conditions (Table 2.4):

Stage	Temp (ºC)	Time	Cycles
Enzyme activation	95	10 min	1
Denaturation	94	30 sec	40
Annealing/extension	**X***	1 min	
Enzyme deactivation	98	10 min	1
Hold	4	∞	1

*assay-dependent. Between 53-62°C.

Table 2.4 – Cycling conditions for ddPCR reactions.

Following PCR, plates were then transferred into the QX200™ Droplet Reader (Bio-Rad) using ddPCR Droplet Reader Oil (Bio-Rad) and data was analysed using the QuantaSoft™ Software v1.7 (Bio-Rad) using either Rare Event Detection settings for mutation detection, or Copy Number Variation (CNV) settings for the analysis of copy number (e.g. deletion).

2.4.2 TapeStation Analysis of DNA

The TapeStation 4200 System (Agilent) was used for the quantitation and fragment profiling of DNA, as well as for the quality control of library preparations for NGS. Depending on the sample to analysed, different kits were used as they have differing sensitivities (Table 2.5):

Sample Type	TapeStation Reagents/ ScreenTapes	Accurate analysis range (ng/µL)	DNA size range (bp)	Ladder/Sample Volume + Sample Buffer Volume (µL)
cfDNA	Agilent High Sensitivity D5000	0.01 – 1	100 – 5,000	2 + 2
FFPE DNA	Agilent Genomic DNA	10 – 100	200 – 60,000	1 + 10
NGS Library	Agilent High Sensitivity D1000	0.01 – 1	35 – 1,000	2 + 2

Table 2.5 – TapeStation reagents by sample type.

For cfDNA, this method was used as it can assess the quality of the plasma isolation procedure from whole blood (see section 2.3.1). Other techniques including qPCR can provide accurate quantitation, yet it cannot detect potential contamination from leukocyte DNA derived from the buffy layer. Leukocyte DNA

is fresh, genomic DNA which has a very low degree of fragmentation. Using TapeStation analysis, it is possible to identify higher molecular weight (1,000-10,000bp) DNA, in addition to the 'typical' cfDNA uniform peak found between 90-150bp in length (Figure 2.2). It is of course worth noting, that cfDNA derived from necrotic cancer cells may also show higher molecular weight DNA.

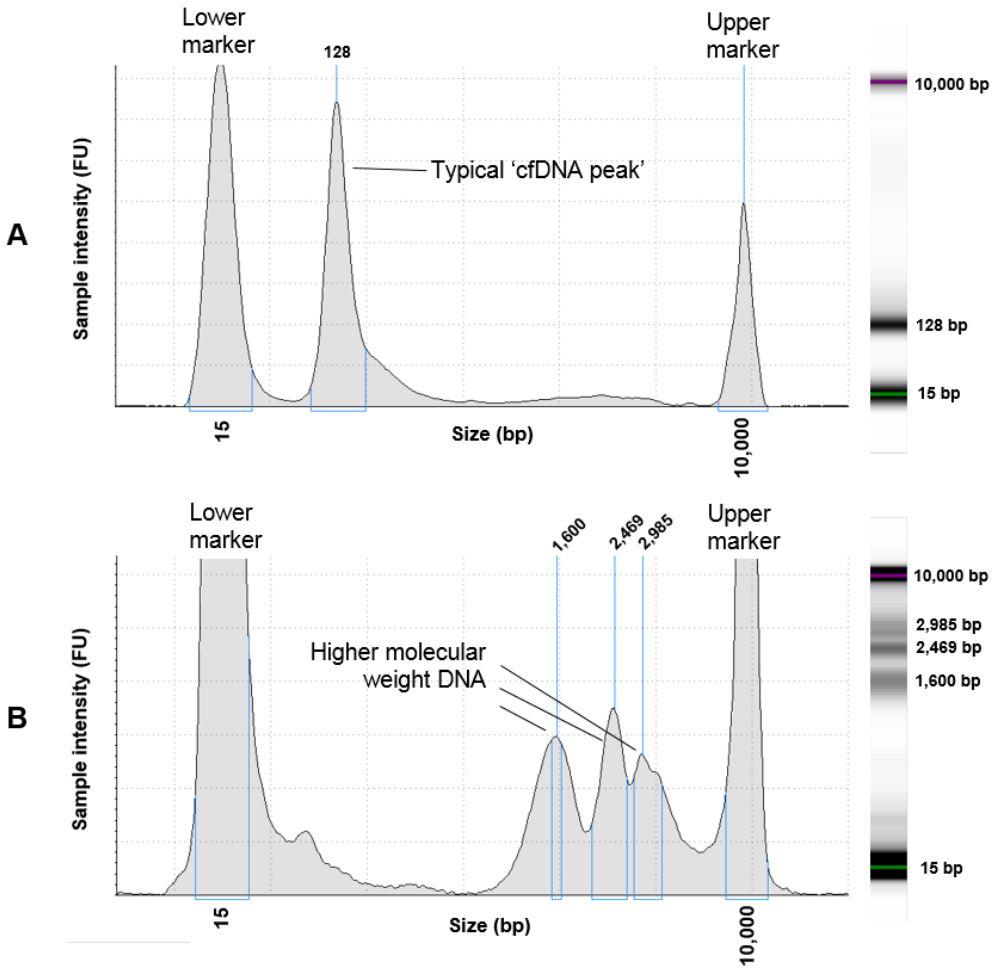


Figure 2.2 – Assessment of blood processing quality by TapeStation analysis of cfDNA. TapeStation electropherograms and gel images for two cfDNA samples, showing a typical cfDNA profile (A), compared to one which may be contaminated with genomic DNA from Leukocytes (B).

All experiments were conducted according to the manufacturer's instructions. For the analysis of cfDNA a region analysis was conducted between both 60-300 bp (cfDNA conc.) and 60-5000bp (total DNA conc.). For NGS Library preparations, a single region analysis of 110-350bp was used, as this is the region at which the amplified products cluster. Region analysis produces both a concentration (pg/µL) and a region molarity for your area of interest, the latter of which was

used for the calculation of 50 pM and 100 pM dilutions for template preparation (see sections 2.4.6.1 and 2.4.6.2).

2.4.3 Qubit quantitation of DNA

The Qubit® 2.0 Fluorometer (Invitrogen) was used to quantify both germline DNA (leukocyte DNA) and DNA derived from fresh-frozen tissues. Qubit® utilises a dye which fluoresces upon intercalation with double-stranded DNA, thereby providing a quantitation of DNA within a sample. The Qubit™ dsDNA HS Assay Kit and Qubit™ dsDNA BR Assay Kits (Invitrogen) were used for the quantitation of fresh-frozen tissue DNA and germline DNA respectively. The protocol was as follows:

A working solution was created using 1µL Qubit® dye and 199µL Qubit® Buffer (per tube to be analysed). Then 2µL sample was combined with 198µL working solution in a Qubit® assay tube. Two reference standards were also used to calibrate the machine before sample quantification, by combining 10µL of each reference standard with 190µL working solution. All tubes were quickly vortexed and incubated for 2 minutes before being placed individually in the Qubit® 2.0 Fluorometer for quantitation. Sample quantitation was recorded in ng/µL.

2.4.4 Statistical analysis

Data from ddPCR/qPCR experiments were imported into MS Excel. Data was presented descriptively as means, medians or proportions. GraphPad Prism 7 (GraphPad Prism Software Inc., CA, USA) was used for statistics tests including Wilcoxon rank-sum test (Mann–Whitney U test). SPSS Statistics 24 (IBM) was used for Kaplan-Meier plots and Cox regression analysis for survival analyses. Any P values in this thesis were considered significant if < 0.05.

2.4.5 Whole Exome Sequencing (WES)

Multiregional WES was conducted on fresh-frozen tumour DNA and matched germline DNA (buffy coat DNA) relating to both the TRACERx (Chapter 3) and the MEDUSA (Chapter 4) studies. TRACERx WES was all conducted externally at the Advanced Sequencing Facility at The Francis Crick Institute, Lincoln's Inn Fields Laboratories. MEDUSA WES was also conducted externally using the following companies; Eurofins AROS, Beijing Genomics Institute (BGI) and Novogene.

2.4.5.1 TRACERx WES

Exome capture was performed on 1-2 µg DNA using a customised version of the SureSelect Human All Exon V5 (Agilent) according to the manufacturer's protocol. Following cluster generation, samples were 100bp paired-end multiplex sequenced on the HiSeq 2500 and HiSeq 4000 (illumina). The data was aligned to the reference human genome (hg19) and median sequencing depths of 431 (range 83-986) for the tumour regions and 415 (range 107-765) for the matched germline was achieved. For further details with respect to sequencing procedure and subsequent analysis, see Supplementary Appendix 1 of the relevant publication from the TRACERx consortium (Jamal-Hanjani *et al.* 2017).

2.4.5.2 MEDUSA WES

Exome capture was performed on ~2 µg DNA using the SureSelect Human All Exon V6 (Agilent) according to the manufacturer's instructions. Samples were 150bp pair-end multiplex sequenced on the HiSeq 4000 (illumina) with a guaranteed Q30 quality score of ≥ 80% and an effective depth of 200. Data was aligned to the human genome (hg19).

2.4.5.3 MEDUSA WES Data Analysis

WES BAM files were generated by Dr Robert Hastings using a workflow that followed GATK best practices (Figure 2.3). In summary this was as follows:

FASTQ files were aligned using BWA (0.7.15) to the human genome (hg19) and variants were called using Mutect2 (v3.6) using filters to exclude low-confidence variants. Only variants that passed the following filters were classified as high-quality and considered in the study: Depth >30, Mutant Reads >5, Mutant VAF >1%, Germline mutant VAF <1%. These BAM files were generated using the University of Leicester HPC services ALICE and SPECTRE. Data was annotated with Variant Effect Predictor (VEP) v86.

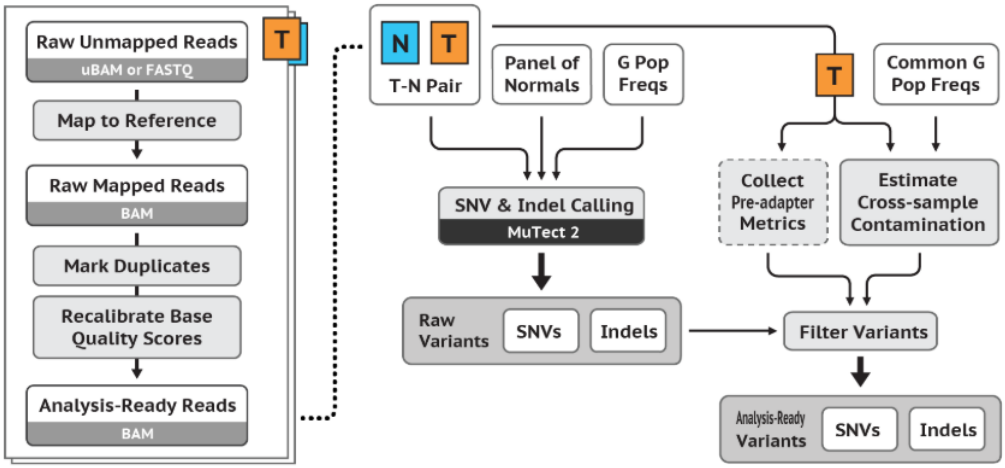


Figure 2.3 – Overview of best practices analysis workflow. Taken from https://software.broadinstitute.org/gatk/best-practices/workflow?id=11146).

SNVs were classified into driver tiers using the Cancer Genome Interpreter (Institute for Research Biomedicine, Barcelona) under the cancer type "PLMeso" (pleural mesothelioma). The Cancer Genome Interpreter uses the OncodriveMUT bioinformatics method to assess the driver status of individual variants, through either validated (known driver events) or predicted means. Each variant was classified as either a Tier 1 driver (high confidence driver event), Tier 2 (potential driver event) or Passenger variant (Tamborero *et al.* 2018).

2.4.6 Targeted next-generation sequencing (NGS)

2.4.6.1 NGS with Oncomine Lung assays

All cfDNA samples in this thesis from the TRACERx and the Metastatic NSCLC Biobank studies were sequenced using the Oncomine Lung cfDNA (OLcfDNA) and Oncomine Lung Cell-Free Total Nucleic Acid Research (OLcfTNAR) assays (Thermo Fisher Scientific) respectively. The latter is an enhanced version of the former, which uses concurrent analysis of cfDNA and RNA to allow for the analysis of gene fusions, exon-skipping mutations and copy number amplification, in addition to 169 hotspots and indels in 11 commonly-mutated genes in lung cancer: *ALK*, *BRAF*, *EGFR*, *ERBB2*, *KRAS*, *MAP2K1*, *MET*, *NRAS*, *PIK3CA*, *ROS1,* and *TP53*. All samples were sequenced with 20ng cfDNA, which would provide ≤ 0.1% limit of detection (LOD) at gene hotspots (Figure 2.4).

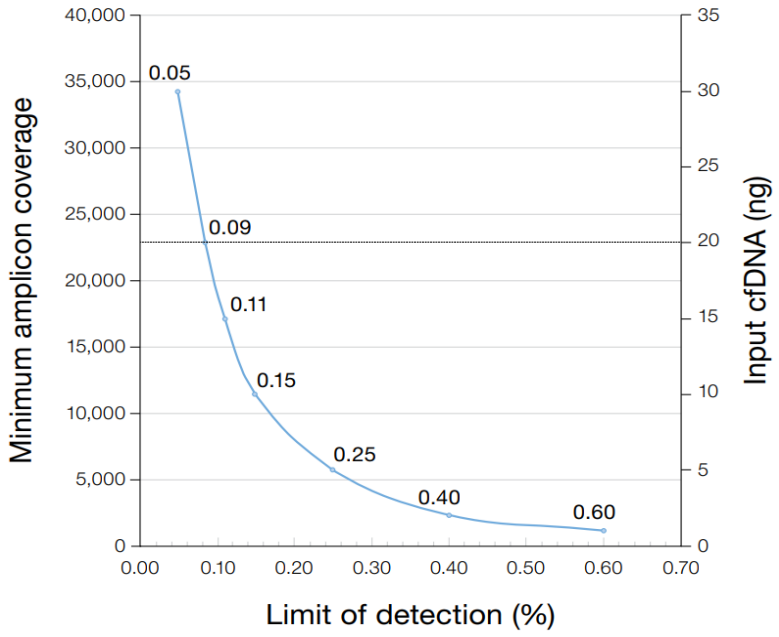


Figure 2.4 – LOD of the OLcfDNA and OLcfTNAR assays based on cfDNA input. Adapted from Thermo Fisher Scientific (https://www.thermofisher.com/order/catalog/product/A31149). 20ng DNA input is indicated.

Library preparation was conducted manually as follows: For the OLcfTNAR assay (Metastatic NSCLC Biobank samples) only, reverse transcription was conducted by combining 20ng cfTNA (in total volume 10.4µL) with 2.6µL SuperScript™ VILO™ Master Mix in a 0.2mL PCR tube and ran using the following cycling conditions: 42°C for 30 minutes, 85°C for 5 minutes, then cooled to 10°C. The rest of the protocol applies for both assays (all samples):

13µL of reverse transcription products (OLcfTNAR) or cfDNA (OLcfDNA) were added to 2µL lung cfDNA panel and 15µL cfDNA Library PCR Master Mix in a 0.2mL tube, creating a total reaction volume of 30µL. Following vortexing and quick centrifugation, samples were ran using the following program (Table 2.6):

Temp (°C)	Time	Cycles
98	2 min	1
98	30 secs	2
64	2 mins	
62	2 mins	
60	4 mins	
58	2 mins	
72	30 secs	
72	2 mins	1
4	∞	1

Table 2.6 – Target amplification cycling conditions using the OLcfTNAR and OLcfDNA assays.

Target amplicons were purified by combining with 45µL (1.5 x sample volume) of Agencourt™ AMPure™ XP Reagent (Beckman Coulter) beads in a 1.5mL Lo-Bind® Eppendorf tube, well-mixed and incubated at RT for 5 minutes. Mixing and incubation was repeated, before the tubes were placed on a DynaMag™-2 magnet, to pellet the beads bound to amplified target DNA. The supernatant was discarded, before the pelleted beads underwent two 80% ethanol washes. The pellet was then air-dried for 5 minutes to prevent residual ethanol inhibiting further PCR. The pellet was subsequently re-suspended using 24µL Low TE Buffer, homogenised well, then incubated for 5 minutes at RT, to allow target DNA to dissociate from the beads. The tube was then returned to the magnet, beads were pelleted, before 23µL of supernatant was transferred to a fresh 0.2mL PCR tube. To this, 1µL Tag Sequencing BC (#1-48), 1µL cfDNA Library Primer P1 and 25µL cfDNA Library PCR Master Mix were added to create a reaction volume of 50µL and ligation of barcodes occurred using the following program (Table 2.7):

Temp (°C)	Time	Cycles
98	2 mins	1
98	15 secs	18
64	15 secs	
72	15 secs	
72	5 mins	1
4	∞	1

Table 2.7 – Cycling conditions for OLcfTNAR and OLcfDNA Barcoding.

Barcoded amplicons were added to 57.5µL (1.15 x sample volume) of Agencourt™ AMPure™ XP Reagent beads in a fresh 1.5mL Lo-Bind® tube (Eppendorf), mixed and then incubated for 5 minutes at RT to allow binding of targets to beads. Subsequently, the tube was placed on a DynaMag™-2 magnet to pellet the beads. The supernatant was discarded, and then the beads/targets were washed using 150µL 80% ethanol (x2). After the second ethanol wash, all ethanol was removed and the pellet was air-dried for 5 minutes. The pellet was subsequently re-suspended (off the magnet) using 50µL Low TE Buffer, mixed and incubated for a further 5 minutes. Following this, the tube was returned to the magnet, a pellet was reformed and then 50µL of supernatant was transferred to a fresh 1.5mL Lo-Bind® tube (Eppendorf) containing 50µL (1.0 x sample volume) Agencourt™ AMPure™ XP Reagent beads to allow size selection of the barcoded library to occur. The tubes were vortexed well and had a further 5 minute incubation at RT. The tubes was then placed on the magnet, to pellet the beads, before the supernatant was discarded. Beads underwent two more 80% ethanol washes, before all ethanol was removed and the pellet was air-dried for 5 minutes. 30µL Low TE Buffer was then used to re-suspend the beads and incubated for 5 minutes at RT, to allow for the dissociation of beads from the purified DNA library. Finally, the homogenised mixture was transferred back to the magnet, a pellet was reformed and 28µL of the final library (supernatant) was transferred to a fresh 1.5mL Lo-Bind® tube (Eppendorf) tube and molarity of the libraries were determined by using the TapeStation 4200 instrument (see section 2.4.2) before being stored at -20°C for later use.

Following manual library preparation, samples were diluted to 50pM and then underwent automated template preparation and chip loading on the Ion Chef system (Thermo Fisher Scientific) according to the manufacturer's instructions. Samples were subsequently sequenced on Ion 530™ chips (Thermo Fisher Scientific) on the Ion S5™ system (Thermo Fisher Scientific).

2.4.6.2 NGS with Ion AmpliSeq panels

For the PRIMINGS study, NGS was conducted manually using Ion AmpliSeq™ chemistry. Library preparation was conducted using the Ion AmpliSeq™ Library Kit 2.0 and Ion AmpliSeq™ Cancer Hotspot Panel v2 (CHPv2) (Thermo Fisher Scientific) reagents respectively, as follows:

10ng of DNA was used to generate libraries using the Ion AmpliSeq™ HiFi Master Mix and the CHPv2 respectively, for a total reaction volume of 20 µL, then cycled using the following conditions (Table 2.8):

Stage	Temp (°C)	Time	Cycles
Enzyme activation	99	2 min	1
Denaturation	99	15 sec	22
Annealing/extension	60	4 min	
Hold	10	∞	1

Table 2.8 – Cycling conditions for initial stage of Ion AmpliSeq™ library preparation.

2µL FuPa Reagent was added and partial digestion of PCR products occurred using the following conditions: 50°C for 10 minutes, 55°C for 10 minutes, and 60°C for 20 minutes. Subsequently, ligation of barcodes and adapters occurred where 4µL Switch solution and 2µL Ion P1 Adaptor/Ion Xpress™ Barcode X (#1-32) mix were added to the digested samples, and the tubes were heated under the following conditions: 22°C for 30 minutes, 68°C for 5 minutes then 72°C for 5 minutes. Subsequently, 30µL of barcoded samples were added to a 1.5mL Lo-Bind® tube (Eppendorf) containing 45µL (1.5 x sample volume) Agencourt™ AMPure™ XP Reagent (Beckman Coulter) beads, mixed and incubated for 5 minutes. Following this, the mixture was placed on a DynaMag™-2 Magnet, to pellet the beads, where the supernatant was discarded. The beads underwent two 200µL 80% ethanol washes whilst on the magnet, were air-dried for 5 minutes to remove residual ethanol, before being re-suspended in 50µL Platinum PCR SuperMix High Fidelity and 2µL Library Amplification Primer Mix. The tube was then returned to the magnet, and 50µL of supernatant was transferred to a 0.2 mL PCR tube and placed in a thermal cycler with the following program (Table 2.9):

Stage	Temp (°C)	Time	Cycles
Enzyme activation	98	2 min	1
Denaturation	98	15 sec	7
Annealing/extension	64	1 min	
Hold	10	∞	1

Table 2.9 – Final cycling conditions for library preparation.

Following this, the samples were added to 25µL (0.5 x sample volume) AMPure™ XP beads, to release the DNA from the beads, before the supernatant (containing DNA) was transferred to a fresh 1.5mL Eppendorf containing 60µL (1.2 x sample volume) beads. Samples were well-mixed and incubated for 5 minutes before being placed on the magnet, with the supernatant then being discarded. Another two 200µL 80% ethanol washes followed, before the pellet was air-dried for 5 minutes. Finally, 50µL Low TE buffer was used to re-suspend the beads, the tube was returned to the magnet, where the purified library supernatant was removed and placed in a new 1.5mL Eppendorf. Libraries immediately underwent quality control/ DNA molarity analysis using the TapeStation 4200 (see section 2.4.2), and samples were frozen at -20°C until later use.

Libraries were diluted to 100pM, then pooled (6 libraries per pool for each chip) and an emulsion PCR step was conducted using the Ion PGM™ Hi-Q™ View OT2 Kit on the Ion OneTouch™ 2 system, according to the manufacturer's instructions. This step involves the clonal amplification of template DNA around Ion Sphere™ Particles (ISPs), in order to correctly configure the samples for sequencing. Following this the ISPs were enriched using the Ion OneTouch™ ES machine, ultimately producing 200µL of enriched, template-positive ISPs. Quality control of each OneTouch™ reaction was conducted using the Ion Sphere™ Quality Control Kit.

Following the initialisation of the Ion PGM™ system, template positive ISPs were loaded onto Ion 316™ chips and sequencing was commenced, according to the manufacturer's instructions.

2.4.6.3 NGS using the Oncomine Solid Tumour DNA Kit

For the "ambiguous" cases in the PRIMINGS study (see section 5.3.4), in addition to NGS being conducted using the CHPv2, these cases were also sequenced with 10ng input DNA using the Oncomine Solid Tumour DNA Kit (OST) (Thermo Fisher Scientific). This was conducted by Leicester Molecular Diagnostics (LMD) according to manufacturer's instructions using user guide (MAN0010935, revision D.0).

Unlike other NGS protocols in this chapter, the OST kit has an equaliser step that normalises concentration to ~100pM, meaning quantitation of prepared libraries is not required (e.g. on the TapeStation as previously described).

Libraries were combined into pools of 12 samples and automated template preparation and chip loading was conducted onto Ion 520™ chips using the Ion Chef Instrument. Loaded chips were subsequently sequenced using the Ion S5™ system (Thermo Fisher Scientific).

2.4.6.4 NGS Data Analysis

All targeted NGS conducted within this thesis (OLcfDNA, OLcfTNAR, CHPv2, OST), was all analysed using the Ion Torrent™ – Torrent Suite™ v4.2.0 – 5.6.0. Reads were aligned against the human genome (hg19) using alignment v4.0-r77189 and variants called using the coverageAnalysis v4.0-r77897 and variantCaller v4.0-r76860, respectively. Variants were manually inspected using the Integrative Genomics Viewer (IGV) 2.3 accessed *via* the University of Leicester High Performance Computing (HPC) services ALICE and SPECTRE. COSMIC IDs were obtained using COSMIC v71.

For all samples not sequenced using the OLcfDNA or the OLcfTNA assays, variant allele frequency (VAF) was calculated as a proportion of the reads at a site, which contained the mutant allele (e.g. if 100 mutant reads out of 1000 total reads, then VAF = (100/1000)*100 = 10%).

For samples "called" (variantCaller) by the OLcfDNA and OLcfTNA assays, a VAF was calculated based on original DNA molecules pre-PCR (termed, Allele Molecular VAF): The Oncomine chemistry involves the molecular barcoding of

individual DNA molecules (unique molecular identifiers (UMI)) prior to library amplification as a means to overcome drawbacks to PCR-based NGS (e.g. DNA polymerase errors and the generation of PCR duplicates as a result of sequencing the same original molecule several times). For tumour-confirmed variants identified by multiregional WES and the bespoke NGS approach (Abbosh *et al.* 2017) that were not “called” by the variant caller were manually identified using IGV. In these cases, VAF was calculated as previously described (previous paragraph), as Allele Molecular VAF was not available for these variants in most cases.

Chapter 3 Non-small cell lung cancer

3.1 Introduction

3.1.1 Non-small cell lung cancer

Lung cancer is the most common cause of cancer-related mortality, responsible for approximately 1.59 million deaths worldwide per year (IARC statistics, GLOBOCAN 2012). An umbrella term, lung cancer is clinically separated into two distinct classes, that differ in prognosis and treatment; small cell lung carcinoma (SCLC) and non-small cell lung carcinoma (NSCLC). NSCLC can be further sub-divided into three main histological types; adenocarcinoma (LUAD), squamous cell carcinoma (LUSC) and large cell carcinoma (LCC) (Figure 3.1). This chapter concerns NSCLC.

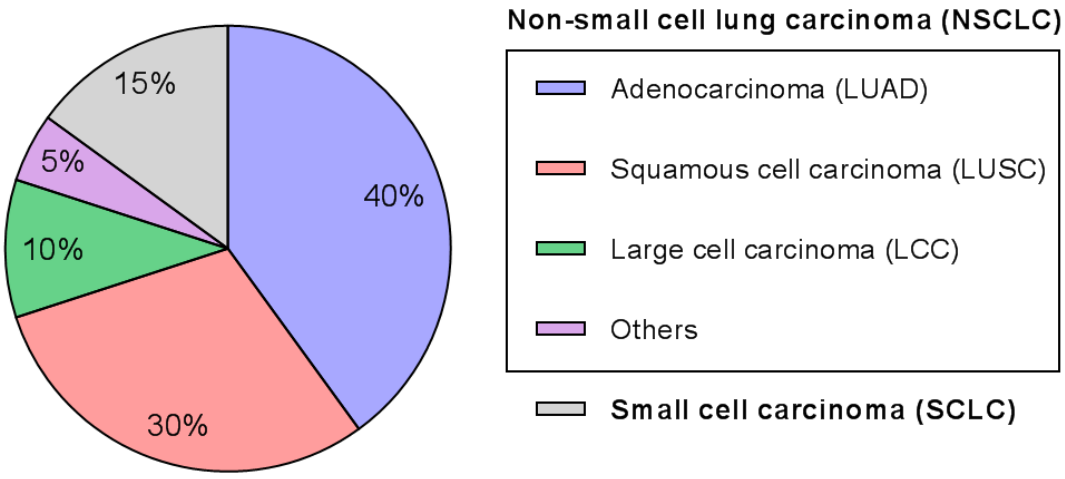


Figure 3.1 – Prevalence of Lung Cancer Subtypes. Adapted from Zappa and Mousa 2016.

NSCLC is recognised as a cancer of unmet need by Cancer Research UK, due to a lack of progress regarding survival rates. Strongly associated with tobacco smoking, NSCLC has very unspecific symptoms meaning that often patients present with metastatic disease (Ellis and Vandermeer 2011). Therefore there are calls for lung cancer screening programs in the NHS and in other parts of the world, likely involving low-dose CT scanning (O'Dowd and Baldwin 2018).

3.1.2 NSCLC Histology – LUAD and LUSCs

NSCLC represents a highly heterogeneous group of lung carcinomas, possessing various epidemiological and pathological factors. It is not uncommon for lung tumours to contain more than one histological subtype, such as adenosquamous tumours and combined-small cell lung cancer (SCLC with NSCLC) (Jamal-Hanjani *et al.* 2014). Even within histological subgroups, particularly LUADs, there is heterogeneity commonly observed.

Due to reduced smoking habits, LUAD has now overtaken LUSC as the most commonly diagnosed NSCLC and accordingly is the most common subtype found in "never smokers". LUAD tumours typically arise as peripheral nodules and are characterised by the presence of glandular differentiation and/or mucin production (Travis *et al.* 2013b). LUADs are thought to arise from alveolar and bronchiolar epithelium (pneumocytes or Clara cells). Classification of LUADs is challenging and the prognosis between different histological patterns is huge. Adenocarcinoma *in situ* (AIS) and minimally invasive adenocarcinoma (MIA) have 5-year survival rates of ~100% if resected (Travis *et al.* 2013a). Intermediate prognosis is seen for patients where papillary and acinar patterns predominate (Russell *et al.* 2011). The worst prognosis is typically seen in those with the presence of predominant solid, micropapillary, invasive mucinous or colloid growth. Large or poorly differentiated LUADs can exhibit regions of necrosis, however typically necrosis is rare in this subtype. For the purposes of this thesis, the term "LUAD" will refer to exclusively invasive adenocarcinoma.

In contrast to LUAD, LUSC typically presents more centrally in the lung, normally within major bronchial epithelium, however peripheral tumours are sometimes reported (Perez-Moreno *et al.* 2012). They have very distinct histological features, including intercellular bridges, individual cell keratinisation and squamous pearl formation (Davidson *et al.* 2013). LUSC arises from a series of preneoplastic lesions ranging from mild, moderate to severe. Unlike LUAD, LUSC growth patterns do not seem to have prognostic or clinical impact. While necrosis is not typical in LUAD, it is not uncommon for LUSCs to exhibit a central comedo-type pattern of necrosis in particularly higher grade lesions (Caruso *et al.* 2012).

3.1.3 NSCLC Staging and Prognosis

NSCLC is staged using the TNM system developed by the American Joint Committee on Cancer (AJCC) and the Union for International Cancer Control (UICC) (Mirsadraee *et al.* 2012, UyBico *et al.* 2010). Since January 2018, the 8th Edition is beginning to be implemented, however for the purpose of this thesis, the earlier 7th Edition has been presented, as this is more applicable to this results section (Goldstraw *et al.* 2016).

The TNM system combines three key criteria; the size and extent of invasion of the primary tumour (T), the involvement of regional lymph nodes (N) and the degree of metastatic spread (M) (Table 3.1). The T, N and M scoring is subsequently combined to assign an overall stage which is used for prognostic information as well as treatment selection (Table 3.2).

NSCLC stage is strongly indicative of prognosis. Unfortunately, approximately 40% of patients with NSCLC at diagnosis have stage IV disease, with one-year survival being below 20% for this group (Zappa and Mousa 2016). Conversely, some studies have shown localised stage I disease to have five-year survival above 70% (Nesbitt *et al.* 1995). Figure 3.2 summarises the one-year survival statistics by stage in the United Kingdom (Cancer Research UK Statistics).

T	Primary Tumour
TX	Primary tumour cannot be assessed, or tumour proven by the presence of malignant cells in sputum or bronchial washings but not visualized by imaging or bronchoscopy
T0	No evidence of primary tumour
Tis	Carcinoma *in situ*
T1	Tumour < 3 cm in greatest dimension, surrounded by lung or visceral pleura, without bronchoscopes evidence of invasion more proximal than the lobar bronchus (i.e., not in the main bronchus)
T1a	Tumour < 2 cm in greatest dimension
T1b	Tumour > 2 cm but < 3 cm in greatest dimension
T2	Tumour > 3 cm but < 7 cm or tumour with any of the following features (T2 tumours with these features are classified T2a if < 5 cm): • Involves main bronchus, > 2 cm distal to the carina • Invades visceral pleura • Associated with atelectasis or obstructive pneumonitis that extends to the hilar region but does not involve the entire lung
T2a	Tumour > 3 cm but < 5 cm in greatest dimension
T2b	Tumour > 5 cm but < 7 cm in greatest dimension
T3	Tumour > 7 cm or one that directly invades any of the following: • Chest wall (including superior sulcus tumours), diaphragm, phrenic nerve, mediastinal pleura, parietal pericardium • Tumour in the main bronchus < 2 cm distal to the carina but without involvement of the carina • Associated atelectasis or obstructive pneumonitis of the entire lung • Separate tumour nodule(s) in the same lobe
T4	Tumour of any size that invades any of the following: • Mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body, carina • Separate tumour nodule(s) in a different ipsilateral lobe
N	**Nodes**
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes, including involvement by direct extension
N2	Metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s)
N3	Metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph node(s)
M	**Metastasis**
MX	**Distant metastasis cannot be assessed**
M0	No distant metastasis
M1	Metastasis
M1a	Separate tumour nodule(s) in a contralateral lobe tumour with pleural nodules or malignant pleural/ pericardial effusion
M1b	Distant metastasis

Table 3.1 – TNM Staging of lung cancer (7th edition). Adapted from Mirsadraee *et al*., 2012

		T	N	M
Occult carcinoma		TX	N0	M0
Stage 0		Tis	N0	M0
Stage I	**A**	T1a, T1b	N0	M0
	B	T2a	N0	M0
Stage II	**A**	T1a,T1b	N1	M0
		T2a	N1	M0
		T2b	N0	M0
	B	T2b	N1	M0
		T3	N0	M0
Stage III	**A**	T1, T2	N2	M0
		T3	N1, N2	M0
		T4	N0, N1	M0
	B	T4	N2	M0
		Any T	N3	M0
Stage IV		Any T	Any N	M1a, M1b

Table 3.2 – TNM stage groupings (7th Edition). Adapted from UyBico *et al.* 2010.

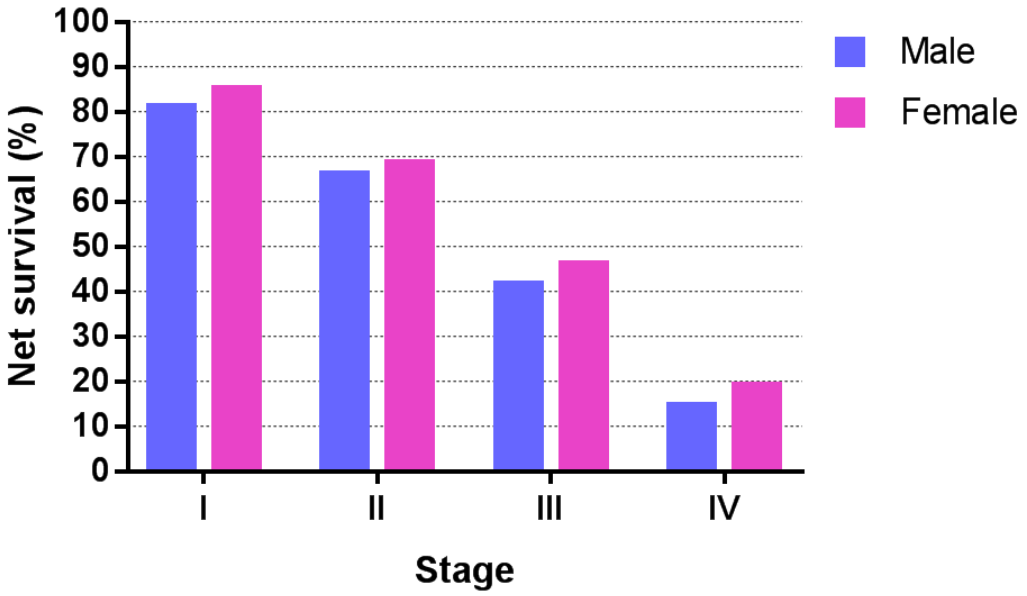


Figure 3.2 – One-year survival for lung cancer by stage. Cancer Research UK statistics.

3.1.4 NSCLC Treatment

NSCLC is relatively insensitive to chemotherapy and radiotherapy. Patients presenting with stage I-IIIA NSCLC are often are applicable to receive surgery as a first line treatment (Scott *et al.* 2007). Lobectomy is generally considered the gold standard treatment, but sublobar resections and pneumonectomy (removal of entire lung) are not uncommon. Sublobar resections (wedge resection or segmentectomy) are typically used when patients are unable to tolerate lobectomy. During surgery, regional lymph nodes are removed and analysed for localised metastatic spread.

Adjuvant chemotherapy is commonly administered for stages IIA, IIB, and IIIA. Patients with N1 and N2 disease (stages II-III) were found to have in a 4-5% five-year survival improvement with adjuvant chemotherapy from meta-analyses (Artal Cortes *et al.* 2015). The use of adjuvant chemotherapy in stage I disease is less clear, with stage IA often showing poorer outcomes, whereas in stage IB it appears to have small benefits, particularly in tumours > 4 cm diameter.

NSCLC Chemotherapy typically involves a combination of platinum (cisplatin/carboplatin) with either a mitotic spindle poison (docetaxel, paclitaxel, vinorelbine), pemetrexed or gemcitabine. One combination has not been shown to be particularly superior to others. For advanced NSCLC stages (IIIB/IV), chemotherapy is often used as first line as this provides modest improvements in median survival from 4.5 to 6 months (Burdett *et al.* 2008). For non-squamous advanced NSCLC, pemetrexed is often used as maintenance therapy following platinum-based chemotherapy (Ciuleanu *et al.* 2009).

Radiotherapy is used as first-line for stage I peripherally located tumours in patients with comorbidities and other reasons for inoperability. Stereotactic ablative radiotherapy (SABR) is the preferred treatment with five-year local control rates of between 79% and 90% (Lindberg *et al.* 2015, Verstegen *et al.* 2015). Radiotherapy is not generally used for completely resected (R0) early stage tumours, but can have utility postoperatively for R1 resections, usually following chemotherapy (Burdett *et al.* 2016). Concurrent chemoradiotherapy is

recommended for the treatment of unresectable, locally-advanced Stage IIIA and IIIB NSCLC for patients that are fit (Bradley *et al.* 2015).

In advanced disease, targeted agents are often adopted in preference to chemotherapy, particularly for NSCLC with mutant *EGFR* or the *EML4-ALK* fusion gene. These therapies typically produce fewer side effects than their chemotherapy counterparts, however their efficacy is typically limited due to the emergence of resistance mechanisms. For tumours possessing activating mutations such as p.L858R and exon 19 deletions in the *EGFR* gene, first and second generation tyrosine kinase inhibitors (TKIs) such as erlotinib, gefitinib (1st) and afatinib (2nd) are commonly offered as first line, as these have been shown to provide longer PFS in a number of studies (Mok *et al.* 2009, Rosell *et al.* 2012). Afatinib as a second generation TKI was developed to overcome the invariable resistance observed in erlotinib and gefitinib, by binding irreversibly to the EGFR kinase domain. There are a number of resistance mechanisms to EGFR TKI, the most common being through the emergence of the *EGFR* p.T790M mutation, occurring in 50 - 60% of cases (Blakely and Bivona 2012, Yu *et al.* 2013). In 2017, a third generation inhibitor targeting the T790M-mutated form of *EGFR*, osimertinib was licenced for use following typical first line resistance to other EGFR TKI therapies (Janne *et al.* 2015).

In NSCLC positive for the *EML4-ALK* fusion, the ALK/MET/ROS1 inhibitor crizotinib has generally been permitted, which extends PFS by several months (Camidge *et al.* 2012). Resistance to ALK inhibition occurs predominantly (~60%) through either secondary *ALK* mutations or indirectly through mutation of *EGFR.* Alike EGFR TKI*,* secondary *ALK* mutations causing crizotinib resistance have prompted development of second generation inhibitors such as ceritinib and alectinib, with the latter now being prioritised as a first line therapy following the results if the phase III ALEX trial (Peters *et al.* 2017). Alectinib has an additional property of being active within the central nervous system (CNS); a factor that is crucial as ALK-positive NSCLC has a higher risk of brain metastases. Accordingly, alectinib has shown significant reductions in events of CNS progression compared to crizotinib (Hida *et al.* 2017, Peters *et al.* 2017).

3.1.5 NSCLC Genomics and Intratumour Heterogeneity

3.1.5.1 Genetic alterations in NSCLC

As an already diverse histological group of lung cancers, NSCLC displays extensive genomic heterogeneity. In addition to single nucleotide variants (SNVs), NSCLC has had somatic copy number alterations (SCNA), genome doubling and translocations all reported, creating a highly convoluted genomic landscape (de Bruin *et al.* 2014, Go *et al.* 2010, Soda *et al.* 2007).

Smoking-related NSCLC is particularly associated with higher mutational burden, with one study reporting a median of 209 point mutations compared to 18 in never-smokers (Govindan *et al.* 2012). Tobacco smoke is associated with C>A transversions genome-wide, consistent with DNA adduct formation exhibited by exposure to PAH (polycyclic aromatic hydrocarbons) and NNK (nicotine-derived nitrosamine ketone) (Gibbons *et al.* 2014). Alternatively, C>T transitions are the most common mutations observed in never-smokers. Interestingly, certain genes are commonly linked with smoking status, with mutations in *KRAS*, *TP53* and *BRAF* being typically found in smokers, while *EGFR* mutations and *ALK* and *ROS1* fusions are enriched in never-smokers (Govindan *et al.* 2012).

The two most commonly mutated proto-oncogenes in NSCLC are *EGFR* and *KRAS*, however these mutations are much more commonly detected in LUAD than in LUSC. Interestingly, *EGFR* and *KRAS* mutations are almost always mutually exclusive in NSCLC, suggesting that the activation of both pathways is redundant in lung tumourigenesis (Shigematsu *et al.* 2005, Sun *et al.* 2007). *EGFR* mutations are very diverse epidemiologically, with these being particularly prevalent among East Asian populations (~38%) and less frequent in Europeans (~14%) (Zhang *et al.* 2016). The “classical” mutations in *EGFR* are p.L858R missense mutation (exon 21) and small in-frame exon 19 deletions; together these constitute ~90% of all *EGFR* mutations in NSCLC (Ladanyi and Pao 2008). The screening of these events is routine as they confer sensitivity to EGFR TKIs (Gazdar 2009).

KRAS is mutated in up to 25% of NSCLC, with mutations typically located in codons 12, 13 and 61 (Johnson *et al.* 2012). Alterations at these codons are

known to induce GDP insensitivity, resulting in constitutively active RAS signalling. Unlike *EGFR*, difficulty in the design of inhibitors targeting *KRAS* mutations have led to the gene being regarded as "untargetable", however recent studies are showing promise, particularly for *KRAS* p.G12C (Janes *et al.* 2018, Ostrem *et al.* 2013). Considering that p.G12C is the most common type of *KRAS* mutation in NSCLC, it is hoped that patients with this mutation may see benefit in the near future. Other genes mutated (point mutations) in NSCLC include *TP53* (~50%), in addition to less common mutations in *BRAF*, *ERBB2*, *FGFR2*, *NF1*, *PIK3CA* and *PTEN* (Table 3.3).

Point mutations are undoubtedly the most studied alterations in NSCLC, however SCNA and translocations are also frequent driver events. Amplification of the *MET* gene is a known alternative mechanism of acquired resistance to EGFR TKIs. Found in 2-4% of TKI-naïve NSCLC, this rises to ~20% in those that have acquired resistance (Bean *et al.* 2007, Cappuzzo *et al.* 2009); emphasis that NSCLC dynamically evolves in response to treatment. Other genes with recurrent reports of SCNA include *CDKN2A*, *EGFR*, *ERBB2*, *FGFR1*, *KRAS*, *PIK3CA*, *SOX2* (Table 3.3).

Gene fusions occur in a subset of NSCLC, with *EML4-ALK* being the most common, occurring in 3 - 6% of patients. Fusions involving *ROS1* and *RET* have also been reported and these were deemed sensitising to existing TKI treatments (Kohno *et al.* 2015, Oxnard *et al.* 2013). Of particular interest, is the observation that *ALK*, *RE*T and *ROS1* fusions are only found in patients with little or no smoking history and are typically exclusive of other drivers (Thunnissen *et al.* 2012).

Gene	Genetic alteration	Frequency (%)		References
		LUAD	LUSC	
BRAF	Activating mutations	~4	Rare	(Cardarella *et al.* 2013)
CDKN2A	Genomic Deletion	20-30	U	(Andujar *et al.* 2010) (Iwakawa *et al.* 2008)
EGFR	Activating mutations	~14 (Caucasian)	Rare	(Gazdar 2009, Zhang *et al.* 2016)
		~38 (East Asian)	Rare	(Tokumo *et al.* 2005, Zhang *et al.* 2016)
	Genomic amplification	4-14	11-15	(Hirsch *et al.* 2003) (Sholl *et al.* 2009)
EML4-ALK	Translocation	3-6	Rare	(Thunnissen *et al.*, 2012)
ERBB2	Activating mutations	2-5	Rare	(Garrido-Castro and Felip 2013)
	Genomic amplification	~1	U	(Takezawa *et al.* 2012)
FGFR1	Genomic amplification	~2	~20	(Bass *et al.*, 2009)(Dutt *et al.* 2011); (Weiss *et al.* 2010)
KRAS	Activating mutations	~15-25	~3	(Boch *et al.* 2013, Brose *et al.* 2002)
	Genomic amplification	~17	~13	(Wagner *et al.* 2011)
MET	Activating mutations	U	U	-
	Genomic amplification	~6 (TKI-naïve)	~6	(Go *et al.* 2010)
		~20 (post-TKI)	U	(Bean *et al.*, 2007c; Chen *et al.*, 2009; Cappuzzo *et al.*, 2009a; Cappuzzo *et al.*, 2009b)
NF1	Inactivating mutations	~6	1-2	(Redig *et al.* 2016)
PIK3CA	Activating mutations	~3	~4	(Yamamoto et al., 2008)
	Genomic amplification	4-6	~33	(Yamamoto *et al.*, 2008; Wang *et al.*, 2012a; Kawano *et al.*, 2006)
PTEN	Inactivating mutations	1-2	~10	(Jin *et al.* 2010) (Lee *et al.* 2010)
SOX2	Genomic amplification	6-8	60-72	(Wilbertz *et al.*, 2011;Brcic *et al.*, 2012) Bass *et al.*, 2009; Hussenet *et al.*, 2010)
TP53	Inactivating mutations	~36	~50	(Ding *et al.* 2008) (Tammemagi *et al.* 1999)

Table 3.3 – Frequency (%) of genetic alterations in LUAD and LUSC. U: Unknown; TKI: Tyrosine Kinase Inhibitor

3.1.5.2 Clonal Evolution in NSCLC

The diverse spectrum of alterations present in NSCLC are not novel findings, yet how these events temporally manifest in the disease has been elusive and limited to small retrospective cohorts. The intratumour heterogeneity present in NSCLC was dissected initially in two independent studies (de Bruin *et al.* 2014, Zhang *et al.* 2014). Through multiregional WES/WGS, the studies phylogenetically reconstructed individual tumours where they found extensive heterogeneity present. However, perhaps encouragingly, the archetypal NSCLC drivers were predominantly acquired early in tumour evolution, being present in every tumour region (clonal) in ~76% of cases. De Bruin and colleagues showed that these driver mutations were clonally enriched compared to passenger variants.

Furthermore, this study illuminated the potential weakness of a single tumour biopsy for optimal treatment selection (de Bruin *et al.* 2014). One patient possessed an activating *BRAF* mutation (G469A) in all tumour regions and a *PIK3CA* mutation (E542K) in only one region (R3). If a biopsy was coincidently obtained from R3, the patient could have been considered a candidate for treatment with an inhibitor targeting the PI3K/mTOR pathway. Conversely, if taken from any other region, a BRAF inhibitor would have been suggested, for which the cancer cells of R3 may be resistant, due to the *PIK3CA* mutation (Shi *et al.* 2014). This emphasises the importance of identifying clonal drivers, yet being aware of sub-clonal drivers capable of orchestrating treatment resistance.

Additionally, the study observed genome doubling events and broad chromosomal instability (de Bruin *et al.* 2014). Three cases exhibited genome doubling events post-acquisition of clonal oncogenic drivers, but prior to sub-clonal divergence. Consistent with findings in colorectal cancer, these events accelerate genome evolution in NSCLC (Dewhurst *et al.* 2014). Insight into endogenous mutational processes responsible for subclonal diversification in NSCLC have also been uncovered, namely through APOBEC deaminase upregulation (de Bruin *et al.* 2014). Detection of this signature suggested inhibition of APOBEC deaminase activity may provide a means of limiting tumour divergence and maintain a less heterogeneous tumour mass that may be easier to treat (Alexandrov *et al.* 2013).

3.1.6 ctDNA in NSCLC

The poor prognosis of NSCLC, coupled with difficulties with the acquisition of solid biopsy has driven an abundance of research aiming to characterise the disease using minimally-invasive methods, particularly through cfDNA and circulating tumour cell (CTC) analysis.

The detection of actionable mutations in plasma in advanced NSCLC (stages IIIB/IV) is becoming increasingly adopted clinically. Of particular note is through the monitoring of treatment response to EGFR TKI. In 2016, the real-time qPCR cobas® EGFR Mutation Test v2 (Roche) became the first cfDNA-based (tissue also) method to be FDA-approved for clinical use. This has enabled the minimally-invasive longitudinal monitoring of *EGFR* mutation status and tumour burden, including the detection of the resistance-conferring T790M mutation; the latter of which could facilitate the transfer of patients from a failing first-line TKI (gefitinib, erlotinib) to the T790M-targeted osimertinib (Thress *et al.* 2015).

Indeed, the prognostic value of ctDNA detection has been previously reported in advanced NSCLC, with the detection of both *EGFR* and *KRAS* mutations in patients receiving treatment being associated with poorer survival (Nygaard *et al.* 2013, Wang *et al.* 2010). For example, recently the quantitative detection of the T790M mutation using digital PCR was shown to be prognostic, with incremental increases in plasma VAF almost tripling the risk of death and progression (Provencio *et al.* 2018).

The detection ctDNA in early-stage NSCLC (I-IIIA) has been less studied. Recently, Chaudhuri and colleagues demonstrated that cfDNA analysis *via* CAPP-seq was capable of detecting MRD in 94% of early stage (I-III) lung cancer patients that experienced relapse(Chaudhuri *et al.* 2017). It is apparent that the detection of MRD *via* cfDNA analysis following curative surgery could be critical in the selection adjuvant therapy for applicable patients, while sparing others that wouldn't receive benefit.

3.1.7 The Lung TRACERx Study

3.1.7.1 Primary objectives

Lung TRACERx (Tracking Cancer Evolution through therapy (Rx)) is the first national study (UK) comprehensively analysing the spatial and temporal evolution of NSCLC over the course of treatment. Recruiting 842 patients across stages I-IIIA from multiple centres (including Aberdeen, Birmingham, Cardiff, Glasgow, Leicester, Liverpool, London, Manchester and Southampton), the study primarily aims to assess the relationship between intratumour heterogeneity and clinical outcome (disease-free survival/overall survival). This includes the association between heterogeneity and clinical disease stage, as well as histological subtype. Additionally, it will assess the impact of adjuvant platinum-based regimens upon intratumour heterogeneity in relapsed disease (Jamal-Hanjani *et al.* 2014).

Multiregional WES/WGS of tumour tissue DNA is conducted following primary surgery, allowing the assembly of phylogenetic trees for each patient. This will allow the discrimination of ubiquitous (truncal, clonal) mutations, from those possessed by only a sub-set of the tumour cells (branch, subclonal) (Figure 3.3). As a result, spatial heterogeneity will be observed within tumours, providing a baseline assessment of the disease state. Additional biopsies will be acquired at recurrence/relapse which will ultimately be compared to baseline trees, to assess the clonal evolution of the tumours. Immunohistochemistry (IHC), Flouresence in situ hydridisation (FISH) and Fluorescence-activated cell sorting (FACS) will be additionally used to study protein expression, chromosomal instability and DNA ploidy respectively. The integration of these techniques with WES/WGS data will provide an unrivalled scale of NSCLC profiling (Jamal-Hanjani *et al.* 2014).

3.1.7.2 Secondary Objectives

Much of the output of the study will focus on the detailed genomic and statistical analysis of tumour tissue obtained by surgery/biopsy. However, the acquisition and sequencing of large amounts of tumour tissue DNA at multiple time-points is not realistically feasible to clinical practice. Therefore, TRACERx has secondary objectives aiming to facilitate translational aspects for NSCLC, namely through

the development of effective, minimally-invasive biomarkers for the disease. This will be accomplished by the longitudinal collection of blood, and subsequent isolation of CTCs and cfDNA for analysis (Figure 3.4) (Jamal-Hanjani *et al.* 2014). cfDNA analysis has a high degree of expertise at the University of Leicester, reflected in its role of being the primary cfDNA biobank for TRACERx.

Although cfDNA has been successful at detecting aberrations in cancer, how accurate and informative it is in the representation of intratumour heterogeneity in NSCLC is currently unknown. Therefore, following the generation of phylogenetic trees from tumour tissue studies, corresponding cfDNA will be analysed for evidence of consistent mutational profiles. An accurate cfDNA-based characterisation of intratumour heterogeneity in such a large cohort of patients (n = 842), could present a strong support for its adoption into a clinical setting. Furthermore the detection of truncal, actionable drivers in cfDNA would provide an efficient and minimally-invasive method of patient stratification and therapeutic intervention.

As blood will be taken at multiple time-points, cfDNA analysis could be used to identify dominant sub-clones of NSCLC, throughout disease progression. The platinum-based chemotherapy received by patients may prove to eliminate certain sub-clones, with others proving more resistant to treatment. As these treatment resistant sub-clones will be ultimately responsible for disease progression and metastasis, their characterisation *via* cfDNA analysis could be informative about drug resistance mechanisms.

Another research question to be addressed is whether, following surgery, cfDNA analysis can successfully detect minimal residual disease (MRD) and hence predict tumour recurrence prior to routine follow-up. Immediate detection of MRD could provide improved prognostic impact, and provide therapeutic insights for these recurring patients.

.

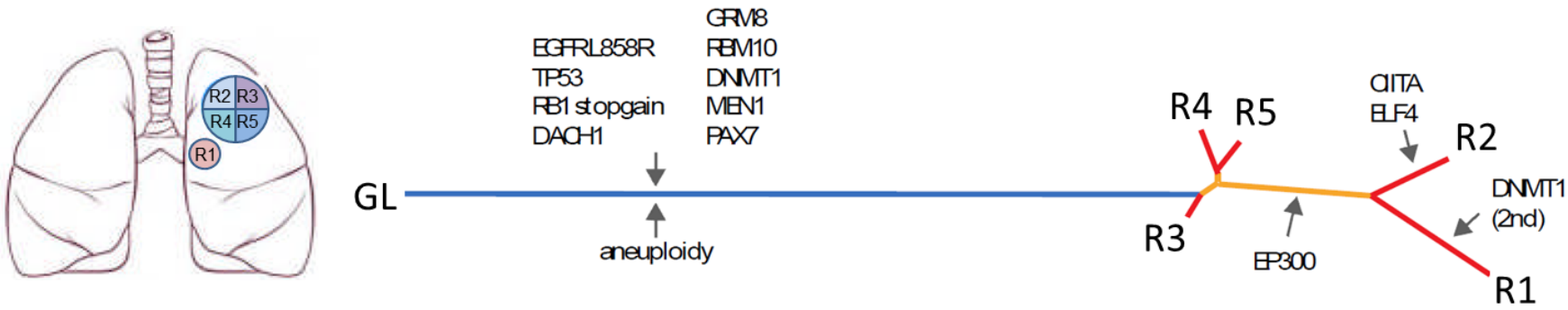


Figure 3.3 – Example of a phylogenetic tree. Blue lines = ubiquitous (truncal) mutations. Yellow lines = shared mutations in a subset of regions (branch). Red lines = private mutations unique to a region (branch). The relative lengths of the lines are proportional to the number of mutations that are ubiquitous, shared or private within each tree. Taken from TRACERx Protocol.

TIME POINT	Surgery	Follow-up (+/- adjuvant chemotherapy	Recurrence	Follow-up	Progression	Completion of all treatment
TISSUE	Primary tumour +/- lymph node(s) (n=842)		Metastatic tumour (n=270)*			
BLOOD	gDNA **cfDNA** CTC Immunology	**cfDNA** 3-monthly (Yr. 1-2) 6-monthly (Yr. 3-5)	**cfDNA** CTC Immunology	**cfDNA** CTC (during Rx)	**cfDNA** (multiple follow-ups) CTC	**cfDNA** CTC

Figure 3.4 – Collection of tissue and blood samples with corresponding time points (TRACERx). Abbreviations: gDNA, germline DNA; cfDNA, circulating-free DNA; CTC, circulating tumour cells; Rx, treatment.*~500 patients expected to have recurrence by 5 years, of which ~270 expected to agree for re-biopsy..

3.1.7.3 Seminal TRACERx findings on NSCLC Clonal Evolution

In 2017, the TRACERx consortium published findings from the first 100 patients (Jamal-Hanjani *et al.* 2017). Building upon data from the two aforementioned studies (de Bruin *et al.* 2014, Zhang *et al.* 2014)(see section 3.1.5.2), the intratumour heterogeneity present in NSCLC was uncovered, where genome doubling events and chromosomal instability (CIS) were shown to be the main indicators of clinical outcome in patients. Additionally, they found that the vast majority of driver mutations (*EGFR*, *MET*, *BRAF*, *TP53*) were almost always clonal (early) events during NSCLC evolution. Meanwhile, mutations in *PIK3CA* and *NF1* were commonly found to be sub-clonal events involved in tumour diversification.

For the purposes of this thesis, a key finding was that without multiregional WES, 76% of sub-clonal mutations could have appeared to be clonal, meaning the targeting of such alterations would not effectively eradicate disease. Therefore for the purposes of ctDNA analysis it was important to predominantly focus on detection of clonal variants for optimal tumour DNA detection in plasma. The results of this study (Jamal-Hanjani *et al.* 2017) and those based on phylogenetic ctDNA analyses (Abbosh *et al.* 2017) will feature in the results and discussion of this chapter.

3.2 Aims and Objectives

The aims of this chapter were to assess the utility of a commercially-available next generation sequencing platform (Oncomine™ Lung assays) for the detection of ctDNA in early-stage NSCLC (Stages I-IIIA) through an assessment of:

i. the detection of ctDNA in pre-surgical TRACERx plasma and whether clinicopathological variables determine this detection; and
ii. longitudinal ctDNA analysis following curative surgery, as a means to detect minimal residual disease (MRD) and/or response to treatment.

The objectives were to:

i. first assess the Oncomine™ platform for the detection of ctDNA in metastatic NSCLC patients to assess sensitivity and specificity of this technology;
ii. compare the Oncomine™ platform to a patient-specific, bespoke NGS approach in the detection of pre-surgical ctDNA from patients with early-stage NSCLC within the TRACERx study; and
iii. conduct longitudinal case studies on TRACERx patients recruited at Leicester, tracking tumour-confirmed variants for up to 3 years following tumour resection.

3.3 Results

3.3.1 Study overview

This chapter is predominantly focussed on the analysis of pre-surgical and follow-up cfDNA isolated from patients with early-stage (I-IIIA) NSCLC within the TRACERx study. All samples analysed were sequenced using one of two different, albeit related NGS assays; the Oncomine Lung Cell-Free Total Nucleic Acid Research Assay (OLcfTNAR) and the Oncomine Lung cfDNA assay (OLcfDNA). The assays are near-identical with regards to SNVs, however the OLcfTNAR also attempts to identify fusion genes (*ALK*, *RET*, *ROS1*) and both exon 14 skipping mutations and copy number amplification in *MET* (Table 3.4).

Assay	Input material	Genes	Selected SNV Hotspots	CNV	Fusions	Other alterations
Oncomine™ Lung Cell-Free Total Nucleic Acid Research Assay (**OLcfTNAR**)	cfDNA **+ cfRNA**	*ALK* *BRAF* *EGFR* *ERBB2* *KRAS* *MAP2K1* *MET* *NRAS* *PIK3CA* ***RET*** *ROS1* *TP53*	>150 hotspots including: ***EGFR***: T790M, C797S, L858R, Exon 19 del ***KRAS***: G12X, G13X, Q61X ***BRAF***: V600E ***ALK***: Exon 21-25 ***PIK3CA***: E542K, E545K, H1047R	***MET***	***ALK*** ***RET*** ***ROS1***	***MET*** **Exon 14 skipping**
Oncomine™ Lung cfDNA assay (**OLcfDNA**)	cfDNA	*ALK* *BRAF* *EGFR* *ERBB2* *KRAS* *MAP2K1* *MET* *NRAS* *PIK3CA* *ROS1* *TP53*	>150 hotspots including: ***EGFR***: T790M, C797S, L858R, Exon 19 del ***KRAS***: G12X, G13X, Q61X ***BRAF***: V600E ***ALK***: Exon 21-25 ***PIK3CA***: E542K, E545K, H1047R	-	-	-

Table 3.4 – Comparison of the two Oncomine™ Lung assays used in this chapter. The extra additions of the OLcfTNAR are highlighted in blue. Adapted from the Thermo Fisher Scientific website.

Prior to analysis of TRACERx samples, the Oncomine™ platform was first assessed in a short study using samples obtained from patients with metastatic NSCLC (IV). An overview of the chapter is shown in Figure 3.5. Other than tumour stage, the main differences between the two phases are the nucleic acid extraction method and the Oncomine™ Lung assay used for analysis.

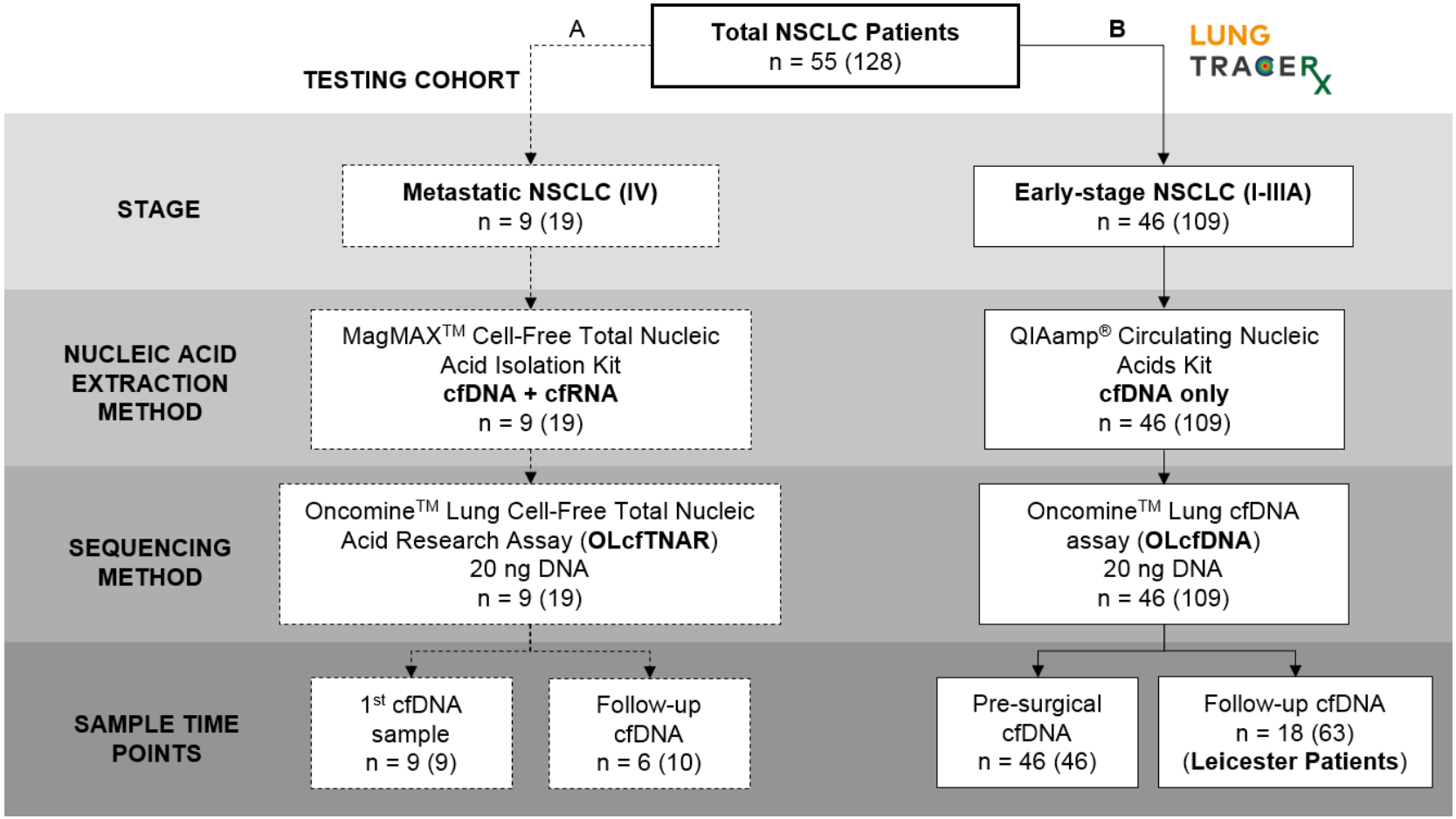


Figure 3.5 – Schematic outlining NSCLC cfDNA analysis. This chapter is comprised of two phases; a pilot study testing the performance of the Oncomine™ platform using the OLcfTNAR assay (A), and analysing Lung TRACERx cfDNA using the OLcfDNA assay (B). The number of patients analysed (n) are indicated, with the number of cfDNA samples analysed shown in brackets. For phase B, follow-up cfDNA was analysed from Leicester patients only, under an ethics amendment where an additional 10ml blood was obtained from each Leicester patient, while maintaining the plasma bank for the rest of the TRACERx Study.

3.3.2 Appraisal of the Oncomine platform with metastatic NSCLC cfTNA

A pilot study was used to assess the Oncomine™ platform through the analysis of cfDNA from 9 patients (19 samples) with stage IV NSCLC, collected by Barbara Ottolini and recruited by Mark Openshaw for the Leicester Cancer Research Biobank. 20 ng DNA (+cfRNA) was analysed with the OLcfTNAR. As this assay requires the isolation of cfRNA in addition to cfDNA in order to assess gene fusions, samples were extracted using the recommended kit; the MagMAX™ Cell-Free Total Nucleic Acid Isolation Kit.

Unlike Ion AmpliSeq™ NGS, which uses the amplified mutant reads/total reads at a position in order to calculate variant allele frequency (VAF), the Oncomine™ chemistry infers the VAF from the number of original DNA molecules pre-PCR by using molecular barcoding (unique molecular identifiers). This overcomes some drawbacks to PCR-based NGS, such as DNA polymerase errors and the generation of PCR duplicates as a result of sequencing the same original molecule several times. Therefore, when called by the variant caller, the “Allele Molecular” metrics are used to calculate VAF of variants (Allele Molecular Coverage/Total Molecular Coverage), but raw “Read” values are still provided. The variant caller will only call a variant if ≥ 2 mutant molecules are detected, as a means to minimise false positives.

If called, variants are also assigned a “LOD (%)” value, based upon the number of interrogated DNA molecules containing the target. Thermo Fisher Scientific use the term “0.1% LOD” to indicate they have data to support a sensitivity and specificity of 90% and 98% respectively at a 0.1% allelic VAF. To further minimise false positives in this chapter, called variants were only considered positive if they had a plasma VAF ≥ LOD value.

Eight out of nine patients were positive (≥ 1 Hotspot variant detected) using the OLcfTNAR, with 21 variants being identified in 14/19 (74%) samples, while five samples had no variants identified (Table 3.5). The variant allele frequency (VAF) ranged between 0.05% – 39.68%, indicating wide variability between metastatic patients regarding ctDNA fraction.

For 6/9 patients, clinical information regarding tumour mutation status was also known. Encouragingly, the assay also identified these listed mutations in 5/6 patients (Table 3.5, green shading). One patient had a known *EML4-ALK* translocation, but this fusion was not detected; a result possibly due to this patient being a long-term responder (>2 years) to crizotinib during this blood sampling period. Unfortunately, there were no other patients that had known fusions or *MET* exon 14 skipping mutation and so no RNA-based variant calls were made.

A recurrent *MAP2K1* p.E203K was found in 5/19 samples at low VAFs (0.06-0.08%), but all were above the limit of detection (LOD) value (0.05%) calculated from the Molecular Coverage values (Table 3.5). Potentially these may represent sequencing artefacts contributing to baseline noise and therefore an important consideration for subsequent analyses.

Altogether, the OLcfTNAR was deemed fit for purpose for the analysis of TRACERx samples. However, considering a high number of TRACERx samples had cfDNA previously isolated using a different extraction method (cfDNA only, not + cfRNA) and that sample throughput was lower than that of the OLcfDNA (6 samples/chip vs 8 samples/chip), the latter was taken forward for TRACERx cfDNA analysis.

Test Case	Time Point	Position	Ref	Var	Gene	AA change	VAF (%)	Allele Mol Cov	Total Mol Cov	LOD (%)	Allele Read Cov	Total Read Cov	Known Mutations
1	0	chr17:7577120	C	T	*TP53*	p.R273H	**0.18**	5	2,767	0.10	21	19,629	**Unknown**
	307 days	no variants identified											
	370 days	chr17:7578190	T	C	*TP53*	p.Y220C	**0.20**	4	1,972	0.10	25	16,457	
		chr15:66774131	G	A	*MAP2K1*	p.E203K	**0.07**	3	4,012	0.05	20	39,161	
2	0	no variants identified											***EML4-ALK* fusion**
	133 days	no variants identified											
	189 days	chr17:7577121	G	A	*TP53*	p.R273C	**0.08**	3	3,608	0.05	31	43,954	
3	0	chr17:7578457	C	A	*TP53*	p.R158L	**0.83**	19	2,277	0.10	121	40,518	**Unknown**
	182 days	no variants identified											
4	0	chr2:29432664	C	T	*ALK*	p.R1275Q	**0.06**	2	3,602	0.05	11	20,716	***EGFR* L858R**
		chr7:55259515	T	G	*EGFR*	p.L858R	**0.07**	3	4,240	0.05	35	50,466	
	112 days	chr15:66774131	G	A	*MAP2K1*	p.E203K	**0.06**	3	5,443	0.05	13	48,520	
	224 days	chr15:66774131	G	A	*MAP2K1*	p.E203K	**0.08**	4	5,287	0.05	10	39,852	
5	0	chr12:25380276	T	A	*KRAS*	p.Q61L	**34.78**	1,428	4,110	0.05	21,466	62,756	***KRAS* Q61L**
		chr17:7578457	C	A	*TP53*	p.R158L	**12.85**	496	3,859	0.05	5,813	44,342	
6	0	chr7:55242465	GGAATTAAGAGAAGC	-	*EGFR*	p.E746_A750delELREA	**4.73**	116	2,450	0.10	1,124	32,874	***EGFR* Exon 19 del**
		chr7:55249071	C	T	*EGFR*	p.T790M	**0.87**	14	1,613	0.10	194	21,248	
	76 days	chr7:55242465	GGAATTAAGAGAAGC	-	*EGFR*	p.E746_A750delELREA	**0.07**	2	2,702	0.10	12	21,676	
	196 days	chr7:55242465	GGAATTAAGAGAAGC	-	*EGFR*	p.E746_A750delELREA	**0.77**	8	1,041	0.15	150	36,976	
7	0	chr7:55242466	GAATTAAGAGAAGCA	-	*EGFR*	p.E746_A750delELREA	**0.44**	14	3,207	0.05	74	26,949	***EGFR* Exon 19 del**
	196 days	chr15:66774131	G	A	*MAP2K1*	p.E203K	**0.08**	3	3,801	0.05	5	19,580	
8	0	chr7:55242465	GGAATTAAGAGAAGC	-	*EGFR*	p.E746_A750delELREA	**39.68**	2,885	7,271	0.05	16,465	51,611	***EGFR* Exon 19 del**
		chr15:66774131	G	A	*MAP2K1*	p.E203K	**0.06**	3	5,093	0.05	6	34,269	
		chr17:7577539	G	A	*TP53*	p.R248W	**0.05**	2	3,700	0.05	28	37,624	
		chr17:7578208	T	C	*TP53*	p.H214R	**6.40**	250	3,904	0.05	1,621	24,552	
9	0	no variants identified											**Unknown**

Table 3.5 – Variants detected by the OLcfTNAR in patients with stage IV NSCLC. A total of 19 samples were analysed in nine patients (Test cases 1-9), with the initial sample labelled “0” and subsequent samples are indicated in the days that had elapsed between blood sampling. Six of these patients had known mutations identified using other clinical methods. Green shading indicates that the OLcfTNAR detected a variant that was concordant with the known mutation status of the patient (5/6 patients). Alternatively, red shading indicates the stated mutation was not identified by the OLcfTNAR (1/6 patients).

3.3.3 Validation of a bespoke NGS method using Oncomine analysis

This analysis contributed to the ctDNA TRACERx study published on the first 100 patients (Abbosh *et al.* 2017). Multiregional WES tumour data was used to inform the design of bespoke multiplex-PCR assay panels targeting a number of clonal and subclonal variants (SNVs) in 100 early-stage NSCLC patients, in an attempt to characterise phylogenetic tree representation within cfDNA (Figure 3.6). This phylogenetic approach provided a detailed annotation of intratumour heterogeneity in NSCLC, particularly with regards to clinicopathological characteristics and their impact upon ctDNA detection. In the study, ≥ 2 clonal variants (SNVs) were required to be detected in cfDNA for a tumour to be deemed "detected" in plasma.

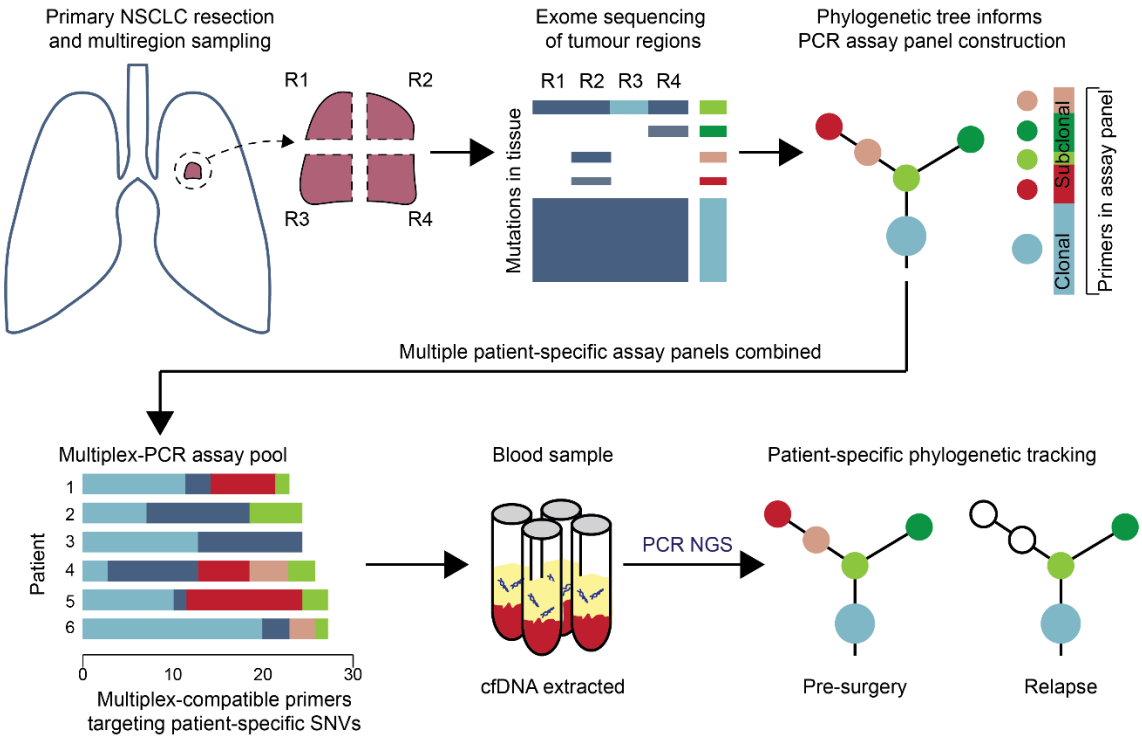


Figure 3.6 – Overview of the Bespoke method of ctDNA detection. Schematic summarising the workflow for the bespoke, phylogenetic ctDNA detection approach featured in the TRACERx ctDNA publication (Abbosh *et al.* 2017). DNA from multiregional tumour tissue was individually sequenced by WES allowing the derivation of phylogeny. Subsequently, PCR assay panels were designed to target clonal and subclonal variants within each patient. These panels were subsequently combined into multiplex assay pools containing primers from up to 10 patients, and cfDNA from pre-surgical and post-operative blood samples were amplified and sequenced before findings were integrated with tumour WES results. Adapted from Abbosh *et al.* 2017.

One key finding was that 97% (30/31) of LUSCs were positive for ctDNA in pre-surgical blood, compared to just 19% (11/58) of LUADs. It was therefore important to investigate whether this disparity was not due to lack of biologically specificity with the bespoke approach, but rather a representation of deficient ctDNA release from LUAD tumours. Therefore, 32 pre-surgical cfDNA samples were also sequenced using the OLcfDNA as cross-platform validation. Bespoke ctDNA "detected" (n = 12) and bespoke "non-detected" (n = 20) NSCLCs were included. 7/12 "bespoke detected" patients had LUSC histology, with 19/20 in the "non-detected" group having LUAD histology. The single "bespoke non-detected" LUSC (Patient 252) from the published study was also sequenced using the OLcfDNA (Abbosh *et al.* 2017).

Unlike the bespoke approach, the OLcfDNA is not patient-tailored and instead targeted >150 hotspots and indels present in 11 genes (Table 3.4). Another difference between these two approaches was that the bespoke method conducted NGS on the total DNA yielded from plasma extraction (various input amounts) and was capped at 50,000 genome equivalents (165 ng). Alternatively, the OLcfDNA conducted analyses on ~20 ng DNA for all cfDNA samples. All 32 of the selected samples had ≥1 tumour-derived clonal variant targeted by the OLcfDNA (44 variants in 32 patients). Subclonal variants could also have been analysed, however there were no tumour-confirmed, subclonal variants covered by the OLcfDNA assay in this cohort.

For the purposes of this study, the detection of tumour-confirmed variants are defined by a few terms outlined in Figure 3.7. "Hotspot" variants denote those that are a registered as such on the OLcfDNA hotspot file, where these represent the most confident genomic positions, as they have been extensively validated by the manufacturer for specificity. "Novel" variants are those that are at a hotspot locus (confident call), but were a different allele to those targeted on the hotspot file. Finally, as this study uses multiregional tumour WES data, it was possible to map all clonal tumour variants that fell within the confines of OLcfDNA amplicons, although these may not be within "Hotspot" regions; these are therefore the least confident variants. These were termed "Covered" variants and all required manual inspection on IGV in order to be deemed positive within cfDNA.

For “Hotspot” and “Novel” variants called by the variant caller, the Molecular metrics (original DNA molecules) were used to calculate the plasma VAF (Allele Molecular Coverage/Total Molecular Coverage). However, for variants that were not called, access to the Molecular VAF was not possible, and therefore in these cases the classical Read VAF (Allele Read Coverage/ Total Read Coverage) was used.

For variants that were not called by the variant caller, the determination of a variant as being positive was based upon a number of criteria: (1) Variant Read VAF of > 0.05% and > 20 mutant reads, (2) variant was not within the first or last 15 bases of an amplicon, (3) variant was called by the bespoke approach at a similar Read VAF (where possible). Bespoke approach read metrics for each variant targeted by the OLcfDNA are shown in Supplementary Table 1.1.

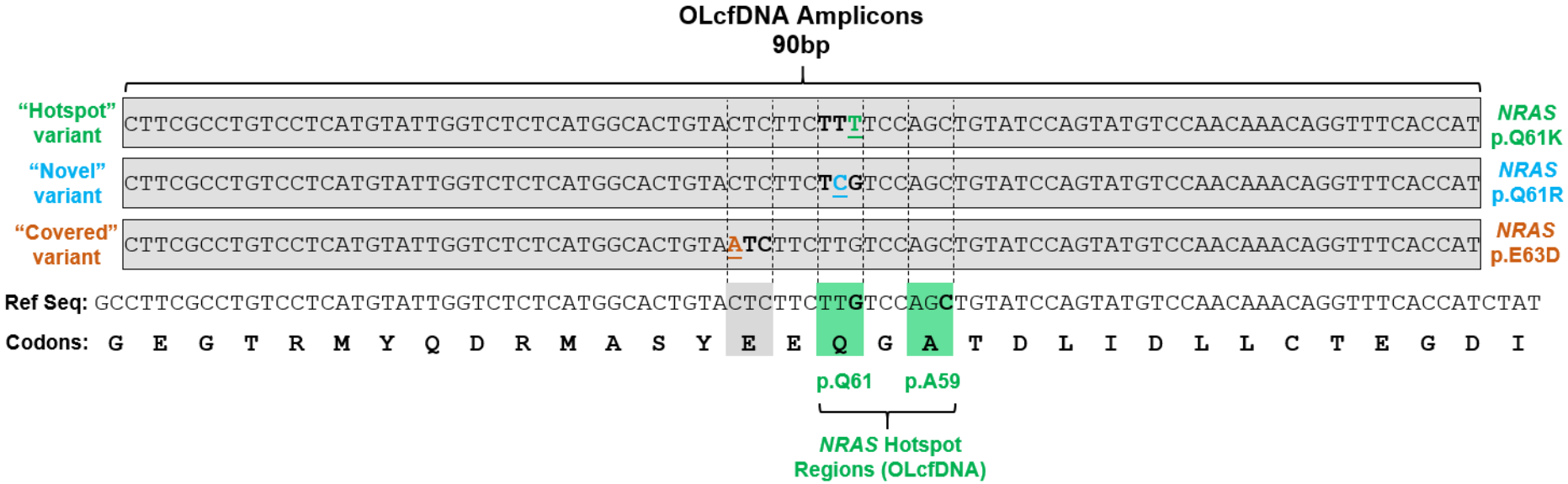


Figure 3.7 – Example of variant terms used in this study. Schematic showing an example of an amplicon targeting a *NRAS* Hotspot region, based on raw NGS data from this chapter. *NRAS* p.Q61K (G>T transversion) is a variant specifically targeted by the OLcfDNA hotspot file ("Hotspot", Green) and therefore will likely be called by the variant caller if 2 ≥ mutant molecular families are present in cfDNA. Although not a Hotspot variant, an *NRAS* p.Q61R (T>C transversion) can be called by the variant caller, as this is located within a Hotspot region, but is termed as a "Novel" variant (Blue). Finally, if a base change occurs away from a Hotspot region (p.E63D), but within the confines of an amplicon targeting a hotspot, these have been termed "Covered" variants (Orange) and represent tumour-confirmed variants (from Multiregional-WES) that were manually identified within TRACERx cfDNA.

In the “bespoke detected” group, 9/12 patients had ≥ 1 clonal tumour variants detected in their pre-surgical cfDNA with the OLcfDNA, either by the variant caller, or by manual inspection on IGV (Table 3.6A). 10/10 (100%) of the tumour-confirmed variants targeted were detected in patients that had a mean clonal VAF of > 0.1% using the bespoke approach (Table 3.6A, green shading).

For patients with a mean clonal VAF of < 0.1% (still detected by the bespoke method), the OLcfDNA was able to detect tumour-derived DNA in only 2/7 variants targeted. This culminated in three patients detected by the bespoke approach (1, 242 and 251) that were not detected by the OLcfDNA. Although these results seem to suggest that the OLcfDNA had poorer sensitivity than the bespoke method, it is important to mention that of the five negative variants, 3/5 were not detected using either method (Table 3.7). As the bespoke approach targeted a median of 11 clonal variants per patient, the “detection” of ctDNA was acquired by the detection of other clonal variants not targeted by the OLcfDNA (Abbosh *et al.* 2017). For example, 242 represented the only LUSC that was not detected by the OLcfDNA in this group, but this was expected, as both approaches could not identify the *TP53* p.V157F variant (Table 3.7); the bespoke approach successfully detected 8/13 other clonal tumour variants and therefore 242 (ID in publication, CRUK0064) was “detected” by the bespoke approach.

There were three variants that were discordant between approaches; two variants were detected by the bespoke approach only and one was found solely by the OLcfDNA. Interestingly in Patient 135, one variant was detected by the OLcfDNA, and the other by the bespoke approach and vice versa (Table 3.7), meaning if a higher quantity of cfDNA was analysed (>20ng), it would have been likely that both variants could have been identified by the OLcfDNA, due to a lower LOD.

Three patients (97, 110 and 175) had two tumour-confirmed clonal variants identified, with the remaining six ctDNA-positive patients having a single variant detected (total of 12 variants detected). With regards to variant terms, 7/12 were “Hotspot” variants, 4/12 were “Covered” variants (not called) and a single variant was “Novel” (97, *NRAS* p.Q61R; called) (Table 3.6A).

A) Detected NSCLCs with Bespoke approach

Patient	Tumour Volume (cm^3)	Histology	Bespoke approach		OLcfDNA assay		
			Mean clonal plasma VAF (%)	ctDNA Positive?	Clonal Tumour variant	Variant detected?	Plasma VAF (%)
33	178.75	LUSC	5.37	Y	*TP53* p.E154K	Y	4.89
97	38.51	LUAD	2.10	Y	*NRAS* p.Q61R	Y	2.05
					TP53 p.Y234C	Y	1.14
175	69.01	LUAD	1.71	Y	*BRAF* p.K601E	Y	1.07
					TP53 p.M237I	Y	2.70
64	58.48	LUSC	1.41	Y	*TP53* p.R273C	Y	0.39
22	16.41	LUSC	0.21	Y	*TP53* p.Y220C	Y	0.19
203	18.71	LUSC	0.18	Y	*TP53* p.T86K	Y	0.07
110	17.39	LUSC	0.16	Y	*PIK3CA* p.E545K	Y	0.28
					ROS1 p.A1955D	Y	0.21
251	17.20	LUAD	0.08	Y	*EGFR* p.L858R	N	x
					TP53 p.E286*	N	x
111	6.64	LUSC	0.07	Y	*TP53* p.R148G	Y	0.09
135	43.69	LUAD	0.06	Y	*KRAS* p.G12V	N	x
					TP53 p.Y220C	Y	0.08
242	9.24	LUSC	0.05	Y	*TP53* p.V157F	N	x
1	10.59	LUAD	0.01	Y	*KRAS* p.G12V	N	x

B) Non-detected NSCLCs with Bespoke approach

Patient	Tumour Volume (cm^3)	Histology	Bespoke approach		OLcfDNA assay		
			Predicted plasma VAF (%)	ctDNA Positive?	Clonal Tumour variant	Variant detected?	Plasma VAF (%)
223	197.42	LUAD	2.96 (1.01 - 8.67)	N	*KRAS* p.Q61L	N	x
41	27.28	LUAD	0.32 (0.21 - 0.49)	N	*KRAS* p.G12C	N	x
46	23.30	LUAD	0.27 (0.18 - 0.41)	N	*EGFR* p.G719A	N	x
55	21.68	LUAD	0.25 (0.16 - 0.38)	N	*KRAS* p.G12C	N	x
183	19.06	LUAD	0.22 (0.14 - 0.33)	N	*KRAS* p.G12C	N	x
					TP53 p.Y220C	N	x
206	17.00	LUAD	0.19 (0.12 - 0.29)	N	*TP53* p.P152L	N	x
					TP53 p.G145C	N	x
					BRAF p.D594N	N	x
163	10.00	LUAD	0.10 (0.06 – 0.18)	N	*TP53* p.R280K	N	x
51	7.45	LUAD	0.08 (0.04 - 0.14)	N	*EGFR* p.L858R	N	x
					TP53 p.F270L	N	x
115	7.28	LUAD	0.07 (0.04 - 0.14)	N	*KRAS* p.G12C	N	x
49	5.95	LUAD	0.06 (0.03 - 0.12)	N	*KRAS* p.G12V	N	x
62	4.65	LUAD	0.04 (0.02 - 0.10)	N	*KRAS* p.G12C	N	x
84	4.61	LUAD	0.04 (0.02 - 0.10)	N	*KRAS* p.G12A	N	x
193	4.18	LUAD	0.04 (0.02 - 0.09)	N	*PIK3CA* p.H1047R	N	x
180	3.61	LUAD	0.03 (0.01 - 0.08)	N	*TP53* p.R158L	N	x
103	3.31	LUAD	0.03 (0.01 - 0.08)	N	*TP53* p.R175H	N	x
142	2.76	LUAD	0.03 (0.01 - 0.07)	N	*TP53* p.R280T	N	x
					EGFR p.L861Q	N	x
75	2.70	LUAD	0.02 (0.01 - 0.07)	N	*EGFR* p.L858R	N	x
					TP53 p.C176F	N	x
60	2.62	LUAD	0.02 (0.01 - 0.07)	N	*TP53* p.A159P	N	x
252	0.73	LUSC	0.01 (0.001 - 0.03)	N	*TP53* p.G154V	N	x
225	0.90	LUAD	0.01 (0.002 - 0.03)	N	*KRAS* p.G12V	N	x
					TP53 p.S215I	N	x

Key

Bespoke panel	>99% sensitivity at ≥ 0.1 VAF
	84% sensitivity at 0.05 – 0.1% VAF
	46% sensitivity at 0.01 – 0.05% VAF
	4.2% sensitivity < 0.01% VAF
OLcfDNA assay	90% sensitivity at ≥ 0.1% VAF

Detected SNVs
"Hotspot", called
"Novel", called
"Hotspot", not called (manually identified)
"Covered", not called (manually identified)

Non-detected SNVs
"Hotspot", not called

Table 3.6 – Cross-platform validation of bespoke approach using the OLcfDNA. Pre-surgical cfDNA detected (A) and non-detected (B) by the bespoke approach, with clonal variants covered by the OLcfDNA shown. 7/12 variants positive by the OLcfDNA were called by the variant caller, with 5/12 not being called (manually identified). All cfDNA non-detected using the bespoke approach was also non-detected using the OLcfDNA.

			Bespoke Approach				OLcfDNA			
Patient	**Histology**	**Tumour variant**	**Detected?**	**VAF (%)**	**Allele Read Coverage**	**Total Read Coverage**	**Detected?**	**VAF (%)**	**Allele Read Coverage**	**Total Read Coverage**
251	LUAD	*EGFR* p.L858R	**Y**	**0.03**	**9**	**28,574**	N	0.007	6	88,363
		TP53 p.E286*	N	0	0	23,796	N	0	0	65,489
111	LUSC	*TP53* p.R148G	**Y**	**0.10**	**22**	**21,494**	**Y**	**0.09**	**57**	**63,514**
135	LUAD	*KRAS* p.G12V	**Y**	**0.10**	**47**	**45,945**	N	0.02	11	55,796
		TP53 p.Y220C	N	0.004	1	28,487	**Y**	**0.15 (0.08*)**	**103**	**69,214**
242	LUSC	*TP53* p.V157F	N	0.15	17	11,588	N	0.002	1	54,944
1	LUAD	*KRAS* p.G12V	N	0.02	14	73,340	N	0.003	2	57,948

Table 3.7 - Comparison of NGS approaches in patients with < 0.1% mean clonal VAFs. For patients that were "bespoke detected" that had mean clonal plasma VAFs of < 0.1%, the variants targeted by the OLcfDNA were compared to the results of the bespoke approach.*Allele Molecular VAF is shown in brackets, as this variant was called by the variant caller, but other variant detected by the OLcfDNA was manually detected (using Read VAF). The NGS variant metrics for the bespoke approach are found in the published article (Abbosh *et al.* 2017).

Of concern, was the frequency of positive variants that were not called by the OLcfDNA software (5/12 positive variants) and required manual identification. A particularly obscure example is in the case of Patient 33, where a *TP53* p.E154K was found using the bespoke approach at 5.82% VAF in cfDNA (see Supplementary Table 1.1). This variant was within an amplicon targeted by the OLcfDNA and yet was not called. Manual inspection showed the variant was present at a Read VAF of 4.89%; a very surprising finding considering this was the highest VAF found amongst the whole TRACERx cohort (Figure 3.8). For IGV images for the other four manually-identified positive variants in the "bespoke detected" group, see Supplementary Figure 1.1.

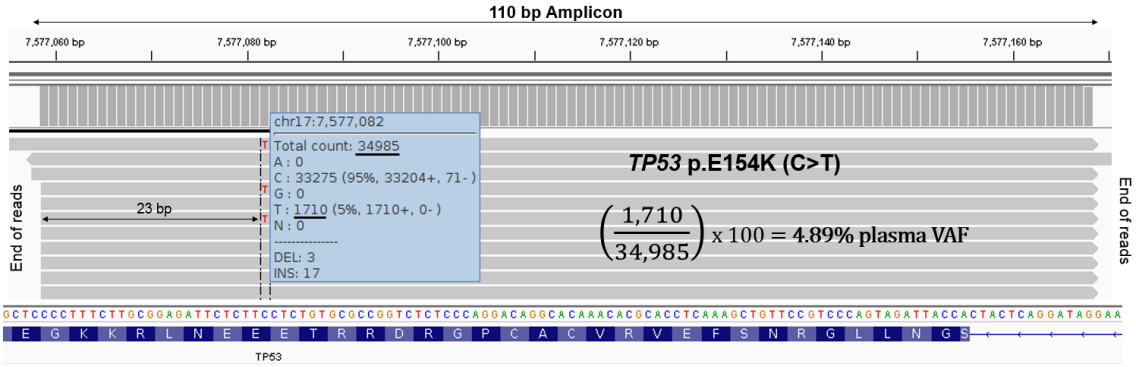


Figure 3.8 - Manual detection of a TP53 variant in Patient 33 on IGV. Image showing the reads of the entire amplicon covering a variant that was not called by the variant caller, despite being present at a read VAF of 4.89%. The variant was located relatively centrally within the amplicon, meaning it was unlikely to be an artefact.

The OLcfDNA was unable to detect each of 27 tumour variants in the "bespoke non-detected" group and therefore the two approaches were in agreement (Table 3.6B). These tumours varied widely in tumour size by CT imaging and despite the study showing that clonal plasma VAF and tumour size correlate, a high proportion of LUADs seemed to be resistant to this trend.

Overall, of the 44 variants targeted by the OLcfDNA in this section, 40 were analysed by both approaches. The four variants not analysed by the bespoke approach were as follows: *PIK3CA* p.E545K (110), *BRAF* p.K601E (175), *KRAS* p.G12C (183) and *KRAS* p.G12V (225). Of the remaining 40 variants, there was a high degree of concordance between approaches, with 37/40 (92.5%) variants

having the same detection status (positive or negative), with the three discordant variants already discussed in this section.

One important consideration for these data was to ascertain whether negative variants were not detected due to a lower read depth and hence lower sensitivity. Therefore, the total read depths covering all 40 variants (amplicon covering each variant) were plotted based on ctDNA-detection status (Figure 3.9). There was no significant difference in variant read depth for the detected and non-detected variants using both approaches (OLcfDNA, P = 0.3307; Bespoke, P = 0.4366). With respect to the three discordant variant calls, all read depths were higher with the OLcfDNA, yet as mentioned 2/3 were detected solely by the bespoke method.

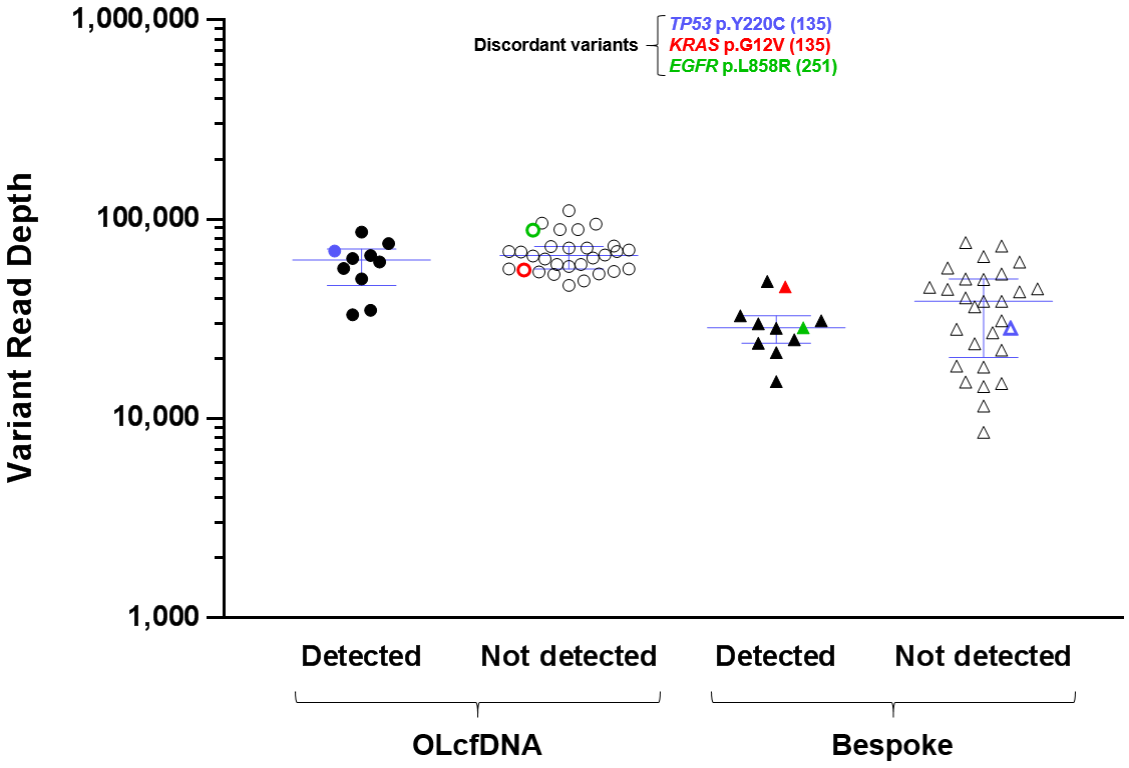


Figure 3.9 - Total variant read depths of both approaches according to detection status. The total read depths at the genomic locations of 40 tumour-confirmed variants analysed by both methods. Data points in black represent variants there were concordant using both approaches (i.e. both detected or both non-detected). Coloured variants represent those that were discordant, with two being detected using the bespoke method only (red and green variants) and a single variant that was detected only by the OLcfDNA (blue). Wilcoxon rank-sum test was conducted between the Detected and Not detected groups of each approach; OLcfDNA, P = 0.3307; Bespoke, P = 0.4366. Median and 25th and 75th quartiles are indicated on the graph.

3.3.4 Pre-surgical ctDNA with clinicopathological variables

Pre-surgical cfDNA analysis was extended to include a further nine patients (n = 41) which all had ≥ 1 clonal tumour variant covered by the OLcfDNA. These additional nine patients were not included in the previous section because they were not analysed using the bespoke approach. Pre-surgical cfDNA from other TRACERx patients (n = 5) have also been sequenced as part of this project, however these were excluded from this section for the following reasons; no tumour WES data (172, 730 and 759) and no clonal tumour variant covered by the OLcfDNA (79 and 445).

In all, this cohort was comprised of 30 (73%) LUADs and 11 (27%) LUSCs. The majority of these patients (21/41) had stage I NSCLC, with 12 and 8 patients being stage II and III respectively. For tumour-confirmed clonal variants, the OLcfDNA covered a total of 54 variants across 41 patients. Of these, one patient (206) had three variants covered by the OLcfDNA, 11 had two and the remaining 29 patients had a single variant covered (For details, see Supplementary Table 1.1). The OLcfDNA detected a total of 16/54 (30%) clonal variants, either through the variant caller (n = 10) or through manual inspection using the IGV tool (n = 6) (see Figure 3.8 and Supplementary Figure 1.1). Altogether, these 16 variants resulted in 13/41 (32%) patients having ctDNA detected pre-surgically, with three patients having two variants and 10 having a single variant detected (Figure 3.10).

Of the 12 patients with ≥ one variant covered by the OLcfDNA, there was concordance regarding ctDNA detection. 10/12 patients displayed 100% concordance, whereby no or all of the variants were detected within each patient cfDNA sample (see Supplementary Table 1.1). The two exceptions were 135 and 429, where only one of two variants were detected using the OLcfDNA; results likely due to these variants falling on the LOD for the assay (~0.05% - 0.1% VAF). The results of 135 have been previously described (see Table 3.7). For 429, where the same sample had not been sequenced using the bespoke approach, the negative variant did have a single mutant molecular family detected (1/2,966; 0.03% VAF), however a minimum of 2 mutant molecular families were required for a positive call and therefore this was deemed negative. It is encouraging that

there was general concordance between clonal variants within individual patient samples, as this provided evidence that phylogenetic reconstruction of these patient tumours is generally represented within cfDNA.

Of note, only 3/21 (14.3%) stage I NSCLCs were detected, and these were all LUSCs (patients 33, 111 and 203) (Figure 3.10). All 14 stage I LUADs were not detected, seemingly indicating that LUAD histology coupled with stage I disease seems to be indicative of poor tumour DNA representation within plasma. The only LUSCs that were not detected by the OLcfDNA were from stage I tumours (3/7 detected). Considering that there were no stage III LUSCs analysed in this dataset, and that all stage II LUSCs were detected (4/4), does indicate that the vast majority of LUSCs across stages II-III are likely to be ctDNA-detectable using current methods possessing 0.1% analytical sensitivity. Across all stages, 7/11 (64%) LUSCs were detected and 6/30 (20%) LUADs; a result that was similar to the findings using the bespoke approach (Abbosh *et al.* 2017).

As expected, the published study found that LUSC tumours were more necrotic than LUADs. As necrosis is likely to contribute to the passive release of ctDNA into circulation (Diaz and Bardelli 2014), it was possible that the higher proportion of LUSCs detected in cfDNA could be due to more abundant necrosis seen in this subtype. In 33/41 patients, necrosis data was available (Abbosh *et al.* 2017). Detected patients (≥1 tumour variant detected = detected) were significantly ($P < 0.001$) more associated with necrosis than non-detected patients using the OLcfDNA (Figure 3.11). Eight out of nine (89%) detected patients had ≥ 40% pathological tumour necrosis, compared to just 3/24 (12.5%) of the non-detected patients.

Considering the lack of detection of earlier stage tumours, tumour size was considered as a potential limiting factor. For 33 patients, volumetric data from CT scans was available for the association with pre-surgical ctDNA detection status. For tumour-confirmed variants, ctDNA positivity was significantly associated with larger tumours (Wilcoxon sum-rank test, $P = 0.0036$) (Figure 3.12).

Regarding clinical outcomes, in order to account for varying patient recruitment dates, clinical relapse or death within two years of surgery was used as a cut-off

for each patient. 7/13 (54%) ctDNA-positive patients experienced clinical relapse or death within 2 years of surgery, compared to 8/28 (29%) of those in the non-detected group (Figure 3.10). This observation is likely due to patients with later stages of disease (Stages II/III vs I) typically being more represented in the ctDNA-detected group. Based on stage, relapse or death (within 2 years) occurred in 4/21 (19%) of stage I, in 5/12 (42%) of stage II and in 6/8 (75%) of stage III NSCLC.

Due to smoking dominating the aetiology of NSCLC, it is difficult to identify trends between smoking history and ctDNA detection status. With the OLcfDNA covering only 11 genes, it was possible that patients with a more substantial smoking history would have had more clonal mutations in their tumours, and hence a higher probability of having one or more of these genes mutated. Clonal mutation burden has been shown to be associated with smoking history in NSCLC (Jamal-Hanjani *et al.* 2017), and therefore it was worth considering the potential that ctDNA detection rates may be improved in heavier smokers using this assay. However this does not seem to be the case in this dataset. For 33 of these 41 patients, pack year data was available. When separated into tertiles, the total number of variants covered by the OLcfDNA was 14 in the heaviest smoker tertile (92 – 48 pack-years), 15 for the middle tertile (35 – 19) and 16 for the lowest smoking tertile (15 – 0 pack years). With regards to ctDNA detection status, 3/11 detected patients were within each tertile, meaning tobacco exposure does not appear to influence ctDNA detection in these patients with the OLcfDNA (Table 3.8)

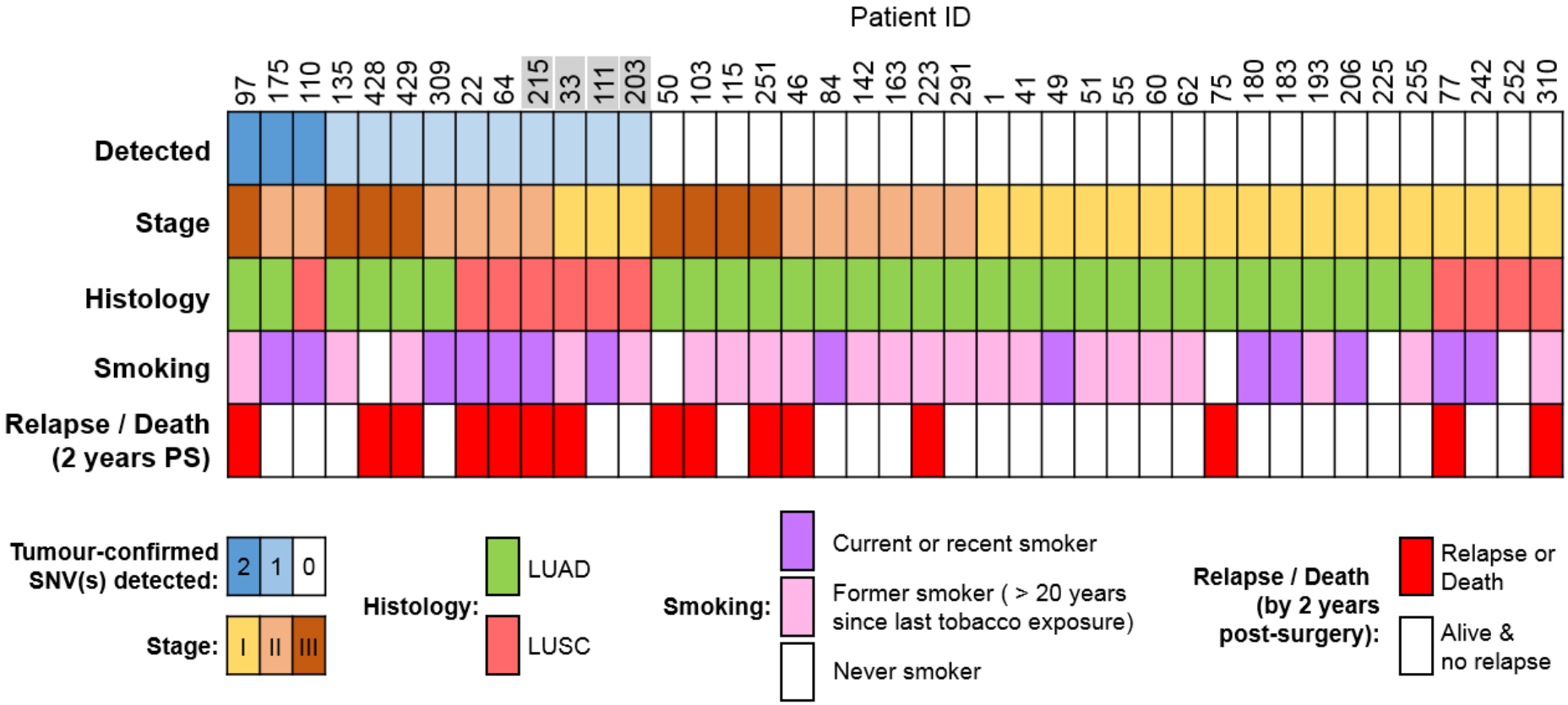


Figure 3.10 - Heatmap showing pre-surgical ctDNA detection status and clinicopathological variables. For tumour-confirmed variants (found on WES of tumours), ctDNA detection is indicated for each of 41 patients, with stage, histology, smoking status and relapse or death (by 2 years post-surgery) indicated. All patients had ≥ 1 variant present in their primary tumours covered by the OLcfDNA assay. Patient IDs in grey indicate the single variant was manually identified (not called by the OLcfDNA software), but were either called using the bespoke method (Patients 33, 111 and 203) or was at a confident read VAF (Patient 215; 0.46% VAF). Patients 110 and 175 also had a single variant that was manually identified, however they also had a second variant that was called.

.

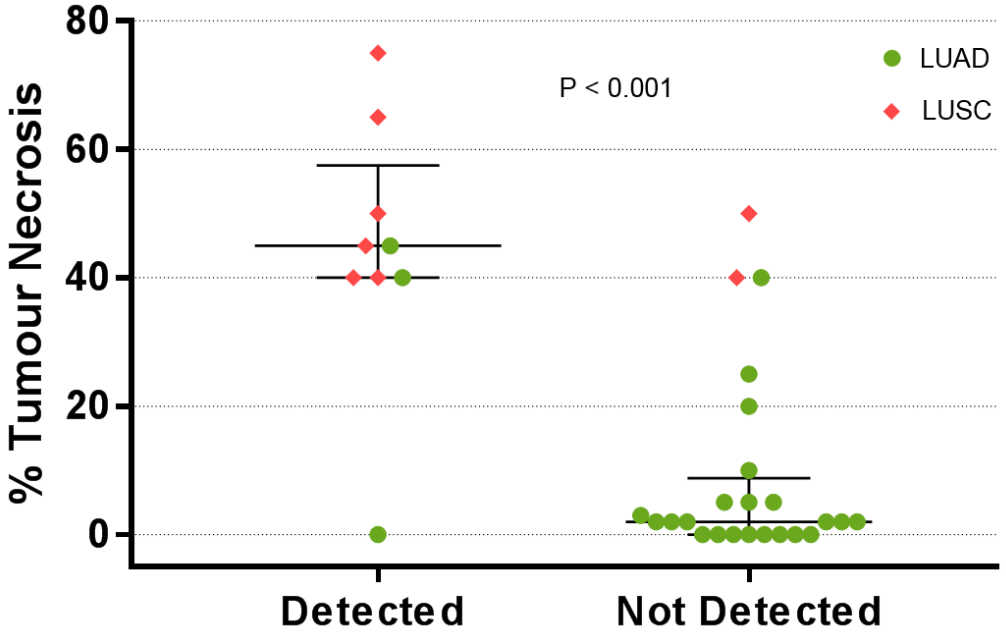


Figure 3.11 - ctDNA detection status vs. tumour necrosis. Pathological examination of tumour tissue conducted by the TRACERx consortium (Abbosh *et al.* 2017) was compared to detection of ctDNA using the OLcfDNA. LUADs (green circles) and LUSCs (red diamonds) are indicated.Wilcoxon rank-sum test, P = 0.0006. Median and 25th and 75th quartiles are indicated.

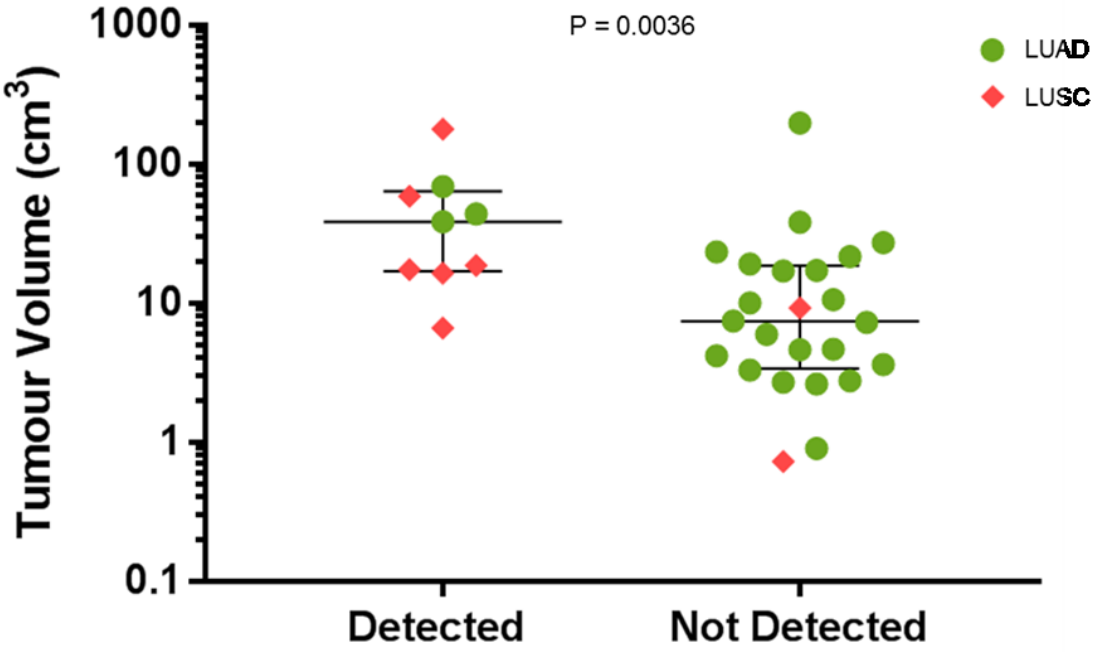


Figure 3.12 - ctDNA detection status vs tumour volume. Tumour volume (cm^3) measured by CT volumetric analysis (Abbosh et al. 2017) was compared to the detection of tumour-confirmed SNVs using the OLcfDNA assay. LUAD (green circles) and LUSC (red diamonds) tumours are specified. Wilcoxon rank-sum test, P = 0.0036. Median and 25th and 75th quartiles are indicated.

Tertile	Pack-years (descending)	Patient	Histology	OLcfDNA targeted variants	Detected OLcfDNA variants	Detected
Heavy smokers	92.00	203	LUSC	1	1	Y
	73.50	41	LUAD	1	0	N
	73.50	60	LUAD	1	0	N
	60.00	115	LUAD	1	0	N
	58.00	33	LUSC	1	1	Y
	55.50	183	LUAD	2	0	N
	53.00	49	LUAD	1	0	N
	52.00	55	LUAD	1	0	N
	50.00	111	LUSC	1	1	Y
	49.00	206	LUAD	3	0	N
	48.00	180	LUAD	1	0	N
Medium smokers	35.00	97	LUAD	2	2	Y
	35.00	135	LUAD	2	1	Y
	34.00	142	LUAD	2	0	N
	33.00	1	LUAD	1	0	N
	32.25	251	LUAD	2	0	N
	31.00	163	LUAD	1	0	N
	30.00	84	LUAD	1	0	N
	28.00	103	LUAD	1	0	N
	27.75	62	LUAD	1	0	N
	26.50	242	LUSC	1	0	N
	19.00	64	LUSC	1	1	Y
Low/no smokers	15.00	110	LUSC	2	2	Y
	13.75	22	LUSC	1	1	Y
	13.50	46	LUAD	1	0	N
	10.25	175	LUAD	2	2	Y
	9.75	223	LUAD	1	0	N
	0.45	193	LUAD	1	0	N
	0.05	51	LUAD	2	0	N
	0.00	50	LUAD	1	0	N
	0.00	75	LUAD	2	0	N
	0.00	225	LUAD	2	0	N
	0.00	252	LUSC	1	0	N

Table 3.8 - OLcfDNA detection status vs. smoking pack-years. 33 patients were separated into three equal tertiles according to pack-years; heavy smokers, medium and low/never smokers. Both the number of variants covered by the OLcfDNA assay and their detection status are indicated.

3.3.5 Tumour-discordant variants present in cfDNA

In addition to tumour-confirmed variants, the OLcfDNA called a high number of discordant “Hotspot” variants of unknown origin in both pre-surgical and longitudinal samples analysed throughout this project. As previously described (Figure 3.7), “Hotspot” variants denote those that are specific targets for the OLcfDNA and therefore if detected by the variant caller are typically more confident variants. Other “Novel” variants were also called, however these were excluded due to decreased confidence. Originally, the number of discordant hotspot calls totalled 46 affecting 33/97 (34%) different cfDNA samples (various time points). These calls involved 21/43 (49%) patients either in pre-surgical cfDNA, in follow-up cfDNA or in both. While in the previous section 16 tumour-confirmed calls were detected in 41 patients, 27 hotspot tumour-discordant calls (variant not found in any tumour region of WES) were made for these samples.

Due to the high number of these events, stringent filtering was subsequently conducted to leave only variants called that had a VAF ≥ 0.1%, a PHRED-quality score of ≥ 15 and a number of unique mutant molecules ≥ 3 (usually ≥ 2), as a means to reduce false positive calls. This left 24 variants affecting 13 patients with VAFs ranging between 0.1% and 4.44% (Table 3.9). *TP53* and *PIK3CA* variants dominated this dataset, responsible for 12 and 10 of these events respectively. All 10 of the *PIK3CA* calls occurred solely in the pre-surgical samples of seven patients, with three of these possessing two independent *PIK3CA* mutations; p.E545K and p.H1047R. Alternatively, recurring (within the same patient) *TP53* variants were identified in two patients; 110 had five confidently called *TP53* p.R273C variants in follow up cfDNA samples (see later section 3.3.7), while 428 had a *TP53* p.Y220C at two independent time points.

Two variants, a *PIK3CA* p.E545K (Patient 62) and a *KRAS* p.G12V (Patient 142) were called at 4.44% and 3.44% VAF respectively (Table 3.9). To put this in perspective, these variants represented the second and third highest VAFs in the entire pre-surgical TRACERx cohort (with the OLcfDNA), and yet they were not detected in any tumour region of their respective NSCLC tumours; a highly surprising finding. It seems unlikely that these “discordant” variants are NSCLC-

derived yet remained undetected by multiregional WES, considering the data available. The OLcfDNA and the bespoke approach were unable to detect a single tumour-confirmed, clonal (ubiquously found/more numerous) variant in these cfDNA samples, which in the case of the latter was the targeting of 23 individual clonal variants (four in Patient 62, 19 in Patient 142) (Abbosh *et al.* 2017). Hence, it is likely that these variants are derived from an alternative source, be it a biological or *via* a technological artefact. For 142, there is a potential explanation; this patient was found to have been previously treated for an oropharyngeal cancer (personal communication, Christopher Abbosh). Although this is conjecture and analysis of the oropharyngeal cancer biopsy has not been conducted (not a Leicester recruit), there is potential that this *KRAS* p.G12V could derive from the formerly diagnosed malignancy and not from the NSCLC subject to WES.

Considering that *PIK3CA* variants were present in the cfDNA of 7/43 patients filtered (10/43 in unfiltered data), when *PIK3CA* mutations are typically reported in just 3% of NSCLC, the regularity of these variants in this group seemed slightly sceptical. It is possible that these calls may represent false calls that could be due to ‘noise’ with OLcfDNA. Ideally, these *PIK3CA*-positive samples could be validated using ddPCR, however unfortunately lack of sample availability meant further verification was not possible.

Another potential source of these tumour-discordant variants could be clonal haematopoiesis of indeterminate potential (CHIP); a source explored with regards to the recurrent discordant variants identified within Patient 110 later in this chapter (see section 3.3.7).

Patient ID	Corresponding CRUK ID*	cfDNA Time Point	Chrom	Position	Ref	Var	Gene	VAF (%)	Allele Mol Cov	Total Mol Cov	Allele Read Cov	Total Read Cov	CDS	AA Change	Quality (PHRED)	LOD (%)
33	CRU0077	Pre-surgical	3	178936091	G	A	*PIK3CA*	**1.16**	57	4,901	489	43,702	c.1633G>A	p.E545K	52	0.05
			17	7578190	T	C	*TP53*	**0.20**	8	3,962	57	41,929	c.659A>G	p.Y220C	32	0.05
46	CRUK0004	Pre-surgical	3	178936091	G	A	*PIK3CA*	**0.59**	9	1,527	531	73,941	c.1633G>A	p.E545K	34	0.1
50	CRUK0003	Pre-surgical	3	178936091	G	A	*PIK3CA*	**0.69**	37	5,375	681	100,290	c.1633G>A	p.E545K	48	0.05
			3	178952085	A	G	*PIK3CA*	**0.11**	5	4,431	136	79,230	c.3140A>G	p.H1047R	28	0.05
			17	7577539	G	A	*TP53*	**0.36**	16	4,422	473	101,301	c.742C>T	p.R248W	39	0.05
60	CRUK0016	Pre-surgical	17	7578190	T	C	*TP53*	**0.68**	18	2,658	687	72,704	c.659A>G	p.Y220C	40	0.1
62	CRUK0018	Pre-surgical	3	178936091	G	A	*PIK3CA*	**4.44**	179	4,030	4,412	99,159	c.1633G>A	p.E545K	63	0.05
			3	178952085	A	G	*PIK3CA*	**0.81**	27	3,352	566	77,806	c.3140A>G	p.H1047R	44	0.05
		182 days	17	7578442	T	C	*TP53*	**0.10**	3	3,093	14	25,765	c.488A>G	p.Y163C	23	0.05
64	CRUK0062	Pre-surgical	3	178936091	G	A	*PIK3CA*	**0.97**	40	4,119	904	89,616	c.1633G>A	p.E545K	48	0.05
84	CRUK0027	Pre-surgical	3	178952085	A	G	*PIK3CA*	**0.11**	4	3,603	121	65,449	c.3140A>G	p.H1047R	25	0.05
110	CRUK0089	96 days	17	7577121	G	A	*TP53*	**0.12**	4	3,391	41	33,309	c.817C>T	p.R273C	25	0.05
		240 days	17	7577121	G	A	*TP53*	**0.25**	6	2,409	140	61,405	c.817C>T	p.R273C	34	0.1
		317 days	17	7577121	G	A	*TP53*	**0.57**	4	702	111	33,623	c.817C>T	p.R273C	34	0.25
		765 days	17	7577121	G	A	*TP53*	**0.55**	7	1,267	383	48,939	c.817C>T	p.R273C	31	0.15
		1,129 days	17	7577121	G	A	*TP53*	**0.14**	3	2,211	107	70,184	c.817C>T	p.R273C	15	0.1
135	CRUK0052	Pre-surgical	3	178936091	G	A	*PIK3CA*	**0.81**	21	2,602	438	60,360	c.1633G>A	p.E545K	42	0.1
			3	178952085	A	G	*PIK3CA*	**0.12**	3	2,406	56	52,325	c.3140A>G	p.H1047R	23	0.1
142	CRUK0058	Pre-surgical	12	25398284	C	A	*KRAS*	**3.44**	137	3,983	2,934	63,090	c.35G>T	p.G12V	61	0.05
215	n/a	Pre-surgical	7	116411990	C	T	*MET*	**0.19**	3	1,613	158	53,795	c.3029C>T	p.T1010I	19	0.1
255	n/a	735 days	17	7577121	G	A	*TP53*	**0.34**	9	2,657	258	73,543	c.817C>T	p.R273C	55	0.1
428	n/a	92 days	17	7578190	T	C	*TP53*	**0.14**	6	4,414	125	105,763	c.659A>G	p.Y220C	21	0.05
		505 days (First Recurrence)	17	7578190	T	C	*TP53*	**0.23**	9	3,965	145	118,203	c.659A>G	p.Y220C	42	0.05

Table 3.9 - Filtered Hotspot variants found using OLcfDNA discordant to tumour WES. Filtered variants (VAF ≥ 0.1%, Quality score ≥ 15, unique mutant molecules ≥ 3) were found in 20 cfDNA samples from 13 TRACERx patients either in pre-surgical cfDNA, follow-up cfDNA or both. Shading indicates the sample was post-surgical (follow-up). Quality (PHRED) score refers to the PHRED-scaled probability of an incorrect call at the respective base call (higher numbers = higher quality). LOD (%) refers to the limit of detection for the specific genomic location estimated based on the number of detected molecules. All tabulated variants possess a VAF exceeding their respective LOD value. *CRUK IDs are the corresponding ID for the patient in the published study (Abbosh *et al.* 2017).

.

3.3.6 Longitudinal ctDNA analysis of Leicester TRACERx recruits

In 18 TRACERx patients recruited at the Leicester site, longitudinal cfDNA analysis was subsequently conducted, to assess whether the OLcfDNA panel can characterise ctDNA with regards to relapse and treatment monitoring in early-stage NSCLC following surgery. In total (including pre-surgical), 81 cfDNA samples were sequenced (mean, 4.5 samples/patient), however a number of these could not detect tumour-confirmed variants at any time point. Three patients have not yet undergone tumour WES and also had no variants found using the OLcfDNA panel (172, 730, 759). One patient (79) had received WES, however no variants were covered by the OLcfDNA panel and no hotspots variants were found. And finally, one patient (445) had only a subclonal variant present on the OLcfDNA but this was not found at any time point.

Of the 13 patients that had a tumour-confirmed clonal variant covered by the OLcfDNA assay, six did not have presence of this variant in any of their cfDNA samples (n = 30) (Table 3.10, no shading). One patient in this group (310) relapsed shortly following surgery, however has survived more than 2 years post-surgery (as of 09/2017). All other patients in this group have not relapsed (median follow-up, 1,120 days). Alternatively, for the seven patients that had a variant detected at one or more time-points, five patients have experienced relapse within 670 days of surgery (Table 3.10, shading). The seven detected cases will be described in the following section (for raw data, see Supplementary Table 1.2).

Patient	Histology	Stage	cfDNA time points analysed (range)	Relapse? (days PS)	OLcfDNA Clonal tumour variant(s)	Variant found at any time point
41	LUAD	IB	5 (Pre-surgical – 986 days PS)	N	*KRAS* p.G12C	N
62	LUAD	IA	6 (Pre-surgical -1,097 days PS)	N	*KRAS* p.G12C	N
115	LUAD	IA	5 (Pre-surgical – 1,104 days PS)	N	*KRAS* p.G12C	N
255	LUAD	IA	5 (Pre-surgical – 917 days PS)	N	*EGFR* p.L861Q	N
291	LUAD	IIB	5 (Pre-surgical - 748 days PS)	N	*KRAS* p.G12C	N
310	LUSC	IA	4 (Pre-surgical - 756 days PS)	Y (<50)	*PIK3CA* p.E542K	N
64	LUSC	IIB	5 (Pre-surgical - 420 days PS)	Y (376)	*TP53* p.R273C	Y
77	LUSC	IA	4 (Pre-surgical - 957 days PS)	Y (669)	*TP53* p.Y234C	Y
110	LUSC	IIA	7 (Pre-surgical - 1,098 days PS)	N	*PIK3CA* p.E545K *ROS1* p.A1955D	Y Y
215	LUSC	IIA	2 (Pre-surgical - 76 days PS)	Y (167)	*TP53* p.K120E	Y
309	LUAD	IIA	4 (Pre-surgical - 916 days PS)	N	*KRAS* p.G12V	Y
428	LUAD	IIIA	4 (Pre-surgical - 505 days PS)	Y (371)	*EGFR* p.M766delinsMASV	Y
429	LUAD	IIIA	5 (Pre-surgical - 435 days PS)	Y (279)	*KRAS* p.G12C *TP53* p.R141L	Y N

Table 3.10 - Patient summary for longitudinal cfDNA analysis. A number of patients not have a clonal tumour-confirmed variant (indicated in each case) found in any of their cfDNA time points (no shading). Alternatively, seven patients (shaded) had tumour variants detected at one or more cfDNA time point. The range of blood sampling time points analysed for each patient is indicated.

3.3.6.1 Patient 428 (Figure 3.13A)

This patient had a stage IIIA LUAD (never-smoker) and received surgery and adjuvant chemotherapy (Cisplatin/Vinorelbine, 2 cycles) however discontinued treatment after 4 weeks due to toxicity. They experienced clinical relapse 371 days post-surgery.

A total of four cfDNA samples were analysed for this patient, including the first blood sample taken post clinical relapse (day 505). Unusually, the pre-surgical cfDNA sample analysed by the OLcfDNA assay was found to have a single molecular family positive for a p.M766delinsMASV insertion in exon 20 of *EGFR* (0.07% VAF). The oncomine variant caller usually automatically filters out variants found with < 2 mutant molecular families present as a means to prevent false positive calls, yet in this case it passed the filters. Communication with the manufacturers (Thermo Fisher Scientific) confirmed that this was due to increased confidence in variants that include multiple nucleotide changes to DNA sequence and therefore being unlikely to be a false positive.

Just 92 days post-surgery the patient was found to possess the same insertion in at a similar VAF (0.06%, 2 mutant molecular families). A call this low would be taken slightly speculatively, however at 218 days there was a six-fold increase in VAF (0.36%), providing strong evidence that earlier findings were true positives. A further 10-fold increase in VAF (3.60%) was found in the first blood sample following relapse. Surgery and an incomplete course of adjuvant chemotherapy was insufficient to clear this clone with the *EGFR* variant being detectable at all time points, possibly indicating an innate resistance to chemotherapy.

Considering clinical relapse was detected at 371 days, cfDNA analysis would have conservatively provided a lead time of 153 days (~5 months). Therefore, this case represents a good example of how cfDNA analysis can assess the success of curative treatment, as well as providing clinicians with crucial time to administer additional adjuvant regimens. Exon 20 insertions in *EGFR* have been shown to be resistant to inhibitors gefitinib, erlotinib and afatinib in a number of studies (Wu *et al.* 2011, Yasuda *et al.* 2012, Yasuda *et al.* 2013) and so additional information could be derived from this data if adopted clinically.

3.3.6.2 Patient 64 (Figure 3.13B)

Alike 428, Patient 64 also experienced relapse within a similar time period (376 days). This patient had a stage IIB LUSC (smoker) and received surgery, but received a full four cycles of adjuvant chemotherapy (Cisplatin/Vinorelbine).

Pre-surgical ctDNA was detected, with a tumour-confirmed *TP53* p.R273C found at 0.39% VAF. This variant was undetectable in both the pre-chemotherapy blood sample (30 days) and at 181 days (post-chemotherapy), seemingly indicating that surgery and adjuvant chemotherapy had succeeded in eradicating disease. However, at 258 days the variant re-emerged at 0.23% VAF, 118 days premature of clinical relapse (376 days). A fourth cfDNA sample after clinical relapse (420 days) also was positive for this variant, however this was manually identified, as this had poor sequencing depth leading to only 1/466 mutant family being detected. Considering this equated to 82/31,693 reads, it was considered likely to be a positive call.

While resistance to therapy in Patient 428 seemed innate, this case shows how surgery and chemotherapy can temporarily suppress tumour growth to undetectable ctDNA levels, without eradication of residual disease. The re-emergence of ctDNA in this patient reflects the metastatic trajectory observed.

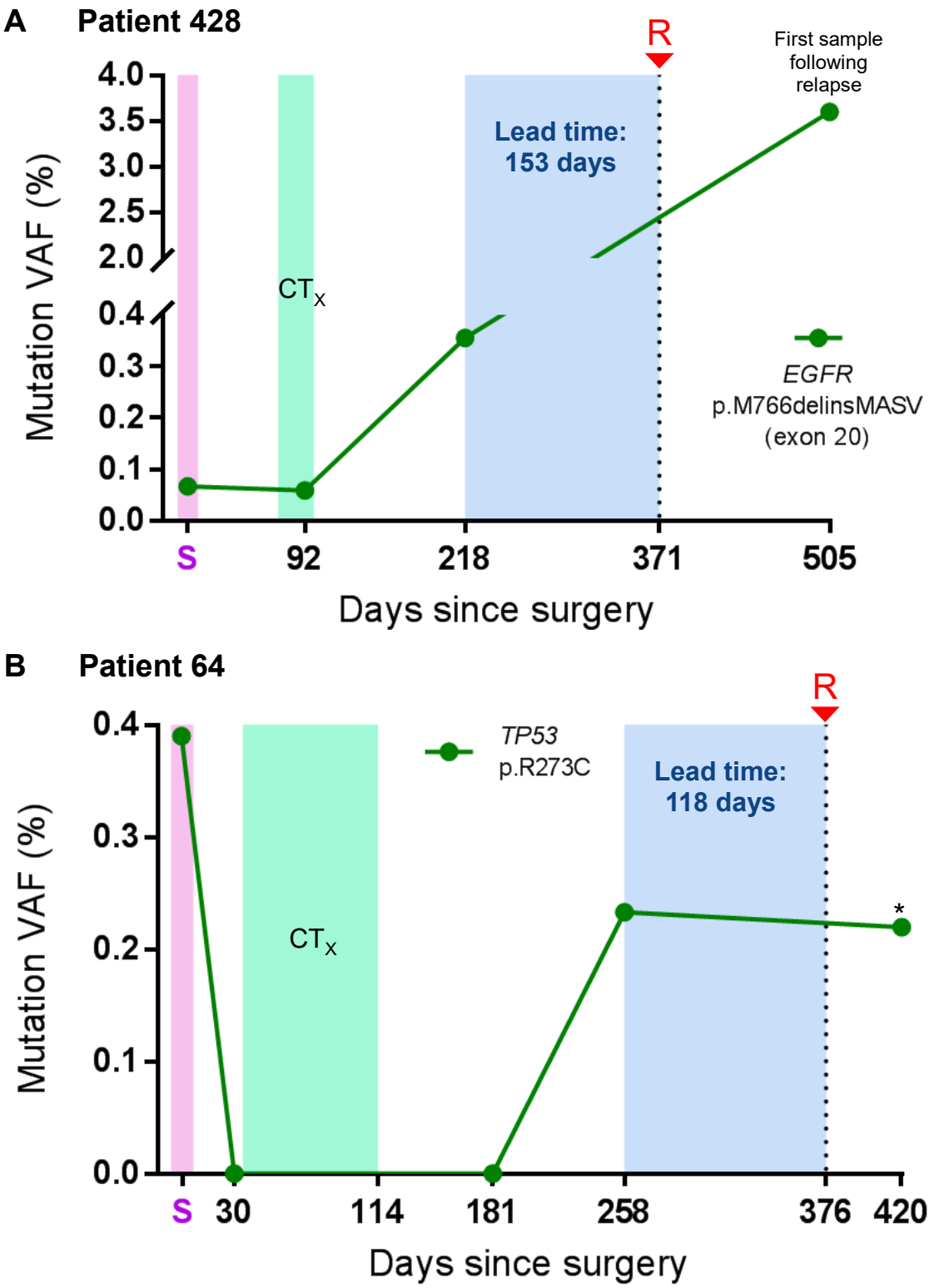


Figure 3.13 - Longitudinal ctDNA profiling in Patients 428 (A) and 64 (B). Graphs displaying ctDNA detection (variant is indicated) across pre-surgical time points (pink shading) and follow-up samples with integrated treatment timelines. Patient 428 had four cfDNA samples analysed, while Patient 64 had five. Green shading indicates the treatment (CTX, Chemotherapy) was adjuvant. Blue shading indicates the lead time period between a positive ctDNA call and clinical relapse date (R). S, surgery (day 0). *cfDNA sample was sequenced however was of poor quality, likely due to high DNA fragmentation, therefore variant was manually identified but only had 1 mutant molecular family positive (not called).

3.3.6.3 Patient 215 (Figure 3.14A)

This patient had a stage IIA LUSC (smoker) that received surgery, however no adjuvant chemotherapy. Relapse occurred just 167 days after surgery whereby the patient received palliative whole brain radiotherapy. This patient finally succumbed to their disease at 254 days.

With regards to cfDNA analysis, this patient was initially considered a ctDNA-negative case, however manual assessment on the IGV tool identified a clonal tumour variant in *TP53* p.K120E at 0.46% VAF in pre-surgical cfDNA. This case represented one of the six manually identified tumour confirmed pre-surgical positive samples (6/16 were manually identified). A single follow-up sample (76 days) was analysed for this patient and it also detected the *TP53* variant at a steadily increasing VAF of 0.67%; a result indicative of failure of surgery to eliminate disease. The lead time for this case was therefore 91 days prior to clinical relapse.

The three aforementioned cases provide examples of how the detection of a clonal variant can prospectively predict NSCLC relapse following surgery.

3.3.6.4 Patient 309 (Figure 3.14B)

While the first three cases provide evidence of impending relapse, Patient 309 represents an example of how cfDNA analysis can provide evidence of more favourable outcomes. This patient had a stage IIA LUAD (recent ex-smoker) and was treated with surgery and adjuvant radiotherapy in a somewhat unorthodox regimen; the radical adjuvant radiotherapy was administered over 2 years post-surgery due to delays regarding the assessment of mediastinal lymphadenopathy (personal communication, Ali Mohamed). This patient has not yet relapsed (1,091 days post-surgery) and is still being followed-up by TRACERx.

The patient was found to have a clonal *KRAS* p.G12V in their primary lung resection, a mutation also present in their pre-surgical cfDNA sample at 0.15% VAF. In the three post-surgical cfDNA samples analysed using OLcfDNA assay (469, 638 and 916 days), the *KRAS* p.G12V mutation was non-detected. This therefore follows the clinical outcome observed in this patient.

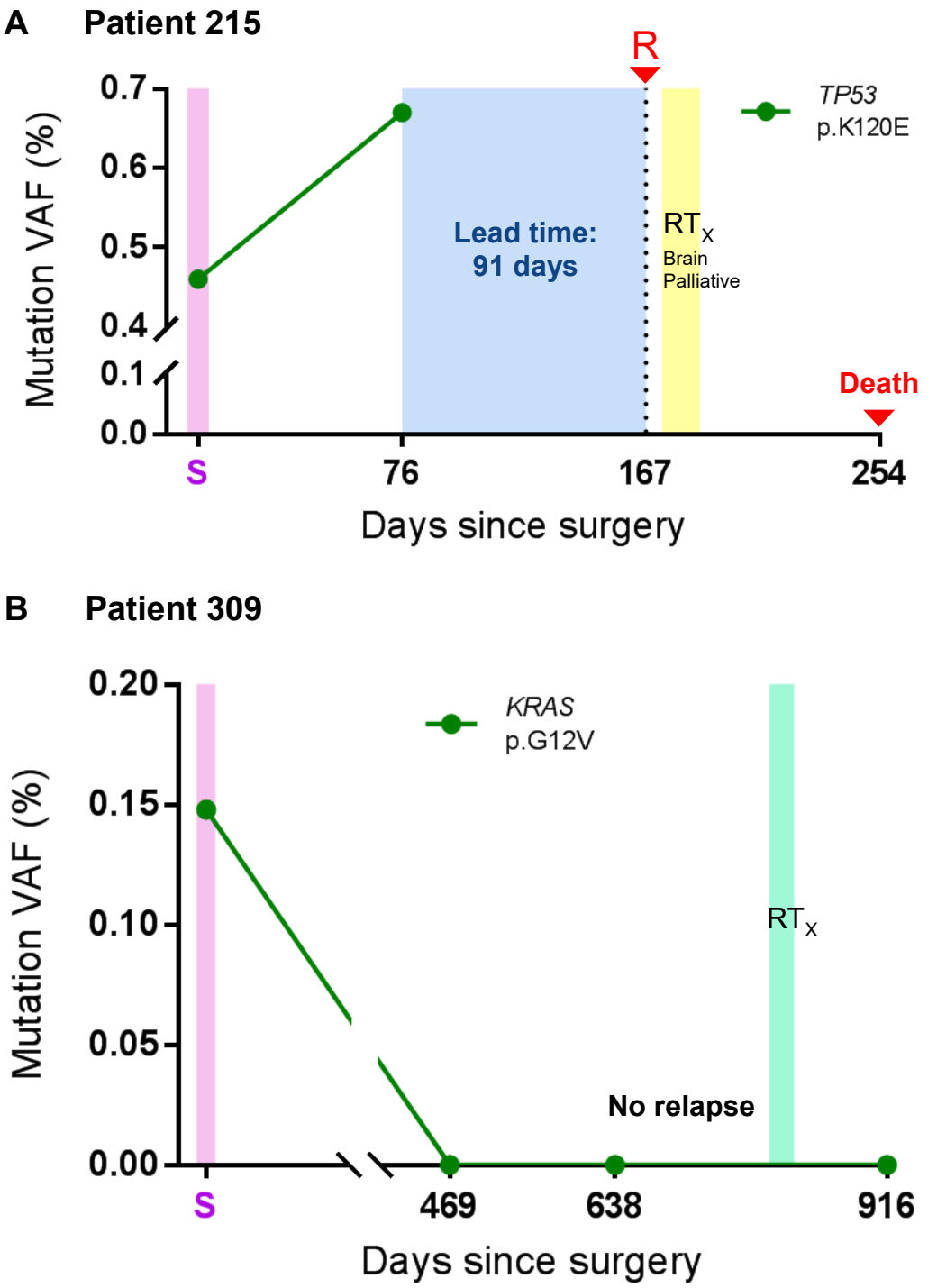


Figure 3.14 - Longitudinal ctDNA profiling in Patients 215 (A) and 309 (B). Graphs displaying ctDNA detection (variant is indicated) across pre-surgical time points (pink shading) and follow-up samples with integrated treatment timelines. Patient 215 had two cfDNA samples analysed, while Patient 309 had four. Patient 215 has evidence of impending relapse post-surgery, whereas the respective mutation remains undetectable following surgery in patient 309. Green shading indicates the treatment (RT_X, Radiotherapy) was adjuvant. Blue shading indicates the lead time period between a positive ctDNA call and clinical relapse date (R). S, surgery (day 0). Yellow shading indicates the treatment was palliative.

3.3.6.5 Patient 77 (Figure 3.15A)

Patient 77 represented an interesting case, whereby ctDNA was not detected in the pre-surgical cfDNA sample, however a clonal variant subsequently emerged upon clinical relapse. This patient had a stage IA LUSC (recent ex-smoker) and therefore was treated primarily with surgery, however they experienced a local recurrence in their right lower lobe at 669 days. Radical radiotherapy was administered to this recurrence and since this timeline, they are receiving immunotherapy and still surviving their disease 1,432 days post-surgery (as of 09/2018).

Unlike Patients 428, 64 and 215, this case was unable to provide a lead time, instead only detecting a tumour-confirmed *TP53* p.Y234C variant post local recurrence at 853 days (0.22% VAF). It is worth mentioning that only a single post-operative, pre-relapse cfDNA sample was analysed using the OLcfDNA assay and potentially the analysis of other samples may have been able to identify this variant. By 957 days, this variant had increased 12-fold in cfDNA to a VAF of 2.65%. It is possible that the lack of ctDNA detection prior to local recurrence could be due to this lesion being detected during a routine follow-up and therefore may have coincidently been a smaller tumour size and not at detectable levels in plasma.

3.3.6.6 Patient 429 (Figure 3.15B)

This patient had a stage IIIA LUAD (ex-smoker) and received surgery and adjuvant chemotherapy, however subsequently presented with a right hilar lymph node local recurrence at 279 days.

From OLcfDNA analysis alone the detection of this recurrence was not possible. Before surgery, a clonal *KRAS* p.G12C was detected, albeit at a low VAF (0.1%). However in four subsequent blood samples this or any other hotspot variant was not detected, including two samples that were taken after clinical relapse. Lack of detection of a clonal variant within cfDNA post-surgery despite relapse could be indicative of the local nature of this recurrence. Although phylogenetic reconstruction based on multiregional analyses can give insight into the timing of mutations, it doesn't necessarily mean that "clonal" variants are present in every

single tumour cell, but rather a dominant population of the tumour. In the case of this *KRAS* mutation, a potential explanation could be that micrometastases seeded in these lymph nodes prior to surgery by a non-mutant *KRAS* clone. Hence, upon surgery all cells containing the *KRAS* mutation were removed and therefore undetectable in follow-up blood samples. An additional explanation could be the presence of concurrent, proximal independent NSCLC; a possibility considering the highly mutagenic aetiology of this disease.

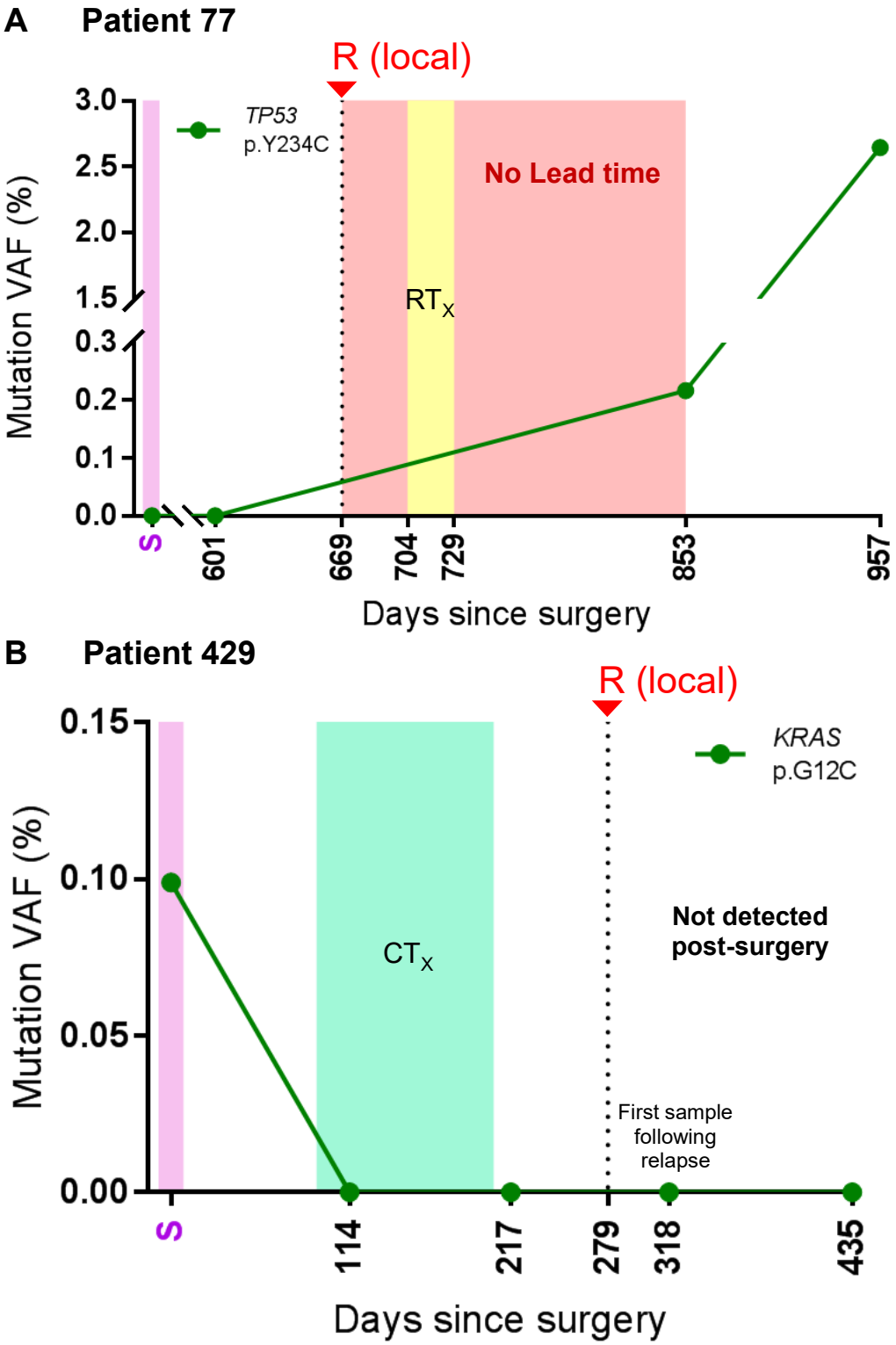


Figure 3.15 - Longitudinal ctDNA profiling in Patients 77 (A) and 429 (B). Graphs displaying ctDNA detection (variant is indicated) across pre-surgical time points (pink shading) and follow-up samples with integrated treatment timelines. 77 had four cfDNA samples analysed, while 429 had five. Pre-surgical cfDNA analysis was unable to detect the variant in 77, however it later emerged post-clinical relapse (R). Red shading indicates the lack of lead time in 77. Yellow shading indicates the treatment (RT_X, Radiotherapy) was palliative. 429 had its variant detected before surgery, however it was not found in any of four follow-up samples despite clinical relapse. Green shading indicates the treatment (CT_X, Chemotherapy) was adjuvant. S, surgery (day 0).

3.3.7 Patient 110 – a case study characterising tumour discordant variants.

This patient had a stage IIA LUSC (recent ex-smoker) and as of 09/2018 has not relapsed 1,273 days (~3.5 years) post-surgery. Multiregional-WES of tumour DNA revealed the presence of a *PIK3CA* p.E545K and a *ROS1* p.A1955D that were both targeted by the OLcfDNA. Analysis of the pre-surgical cfDNA also detected the presence of both of these variants, at 0.28% and 0.21% respectively. In serial plasma samples, both were absent in each of six longitudinal cfDNA samples (range 65-1,098 days); a result consistent with curative treatment.

However, a *TP53* p.R273C hotspot mutation absent from WES was also found in all but one cfDNA samples at low but varying VAFs (Figure 3.16). As this finding did not seem to follow the clinical outcome (no relapse) or WES data, it was important to characterise this mutation further. This likely represented either a tumour-specific minor clone that was not detected due to the lower sequencing depth associated with WES (~200X) or potentially the presence of Clonal Haematopoiesis of Indeterminate Potential (CHIP); a somatic, ageing-related event in haematopoietic stem cells (HSC) unrelated to the LUSC tumour.

As DNA derived from patient buffy coats were not sequenced using the OLcfDNA assay, it was impossible to determine the presence or absence of CHIP through these means. Therefore ddPCR was conducted targeting this variant in serial buffy coat DNA samples and a single cfDNA sample with the appropriate positive and negative controls (Figure 3.17). Unfortunately due to lack of DNA availability, the cfDNA sample (734 days post-surgery) was unable to detect *TP53* p.R273C mutant droplets, with only 853 WT droplets being detected thereby hindering sensitivity. In contrast, all three of the buffy coat DNA samples at 209, 286 and 734 days were found to possess the mutation at 0.26%, 0.40% and 0.42% VAF respectively, indicating the presence of CHIP in this patient.

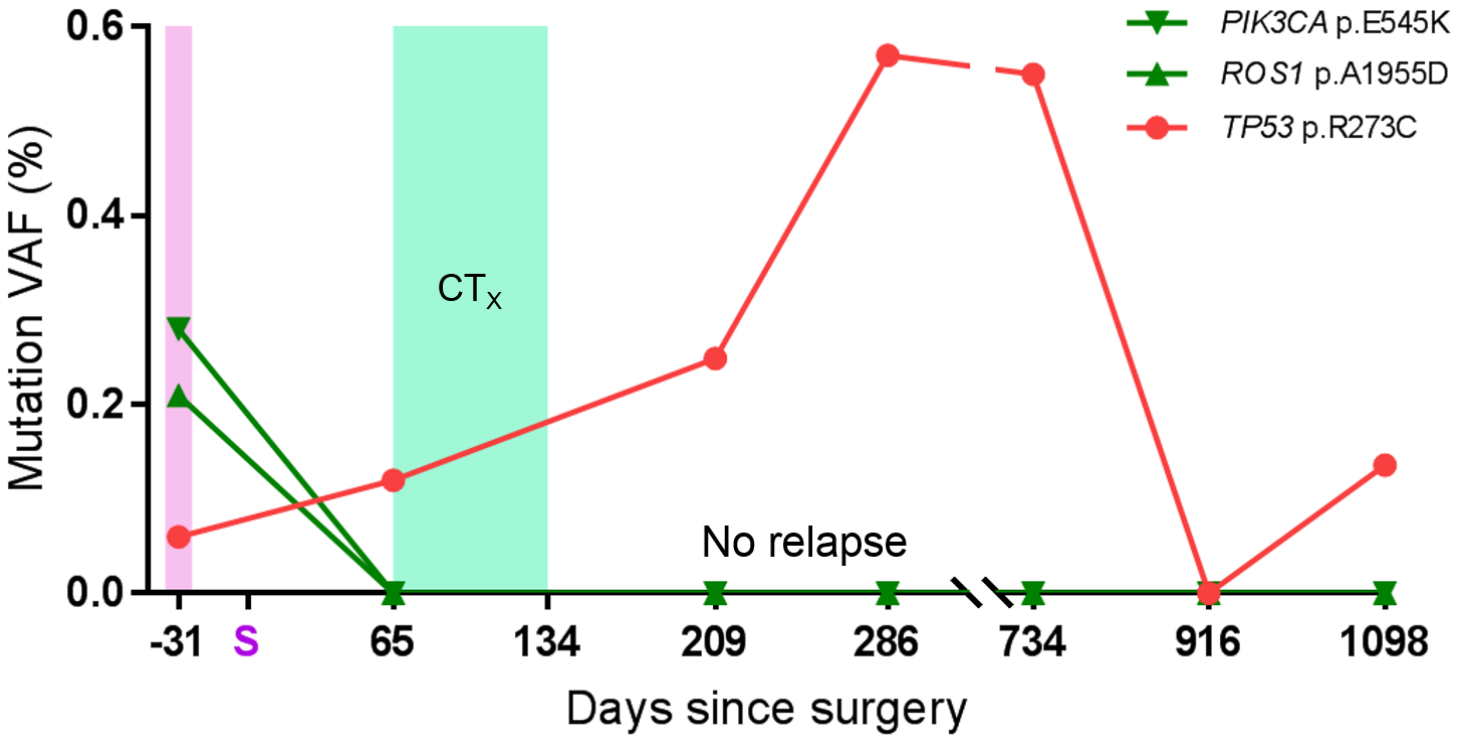


Figure 3.16 - Longitudinal cfDNA analysis of Patient 110. A total seven cfDNA samples were analysed using the OLcfDNA assay, where tumour-confirmed variants *PIK3CA* p.E545K and *ROS1* p.A1955D (green triangles) and a discordant variant, *TP53* p.R273C (red circles) were identified in pre-surgical cfDNA (purple shading). In contrast to the tumour-confirmed variants that are not detected following surgery/adjuvant chemotherapy (CT_X; green shading), the TP53 variant is found in 5/6 follow-up samples at varying VAFs. The profiles of the tumour-confirmed variants is consistent with the positive outcome observed in this patient (no relapse), however the fluctuation of the discordant variant warranted further investigation.

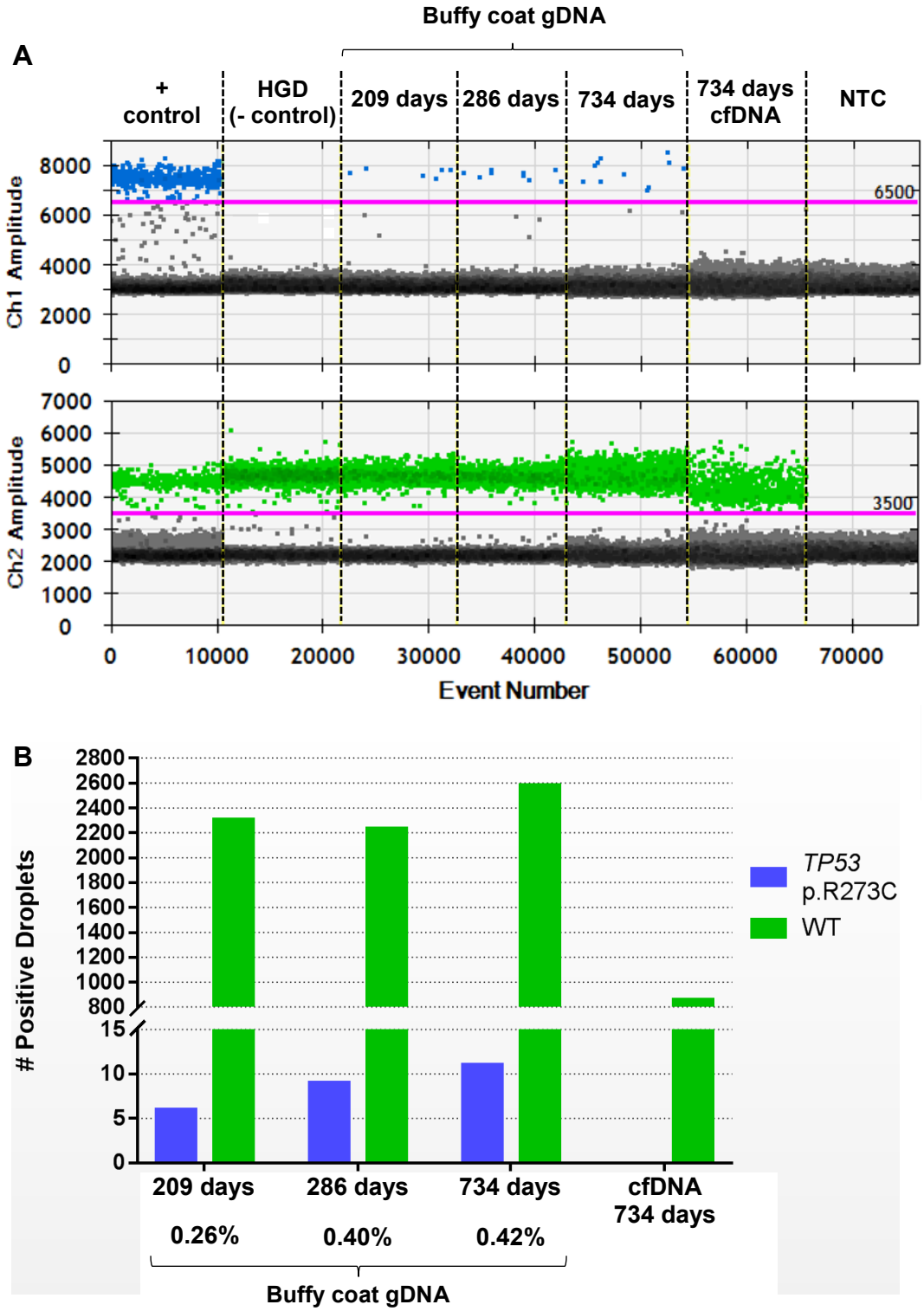


Figure 3.17 - Longitudinal evidence of CHIP. ddPCR targeting a *TP53* p.R273C mutation in Patient 110 in three longitudinal buffy coat DNA samples (240, 317 and 765 days) and a cfDNA sample (765 days). A) 1D amplitudes of mutant and WT channels, showing detection of mutant droplets in all three buffy coat DNA samples, but not in cfDNA (lack of input DNA). B) Bar chart showing number of droplets for mutant and WT for serial buffy coat samples and a cfDNA sample. VAFs are displayed below.+ control, tumour DNA sample with known *TP53* p.R273C from NGS (source, Mark Openshaw); HGD, human genomic DNA; NTC, No template control.

3.4 Discussion

3.4.1 Pre-surgical ctDNA detection in early-stage NSCLC

A significant proportion of patients with NSCLC (~40%) present at an advanced stage, where curative treatment is not possible (Zappa and Mousa 2016). cfDNA analysis in advanced NSCLC is well-characterised and is being utilised in the monitoring of treatments such as the detection of *EGFR* mutations associated with resistance to TKI therapy (e.g. T790M). However, when this PhD commenced, little had been found regarding the characterisation of early-stage NSCLC with regards to cfDNA analysis. Early-stage NSCLC has more favourable outcomes, yet the majority of patients ultimately relapse, meaning the use of minimally-invasive methods to both detect disease earlier and direct treatment is warranted (Heitzer and Perakis 2017, Liang *et al.* 2018).

This chapter has characterised ctDNA in early-stage NSCLC within the confines of a commercially available NGS panel (OLcfDNA) in patients that received curative (radical) treatment as part of the TRACERx study. Using multiregional tumour WES as a reference, the representation of tumour-confirmed, clonal variants within patient pre-surgical cfDNA was made possible. These clonal variants represent those that have arisen early during NSCLC evolution and are typically present in the majority, if not all tumour cells. Hence these are likely to represent the most robust targets for ctDNA detection and subsequent longitudinal profiling. The majority of studies focusing on cfDNA in NSCLC typically use the analysis of a single region of tumour tissue, be that a biopsy or an area of resected tumour as a reference point (Couraud *et al.* 2014, Guo *et al.* 2016). However this method presumes that all identified variants are equally valuable, not considering the impact of intratumour heterogeneity on findings.

The bespoke approach conducted phylogenetic ctDNA profiling on pre-surgical plasma from 100 TRACERx patients, targeting a median of 11 clonal variants (Abbosh *et al.* 2017). However, there was a huge disparity between patients with LUSC and LUAD histology in terms of ctDNA detection status (97% vs. 19% detected). The OLcfDNA was subsequently used as validation of the bespoke method, where both approaches showed a high degree of concordance regarding

variant detection status (92.5%). The OLcfDNA could not detect ctDNA in all (20/20) of the “bespoke non-detected” NSCLCs; 19 were LUADs, with the single “bespoke non-detected” LUSC (252) in the 100 patients also negative. These results indicate that the bespoke method had sufficient sensitivity and it is likely that the lack of LUAD detection in plasma is due to biological factors associated with this subtype. Indeed, a potential contributor could be a lower frequency of necrosis typically observed in LUAD tumours; both the bespoke and OLcfDNA approaches typically detected more necrotic NSCLCs. Nevertheless, by multivariable analyses, pathological necrosis was not determined as an independent predictor of ctDNA release in the published paper, with non-adenocarcinoma histology being a more significant indicator. Other independent clinicopathological predictors of ctDNA detection were a high Ki67 proliferation index and lymphovascular invasion (Abbosh *et al.* 2017).

When stage was considered, the OLcfDNA could not detect a single stage I LUAD (0/14) in pre-surgical blood. Considering LUAD now represents the most common subtype of NSCLC, and that stage I disease has much more favourable outcomes than other stages, highlights a significant issue with regards to the early detection of NSCLC using current cfDNA technologies (Yang 2009). For patients with LUSC histology, early detection of disease *via* cfDNA analysis may be more promising. Although no stage III LUSCs were analysed by the OLcfDNA, it seems likely that ctDNA detection in this group will be widespread, considering all four of the stage II LUSC analysed were ctDNA-positive and these were comfortably above the assay LOD (range 0.19% - 0.46% VAF). The only LUSCs that were not detected were from stage I tumours (3/7 detected), although this is likely due to the lack of tumour variants covered by the OLcfDNA (only 1 variant covered in each of the 4 negative LUSC patients) and not necessarily due to a the lack of ctDNA in the blood of these patients. Take the bespoke approach as an indicator, where 16/17 stage I LUSCs had 2 ≥ tumour variants detected within cfDNA (Abbosh *et al.* 2017).

Another characteristic typically coinciding with earlier stage disease are smaller tumours. It is clear that tumour size provides further difficulty with regards to ctDNA-positivity. Although this study had a smaller cohort than the one using the

bespoke method, unsurprisingly there was an association between ctDNA detection and tumour size. For patients where CT volumetric data was available (9/13 ctDNA positive patients), the smallest tumour that had ctDNA detected using the OLcfDNA was 6.64 cm^3 (Patient 111, 0.09% plasma VAF). The smallest tumour the bespoke approach detected (≥ 2 variants detected) was 1.02 cm^3 (Abbosh *et al.* 2017). Considering low-dose CT screening can identify nodules from around 3 or 4 mm in diameter, which if spherical would equate to a volume of between 0.014 cm^3 and 0.034 cm^3, shows that the use of this assay and others to help identify smaller tumours may be hindered (Aberle *et al.* 2011).

The advancements in CT screening with not only regards to tumour detection but also the deciphering of molecular phenotypes using artificial intelligence algorithms, could surpass cfDNA analysis with regards to early diagnosis (Rios Velazquez *et al.* 2017). It seems unlikely however, that imaging will be able to provide information on precise nucleotide changes, for example deciphering a *KRAS* p.G12C from a *KRAS* G12V mutation. The requirement to identify and monitor the presence of specific mutations is likely to only increase, as mutation-specific drug design is becoming more mainstream. In the case of *KRAS*, G12C-specific inhibitors have been developed and are currently undergoing *in vitro* and *in vivo* testing (Janes *et al.* 2018). Hence, the use of cfDNA analysis in measuring the prolonged success of these therapies is something imaging would be unable to ascertain.

3.4.2 Longitudinal monitoring of NSCLC using cfDNA

A key advantage to cfDNA analysis for treatment monitoring is rooted in its minimally-invasive and cost-effective qualities, while providing evidence of MRD, response to treatment and impending relapse. This chapter followed a number of TRACERx patients recruited at Leicester, assessing ctDNA within plasma prior and following curative surgery.

The detection of tumour-confirmed variants following surgery was strongly associated with impending relapse in these patients, with 4/5 patients (exception Patient 429) having ctDNA detected post-operatively who experienced this outcome. For three of these patients, this was in advance of clinical relapse, with lead times falling between 3 months (215) and 5 months (428) prior to clinical presentation. These cases exemplify the benefit of cfDNA analysis, as they may allow for a more proactive approach to be taken regarding patient treatment, should such treatment be available. In cases of local recurrence, the early detection of these variants may represent a crucial treatment window to locate residual disease and treat it accordingly, potentially with curative intent. For 2/3 of these cases, the mutation detected was in the currently "undruggable" *TP53* (Kastenhuber and Lowe 2017), and therefore there would be no direct therapeutic benefit for these patients, however the detection of such variants could provide cues to clinicians for the administering of more rigorous adjuvant regimens.

For cases like Patient 428, where an *EGFR* exon 20 insertion was found at every time point, it begs the question whether targeted treatment may be appropriate in the adjuvant setting. Despite radical surgery and a short period of adjuvant chemotherapy, this *EGFR* variant was maintained in plasma, indicating both a failure to eradicate disease and potentially an innate resistance chemotherapy of this clone. The clinical use of cfDNA analysis therefore could be vital in the switching to other therapies that could bring better patient outcomes. Although there are EGFR TKI therapies available, exon 20 insertions are known to confer resistance to such treatment. There are currently no approved therapies for this type of mutation, however inhibitors are showing promising preclinical efficacy (Hasako *et al.* 2018). Therefore in the future, the use of cfDNA for the direction to more efficacious modalities could be realised.

Detection of clinical relapse was not true of every case; Patients 77 and 429 had less consistent findings. For 77, the detection of relapse *via* ctDNA was not possible until after local clinical recurrence, and therefore other factors must be considered to explain this delay in detection. As most patients will have relapse detected during a routine follow-up (e.g. one year, two year etc.), coincidental detection of this relapse may have occurred whilst residual disease was less substantial and hence less likely to be at detectable levels in plasma. It would be interesting to compare the tumour size of recurrences between these patients, to characterise ctDNA detection status further.

Meanwhile in Patient 429, the lack of representation of a clonal driver mutation at any time point despite relapse was unusual. Considering the OLcfDNA assay only targeted a single variant (*KRAS* p.G12C), it would be useful to see whether a bespoke approach targeting several variants would provide different findings. A potential explanation regarding the delay/lack of ctDNA detection in 77 and 429 could be related to the local nature of each clinical recurrence; in longitudinal ctDNA profiling conducted in the publication (different patients to 77 and 429), there was a reduction in lead time for patients that presented with solely loco-regional recurrences and therefore the OLcfDNA could be reflecting these findings (Abbosh *et al.* 2017). This agrees with other studies, where it has been shown that M1b (extrathoracic) metastases are much more commonly detected in plasma than in patients with those with M1a/M0 disease (Karlovich *et al.* 2016).

As well as characterising relapse, the lack of detection of ctDNA post-surgery in 110 and 309 was consistent with the positive outcomes observed. Thus far, these patients have not relapsed 3.5 and 3 years following surgery respectively and therefore are in long-term remission, if not cured of their disease. Key questions to address in these patients is the time at which blood sampling and associated cfDNA analysis is no longer required, and if these patients do subsequently relapse, what are the explanations for this long-term dormant state in the blood.

Altogether, these findings suggest that ctDNA can be used to inform success following surgery, emerging resistance mechanisms and potentially identify personalised therapy for patients to maximise positive outcomes.

3.4.3 Methodological limitations regarding ctDNA detection

There is no doubt that a bespoke, patient-tailored approach is the most efficient method for ctDNA detection, as a rigorously-designed panel can maximise its chances to identify tumour-derived DNA within plasma. For example, in the bespoke study cited in this chapter, a median of 11 clonal variants were targeted (Abbosh *et al.* 2017). In contrast, the OLcfDNA typically had only a single tumour-confirmed variant covered, meaning it is unsurprising that non-detection was more common using this assay. However, since the majority of NSCLCs analysed in this chapter that had undergone multiregional-WES (41/43) had ≥ 1 clonal variants covered by the OLcfDNA it is encouraging, especially considering these variants are the archetypal drivers of NSCLC. The detection of these variants are therefore likely to represent the most informative targets, rather than passenger variants that may not contribute to disease progression. It is these drivers after all that are the typical subjects of drug design and potential stratification methods.

One caveat regarding the high number of patients targeted by the OLcfDNA is that 6/16 (37.5%) of tumour-confirmed variants that were identified within pre-surgical cfDNA had to be found using manual analyses and were not automatically identified by the variant caller software. Targeted variants included not only “Hotspots” and “Novel” variants of the OLcfDNA assay, but also tumour-confirmed variants that fell within the confines of amplicons targeting these hotspot regions (“Covered”). If these six manually identified variants were excluded, three patients in the “detected” group would have been lost; a 23% drop in ctDNA-positive pre-surgical patients. It is therefore important to extend target/hotspot files to maximise automated ctDNA detection capacity, especially if clinical implemented; an environment where process and standardisation would be necessity. This issue is clearly not one confronted by a patient-tailored, bespoke approach, due to its conscious nature of variant targeting.

Two unavoidable limitations of a bespoke approach is cost and logistics within a clinical environment. The labour intensive nature of patient-tailored approaches, not only requiring prior WES data to inform panel design, but also the pooling of these assays to create efficient multiplexing seems far from being feasible on a large scale. A distinct advantage of generic panels such as the OLcfDNA assay

is that these can be used in a time-efficient manner and as previously mentioned, may provide details on the dominant driver clones present within each patient. It is hoped that these elegant tailored approaches can either create a knowledgebase that leads to more informed “generic” panel design, or that automation of the process becomes reality.

A key consideration before cfDNA is adopted to the clinic is to circumvent the presence of confounding variants that may distract rather than inform. This chapter has highlighted one of these factors; clonal haematopoiesis of indeterminate potential (CHIP). This phenomenon is likely to be rife in NSCLC patients, as this is not only a disease of older age, but one with a highly mutagenic aetiology (smoking). Studies have shown that smoking has an association with CHIP, and so the analysis of cfDNA in NSCLC needs to account for these variants (Acuna-Hidalgo *et al.* 2017, Genovese *et al.* 2015, Hu *et al.* 2018). If not, then a likely scenario is that these discordant variants are incorrectly attributed to intratumour heterogeneity and accordingly impact treatment decisions. Such scenarios are more likely when using generic sequencing approaches (e.g. OLcfDNA) that do not intentionally target patient-specific variants, but rather a panel of hotspot genes commonly associated with a malignancy (i.e. NSCLC).

A simple way to account for CHIP is to sequence DNA from the buffy coat; a common practice in the cfDNA field. However, the cost of performing this has meant that increasingly the typical methodology is to perform WES on buffy coat DNA when analysing tumour tissue DNA (as a germline filter), and not alongside the deep sequencing of cfDNA. The problem with this methodology, is that CHIP is often found in the buffy coat in a minor fraction (<1%) and therefore WES will often miss these variants as typical WES depth is around 200X. Therefore it is worth considering performing targeted deep sequencing on a single buffy coat DNA sample for each patient to filter out these variants. With NGS technologies having increasingly high sequencing depths, the detection of CHIP variants is becoming more widespread (Acuna-Hidalgo *et al.* 2017), meaning they must be accounted for when assessing data.

Of course, there is plenty of evidence to suggest that CHIP may have an independent biological role of its own within patients. CHIP is associated with increased mortality and therefore there may be additional utility in identifying these variants (Jaiswal *et al.* 2014). It was not possible to characterise each of the tumour discordant variants found in this chapter as CHIP variants, however it is hoped that from the case study of Patient 110, our attention is raised.

3.5 Conclusions

The work in this chapter has highlighted both the potential benefits and the hurdles to overcome when analysing cfDNA in patients with early-stage NSCLC (I-IIIA). Here, it has been shown that when using a generic approach (OLcfDNA), the detection of ctDNA is generally impaired for stage I NSCLCs, particularly in patients with adenocarcinoma histology (LUAD). Therefore, despite this method representing a highly-sensitive NGS platform, the routine early detection of these patients will be difficult using current NGS technologies.

Patient-tailored approaches clearly provide an increased detection rate, as these intentionally target multiple variants found within tumours. However this is costly, logistically challenging and hence difficult to be clinically feasible. Perhaps results of these approaches can lead to the informed design of more comprehensive generic methods that satisfy both clinical and ctDNA sensitivity requirements.

Finally, the OLcfDNA has described how longitudinal ctDNA profiling can be used to determine the success of curative surgery, but also has uncovered potential confounders to cfDNA analysis (e.g. CHIP) that could distract rather than inform.

3.6 Future direction

Considering detection of SNVs in pre-surgical plasma was generally impaired in patients with stage I NSCLC, a future avenue to explore would be to investigate other genomic alterations, namely epigenetic changes such as methylation patterns, to determine whether these tumour-specific alterations can be detected within cfDNA from patients in this setting. This could increase the sensitivity of cfDNA-based analyses and thereby allow this analyte to compete with other screening methods (e.g. low dose CT screening) for early detection of NSCLC.

Chapter 4 Malignant pleural mesothelioma

4.1 Introduction

4.1.1 Malignant Pleural Mesothelioma

Mesothelioma is an aggressive tumour of mesothelium, predominantly those of the pleura and peritoneum. The most common mesothelioma is malignant pleural mesothelioma (MPM), accounting for approximately 80% of all mesothelioma cases worldwide. Although rare (~1% of all annual cancer diagnoses), MPM incidence has increased by around 67% since the early 1990s and is responsible for ~2,700 diagnoses each year in the UK alone (Figure 4.1).

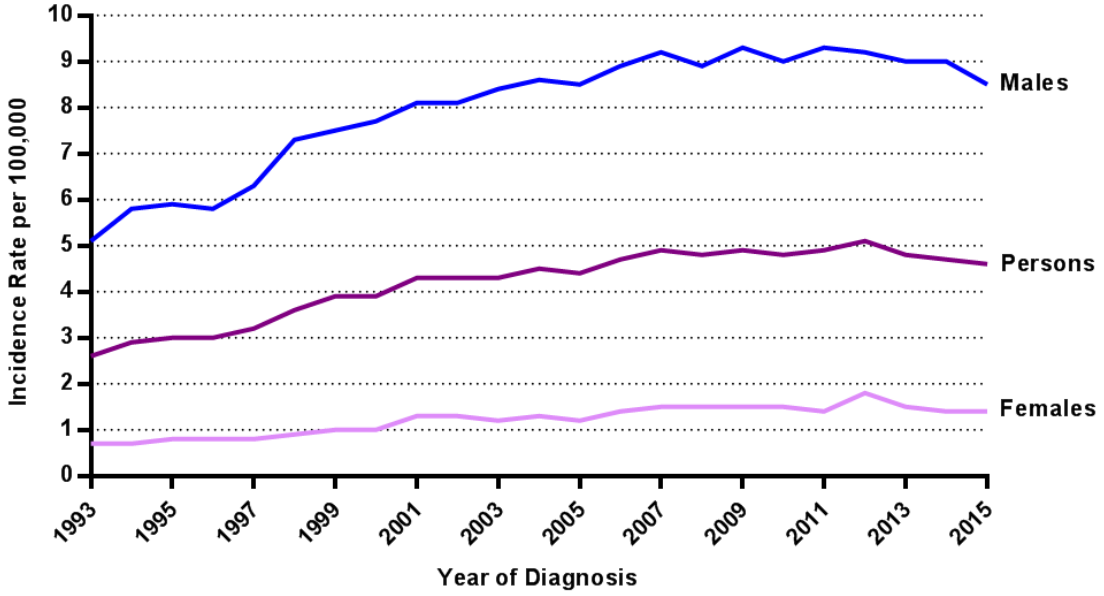


Figure 4.1 - Standardised incidence of mesothelioma, UK. Adapted from Cancer Research UK.

MPM is a disease strongly associated with exposure to asbestos, with over 80% of patients exhibiting some evidence of prior asbestos inhalation at diagnosis (Robinson 2012). The development of MPM is often characterised by a long latency period, often between 20 and 50 years post-exposure to asbestos (Frost 2013). Therefore, the sequence of events leading to MPM formation are highly enigmatic. Despite being banned in over 50 countries (including all EU countries), the mining of asbestos is still conducted worldwide with China, Russia and India being the largest global consumers (Schuz *et al.* 2013). Thus, this combination of long latency and continued asbestos mining means that MPM diagnoses will continue to persist for many decades to come.

One observation in MPM is that patients rarely show evidence of distant metastasis at diagnosis. Predominantly presenting around one lung, MPM often progresses through local invasion to the chest wall, pericardium, diaphragm and lungs (Servais *et al.* 2012). Because of this, MPM is often perceived as a locally aggressive disease, however a post mortem study found extrathoracic dissemination in 176/318 (55.4%) cases, with common sites of metastasis being the liver (31.9%), spleen (10.8%), thyroid (6.9%), and the brain (3%) (Finn *et al.* 2012). Interestingly, the precise cause of death was only identified in 19.8% of these cases, which may point towards physiological and metabolic causes of death.

The prognosis for patients with MPM is dismal, with median overall survival (OS) ranging between 8 – 36 months, depending on stage of presentation (Christoph and Eberhardt 2014).

4.1.2 Mechanisms of asbestos-induced carcinogenesis

Paradoxically, *in vitro* experiments have shown that asbestos fibres are extremely cytotoxic to mesothelium and do not alone induce the transformation of these cells, prompting research to focus on more nuanced causes of carcinogenesis (Liu *et al.* 2000). There are a number of potential mechanisms of asbestos-induced carcinogenesis that may all contribute to the strong causal link observed between asbestos exposure and MPM development.

Firstly, the “oxidative stress theory” stipulates that asbestos fibres stimulate cells to produce reactive oxygen species (ROS), which induces DNA damage ultimately leading to strand breaks in mesothelium (Heintz *et al.* 2010, Sekido 2013). Additionally, the inability of macrophage to digest asbestos fibres results in frustrated phagocytosis, creating further accumulation of free radicals, ultimately producing a pro-inflammatory environment.

Secondly, the “chromosome tangling theory” is based upon findings that asbestos fibres can penetrate mesothelial cells and disrupt chromosomes during cell division, leading to widespread genome instability. Such interactions would

explain the abundant chromosomal structural abnormalities and aneuploidy observed in MPM (Sekido 2013).

Thirdly, it has been shown that asbestos fibres can absorb proteins and chemicals which potentially lead to an accrual of hazardous molecules (e.g. carcinogens). Furthermore, this sequestering of cellular proteins can cause key deficiencies in otherwise healthy mesothelial cells (Sekido 2013).

Finally, mesothelium and macrophages incubated with asbestos fibres release of a variety of cytokines and growth factors, including tumour necrosis factor-alpha (TNF-α), interleukin-1β and transforming growth factor-β (Yang *et al.* 2006). In the case of the former, TNF-α release was shown by one study to lead to the inhibition of asbestos-induced cytotoxicity *via* NF-κB signalling. Such examples provide explanations for how the tumour microenvironment can provide a nourishing setting to support MPM growth.

Although asbestos regulation is becoming more widespread, there is concern that manmade materials, namely carbon nanomaterials may induce similar carcinogenic effects. Carbon nanotubes, particularly those with high aspect ratios are currently subject to *in vitro* and *in vivo* study, where remarkably similar effects to asbestos inhalation are being exhibited (Xu *et al.* 2014). Chernova and colleagues showed that these nanotubes replicate asbestos-induced MPM *via* disruption of tumour suppressor gene *CDKN2A* (Chernova *et al.* 2017). Considering nanomaterials are becoming increasingly adopted for building materials, medical devices and computer manufacturing, emerging evidence suggests that regulation of these materials is potentially warranted.

4.1.3 MPM Histology

There are three major histological MPM subtypes; epithelioid, sarcomatoid and biphasic, accounting for 60%, 20% and 20% of MPM respectively (Figure 4.2) (Inai 2008).

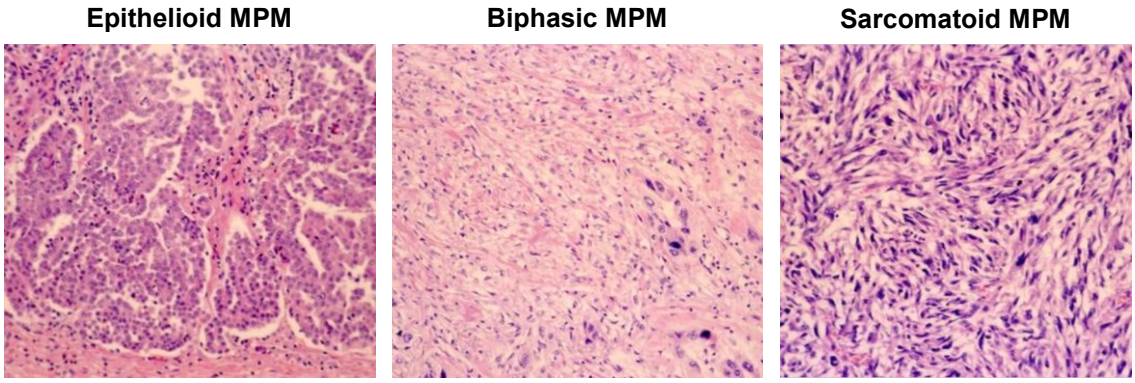


Figure 4.2 - Histological subtypes of MPM. H&E images showing Epithelioid MPM with a prominent papillo-tubular structure. Sarcomatoid MPM has spindle cell proliferation, alike sarcomas. Biphasic MPM shares both epithelioid and sarcomatoid MPM histology. Adapted from Inai, 2008.

Epithelioid MPM is not only the most common subtype, but also has the best prognosis. Studies have shown that median survival of epithelioid MPM was ~15 months, compared to just five months for sarcomatoid MPM (Balduyck *et al.* 2010, Meyerhoff *et al.* 2015). Interestingly, Meyerhoff and colleagues' data also suggested that histology is a much more robust prognostic indicator than stage, as stage within each histological subtype had no significant impact on survival (Meyerhoff *et al.* 2015). Epithelioid MPM displays a variety of histological patterns, some of which may have clinical and prognostic significance. Of these, a pleomorphic growth, when present in >10% tumour has a particularly poor prognosis with median survival being just eight months and thereby more resembling that seen in sarcomatoid MPM (Kadota *et al.* 2011, Ordonez 2012).

Sarcomatoid MPM has the poorest prognosis. Studies have also shown that patients with sarcomatoid MPM are often not applicable for surgical resection, as there is little evidence suggesting survival benefit, compared to the other two subtypes (Ricciardi *et al.* 2018). The diagnosis of sarcomatoid MPM can be difficult using routine methods. As a common symptom of MPM is pleural

effusion, cytology is often conducted by aspirating pleural fluid for diagnosis. However this practice can lead to misleading conclusions; epithelioid MPM often sheds cells into fluid, whereas the same is often not true for sarcomatoid disease (van Zandwijk *et al.* 2013).

Biphasic MPM contains a mixture of epithelioid and sarcomatoid histology within the same tumour (Ali *et al.* 2018). To be diagnosed as biphasic, there must be at least 10% of each pattern (epithelioid and sarcomatoid) within the neoplasm; for cases where there is less of either, these are usually designated as predominantly epithelioid or sarcomatoid MPM (Travis *et al.* 2015).

4.1.4 MPM Staging

Like the majority of solid tumours, MPM is staged using the TNM system, however the tailoring of this system to describe the progression of this disease has been extremely challenging. In particular, the T-component represents a significant staging issue, due to the highly unusual growth pattern that is observed in MPM. Unlike the majority of tumours which tend to develop concentrically, MPM develops as a “rind” around the pleural cavity, meaning an accurate assessment of tumour burden has been challenging and has historically been largely descriptive (Nowak *et al.* 2016).

To address this issue, the International Association for the Study of Lung Cancer (IASLC) and International Mesothelioma Interest Group (IMIG) created an international database, leading to a data-driven revision of the MPM staging system (Rusch *et al.* 2016). This edition (8^{th} edition) is now the current consensus on MPM staging and does show some, albeit minor prognostic significance (Tables 4.1, 4.2 and 4.3). Of note, recent findings have suggested that the measurement of pleural thickness or using computed tomography–based calculations of tumour volume may be a more accurate representation of the T-category for any subsequent revisions.

T	Tumour
T1	Tumour involving ipsilateral parietal or visceral pleura only
T2	Tumour involving ipsilateral pleura (parietal or visceral pleura) with invasion involving at least one of the following: • diaphragmatic muscle • pulmonary parenchyma
T3[1]	Tumour involving ipsilateral pleura (parietal or visceral pleura) with invasion involving at least one of the following: • endothoracic fascia • mediastinal fat • chest wall, with or without associated rib destruction (solitary, resectable) • pericardium (non-transmural invasion)
T4[2]	Tumour involving ipsilateral pleura (parietal or visceral pleura) with invasion involving at least one of the following: • chest wall, with or without associated rib destruction (diffuse or multifocal, unresectable) • peritoneum (via direct transdiaphragmatic extension) • contralateral pleura • mediastinal organs (oesophagus, trachea, heart, great vessels) • vertebra, neuroforamen, spinal cord or brachial plexus • pericardium (transmural invasion with or without a pericardial effusion)
N	**Nodes**
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastases
N1	Metastases to ipsilateral intrathoracic lymph nodes (includes ipsilateral bronchopulmonary, hilar, subcarinal, paratracheal, aortopulmonary, paraoesophageal, peridiaphragmatic, pericardial, intercostal and internal mammary nodes)
N2	Metastases to contralateral intrathoracic lymph nodes. Metastases to ipsilateral or contralateral supraclavicular lymph nodes
M	**Metastasis**
M0	No distant metastasis
M1	Distant metastasis present

Table 4.1 - TNM definitions (8th Edition). [1]T3 describes locally advanced, but potentially resectable tumour. [2]T4 describes locally advanced, technically unresectable tumour. Adapted from Nowak *et al*., 2016.

Stage	T	N	M
IA	T1	N0	M0
IB	T2, T3	N0	M0
II	T1, T2	N1	M0
IIIA	T3	N1	M0
IIIB	T1, T2, T3	N2	M0
	T4	N0, N1, N2	M0
IV	Any T	Any N	M1

Table 4.2 - TNM stage groupings (8th Edition). Adapted from Rusch *et al*., 2016.

Tumour Stage	Two-year survival (%)	Five-year survival (%)
T1	44	17
T2	40	13
T3	40	13
T4	28	3

Table 4.3 - MPM pathological stage and overall survival based on T-category (8th Edition). Taken from Nowak *et al*., 2016**.**

4.1.5 MPM Treatment: One size fits all

Treatment of MPM has remained relatively stagnant for the past 15 years, with cytotoxic chemotherapy being the main option for patients (Nowak 2012). This lack of progress is symptomatic of an absence of understanding the genomics, transcriptomics and proteomics underlying MPM biology. The long latency period (20-50 years) of MPM, along with its relatively unspecific symptoms (e.g. shortness of breath, coughing, chest pain), means that diagnosis of often occurs at an advanced stage, meaning curative treatment (surgery) is exceedingly rare.

First line chemotherapy usually involves a combination of cisplatin or carboplatin with the anti-folate pemetrexed, the latter of which was adopted following the results of the 'Emphacis' trial, where the addition of pemetrexed provided an further 3-month survival benefit compared to that of cisplatin alone (Vogelzang *et al.* 2003). Gemcitabine and vinorelbine are sometimes administered as second-line therapies, following invariable first-line resistance. The efficacy of gemcitabine and vinorelbine is generally considered poor, however a lack of randomised clinical trials and biomarkers for drug response, has meant patient stratification is currently deficient (Stahel *et al.* 2010, Zauderer *et al.* 2014).

Like many cancers in the advanced setting, MPM is currently subject to immunotherapy clinical trials. Pembroluzumab was shown in a Phase I/II trial to have a response rate of 20%, with a stable disease rate of 52% (Alley *et al.* 2017). Nivolumab is now being studied in the Cancer Research UK Phase III CONFIRM trial (NCT03063450) in patients with relapsed MPM (Fennell *et al.* 2018).

4.1.5.1 Surgery in MPM

Surgery is used as both a curative measure and as a de-bulking treatment for MPM, however its practice is controversial, largely due to some studies showing the risk-benefit ratio being marginal (Kaufman and Flores 2011). Two main procedures are currently practiced; extended pleurectomy/decortication (EPD) and extrapleural pneumonectomy (EPP). EPD involves the resection of parietal and visceral pleura, in addition to the hemidiaphragm +/- pericardium (Figure 4.3).

A more radical surgery, EPP additionally involves the resection of one lung (Rice 2012). The practice of EPP has generally decreased following the results of the

Mesothelioma and Radical Surgery (MARS) trial; a randomised feasibility study comparing EPP & chemotherapy vs. chemotherapy alone in MPM. MARS summarised that it was feasible to randomly assign patients to surgery/no surgery, however the high morbidity associated with EPP, meant a larger study was not feasible (Treasure *et al.* 2011). Their findings suggested that EPP "offers no benefit and possibly harms patients". Still, advocates of EPP maintain the rationale that it has improved curative potential over EPD and/or chemotherapy.

Currently the efficacy of EPD is subject to testing in a follow-on trial, MARS2; EPD & chemotherapy vs. chemotherapy alone (ClinicalTrials.gov Identifier: NCT02040272). Providing its success, phase III clinical trials are expected to follow. Overall, multimodality therapy (EPD + Chemotherapy) has been shown to provide best overall survival from meta-analyses (~16 months) (Cao *et al.* 2014).

It is clear that while many malignancies have entered into an era of targeted therapy, MPM has been neglected. A lack of exploitation of the common genomic events observed in the disease has certainly hindered progress, requiring comprehensive analyses to further characterise the disease (Bronte *et al.* 2016).

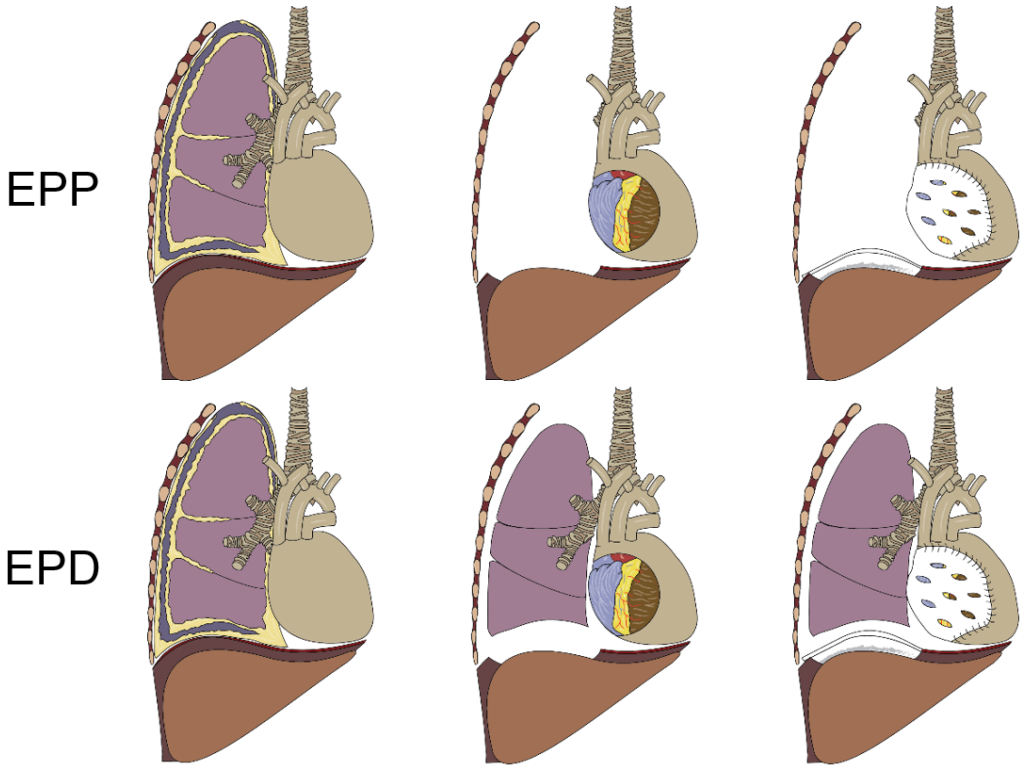


Figure 4.3 - Schematic describing the difference between EPP and EPD. EPP involves resection of the affected lung, diaphragm and pericardium. Extended pleurectomy on the other hand preserves the affected lung, but removes other structures. Adapted from Rice, 2012.

4.1.6 Attempts to move into the targeted era: MPM Genomics

Although asbestos exposure is the main cause of MPM, the genomic architecture of the disease and the extent of intratumour heterogeneity is unclear. Unlike the majority of epithelial-derived malignancies (e.g. NSCLC), MPM does not typically exhibit mutations of well-established proto-oncogenes such as *EGFR*, *KRAS* and *BRAF* (Bueno *et al.* 2016). Instead, previous studies have almost exclusively observed the loss of three tumour suppressor genes as the main genetic features of the disease; *CDKN2A*, *BAP1* and *NF2* (Guo *et al.* 2015).

Apart from these three genes, studies have shown wide variation with regards to other candidate drivers; a factor likely due to comparatively low sample sizes in MPM research. Even with *BAP1* and *NF2* there is often conflicting data regarding the frequencies of each within MPM, as both exhibit somatic copy number losses, as well as inactivating mutations and gene fusions, meaning different analytical methods underrepresent the genetic alterations present among large MPM patient populations. One of the largest genomic analyses in MPM was conducted by Bueno and colleagues, which uncovered 10 significantly-mutated genes in MPM, including *TP53*, *SETD2* and *DDX3X* (Bueno *et al.* 2016). Table 4.4 summarises the most commonly altered genes in MPM.

Gene	Alteration type	Frequency (%)	References
CDKN2A	Copy number loss	60 - 72	(Bott *et al.* 2011, Illei *et al.* 2003)
BAP1	Somatic mutations, gene fusions and/or copy number loss	22 – 61	(Bott *et al.* 2011, Bueno *et al.* 2016, Yoshikawa *et al.* 2012)
NF2	Somatic mutations, gene fusions and/or copy number loss	30 - 50	(Bott *et al.* 2011, Petrilli and Fernandez-Valle 2015, Thurneysen *et al.* 2009)
TERT	Somatic mutations and/or copy number gain.	15 - 27	(Hylebos *et al.* 2017, Tallet *et al.* 2014)
LATS2	Somatic mutations	4 – 22	(Bott *et al.* 2011, Murakami *et al.* 2011)
SETD2	Somatic mutations, gene fusions and/or copy number loss	~8	(Bueno *et al.* 2016)
TP53	Somatic mutations	6 - 8	(Bott *et al.* 2011, Bueno *et al.* 2016, Guo *et al.* 2015)
DDX3X	Somatic mutations	4	(Bueno *et al.* 2016)
TRAF7	Somatic mutations	2	(Bueno *et al.* 2016)

Table 4.4 - The most common genomic alterations in MPM. Green shading highlights the genes which the most comprehensive evidence of involvement in MPM.

4.1.6.1 *CDKN2A*

Cyclin-dependent kinase inhibitor 2A (*CDKN2A*), a gene encoding two tumour suppressor proteins (p16 and $p14^{ARF}$), is the most commonly deleted gene in MPM; found in over 60% of patients (Bott *et al.* 2011). p16 normally functions as an inhibitor of CDK4 and CDK6, which in turn restricts the progression of the cell cycle from G1 to S phase. Alternatively, $p14^{ARF}$ is induced by increased mitogenic signalling (e.g. MYC), where it forms stable complexes with Mdm2, thereby facilitating p53-dependent cell cycle arrest, as well as inhibition of ribosome biogenesis (Sekido 2013).

Homozygous deletion of *CDKN2A* at the 9p21 locus, therefore results in an impaired ability to control proliferation, leading to the 'immortal' cancer phenotype. *CDKN2A* deletion is associated with a poorer prognosis in patients (Hylebos *et al.* 2017). As the most prominent genetic alteration in MPM, the exploitation of *CDKN2A* deletion through a synthetic lethal mechanism could become a key therapeutic option for MPM treatment. CDK4/CDK6 inhibitors such as palbociclib (PD-033299), abemaciclib (LY2853219) and ribociclib (LEE 011) would represent good candidates in MPM. These have all been FDA-approved for the treatment of oestrogen receptor-positive advanced breast cancer (Finn *et al.* 2015). Palbociclib has recently been shown to have synergistic antitumour affects when combined with PI3K inhibitors in MPM cells. Such studies provide optimism that exploiting *CDKN2A* loss in MPM may be subject to clinical trials in the coming years (Bonelli *et al.* 2017).

4.1.6.2 *BAP1*

A member of the polycomb-group proteins (PgG), BRCA1 associated protein-1 (*BAP1*) encodes a nuclear deubiquitinase that has diverse role in many cellular processes, including transcriptional repression and cell cycle progression. Amongst other roles, BAP1 modulates the E3 ubiquitin ligase activity of the BRCA1-BARD1 complex, initially postulating a role in the regulation of DNA damage repair (DDR) (Carbone *et al.* 2013). However, *in vitro* experiments have indicated *BAP1* mutation/loss is not strongly associated with this assumption. Additionally, *BAP1*-mutant cell lines were not shown to be increasingly sensitive to PARP inhibition, compared to that of wild-type *BAP1* cell lines (Bott *et al.* 2011). Therefore, loss of functional BAP1 is likely to contribute to MPM *via* multiple mechanisms, including the induction of a more permissive G1/S checkpoint and the deregulation of histone ubiquitination patterns.

There is evidence of both somatic and germline loss of the tumour suppressor BAP1 in MPM. Germline mutations account for a predisposition syndrome observed in melanocytic skin cancer, in addition to MPM. Two families studied with a history of MPM (despite only modest asbestos exposure), were shown to exhibit truncating germline mutations in the C-terminus of *BAP1*, particularly affecting the nuclear localization sequence (NLS) (Testa *et al.* 2011). Loss of the NLS has been shown to radically impair BAP1 function, emphasising its deubiquitinating activity within the nucleus is critical in its role as a tumour suppressor (Testa *et al.* 2011, Ventii *et al.* 2008). Recently it was shown that germline *BAP1* mutations induce a Warburg effect, meaning this phenomenon is not necessarily an adaptive process, but one that may predate and potentially facilitate carcinogenesis (Bononi *et al.* 2017). The true prevalence of germline loss of *BAP1* is currently unknown due to a lack of large GWAS data due to the relatively low incidence of MPM.

Somatic *BAP1* mutations have also been observed in MPM, with one study showing a frequency of 23%. When coupled with cases of recurrent 3p21.1 loss (containing *BAP1*) observed in sporadic MPM, the frequency of functional *BAP1*

loss may be over 40% (Bott *et al.* 2011). The complex role of *BAP1* in MPM undoubtedly warrants an array of multidisciplinary study.

4.1.6.3 *NF2*

The neurofibromatosis type 2 (*NF2*) gene encodes a tumour suppressor protein known as merlin. Originally identified in its namesake autosomal dominant disorder, the loss of function of *NF2* has been implicated in many cancers, predominantly mesothelioma (pleural and peritoneal) and glioma. Merlin is a scaffolding protein involved in cell-cell adhesion, where it contributes to maintained contact inhibition of cell growth. Unlike most tumour suppressors, merlin does not directly influence the cell cycle, but rather acts a link between cell signaling and the extracellular matrix (Petrilli and Fernandez-Valle 2015). Nonetheless, the precise role by which merlin suppresses tumour growth is unrefined.

Loss of *NF2* has been shown to potentiate mitogenic signaling, including the HER1/2, mTOR, ERK, and FAK pathways (Cooper and Giancotti 2014). Initially, this prompted the belief that merlin usually functions as a negative regulator by inhibiting cell surface receptors. However, recent studies have revealed the closed (active) form of merlin is not enriched at the cell membrane, but rather inside the nucleus (Ladanyi *et al.* 2012). Merlin accumulates and inhibits the E3 ubiquitin ligase CRL4-DCAF1 in the nucleus, thereby reducing the ubiquitination of transcription factors and chromatin remodeling enzymes. This finding has alluded to possible parallels to the mechanisms described of *BAP1* in MPM (Li *et al.* 2010). It questions whether combined *NF2* and *BAP1* loss are synergistic in their effects by introducing vast epigenetic deregulation in MPM.

Approximately 35-40% of all MPM cases exhibit loss of *NF2*, either through deletion at the 22q12 locus or *via* truncating mutations (Petrilli and Fernandez-Valle 2015). *NF2* mutations have been found to be most frequent in sarcomatoid MPM and least in epithelioid MPM (Bueno *et al.* 2016, Sato and Sekido 2018).

MPM cell lines and tumour xenograft models lacking *NF2* were shown to be more sensitive to FAK inhibition. This seemingly synthetic lethal relationship is expected to occur due to an increased dependence on cell-ECM-induced FAK signaling, to compensate for weaker cell-cell adhesion (Shapiro *et al.* 2014). Furthermore, the FAK inhibitor, defactinib (VS-6063) seems to preferentially eliminate 'stem-like' MPM cells, in great contrast to the current chemotherapy agents, cisplatin and pemtrexed (Sulzmaier *et al.* 2014). In light of these findings, defactinib was subject to phase II clinical trial (COMMAND; NCT01870609), however as a single agent promising pre-clinical findings were not translated clinically. Defactinib is now subject to a trial as a combination therapy with the PD-1 monoclonal antibody, pembrolizumab (FAK-PD1; NCT02758587) in multiple solid tumours including MPM. Other protein interactions with merlin are also being explored, with the scaffold protein *KSR1,* being found to be a novel target in merlin-deficient tumours (Zhou *et al.* 2016).

4.1.7 Circulating Biomarkers in MPM

The search for effective and accurate biomarkers for MPM has been highly problematic. The three genes previously discussed (*CDKN2A*, *BAP1* and *NF2*) are prevalent in MPM, however their presence with regards to patient stratification is still seeking validation. With recent development in drugs with potential efficacy, whether MPM can be efficiently monitored over the course of their disease is still unknown.

The identification of MPM biomarkers in the blood has been visited. The glycoprotein, mesothelin is highly expressed on the cell surface of some cancer cells, notably MPM, but also pancreatic and ovarian adenocarcinomas. Robinson and colleagues demonstrated that the three isoforms of mesothelin, known collectively as soluble mesothelin-related proteins (SMRP), are found at elevated levels in the blood of MPM patients. The detection of SMRP was shown to provide 83% sensitivity and 95% specificity for MPM, over other malignancies (Robinson *et al.* 2005). However, the use SMRP in studying the clonal evolution of MPM is not applicable, as it unable of elucidating genetic heterogeneity of the disease.

Still, mesothelin is the most successful blood-based biomarker for MPM at present, reflected by its FDA-approval in the US (Creaney *et al.* 2015). It is predominantly used as a monitoring tool of disease response/progression, alongside traditional means.

The analysis of mutations in cells obtained from pleural fluid and/or ctDNA analysis in MPM could eclipse mesothelin-based methods, particularly with regards to therapeutic intervention and patient stratification. As it is likely that targeted therapies will be developed to target specific genomic alterations, the stratification of patients using blood-based analyses could potentially be achieved. However, at present the most basic of questions regarding ctDNA analysis are yet to be addressed in MPM. Providing ctDNA is detectable in MPM, larger studies aiming to inform the design of targeted NGS panels to routinely detect MPM-derived variants in cfDNA would be a next logical step. The apparent lack of recurrent “hotspot” variants in MPM mean that the detection of ctDNA may require a patient-tailored approach, at least in the short-term.

One potential technical problem for ctDNA detection in MPM is centered on the type of genomic alterations it aims to assess. As described, the vast majority of literature points towards the deletion of genes such as *CDKN2A*, *BAP1* and *NF2*, meaning these would be a priority for detection in plasma. Although somatic copy number gain has been shown to be detectable in cfDNA in a variety of different malignancies (Shaw *et al.* 2012, Ulz *et al.* 2016), the detection of copy number loss in plasma is more challenging using methods such as ddPCR. Due to the heavily restricted fold-change with deletion (versus amplification) combined with the high healthy DNA background “diluting” this fold change, in patients with early-stage disease (lower ctDNA fraction) ddPCR is incapable of such detection (Figure 4.4).

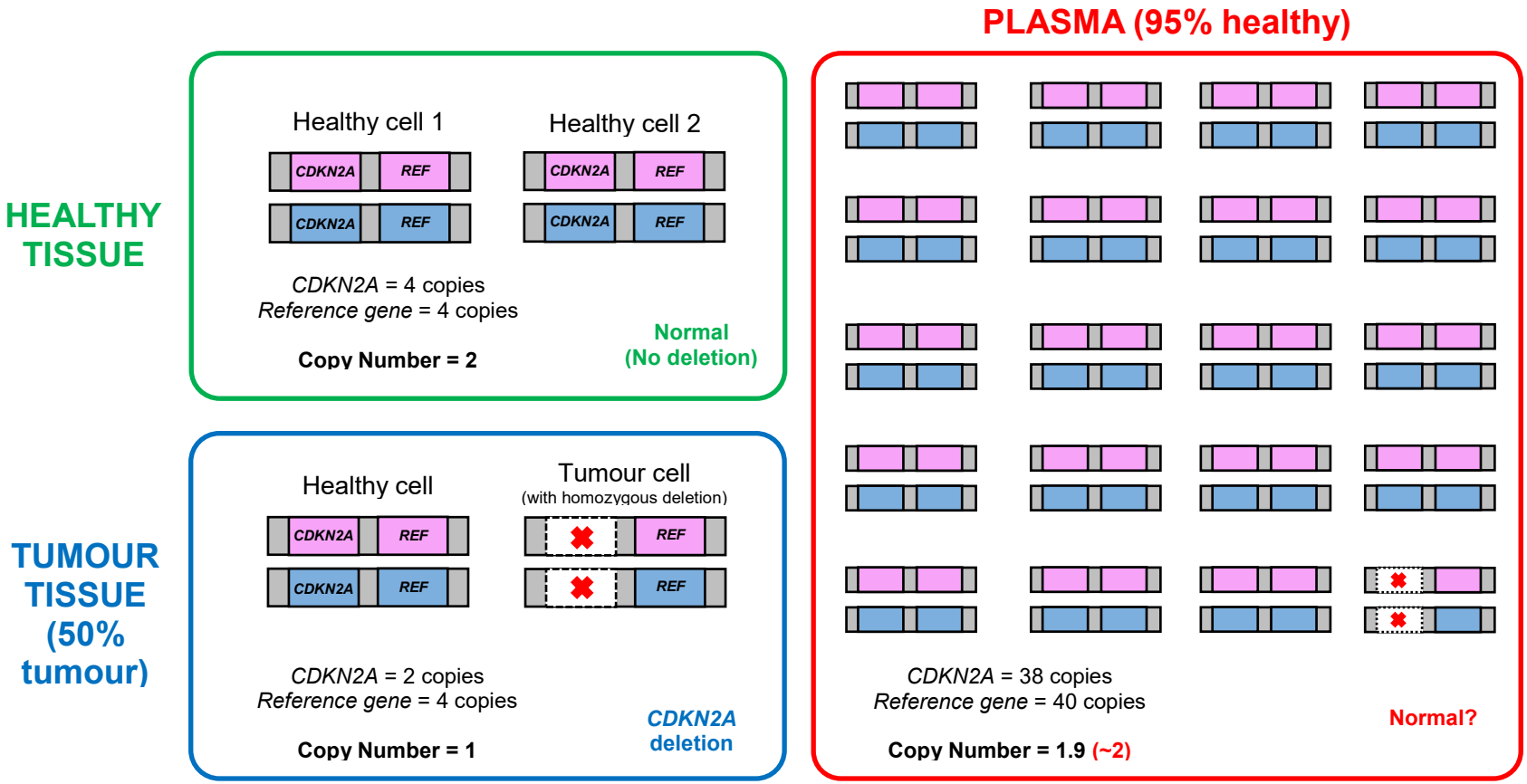


Figure 4.4 - The difficulty of somatic copy number loss detection using cfDNA analysis. Comparison of copy number loss of a target in different tissue types. With methods such as ddPCR, the copy number of *CDKN2A* would compared to that of a reference gene that does not typically exhibit copy number alteration (2 copies per cell, one maternal (pink), one paternal (blue)). A normal copy number (2) is exhibited in healthy tissue, while in a tumour tissue there will be a mixture of tumour cells (with *CDKN2A* deletion) and healthy cells (stroma etc.), providing a copy number of 1, indicating either heterozygous or homozygous *CDKN2A* loss. In this example, ctDNA constitutes 5% of total cfDNA within plasma, meaning that a copy number would be very similar to that of healthy tissue.

4.1.8 The MEDUSA Study

MEDUSA (Mesothelioma Evolution: Deciphering Druggable Somatic Alterations) aims to access intratumour heterogeneity and clonal evolution in MPM through phylogenetic analyses. Using similar methodologies to that of TRACERx, MEDUSA enrols patients who are applicable for extended pleurectomy/ decortication (EPD) as their first-line modality, allowing access to an abundance of fresh tumour material. Following EPD, multiregional tumour sections (five regions per patient) are isolated. One tumour region per patient is used to generate primary cell lines/explants. Fresh multiregional tumour tissue is subsequently frozen for later analyses. As of 09/2018, MEDUSA has tumour tissue stored for 108 patients (> 540 tumour regions), making it one of the largest known MPM tissue banks (recruitment ongoing).

In contrast to NSCLC which are usually spherical, compact tumours, MPM usually presents as a dense tumour mass enveloping ("rind") a single lung, meaning the selection of tumour regions requires a more systematic approach. The five tumour regions (R1-R5) were selected based upon anatomical location around the lung (Figure 4.5).

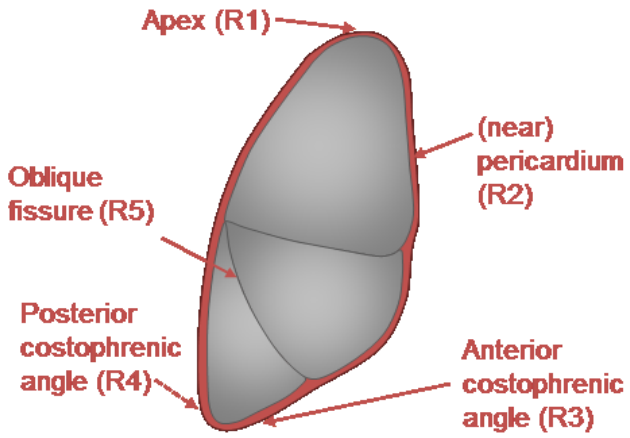


Figure 4.5 - Multiregional tumour tissue analysis. Approximate anatomical location of studied tumour regions, relative to the right lung. Each region is assigned an 'R' number. R2 represents either pericardium-adjacent tissue, or pericardial tissue, depending on whether this tissue was removed during EPD.

Following an assessment of tumour content *via* H&E staining analysis and the propensity of these tumours to establish primary cell lines, multiregional tumour samples were selected for whole exome sequencing (WES). This data is used to generate phylogenetic trees, to assess the extent of intratumour and interpatient heterogeneity in MPM. This represents an important starting point for subsequent analyses. Multiregional-WES will aim to shed light upon the evolutionary framework present in MPM and potentially uncover key early events (clonal) that could be exploited through the next generation of therapeutics.

As loss of heterozygosity (LOH) and homozygous deletions are common somatic alterations in MPM (e.g. *CDKN2A*, *BAP1* and *NF2*), accurate methods somatic copy number alteration (SCNA) are warranted. Therefore, OncoScan™ (Thermo Fisher Scientific) was used to assess genome-wide SNCA for a number of patient samples. MEDUSA will also infer SCNA from WES data to describe intratumour heterogeneity on both a SNV and copy number level.

MEDUSA also aims to assess the utility of cfDNA analysis in MPM. All patients prior to surgery will donate a blood sample, allowing for a blood-based assessment of each patient when they have their highest tumour burden. As late diagnosis of MPM is an all too frequent occurrence, it is hoped that preliminary cfDNA analysis when MPM tumours are *in situ* will ultimately provide the foundation for the development of blood-based detection methods to provide more favourable outcomes for patients with MPM. Following phylogeny being derived, variants found in every region of tumour tissue (clonal variants) will be targeted in cfDNA for optimal ctDNA detection. After all, the detection of clonal variants would be of particular use for the detection of minimal residual disease (MRD) post-surgery.

4.2 Aims and Objectives

The aims of this chapter were to:

i. determine the utility of ddPCR for the detection of *CDKN2A* deletion in MEDUSA patient tumours.
ii. determine whether ctDNA can be detected in MPM.
iii. investigate any potential prognostic significance of ctDNA status in MPM and whether cfDNA analysis could be used for patient stratification.

The objectives were to:

i. develop a ddPCR assay for the detection of copy number loss of *CDKN2A* and compare to the commercially-available OncoScan™ method;
ii. identify clonal variants in MPM tumours from multiregional-WES data to allow for assay design;
iii. develop patient-specific ddPCR assays targeting clonal variants in 11 MEDUSA patients;
iv. conduct ddPCR on MEDUSA patient cfDNA;
v. compile clinical data with ctDNA detection status.

4.3 Results

4.3.1 *CDKN2A* analysis in MPM tumour tissue

4.3.1.1 Detection of *CDKN2A* deletion in MEDUSA tumour tissue

As *CDKNA* deletion is the most common genomic event observed in MPM and has been shown to be associated with a poorer prognosis (Hylebos *et al.* 2017), efficient methods for the detection of this event are required as they could potentially aid in the stratification of patients for treatment in the future. Although somatic copy number gain and loss can be interpreted from WES data, the use of this method for the detection of *CDKN2A* loss is inefficient, and not likely to be applied in the clinical setting to all MPM patients. Therefore, it was appropriate to establish whether a ddPCR assay was capable of providing fast and accurate *CDKN2A* status on patient tumour material. From 10 patients within the MEDUSA study, a copy number assay targeting the *CDKN2A* gene (83bp) was multiplexed with a reference assay targeting *RPPH1* (62bp) on 5 ng multiregional tumour tissue DNA. The reference is assumed to have two copies per diploid genome, and hence provides a means for relative quantitation of copy number gain or loss. For oligonucleotide sequences for the assays, see Materials and Methods, section 2.4.1.1.

All tumour regions for the 10 patients were first assessed for approximate tumour cellularity (%) from H&E tissue sections by Dr David Moore (consultant histopathologist), as large contamination from stromal (healthy) cells would impair the accuracy of the assay to detect *CDKN2A* deletion. Overall, 40/49 (81.6%) of tumour regions contained ≥50% tumour cellularity. As there would be stromal cells present within each tumour region, the differentiation between homozygous and heterozygous deletion of *CDKN2A* would not be able to be strictly defined.

Examples of the ddPCR results are shown for four patients; 3 and 24 (Figure 4.6) and 7 and 23 (Figure 4.7), with results for all patients summarised in Table 4.5. Patients 1, 6, 12, 18 and 24 all showed either homozygous or heterozygous *CDKN2A* deletion. Amongst these patients, all regions exhibited largely comparative *CDKN2A* loss. A ratio of ~0.5 would have suggested heterozygous

deletion, providing 100% tumour was present within each respective region and that *CDKN2A* loss was present in every MPM tumour cell. In reality, as tumour tissue always has stromal contamination, it could be presumed that at least some of these patients could have homozygous deletion of *CDKN2A*. However, from these data alone, this assumption would be speculative. Alternatively, Patients 7 (Figure 4.7) and 16 seemed to exhibit a copy number consistent with the reference assay (~2), indicating diploid *CDKN2A*.

Unfortunately, Patient 3 was found to contain zero tumour in four out of five regions (Figure 4.6). Therefore, as expected these regions exhibited no evidence of *CDKN2A* deletion (ratio of ~1.00). However, region 4 (R4) contained >70% tumour and did detect loss of *CDKN2A,* with a ratio of 0.41 (Copy number, 0.82). This case therefore provided evidence of concordance between histopathological tumour cellularity and the detection of *CDKN2A* loss using this assay.

Patients 8 and 23 exhibited more obscure findings. Patient 23 had a ratio between 0.69 - 0.90 (1.38 – 1.8 Copy number) across the five tumour regions (Figure 4.7). Considering the high tumour cellularity within these regions (>70%), it is debatable whether this patient has heterozygous deletion or wild-type *CDKN2A* using this technology. Similarly for Patient 8, the assay was not informative of *CDKN2A* status (Table 4.5).

9/10 of these patients were also analysed for genome-wide copy number using OncoScan™ platform (Thermo Fisher Scientific) by Dr Annabel Sharkey (see Annabel Sharkey thesis). 8/9 patients analysed by OncoScan™ were found to have either heterozygous (n = 4) or homozygous loss of *CDKN2A* (n = 4), emphasising the frequency of this event in MPM (Table 4.5). For samples histologically containing tumour cells (cellularity > 0%), patients with homozygous deletion of *CDKN2A* by OncoScan™ had a mean ratio of 0.40 (copy number, 0.80) by ddPCR, while heterozygous deleted patients had 0.79 (copy number 1.58). These results indicate that the ddPCR assay was generally robust for the detection of *CDKN2A* loss in DNA derived from fresh-frozen tumour tissue.

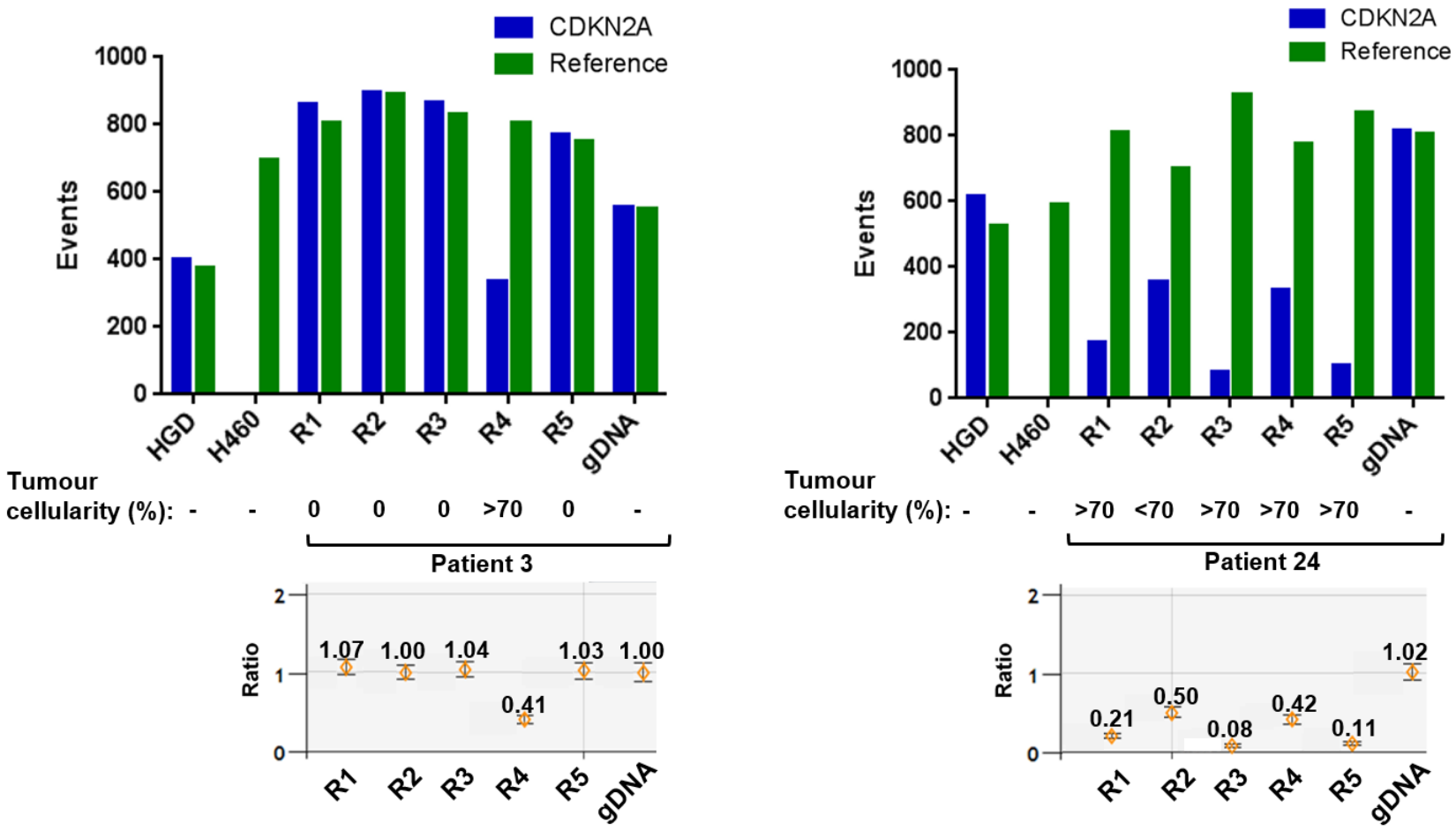


Figure 4.6 - *CDKN2A* deletion assay in Patients 3 and 24. The top panel shows the number of positive droplets for both the *CDKN2A* assay (blue) and the *RPPHI* reference assay (green). The middle panel gives the approximate tumour cellularity (%) found in corresponding H&E stained sections for each region (R1-R5). The bottom panel is showing the ratio between *CDKN2A*-positive and *RPPH1*-positive calls. Human genomic DNA (HGD) (Roche) and matching leukocyte DNA (gDNA) are negative controls (no *CDKN2A* deletion). H460 cell line DNA is acting as a positive control (*CDKN2A* homozygous deletion).

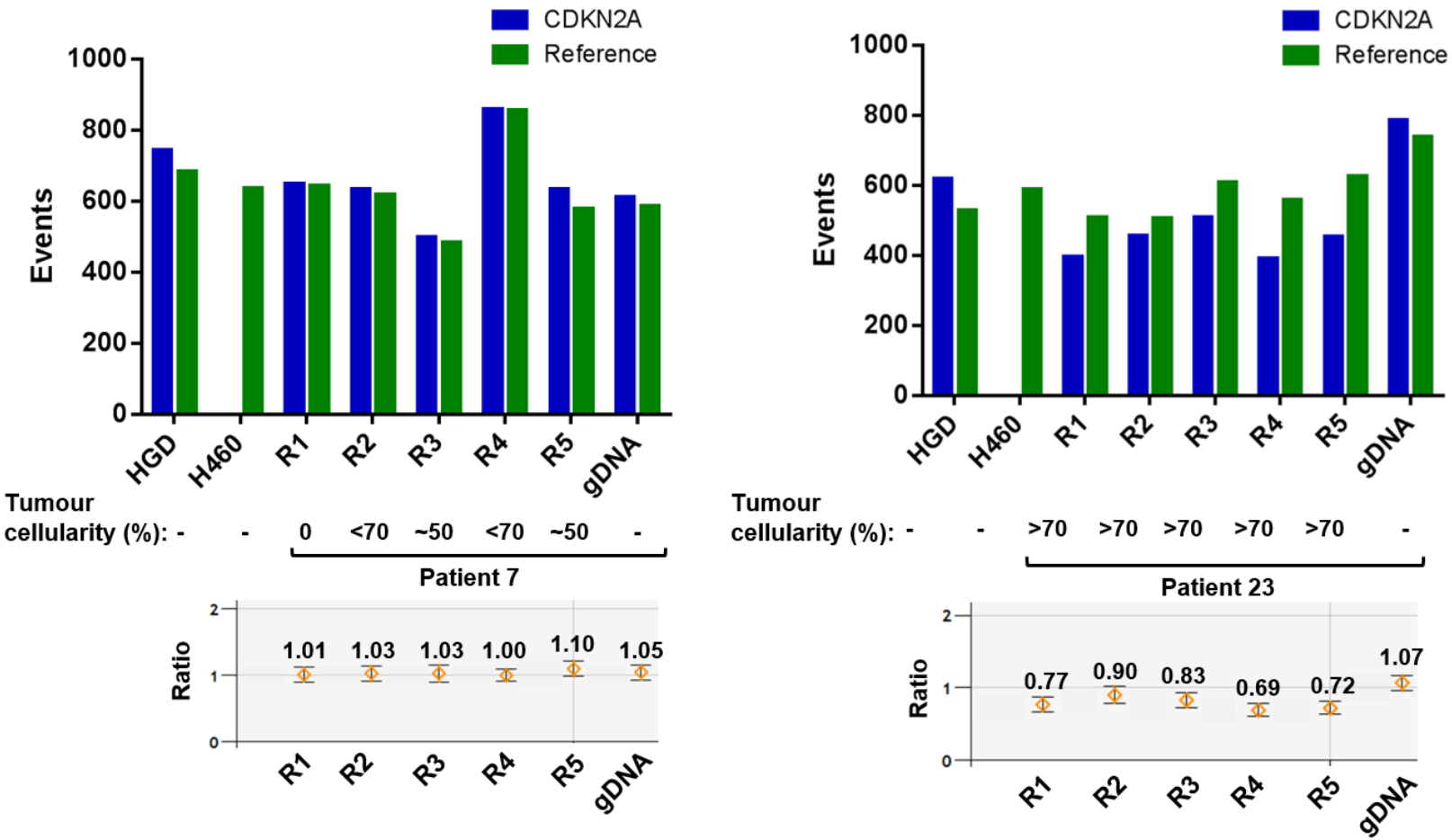


Figure 4.7 - *CDKN2A* deletion assay in Patients 7 and 23. The top panel shows the number of positive droplets for both the *CDKN2A* assay (blue) and the *RPPHI* reference assay (green). The middle panel gives the approximate tumour cellularity (%) found in corresponding H&E stained sections for each region (R1-R5). The bottom panel is showing the ratio between *CDKN2A*-positive and *RPPH1*-positive calls. Human genomic DNA (HGD) (Roche) and matching leukocyte DNA (gDNA) are negative controls (no *CDKN2A* deletion). H460 cell line DNA is acting as a positive control (*CDKN2A* homozygous deletion).

Patient	Sample	Tumour cellularity (%)	ddPCR		*CDKN2A* Status by OncoScan™
			Ratio	Copy Number	
Patient 1	R1	<70	0.55	1.10	Homozygous deletion
	R2	>70	0.36	0.72	
	R3	<70	0.53	1.06	
	R4	>70	0.38	0.76	
	R5	>70	0.40	0.80	
	gDNA	n/a	1.02	2.04	
Patient 3	R1	0	1.07	2.14	Homozygous deletion
	R2	0	1.00	2.00	
	R3	0	1.04	2.08	
	R4	>70	0.41	0.82	
	R5	0	1.03	2.06	
	gDNA	n/a	1.00	2.00	
Patient 6	R1	>70	0.53	1.06	Homozygous deletion
	R2	>70	0.59	1.18	
	R3	>70	0.44	0.88	
	R4	>70	0.57	1.14	
	R5	>70	0.38	0.76	
	gDNA	n/a	0.93	1.86	
Patient 7	R1	0	1.01	2.02	WT
	R2	<70	1.03	2.06	
	R3	~50	1.03	2.06	
	R4	<70	1.00	2.00	
	R5	~50	1.10	2.20	
	gDNA	n/a	1.05	2.10	
Patient 8	R1	<70	0.77	1.54	Heterozygous deletion
	R2	<70	0.97	1.94	
	R3	0	0.96	1.92	
	R4	0	0.91	1.82	
	R5	10	0.76	1.52	
	gDNA	n/a	1.03	2.06	
Patient 12	R2	>70	0.60	1.20	Heterozygous deletion
	R3	<70	0.74	1.48	
	R4	>70	0.57	1.14	
	R5	>70	0.51	1.02	
	gDNA	n/a	1.07	2.14	
Patient 16	R1	0	1.04	2.08	Heterozygous deletion
	R2	<70	0.92	1.84	
	R3	<70	1.05	2.10	
	R4	<70	1.01	2.02	
	R5	<70	0.75	1.50	
	gDNA	n/a	1.11	2.22	
Patient 18	R1	>70	0.39	0.78	n/a
	R2	>70	0.47	0.94	
	R3	~50	0.75	1.50	
	R4	<70	0.56	1.12	
	R5	>70	0.37	0.73	
	gDNA	n/a	0.90	1.80	
Patient 23	R1	>70	0.77	1.54	Heterozygous deletion
	R2	>70	0.90	1.80	
	R3	>70	0.83	1.66	
	R4	>70	0.69	1.38	
	R5	>70	0.72	1.44	
	gDNA	n/a	1.07	2.14	
Patient 24	R1	>70	0.21	0.41	Homozygous deletion
	R2	<70	0.50	1.00	
	R3	>70	0.08	0.16	
	R4	>70	0.42	0.84	
	R5	>70	0.11	0.22	
	gDNA	n/a	1.02	2.04	

Table 4.5 - Summary of *CDKN2A* ddPCR results including OncoScan™ *CDKN2A* status.

4.3.1.2 *CDKN2A* assay refinements

Despite the ddPCR assay being relatively informative on high molecular weight DNA (fresh-frozen), it was important to test whether the same was true when using more degraded DNA derived from FFPE tissues. Unlike fresh-frozen DNA, FFPE-derived DNA is usually heavily fragmented, due to the formalin fixation process introducing DNA crosslinks. If a DNA test is to be clinically applicable, it typically needs to interrogate FFPE DNA as this is the standard tissue preservation method. Therefore, this assay was subsequently tested on 5 ng of FFPE DNA from 40 MPM sourced by Dr Annabel Sharkey, all of which have also had *CDKN2A* status identified via OncoScan™ analysis.

Upon comparison of ddPCR and OncoScan™ methods, there was disparity between findings (Table 4.6A). Detection of normal *CDKN2A* would have an expected ratio value of 1 (1 copy *CDKN2A*/1 copy of *RPPH1*), however a mean ratio of 0.698 was found for this group. Mean ratios for heterozygous and homozygous deletion were 0.440 and 0.379 respectively.

One explanation for a lower than expected ratio, could be due to the fragmented nature of this DNA, resulting in *RPPH1* templates being overrepresented due to a smaller product size; the *RPPH1* assay amplicon was 62bp vs. 83bp for the *CDKN2A* assay. Therefore, a shorter *CDKN2A* assay was devised (61bp) to ensure the target and reference were approximately equal in length, meaning the ratio would be theoretically unaffected by DNA fragmentation. This shorter assay had a common forward primer and probe to the 83bp assay, but with a different reverse primer in order to reduce amplicon size (for oligonucleotide sequences, see section 2.4.1.1). To test this hypothesis, the same 40 FFPE samples were analysed with this new assay, in addition to multiregional fresh-frozen DNA from six MEDUSA patients as a fragmentation control (Figure 4.8).

A

Sample	*CDKN2A* 83bp / *RPPH1* ratio (ddPCR) ▼	OncoScan™ *CDKN2A* status
F1	0.91	WT
F2	**0.89**	**WT**
F3	0.836	WT
F4	**0.82**	**Homozygous deletion**
F5	0.79	Heterozygous deletion
F6	**0.784**	**Homozygous deletion**
F7	0.782	WT
F8	**0.779**	**WT**
F9	0.766	WT
F10	**0.733**	**WT**
F11	0.732	WT
F12	**0.726**	**WT**
F13	0.702	WT
F14	**0.687**	**WT**
F15	0.681	WT
F16	**0.637**	**WT**
F17	0.627	WT
F18	**0.631**	**WT**
F19	0.629	WT
F20	**0.616**	**Homozygous deletion**
F21	0.605	WT
F22	**0.575**	**WT**
F23	0.561	Heterozygous deletion
F24	**0.536**	**Heterozygous deletion**
F25	0.528	WT
F26	**0.498**	**Heterozygous deletion**
F27	0.496	WT
F28	**0.443**	**Heterozygous deletion**
F29	0.388	Homozygous deletion
F30	**0.312**	**Heterozygous deletion**
F31	0.289	Homozygous deletion
F32	**0.285**	**Homozygous deletion**
F33	0.28	Homozygous deletion
F34	**0.257**	**Homozygous deletion**
F35	0.252	Homozygous deletion
F36	**0.231**	**Homozygous deletion**
F37	0.205	Heterozygous deletion
F38	**0.197**	**Homozygous deletion**
F39	0.172	Heterozygous deletion
F40	**0.148**	**Homozygous deletion**

B

Sample	*CDKN2A* 61bp / *RPPH1* ratio (ddPCR) ▼	OncoScan™ *CDKN2A* status
F7	1.160	WT
F21	**1.150**	**WT**
F19	1.133	WT
F2	**1.111**	**WT**
F16	1.110	WT
F3	**1.100**	**WT**
F11	1.076	WT
F6	**1.035**	**Homozygous deletion**
F4	1.000	Homozygous deletion
F1	**1.000**	**WT**
F12	0.995	WT
F18	**0.975**	**WT**
F17	0.972	WT
F10	**0.970**	**WT**
F13	0.960	WT
F8	**0.940**	**WT**
F9	0.926	WT
F15	**0.920**	**WT**
F22	0.900	WT
F5	0.880	Heterozygous deletion
F25	0.854	WT
F14	**0.850**	**WT**
F27	0.804	WT
F20	**0.797**	**Homozygous deletion**
F23	0.790	Heterozygous deletion
F24	**0.742**	**Heterozygous deletion**
F26	0.741	Heterozygous deletion
F28	**0.692**	**Heterozygous deletion**
F29	0.655	Homozygous deletion
F30	**0.517**	**Heterozygous deletion**
F32	0.451	Homozygous deletion
F31	**0.450**	**Homozygous deletion**
F35	0.440	Homozygous deletion
F34	**0.419**	**Homozygous deletion**
F33	0.378	Homozygous deletion
F38	**0.366**	**Homozygous deletion**
F37	0.355	Heterozygous deletion
F39	**0.293**	**Heterozygous deletion**
F36	0.232	Homozygous deletion
F40	**0.209**	**Homozygous deletion**

Table 4.6 - ddPCR using the 83bp (A) and 61bp (B) *CDKN2A* assays with OncoScan™ status. 40 FFPE MPM samples are listed in descending order by ratio.

The mean ratio change between the 83bp and 61bp *CDKN2A* assay in FFPE DNA and Fresh/Frozen DNA was 0.233 and -0.018 respectively, indicating that indeed DNA fragmentation was the likely proponent (Figure 4.8). By shortening the assay, the WT samples (as defined by OncoScan™) subsequently had an average ddPCR ratio of 0.995; much more in line with the expected ratio of 1. However, F4 and F6 remained as outliers with ratios of 1.00 and 1.035 respectively, despite being listed as homozygous deletion *via* OncoScan™ (Table 4.6B). One possible explanation for these findings could be that these samples had concurrent deletion of *RPPH1* and *CDKN2A*, hence keeping the ratio 1. Indeed, deletions at 14q have been reported in MPM (*RPPH1* is located at 14q) (Taniguchi *et al.* 2007). This therefore warrants to further testing of the *CDKN2A* assay in conjunction multiple reference assays, which will hopefully resolve ambiguities. Nevertheless, these results show that even narrow differences between target and reference amplicon lengths can have a large impact on assay performance when interrogating fragmented DNA.

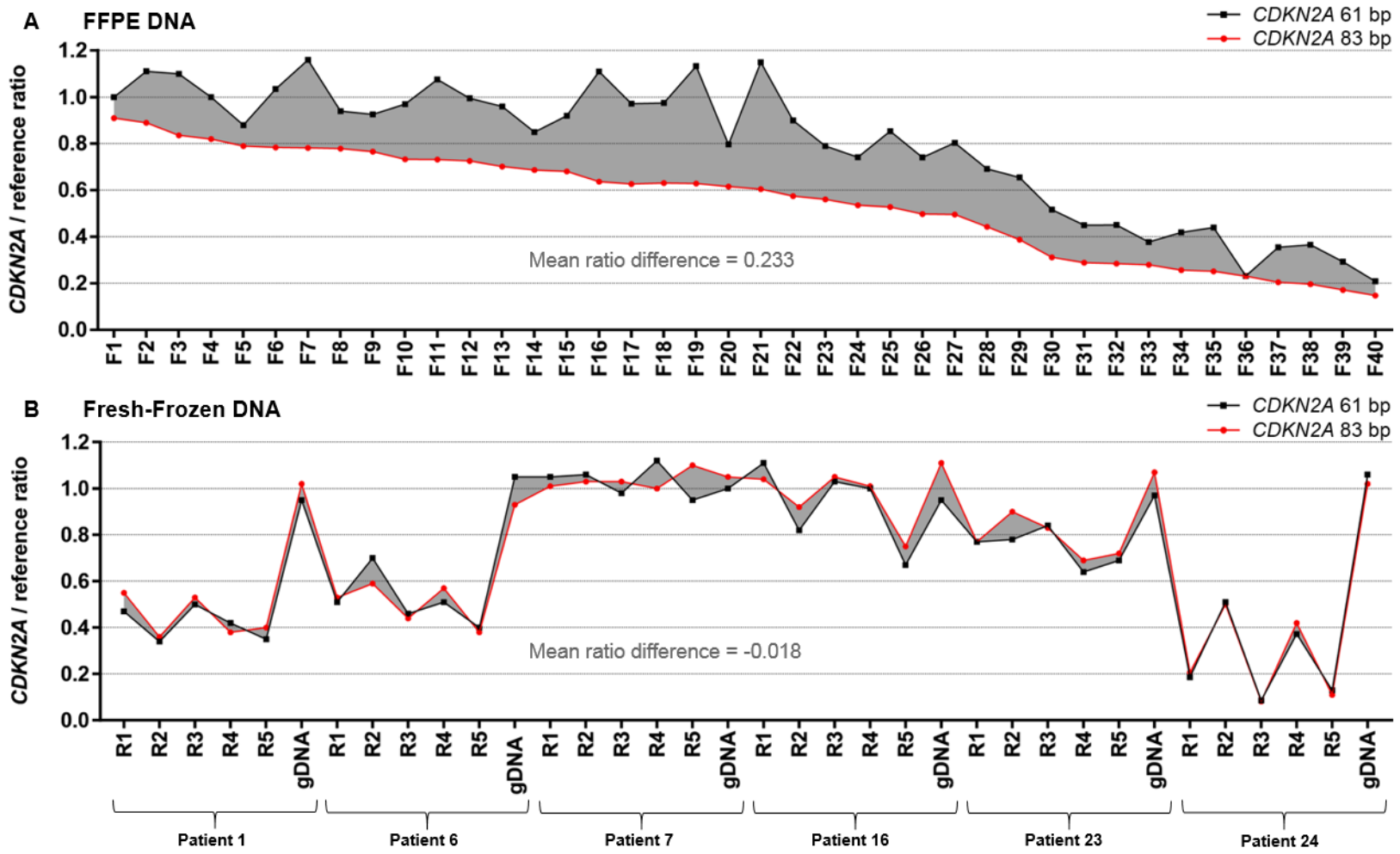


Figure 4.8 - DNA fragmentation affects *CDKN2A/RPPH1* ratio.Ratios for the 61bp & 83bp *CDKN2A* assays on FFPE DNA (A) & Fresh-Frozen DNA (B). The shading indicates the difference between the ratios for each assay. The mean ratio difference (61bp – 83 bp ratio per sample) for each DNA source is indicated.

4.3.2 ctDNA analysis

4.3.2.1 Identification of clonal variants from Multiregional WES data

The phylogenetic analyses on individual MPM tumours made possible by MEDUSA would enable the identification of clonal (early) tumour variants for optimal ctDNA detection within plasma. When this work began there were no publications describing ctDNA analysis in MPM and therefore it was important to first determine whether this was feasible. Although somatic copy number analysis (SCNA) of tumour DNA is currently being conducted within MEDUSA and is likely to yield particularly informative results, the detection of tumour-confirmed variants (SNVs) currently represent the most robust targets for ctDNA detection. Therefore, this section focusses on the analysis of SNVs and their detection within cfDNA.

Multiregional WES of MPM tumours was conducted by an external service provider (Eurofins AROS, BGI and Novogene) in multiple batches. Initially, five tumour regions (R1 – R5), plus germline DNA were sequenced for Patients 12, 23, 24, 27 and 34. Following this batch, it was deemed unnecessary to sequence R2 (pericardium), as not all patients had pericardium removed, instead having pericardium-adjacent pleura removed. Therefore in later batches, MEDUSA restricted multiregional WES to four regions (R1, R3, R4 and R5) in Patients 6, 9, 18, 33, 37 and 64.

For the initial batch of samples (5 regions), expertise regarding phylogenetic analyses was sought through the TRACERx consortium, whereby these samples were analysed using their refined pipeline. Following data processing, the assembly of phylogenetic evolutionary trees for the first five patients was conducted by Dr Gareth Wilson (Jamal-Hanjani *et al.* 2017). On an SNV level the phylogenetic trees showed a high degree of interpatient heterogeneity, with largely dominant truncal regions (Figure 4.9).

The remaining patients featured in this thesis had their multiregional WES data processed using a similar methodology developed by Dr Robert Hastings (see section 2.4.5.3). At this time, phylogenetic trees have not yet been developed for

these subsequent tumours, but through variant filtering, clonal variant (present in all tumour regions) identification was achieved.

The identification of clonal variants for assay design was determined using the following criteria: (1) Absent from germline DNA, (2) Found in all tumour DNA regions (4 or 5, depending on patient), (3) “Driver mutations” were prioritised (4) Highest mean VAF amongst tumour regions.

In order to address criteria (3), variants were categorised into “Driver Tiers” using the Cancer Genome Interpreter (Institute for Research Biomedicine, Barcelona), whereby variants were assigned Tier 1 (high confidence driver event), Tier 2 (potential driver event) or Passenger status (Tamborero *et al.* 2018).

In these 11 patients, six were found to have only “passenger” clonal variants in their tumours (Table 4.7). Of the remaining five, Patients 27 and 34 had two Tier 1 clonal variants each, Patient 33 had one, while Patients 9 and 12 only had a single Tier 2 clonal variant identified.

Patient	Clonal Tier 1	Clonal Tier 2
6	0	0
9	0	***SOX9* p.I397M**
12	0	***TRAF7* p.C388G**
18	***NF2* p.E372***	*JAK1* p.L396V
23	0	0
24	0	0
27	*NF2* p.Q178* *FBXW7* p.R479Q	0
33	***NUP98* p.W1407C**	0
34	***BAP1* p.L635Sfs*2** *NF2* p.F118Lfs*11	0
37	0	0
64	0	0

Table 4.7 - Clonal variants found in 11 MPM patients classified as Tier 1 and Tier 2. Variants in bold indicate that an efficient assay was ultimately used for ctDNA detection (see later section).

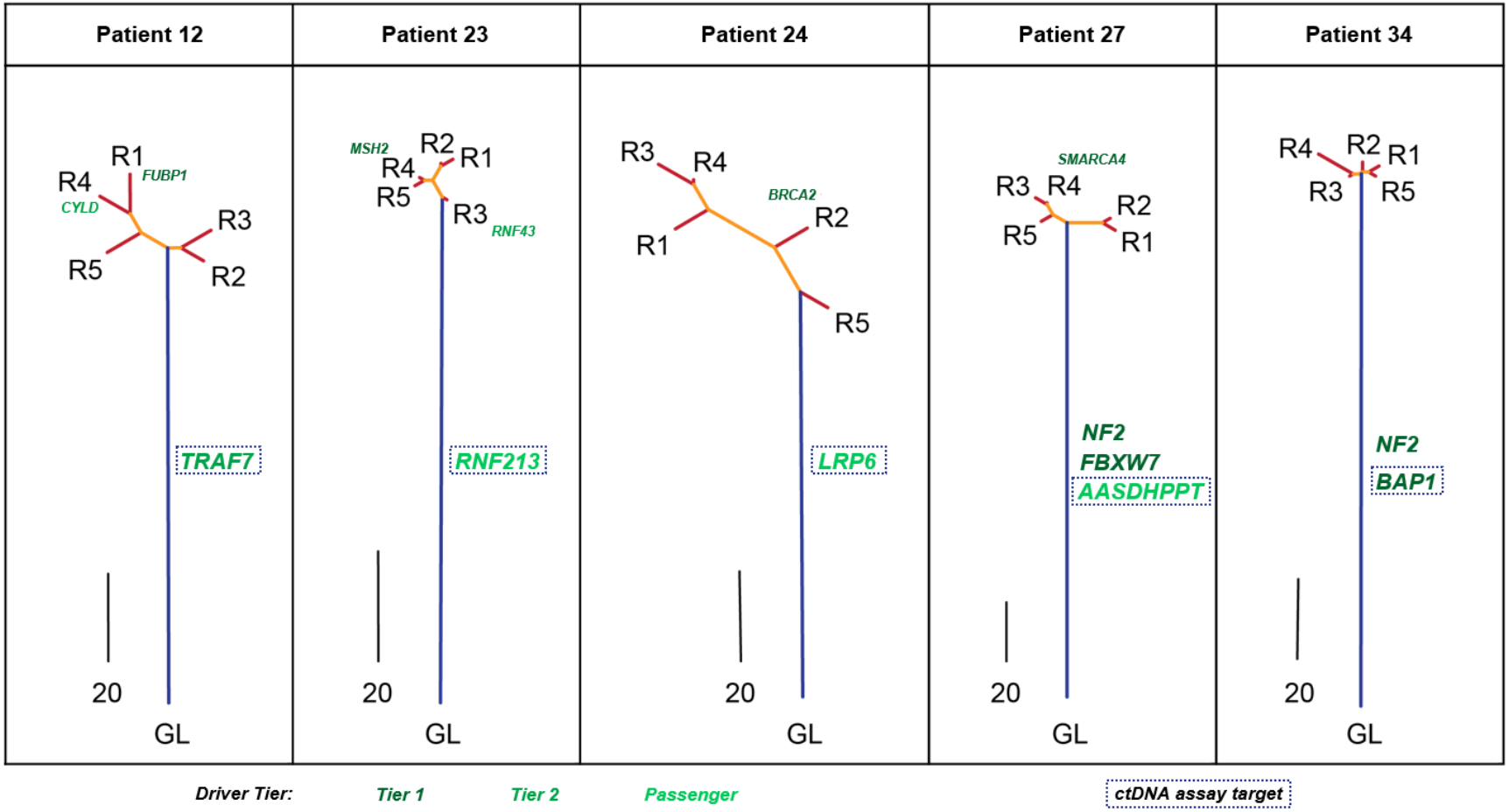


Figure 4.9 - SNV phylogenetic Trees of 5 MEDUSA tumours. Blue lines represent clonal variants, yellow represent shared variants within a subset of tumour regions, with red lines indicating private variants. A scale bar is provided for each tree to indicate the number of variants identified. Variants were categorised according to driver status (Tier 1, 2, passenger).Genes of clonal variants are annotated if predicted to be Tier 1 or 2 drivers, or were variants ultimately targeted for ctDNA detection (blue boxes). Genes of subclonal variants were only plotted if predicted as Tier 1/2 drivers. GL, germline DNA; R1-R5, tumour region.

4.3.2.2 Selection, design and optimisation of patient-specific assays

Providing there were clonal variants predicted to be Tier 1/2 drivers, these were the initial targets for primer/probe design, however the development of efficient assays targeting these were not always possible. Only in Patient 27 did assay design for clonal Tier 1/2 variants fail. The other 10 patients either had an assay successfully designed targeting a Tier 1/2 variant (n = 5) or only had passenger clonal variants in their tumours (n = 5).

Patient 27 had two Tier 1 clonal drivers in *NF2* and *FBXW7* respectively (Table 4.7). However the final assay designed and used for ctDNA detection targeted a passenger clonal variant in *AASDHPPT*. For the *NF2* variant, wild-type and mutant probes were predicted to have high hairpin stability which would have likely hindered assay performance (Figure 4.10A). Additionally the 100 bp sequence including the variant had a highly repetitive sequence (consecutive thymine bases) upstream of the variant site, which made the designing of primers with a similar T_m within such as short amplicon size highly problematic (Figure 4.10A). Whereas for the *FBXW7* variant, the WT probe was predicted to contain an extremely high self-complementarity precisely in the position of the variant site and hence would be likely to form self-dimers (Figure 4.10B).

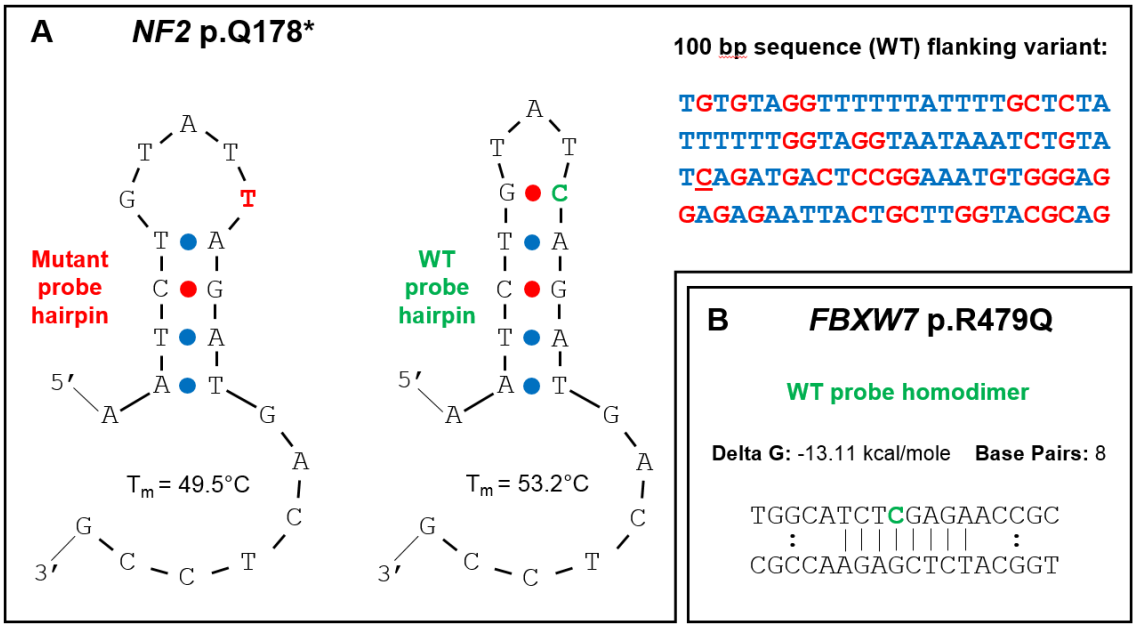
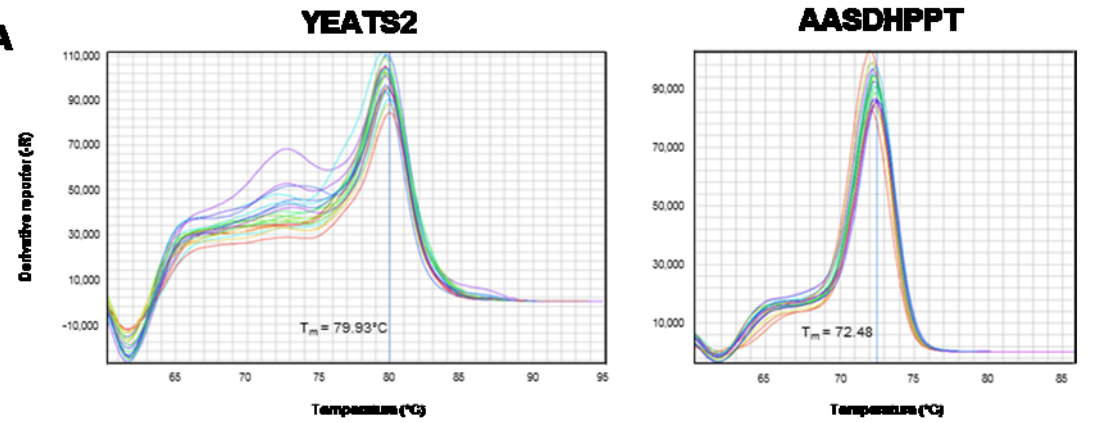


Figure 4.10 - Assay design issues for *NF2* (A) and *FBXW7* (B) Tier 1 variants in Patient 27. A) High hairpin stability in particularly both WT and mutant probes, but also an AT rich region directly upstream of the p.Q178* variant site ("C" underlined). B) The WT probe was predicted to have a highly detrimental homodimer predicted at the variant locus ("C" in green).

Once primers and probes passed the aforementioned design checks, the primers flanking the respective variant sites were ordered and tested for amplification efficiency on human genomic DNA (hgDNA) and their propensity to form secondary structures using SYBR® green melt curves, prior to the ordering of probes (mutant and WT). In Patient 27, two different assays (primers only) were compared and the optimal assay would ultimately be chosen for further optimisation steps. The two assays targeted a clonal passenger variant found in *YEATS2* and *AASDHPPT* respectively, which had similar VAFs from multiregional WES. Through SYBR-Green melt curve analysis, it was found that *AASDHPPT* primers showed optimal performance, with no secondary product formation, compared to primer dimers being found in *YEATS2*, as shown by the presence of a 'shoulder' (Figure 4.11). Therefore the *AASDHPPT* assay was consequently selected for further optimisation and validation.

A



B

		R1	R2	R3	R4	R5	gDNA
WES VAF (%)	YEATS2	23.19	36.09	19.57	22.62	21.72	0.00
	AASDHPPT	26.49	36.21	26.68	27.18	29.91	0.00

Figure 4.11 - Example of target selection using SYBR-Green Melt curve analysis. A) Melt curves of primers targeting the area of interest in *YEATS2* and *AASDHPPT* respectively, with melting temperature (T_m) of the product indicated. *YEATS2* primers show evidence of primer dimers, whereas *AASDHPPT* show no secondary product formation. B) The variant allele frequency (VAF) of the two variants from multiregional-WES of tumour tissue.

Probes were subsequently ordered and incorporated into each assay. Either MGB-NFQ or LNA-BHQ-1 probes were used, depending on certain assay

requirements. For example, in sequences with higher GC contents, a smaller probe was considered ideal as mismatch discrimination is higher in shorter probes, meaning an MGB probe was used as it increases T_m by ~10°C. Alternatively, in sequences that are more AT rich and therefore requiring longer probes, LNA were incorporated to improve mismatch discrimination. A list of the final primers and probes used for the patient-specific assays are shown in Table 4.8.

Patient	Assay Target	Genomic Position	Oligonucleotide	Oligonucleotide sequence	Length (nt)	Amplicon Length (bp)
6	*RNF187*	chr1:228,680,744 - 228,680,827	Forward	CCGTTTGCAGGGACACG	17	84
			Reverse	TCCTCCTCCACAAAGAAGGC	20	
			WT probe	CGACTACAAGAAGCTGCG	18	
			Mut probe	CGACTACAAGTAGCTGCG	18	
9	*SOX9*	chr17:70,120,142-70,120,220	Forward	CTGACCACGCTGAGCA	16	79
			Reverse	TGTAGTGGCTGGGGCT	16	
			WT probe	ACGCACATCAAGACGGA	17	
			Mut probe	ACGCACATGAAGACGGA	17	
12	*TRAF7*	chr16:2,223,921-2,224,002	Forward	GCCTACGACCCTCAGCA	17	82
			Reverse	AGTAGACGCAGAGACACCAC	20	
			WT probe	TCC[C]TTTG[C][A][C]TTGAAGAT	19	
			Mut probe	TCCCTTTG[C][C][C]TTGAAGAT	19	
18	*NF2*	chr22:30,067,898-30,067,983	Forward	GCAGATGAAAGAAGAAGCAAC	21	86
			Reverse	GAAGTCCCCAAGTAGCCTC	19	
			WT probe	CCAACGAAGCACTGGTG	17	
			Mut probe	CCAACTAAGCACTGGTG	17	
23	*RNF213*	chr17:78,265,561-78,265,637	Forward	AGCAGAAGAAGGGCGAGTAC	20	77
			Reverse	CCACGACTTACCTCCTGAGC	20	
			WT probe	TCAA[C]CGC[T][G][T]CTGTTCAT	19	
			Mut probe	TCAA[C]CGC[T][A][T]CTGTTCAT	19	
24	*LRP6*	chr12:12,291,270-12,291,351	Forward	AGGTTCAGCTCCTTTACTGC	20	82
			Reverse	TGACAGGTCGAGAGGGTAG	19	
			WT probe	CAAGCTCG[A][A][T]TGCCCAGC	19	
			Mut probe	AAGCTCG[A][C][T]TGCCCAG	17	
27	*AASDHPPT*	chr11:105,967,498-105,967,586	Forward	CCCAGAGGCAATTTACTATTCTCA	24	89
			Reverse	ACAGTCCCAAAATGAAGGATCT	22	
			WT probe	TCATCTG[C][G][T]TC[C]CATGA	18	
			Mut probe	TCATCTG[C][A][T]TC[C]CATGA	18	
33	*NUP98*	chr11:3,714,511-3,714,582	Forward	GGTGGAAGCAAATACCAAAGAT	22	72
			Reverse	GCAAATAAACGTGTGCTCCC	20	
			WT probe	AGCGTTT[C][C]A[A]TCCAACT	18	
			Mut probe	AGCGTT[T][A][C]AAT[C]CAACT	18	
34	*BAP1*	chr3:52,436,846-52,436,919	Forward	TGCAATCTCAGCCTCCACA	19	74
			Reverse	CCGATCAGAGGTGCAATGC	19	
			WT probe	ACTTCAGC[A][G]TGCCAGCAG	19	
			Mut probe	CACTTCAG[C]T[T]GCCAGCAG	19	
37	*MYBPC1*	chr12:102,056,108-102,056,185	Forward	CCTTAGTTTTCATTTTGCATACCTG	25	78
			Reverse	CAGCCTCATCCACCTGG	17	
			WT probe	CTTTGTTTCTTCTTCCTCTCA	21	
			Mut probe	CTTTGTTTCCTCTTCCTCTCA	21	
64	*BRCA1*	chr17:41,251,764-41,251,836	Forward	AAATACTTAAAAAACCTGAGACCCT	25	73
			Reverse	AGCAGCGGATACAACCTC	18	
			WT probe	CCCAATTCAATGTAGACAGA	20	
			Mut probe	CCCAATTCAAGGTAGACAGA	20	

Table 4.8 - Assay details for patient-specific ctDNA testing in 11 MEDUSA patients. The final primer and probe sequences for assays targeting a single variant in cfDNA in each patient. Square brackets "[x]" denote the presence of a locked nucleic acid (LNA) attached to the respective base. For probes containing LNAs, these were either FAM (mutant) or HEX (WT) labelled with a 3' BHQ-1. For the remaining probes, these were either FAM (mutant) or VIC (WT) labelled with a 3' MGB-NFQ.

Once probes were incorporated, assay optimisation was achieved using a ddPCR temperature gradient, in order to select the optimal annealing temperature (T_a) for each assay. Typically, this gradient ranged between 62°C and 55°C as during primer/probe design, programs estimated these oligonucleotides would be anneal within this range. For ddPCR, the optimal T_a was one that resulted in the largest fluorescence amplitude discrimination between positive and negative events while avoiding non-specific amplification (see Bio Rad ddPCR applications guide). For each assay a temperature gradient was conducted using two separate DNA samples; one tumour tissue DNA sample positive for the respective mutant variant (from WES) and human genomic DNA (HGD). The latter was included as this will not contain mutant DNA, thereby acting as a test for non-specific amplification (mutant probe binding WT template), hence determining the false positive rate.

An example of a temperature gradient for an assay targeting an *NF2* p.E372* using tumour tissue DNA from a single region (R1) from Patient 18 is shown (Figure 4.12). Based on this, a T_a of 57°C was considered optimal for this assay, as this was the highest approximate temperature at which discrimination between WT and mutant droplets was most pronounced.

Following an optimal T_a being determined for each assay, DNA from all five tumour regions (R1 - R5) along with germline DNA for each patient was analysed for the presence (R1 - R5) and absence (germline DNA) of each respective variant. For the six patients that only had four regions sent for WES, R2 (pericardium-adjacent) was also assessed for the presence of the variant allele. Interestingly, in all of these patients, the variant was found within all five tumour regions, confirming these as clonal variants. There were no instances where germline DNA was found to contain any mutant positive droplets and so each assay was prepared for cfDNA analysis.

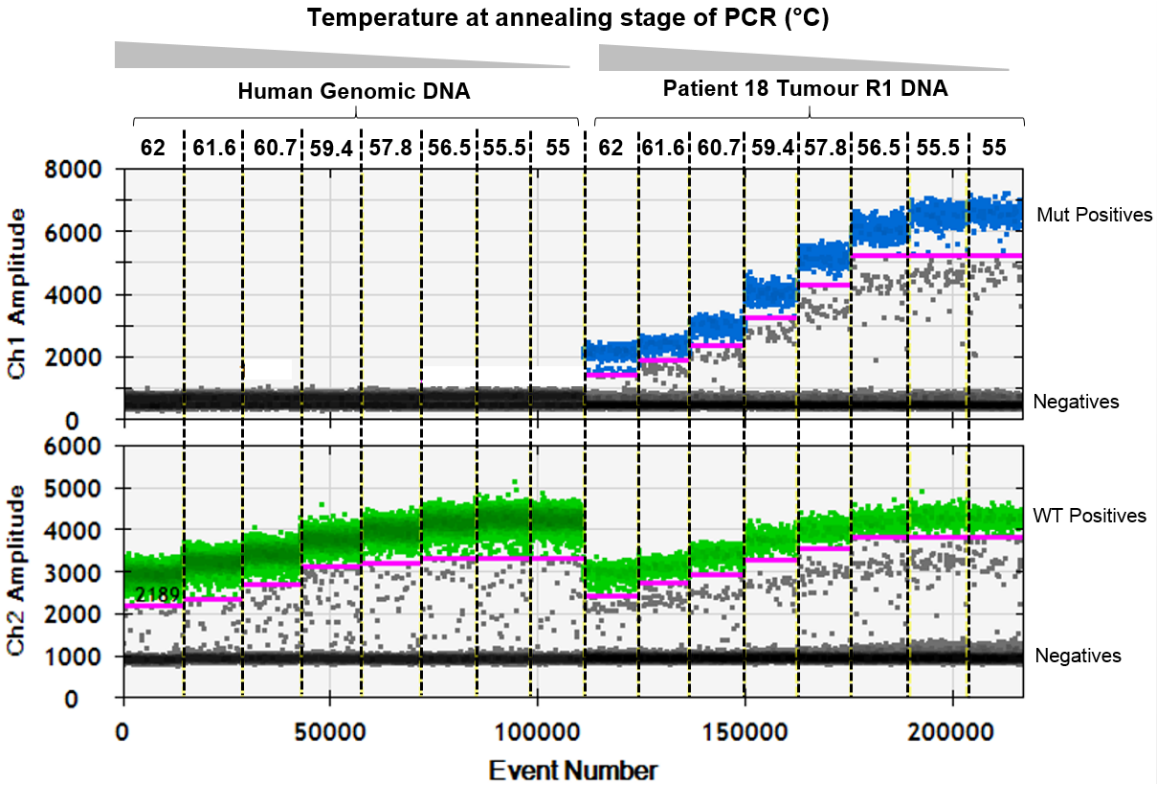


Figure 4.12 - ddPCR temperature gradient for assessing optimal assay T_a for assay targeting a variant in Patient 18. 1D amplitudes of mutant (*NF2* p.E372*; blue) and WT (green) droplets in 10ng human genomic DNA (HGD; left) and tumour tissue DNA from Patient 18 (R1; right). The annealing temperature results in increased yield, but decreased specificity, thereby increasing the potential for non-specific amplification. Typically, at lower temperatures the amplitudes of mutant and WT channels increase, providing a larger discrimination between positive and negative droplets. In this example there were no false positive mutant droplets in HGD at any annealing temperature (62 – 55°C). For this assay, the optimal annealing temperature was determined as 56.5°C, and so the 57°C was decided as the T_a for this assay.

4.3.2.3 Patient-specific mutation detection in cfDNA

Patient-specific assays were then tested on pre-surgical cfDNA from the 11 patients. Table 4.9 summarises the clinical information for these patients. Histologically, seven patients were epithelioid MPM, with the remaining four patients having biphasic disease. As mentioned previously, it is unsurprising that sarcomatoid MPM was not represented in this pilot cohort, as patients with this subtype are rarely applicable for surgical resection and hence did not receive multiregional WES within MEDUSA.

A single patient (Patient 27) had neoadjuvant therapy, which included the standard of care (pemetrexed and cisplatin), in addition to the vaccine TroVax® as part of the SKOPOS trial (NCT01569919). The other patients were chemotherapy naïve at the time of surgery. Patients were largely homogeneous with regards to stage, with 9/11 patients being stage IIIB, which highlights the frequency of late diagnosis in MPM.

Based on quantification using the TapeStation 4200 (Agilent Technologies), 10 ng of cfDNA was used for initial ddPCR runs. As a means to reduce false positive calls, stringent thresholds were applied whereby a patient was only deemed ctDNA positive if ≥ 3 mutant droplets and had a VAF ≥ 0.1%.

Patient	Age at Surgery	Sex	Histology	TNM	IMIG Stage (8th Edition)	Neoadjuvant Chemotherapy
6	72	M	epithelioid	T3, N2, M0	IIIB	n/a
9	78	F	biphasic	T3, N0, M0	IB	n/a
12	63	M	epithelioid	T3, N2, M0	IIIB	n/a
18	67	M	biphasic	T2, N0, M0	IB	n/a
23	63	M	epithelioid	T3, N2, M0	IIIB	n/a
24	71	M	biphasic	T3, N2, M0	IIIB	n/a
27	53	M	epithelioid	T4, N0, M0	IIIB	pem-cis 6 cycles + TroVax®
33	71	M	epithelioid	T3, N2, M0	IIIB	n/a
34	61	M	biphasic	T3, N2, M0	IIIB	n/a
37	74	M	epithelioid	T4, N2, M0	IIIB	n/a
64	69	M	epithelioid	T3, N2, M0	IIIB	n/a

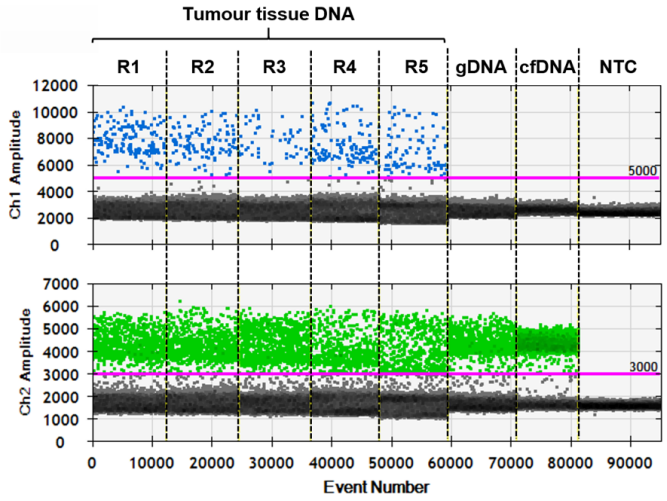
Table 4.9 - Clinicopathological characteristics of the 11 MEDUSA patients for cfDNA analysis.

4.3.2.3.1 ctDNA-negative samples

5/11 patients had no mutant droplets detected in their cfDNA samples; Patients 12, 23, 33, 37 and 64 (Figure 4.13 and 4.14). Four of these (12, 33, 37 and 64) had between 1,795 and 4,720 WT droplets identified (mean 2,844), indicating that DNA input was not a likely limiting factor in the non-detection of ctDNA. Therefore, it is likely that if tumour-derived DNA was present in these cfDNA samples, it was at very low VAFs (< 0.06%) that would represent very unconfident positive calls. For example, Patient 37 had the lowest number of WT droplets in this group (1,795), meaning if one mutant positive droplet was identified, it would have a VAF of ~0.056% VAF (1/1,796).

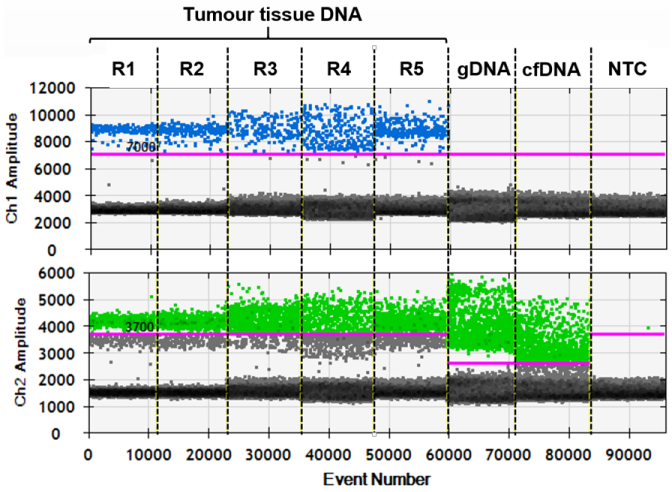
The final patient in this group, Patient 23 had 1,012 WT droplets identified, and therefore was a less confident ctDNA-negative result (Figure 4.13B). Due to low sample concentration (0.12 ng/μL), re-extraction (plasma stock depleted), quantitation and lyophilisation was conducted on more cfDNA from this patient in order to increase DNA input and potentially detect the *RNF213* p.C479Y variant. Unfortunately however, total droplet generation failed during this rerun and so analysis was not possible. A potential explanation for this failure may be that the lyophilisation of high volumes of low concentration cfDNA (0.12 ng/μL) concentrates the salts derived from cfDNA extraction buffers. The reconstituting of lyophilised DNA in a smaller volume of TE buffer results in this crystalline salt persisting as particulate matter, blocking the microfluidic channels during droplet generation. Nevertheless, considering ≥ 3 mutant droplets are required for a sample to be deemed ctDNA-positive, it seems likely that Patient 23 is a ctDNA-negative case.

A Patient 12 – *TRAF7 p.C388G*



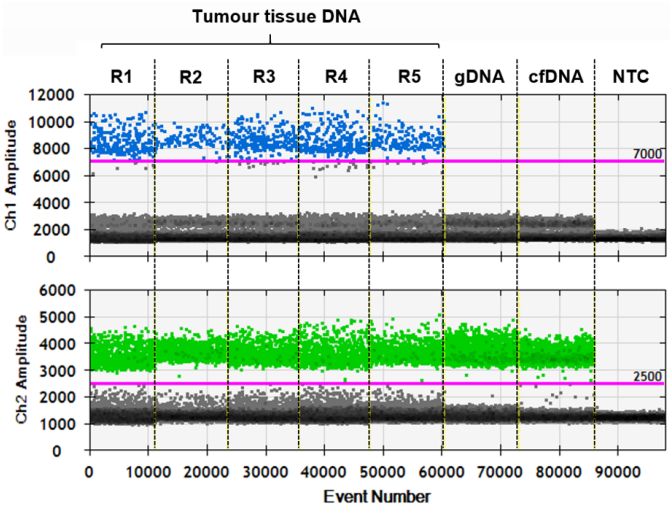
	VAF (%)		# positive droplets	
	WES	ddPCR	Mut	WT
R1	12.14	14.50	164	967
R2	11.89	11.70	119	898
R3	3.47	4.50	50	1,060
R4	15.75	17.98	135	616
R5	13.91	13.62	96	609
gDNA	0.00	0.00	0	1,002
cfDNA	n/a	0.00	0	4,720

B Patient 23 - *RNF213* p.C479Y



	VAF (%)		# positive droplets	
	WES	ddPCR	Mut	WT
R1	18.43	20.30	215	844
R2	19.80	18.88	235	1,010
R3	14.36	16.36	221	1,130
R4	28.19	34.67	295	556
R5	24.53	23.91	308	980
gDNA	0.00	0.00	0	1,577
cfDNA	n/a	0.00	0	1,012

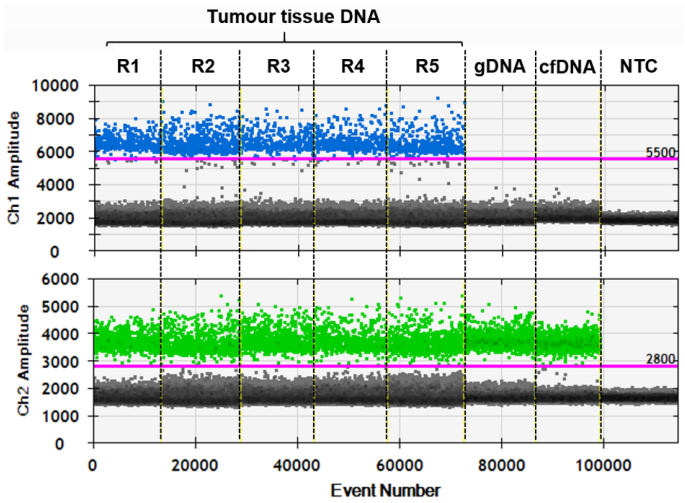
C Patient 33 - *NUP98* p.W1407C



	VAF (%)		# positive droplets	
	WES	ddPCR	Mut	WT
R1	18.40	23.60	269	871
R2	n/a	9.41	117	1,126
R3	23.80	19.90	277	1,115
R4	33.70	25.02	325	974
R5	10.7	14.49	183	1,080
gDNA	0.00	0.00	0	2,453
cfDNA	n/a	0.00	0	2,625

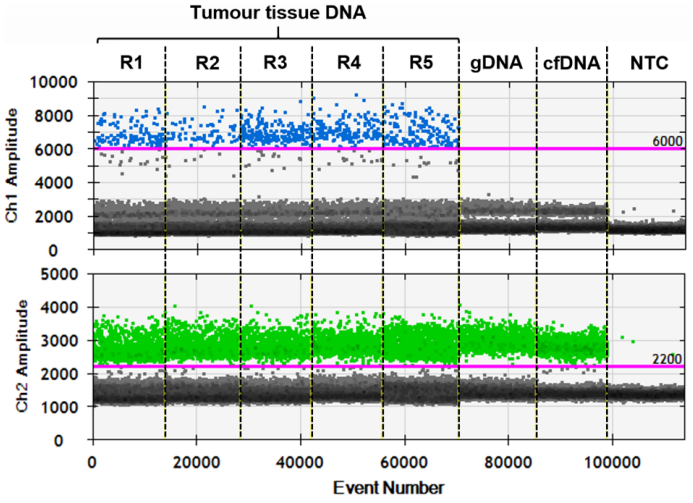
Figure 4.13 - cfDNA analysis in Patients 12 (A), 23 (B) and 33 (C). 1D amplitudes of mutant (blue) and WT (green) channels showing positive droplets. Also present are tables providing the ddPCR VAF and number of positive droplets for each respective run. For samples that had undergone WES, the VAFs were also shown.

A Patient 37 – *MYBPC1* p.K660R



	VAF (%)		# positive droplets	
	WES	ddPCR	Mut	WT
R1	23.90	32.98	399	811
R2	n/a	39.53	593	907
R3	30.90	29.04	377	921
R4	42.90	39.35	484	746
R5	43.40	30.55	435	989
gDNA	0.00	0.00	0	1,810
cfDNA	n/a	0.00	0	1,795

B Patient 64 – *BRCA1* p.I180L



	VAF (%)		# positive droplets	
	WES	ddPCR	Mut	WT
R1	13.30	8.34	126	1,385
R2	n/a	4.87	77	1,504
R3	10.80	13.68	221	1,395
R4	13.40	11.55	177	1,355
R5	16.10	10.90	175	1,430
gDNA	0.00	0.00	0	1,607
cfDNA	n/a	0.00	0	2,235

Figure 4.14 - cfDNA analysis in Patients 37 (A) and 64 (B). 1D amplitudes of mutant (blue) and WT (green) channels showing positive droplets. Also present are tables providing the ddPCR VAF and number of positive droplets for each respective run. For samples that had undergone WES, the VAFs were also shown.

4.3.2.3.2 ctDNA-positive samples

4/11 patients had their respective variants identified within cfDNA when using the aforementioned thresholds (ctDNA positive = ≥ 3 mutant droplets + VAF ≥ 0.1%); Patients 6, 9, 18, 27 (Figure 4.15 and 4.16). These ranged in VAFs between 0.16% and 2.96%. Of these patients, Patient 27 had the highest ctDNA VAF (2.96%) and was also only patient studied that had received neoadjuvant chemotherapy, in addition to the TroVax® vaccine (SKOPOS trial).

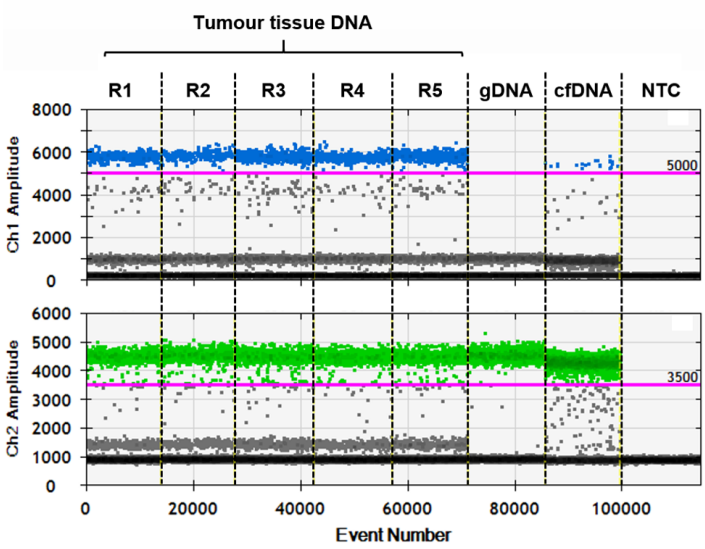
When compared to the ctDNA-negative patients, there did not seem to be a relationship between Driver Tier and ctDNA detection, indicating that the targeting of passenger variants was not associated with a decreased detection rate in plasma (Table 4.10).

Considering a higher number of WT positive droplets would mean a higher sensitivity (lower LOD) and potentially improved ctDNA detection, it was important to compare the number of positive WT droplets between the ctDNA-positive and ctDNA-negative groups. There did not appear to be a significant difference in WT droplets between the ctDNA-positive (median 2,946; mean 3,284) and ctDNA-negative (median 2,235; mean 2,477), even when the results of Patient 23 were included (Table 4.10). This therefore indicated that ctDNA detection status was likely due to factors independent of sample input.

	ctDNA assay variant classification			# Mutant droplets	# WT droplets
Patient	**Tier 1**	**Tier 2**	**Passenger**		
6			*RNF187*	22	5,966
9		*SOX9*		9	3,463
18	*NF2*			4	2,428
27			*AASDHPPT*	39	1,279
12		*TRAF7*		0	4,720
23			*RNF213*	0	1,012
33	*NUP98*			0	2,625
37			*MYBPC1*	0	1,795
64			*BRCA1*	0	2,235

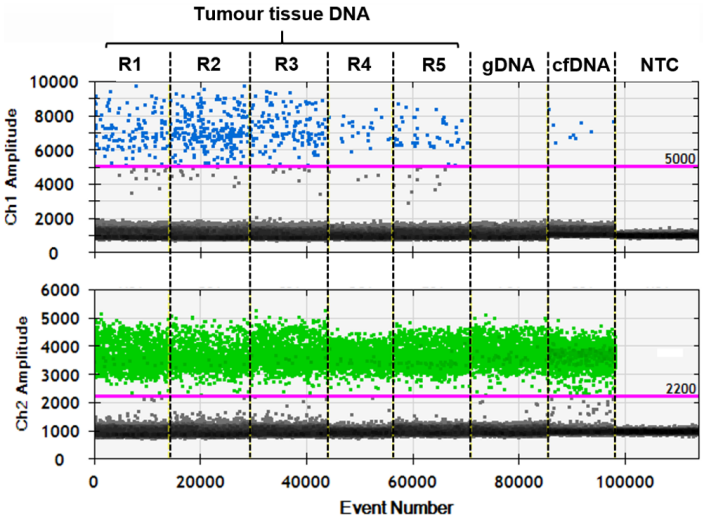
Table 4.10 - ctDNA positive vs ctDNA-negative patients with respect to driver classification and droplet values. The gene target is shown as either a Tier 1, Tier 2 or passenger variant, with the number of mutant and WT droplets shown for each ddPCR. Shading indicates the ctDNA-positive patients, the remaining patients were ctDNA-negative.

A Patient 6 – *RNF187* p.K68*



	VAF (%)		# positive droplets	
	WES	ddPCR	Mut	WT
R1	26.10	22.71	384	1,307
R2	n/a	14.79	233	1,342
R3	25.10	27.03	503	1,358
R4	20.10	21.66	354	1,280
R5	30.10	30.93	510	1,139
gDNA	0.00	0.00	0	2,153
cfDNA	n/a	0.37	22	5,966

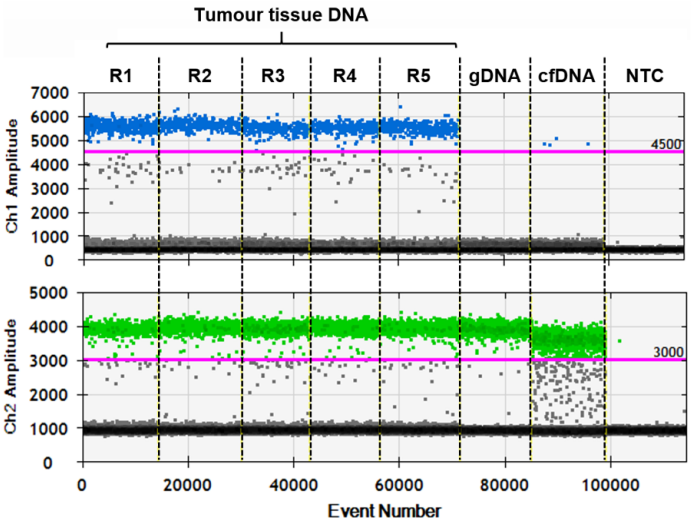
B Patient 9 – *SOX9* p.I397M



	VAF (%)		# positive droplets	
	WES	ddPCR	Mut	WT
R1	13.4	7.80	128	1,514
R2	n/a	16.57	239	1,203
R3	11.00	8.75	135	1,408
R4	2.80	1.82	24	1,294
R5	2.70	2.85	52	1,772
gDNA	0.00	0.00	0	1,661
cfDNA	n/a	0.26	9	3,463

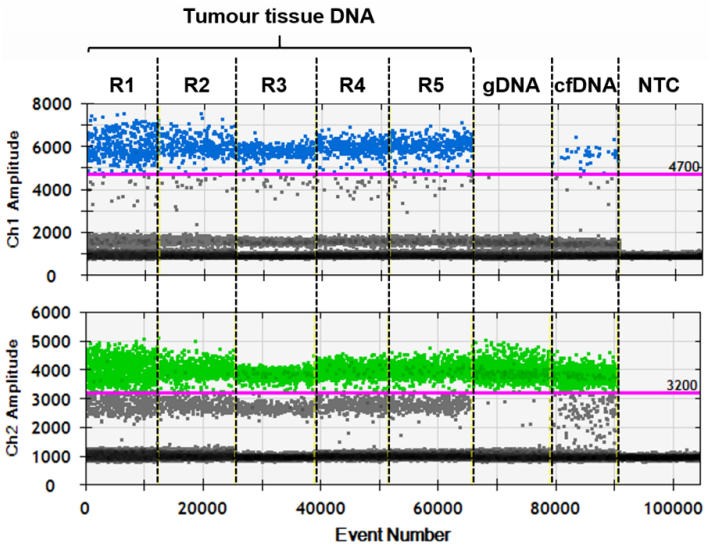
Figure 4.15 - ctDNA detection in Patients 6 (A) and 9 (B). 1D amplitudes of mutant (blue) and WT (green) channels showing positive droplets. Also present are tables providing the ddPCR VAF and number of positive droplets for each respective run. For samples that had undergone WES, the VAFs were also shown.

A Patient 18 – *NF2* p.E372*



	VAF (%)		# positive droplets	
	WES	ddPCR	Mut	WT
R1	37.90	40.80	481	698
R2	n/a	25.46	317	928
R3	11.70	15.65	237	1,277
R4	24.40	24.26	272	849
R5	24.70	28.05	373	957
gDNA	0.00	0.00	0	1,785
cfDNA	n/a	0.16	4	2,428

B Patient 27 – *AASDHPPT* p.C479Y



	VAF (%)		# positive droplets	
	WES	ddPCR	Mut	WT
R1	26.49	26.83	414	1,129
R2	36.21	31.69	308	664
R3	26.68	23.40	340	1,113
R4	27.18	26.25	358	1,006
R5	29.91	22.95	363	1,219
gDNA	0.00	0.00	0	2,086
cfDNA	n/a	2.96	39	1,279

Figure 4.16 - ctDNA detection in Patients 18 (A) and 27 (B). 1D amplitudes of mutant (blue) and WT (green) channels showing positive droplets. Also present are tables providing the ddPCR VAF and number of positive droplets for each respective run. For samples that had undergone WES, the VAFs were also shown.

4.3.2.3.3 Borderline ctDNA samples

The remaining two patients (24 and 34) did have one or more positive mutant droplets identified, but did not pass the thresholds required to be confident positives.

Patient 24 had 1 mutant/2,321 total positive droplets (0.04% VAF) in an initial run (see Supplementary Figure 1.2). There was enough remaining cfDNA/plasma for this cfDNA to be repeated with a higher amount of DNA. The rationale was that the quantity of mutant droplets would increase with an increase in template DNA, if the sample was a true positive. No change or decreased mutant positives would indicate these mutant droplets may represent false positive calls. The subsequent run for this patient had 1 mutant/3,227 total positives (0.03% VAF). Therefore this was ultimately deemed a ctDNA-negative patient (Figure 4.18A).

The assay for Patient 34 targeted an unusual variant found in *BAP1*. WES listed two independent variants at neighbouring base positions; a single deletion of an adenine and a G>T substitution. However upon inspection on IGV these two events likely represented a single genomic event, as there were no reads where the two variants were private (Figure 4.17).

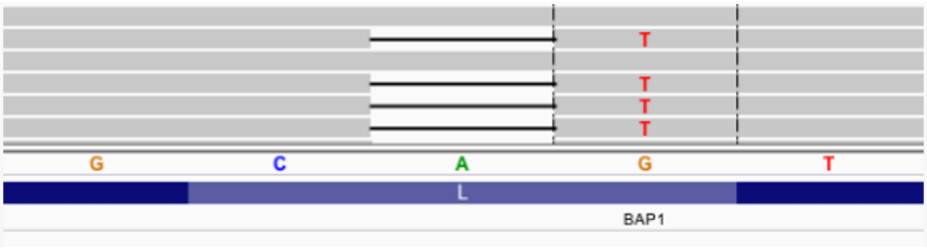
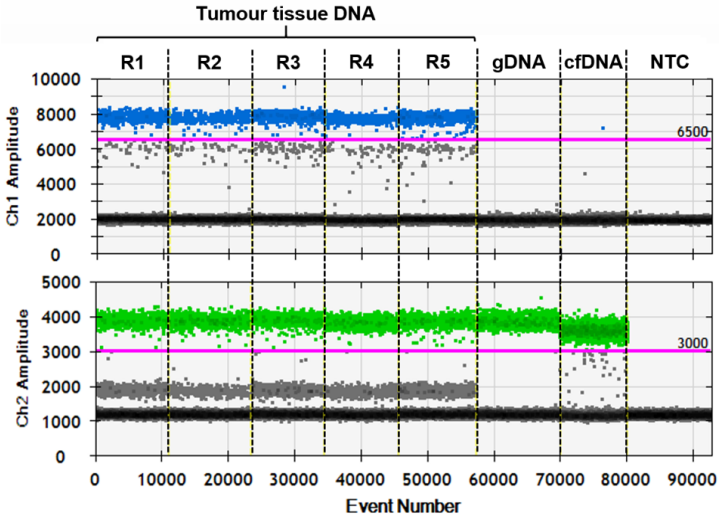


Figure 4.17 - Screenshot of *BAP1* variant on IGV. The grey horizontal lines represent individual sequencing reads, which display that the independently called events are coexisting and therefore likely to represent a single genomic event.

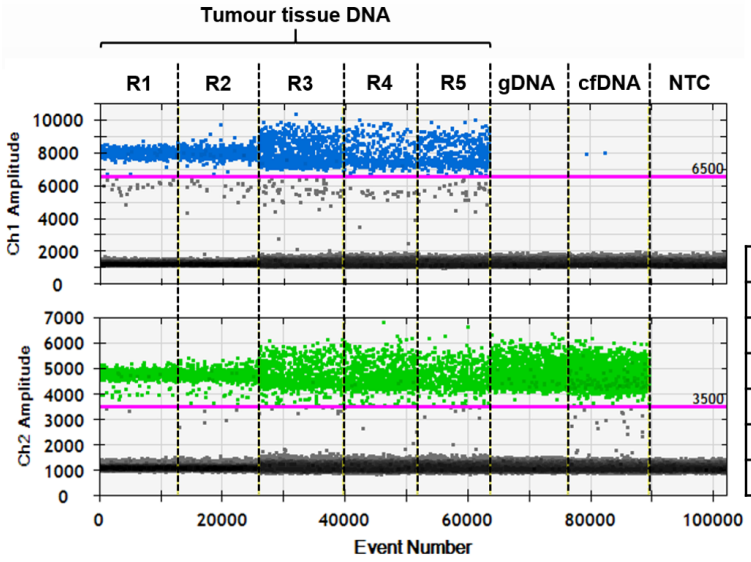
This frameshift was predicted as a Tier 1 driver event, but within plasma this variant did not pass the 3 mutant droplet threshold, having only 2 mutant droplets/2,353 total droplets (0.08% VAF) (Figure 4.18B). Alike Patient 24, there was enough remaining cfDNA for a repeat for this sample. Despite this second run having 87% more WT droplets (4,388), the number of mutant droplets decreased to a single positive mutant droplet (0.02% VAF) and likewise this was regarded as a ctDNA-negative sample (see Supplementary Figure 1.2).

A Patient 24 – *LRP6* p.I1185L



	VAF (%)		# positive droplets	
	WES	ddPCR	Mut	WT
R1	40.88	38.01	558	910
R2	25.00	21.87	376	1,343
R3	46.26	41.44	762	1,077
R4	30.00	27.33	461	1,226
R5	35.71	41.45	766	1,082
gDNA	0.00	0.00	0	1,754
cfDNA	n/a	0.03	1	3,226

B Patient 34 – *BAP1* p.L635Sfs*2



	VAF (%)		# positive droplets	
	WES	ddPCR	Mut	WT
R1	45.21	48.19	585	629
R2	53.41	53.26	629	552
R3	47.43	53.31	895	784
R4	27.98	30.11	349	810
R5	44.97	47.03	460	518
gDNA	0.00	0.00	0	1,646
cfDNA	n/a	0.08	2	2,351

Figure 4.18 - cfDNA analysis in Patients 24 (A) and 34 (B). 1D amplitudes of mutant (blue) and WT (green) channels showing positive droplets. Also present are tables providing the ddPCR VAF and number of positive droplets for each respective run. For samples that had undergone WES, the VAFs were also shown.

4.3.2.4 ctDNA status and survival

In summary, when the thresholding was implemented, this resulted in four patients being ctDNA-positive with the other seven being ctDNA-negative. Considering the poor survival observed in MPM, it is unsurprising that all 11 of the patients studied have since succumbed to their disease, with the longest time from surgery date to death being just 772 days (25.4 months) (Table 4.11).

Patient	Assay target	# generated droplets	# positive droplets		VAF (%)	Days between surgery and death
			Mut	WT		
6	***RNF187***	14,044	22	5,966	0.37	174 (5.7 m)
9	***SOX9***	12,676	9	3,463	0.26	124 (4.1 m)
12	***TRAF7***	10,397	0	4,720	/	622 (20.4 m)
18	***NF2***	14,043	4	2,428	0.16	460 (15.1 m)
23	***RNF213***	13,146	0	1,012	/	578 (19.0 m)
24	***LRP6***	10,093	1	3,226	0.03	389 (12.8 m)
27	***AASDHPPT***	11,356	39	1,279	2.96	186 (6.1 m)
33	***NUP98***	12,916	0	2,625	/	213 (7.0 m)
34	***BAP1***	13,146	2	2,351	0.09	616 (20.3 m)
37	***MYBPC1***	12,844	0	1,795	/	196 (6.4 m)
64	***BRCA1***	13,729	0	2,235	/	772 (25.4 m)

Table 4.11 - Summary of ddPCR cfDNA results with patient survival following surgery. Patients that were ctDNA-positive are shaded in green with the ctDNA-negative patients in red.

In order to assess whether ctDNA positivity was informative of survival in these patients, the time from surgery until death was investigated based on ctDNA status. The median survival of four ctDNA-positive patients was significantly shorter than that of seven ctDNA-negative patients (P = 0.047)(Figure 4.19). Cox regression analysis was also conducted, however this did not show significant findings (HR, 4.226; 95% CI 0.91-19.55; P = 0.065), due to a small sample size. In the four ctDNA-positive patients, VAF ranged from 0.15% to 2.76% (mean, 0.85%). In this small cohort, there was no clear link between ctDNA VAF and survival.

Accounting for any potential confounder variables in survival analyses was subsequently considered. As *CDKN2A* deletion is known to be associated with shorter survival in MPM (Hylebos *et al.* 2017), it was important to ensure the

shorter survival observed in the ctDNA-positive group was not simply due to these patients having *CDKN2A* deletion, but rather ctDNA detection being an independent prognostic marker. *CDKN2A* status was provided by Lee Brannan who is currently conducting phylogenetic analyses on MEDUSA WES (Figure 4.20A). *CDKN2A* status did not seem to be associated with ctDNA detection in this data, with 3/4 ctDNA-positive patients having *CDKN2A* deletion compared to 5/7 of the ctDNA-negative patients (Fisher's Exact test, P = 0.721) (Figure 4.20B). Likewise, when *CDKN2A* zygosity was considered there seemed no relationship between these factors (Figure 4.20C). When survival is assessed based on *CDKN2A* status, it indicates that ctDNA-positivity may be a stronger prognostic marker in these patients (Figure 4.20D).

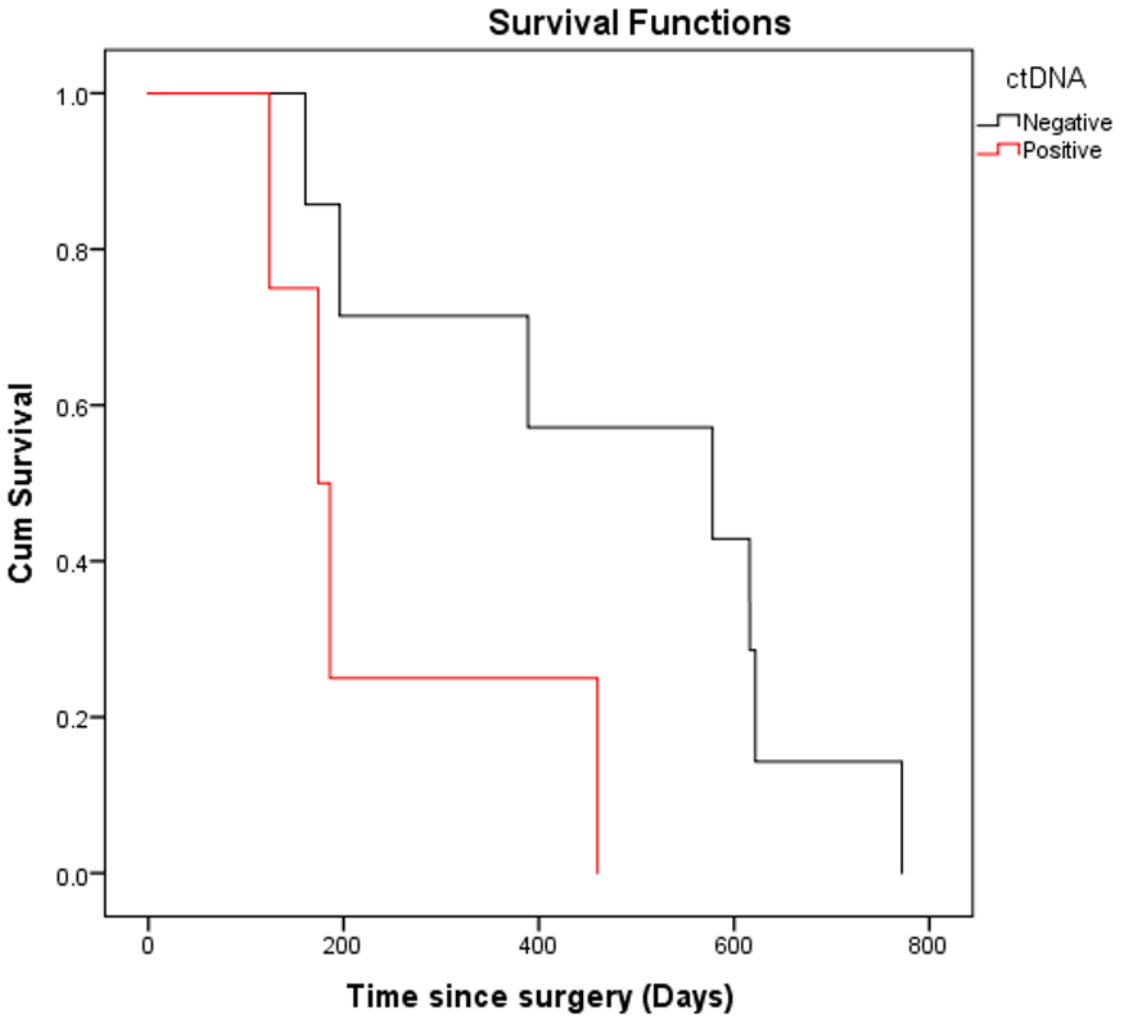


Figure 4.19 - Kaplan–Meier survival based on ctDNA status. Based on ddPCR ctDNA detection status, the survival in time since surgery (days) is shown for the 11 MEDUSA patients. ctDNA-positive patients are in red, ctDNA-negative patients are in black.

A

Patient	*CDKN2A* Status
6	Heterozygous deletion
9	Normal
12	Heterozygous deletion
18	Homozygous deletion
23	Heterozygous deletion
24	Homozygous deletion
27	Heterozygous deletion
33	Normal
34	Homozygous deletion
37	Homozygous deletion
64	Normal

B

		ctDNA +	ctDNA -	
CDKN2A	Deleted	3	5	8
	Normal	1	2	3
		4	7	11

Fisher's Exact test P = 0.721

C

		ctDNA +	ctDNA -
CDKN2A	Homozygous deletion	1	3
	Heterozygous deletion	2	2
	Normal	1	2

D

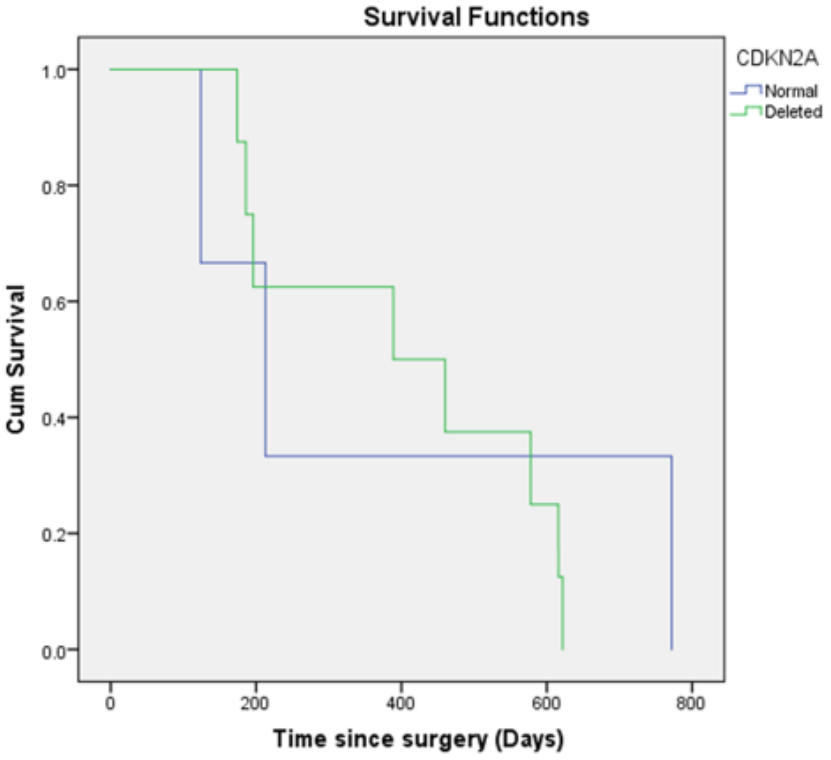


Figure 4.20 - *CDKN2A* vs ctDNA detection status. A) Table showing the *CDKN2A* status of the 11 MEDUSA patients, with green shading indicating ctDNA-positive and red indicating ctDNA-negative patients (sourced by Lee Brannan). B) Fishers Exact test analysis of ctDNA vs *CDKN2A* status. C) Table showing zygosity of *CDKN2A* vs ctDNA detection status. D) Kaplan-Meier based on *CDKN2A* status.

4.3.2.5 Factors that may influence ctDNA detection in MPM

Due to low sample numbers, the association of ctDNA detection is with clinicopathological variables was difficult. As 9/11 patients had stage IIIB disease, it was not possible to identify any potential link between stage and ctDNA detection.

Surprisingly, all of the ctDNA-negative patients (7/7) had lymph node involvement (all N2), while just one of the four ctDNA-positive patients (Table 4.12). This is potentially an interesting finding, but needs validating in a larger series of patients. Two out of four ctDNA-positive patients had a biphasic histology, compared to 2/7 of the ctDNA-negative patients.

Patient	Age at Surgery	Sex	Histology	TNM	IMIG Stage (8th Edition)	Neoadjuvant Chemotherapy	Days between surgery and death
6	72	M	epithelioid	T3, N2, M0	IIIB	n/a	174 (5.7 m)
9	78	F	biphasic	T3, N0, M0	IB	n/a	124 (4.1 m)
12	63	M	epithelioid	T3, N2, M0	IIIB	n/a	622 (20.4 m)
18	67	M	biphasic	T2, N0, M0	IB	n/a	460 (15.1 m)
23	63	M	epithelioid	T3, N2, M0	IIIB	n/a	578 (19.0 m)
24	71	M	biphasic	T3, N2, M0	IIIB	n/a	389 (12.8 m)
27	53	M	epithelioid	T4, N0, M0	IIIB	pem-cis 6 cycles + TroVax®	186 (6.1 m)
33	71	M	epithelioid	T3, N2, M0	IIIB	n/a	213 (7.0 m)
34	61	M	biphasic	T3, N2, M0	IIIB	n/a	616 (20.3 m)
37	74	M	epithelioid	T4, N2, M0	IIIB	n/a	196 (6.4 m)
64	69	M	epithelioid	T3, N2, M0	IIIB	n/a	772 (25.4 m)

Table 4.12 - Comparison between ctDNA detection status and clinicopathological variables. ctDNA positive (green) and negative patients (red) are indicated.

4.4 Discussion

Malignant pleural mesothelioma (MPM) is an unusual malignancy, characterised by localised invasion and a strong single association with an environmental carcinogen (asbestos) (Robinson 2012). A combination of lack of research characterising MPM and the resulting lack of targeted therapy successes have maintained dismal survival rates. Unlike better understood malignancies, MPM currently lacks significant means of patient stratification, with largely all patients receiving near-identical treatment; platinum and pemetrexed chemotherapy and/or surgical resection.

4.4.1 *CDKN2A* analysis using ddPCR

Homozygous deletion of *CDKN2A* is currently the main independent genomic prognostic marker in MPM, meaning that efficient detection of this alteration could stratify patients. In this section the differentiation between heterozygous and homozygous *CDKN2A* loss was difficult using ddPCR, due to stromal contamination affecting the *CDKN2A*/*RPPH1* ratios. As each sample has different fraction of healthy stromal cells, this assay was limited to defining between simply normal *CDKN2A* and deleted. If tumour microdissection was conducted that enriched for tumour cells, it is possible that ddPCR may be able to detect homozygous from heterozygous *CDKN2A* loss in MPM tumours. The problem with this methodology however, is that it is a time-consuming process and requires FFPE tissue as a starting material.

The 83bp *CDKN2A* assay was relatively successful at detecting *CDKN2A* deletions in the Fresh-Frozen DNA from MEDUSA patients, however for FFPE DNA its success faltered due to increased DNA fragmentation. The re-designing of the *CDKN2A* assay to a shorter length of 61bp improved the *CDKN2A*/*RPPH1* ratios for most patients, whereby normal *CDKN2A* status had a ratio of ~1.00, as desired. There were some outliers remaining with this shorter assay, whereby OncoScan™ detected homozygous deletion of *CDKN2A* in F4 and F6, but with the 61bp assay had ratios of ~1.00. These results emphasise how the use of multiple reference assays are required to improve assay accuracy. The OncoScan™ platform uses molecular inversion probes that targets across the genome, meaning that in essence this method has thousands of references for

each copy number assessment. In ddPCR, a single reference assay is multiplexed with the target assay, which relies on the assumption that the reference of choice does not undergo copy number changes. As MPM is characterised by copy number loss in *CDKN2A*, *BAP1* and *NF2*, potentially due to asbestos fibres interrupting chromosomes during cell division, it is possible that other genes may also be subject to copy number alteration, including "reference" genes. As previously mentioned, deletions at 14q have been reported in MPM, and therefore the results found in samples F4 and F6 may be explained by concurrent deletion of *CDKN2A* and *RPPH1* in these samples (Taniguchi *et al.* 2007). A future direction will be to run samples with additional reference assays to attempt to improve the accuracy of this technology.

Providing ddPCR results improve, one potential study could involve the detection of *CDKN2A* deletion in cells obtained from pleural fluid of patients with MPM. Pleural effusion is a common event observed in patients with MPM and is often present at diagnosis (Sneddon *et al.* 2018). Furthermore the analysis of pleural fluid for diagnosis is not uncommon in MPM, therefore an assay such as this could provide fast and accurate information on *CDKN2A* status. It is hoped that therapies exploiting *CDKN2A* loss will soon be available which could benefit the ~60-70% patients with this alteration (Bott *et al.* 2011, Illei *et al.* 2003).

4.4.2 ctDNA detection in MPM

This chapter investigated whether ctDNA can be detected in patients with MPM using a patient-specific targeted approach. cfDNA analysis is emerging as a vital tool for translational research in a number of malignancies, however at the time of conducting this work, there were no studies addressing its utility in MPM. Pre-surgical plasma obtained from eleven MPM patients was assessed for the presence of tumour-specific variants. By targeting a single clonal variant (found in every region of tumour tissue), this maximised the potential of detecting ctDNA within individual patient plasma.

Unfortunately, there were no shared clonal variants between the eleven patients by multiregional-WES, meaning that patient-specific assays were required to be designed. It remains to be seen whether the extent of interpatient heterogeneity

is maintained when phylogenetic analyses are conducted in more MEDUSA patients.

Four of eleven patients were ctDNA-positive, when using stringent thresholds (ctDNA positive = ≥ 3 mutant droplets + VAF ≥ 0.1%). At the time of writing this thesis (summer 2018), there has been a single publication also investigating ctDNA in MPM. The group reported ctDNA in 3/10 patients, either in one or both independent ddPCR runs, with VAFs ranging from 0.9% - 0.28% (Hylebos *et al.* 2018). However, the group's definition of a ctDNA-positive sample was simply the presence of one or more mutant droplets in at least one independent ddPCR run. If thresholds featured in this chapter were loosened in line with those of the publication, there would have been 6/11 ctDNA-positive patients found, with Patients 24 and 34 being deemed ctDNA-positive cases. Despite this, considering false positives are not infrequent using ddPCR, a three droplet threshold was considered the appropriate cut-off if taken forward clinically.

Another difference between the two studies was that Hylebos and colleagues searched for a variants based on WES of a single region of tumour tissue and therefore it is unknown whether these variants were clonal or subclonal (Hylebos *et al.* 2018). From studies in other malignancies, it has been shown that subclonal variants are more difficult to detect in cfDNA and are at lower VAFs in comparison to their clonal counterparts (Abbosh *et al.* 2017, Jamal-Hanjani *et al.* 2016). Yet, It is also worth mentioning that there seemed to be a lower amount of initial template DNA in the samples from the publication, with 13/20 runs having <1,000 WT positive droplets and therefore their limit of detection will have been compromised. With higher template DNA input, they may have detected more ctDNA-positive patients in their cohort.

Nevertheless, it is encouraging that two independent studies have shown similar findings; a majority of ctDNA-negative patients and low VAFs in those that were ctDNA-positive when tracking a single variant. There are now bespoke approaches that allow for the interrogation of several patient-specific variants, which may improve ctDNA detection rates in MPM (Abbosh *et al.* 2017).

Considering 9/11 patients analysed in this cohort had stage IIIB disease, of which only two of which had ctDNA detected, indicates that MPM may represent a more

difficult malignancy to examine using ctDNA analyses compared to other solid tumours, at least using current technologies.

4.4.3 Is ctDNA prognostic in MPM?

These preliminary data indicate that ctDNA-positivity may have prognostic significance in MPM. Although cox regression analysis was not significant, the median survival of ctDNA-positive patients was significantly shorter than those of ctDNA-negative patients, despite small sample numbers (n = 11). If this trend is maintained when the analysis is extended to more patients, there is potential value of cfDNA analysis in MPM as a method of patient stratification. Although sarcomatoid MPM were not a studied histology in this dataset, as this subtype has the poorest prognosis, the VAF of ctDNA detection in this group may be worthy of investigation (Ricciardi *et al.* 2018).

Considering the short survival observed in MPM, combined with the risk and associated recovery time with resection, ctDNA-positive patients may not see benefit from surgery and may be better served by receiving alternative treatment modalities that have improved risk-benefit. The majority of patients diagnosed with MPM are over 70 where surgery typically has an increased risk of complications (Sharkey *et al.* 2017, Williams *et al.* 2015).

One important consideration for this survival data was tumour *CDKN2A* status. Homozygous deletion of *CDKN2A* is associated with shorter survival in MPM and therefore it was important to assess whether the shorter survival observed for the ctDNA-positive group was not simply due to these patients also possessing *CDKN2A* deletion in their primary tumour (Hylebos *et al.* 2017). Assessment of *CDKN2A* status was inferred from multiregional tumour WES data (plus OncoScan™ in some cases) on these eleven patients and there does not seem to be a relationship between ctDNA detection and *CDKN2A* status, indicating that these variables may be independent prognostic factors in MPM. It will be interesting to see in larger sample numbers whether ctDNA detection is more prognostic than other markers such as tumour *CDKN2A* status.

Regarding clinicopathological variables, a surprising finding was that ctDNA-positivity may be more associated with a lack of lymph node involvement, rather than the opposite. ctDNA detection is typically associated with lymph node

metastasis in solid tumours, consistent with the increased metastatic potential of such cases (Abbosh *et al.* 2017, R. Wang *et al.* 2017). Yet, MPM may be an exception to the rule. All patients that were ctDNA-negative had N2 status. Studies have shown that N2 (and N1) nodal status are significantly associated with poorer survival in MPM patients following surgery and therefore it will be interesting to see whether this seemingly paradoxical trend continues in a larger cohort (Edwards *et al.* 2006, Friedberg *et al.* 2017, Sugarbaker *et al.* 1993).

The understanding of metastasis in MPM has been largely elusive. MPM is widely perceived as a locally invasive cancer and the precise cause of death is poorly understood. Despite this, a post-mortem study of 318 patients has shown that extrathoracic metastases are actually very common in MPM, with the liver (31.9%) and spleen (10.8%) being the most common metastatic sites (Finn *et al.* 2012). Whether ctDNA detection is more associated with patients that possess metastases could be another avenue to investigate. With national studies such as PEACE (Posthumous Evaluation of Advanced Cancer Environment) that conduct detailed analyses of multiple metastatic tumour sites in patients at port-mortem, the potential for mapping metastatic spread using ctDNA detection may be of interest. Alternatively, local recurrences such as those typically observed in MPM (e.g. contralateral lung pleura) may represent MPMs that are more difficult to detect in plasma.

4.4.4 Methodological considerations for ctDNA detection in MPM

Nevertheless, even if ctDNA-positivity was shown to be prognostic in MPM, there is an unavoidable hurdle that is needed to be overcome; the seeming lack of typical point mutation hotspots in MPM, coupled with high interpatient heterogeneity means that the design of a sequencing panel to generically (non-patient-specific) identify ctDNA across multiple patients would be difficult. Of course in this study, WES data of tumour tissue was required in order to identify a single variant to detect in plasma. However, clinically a broader approach that can be applied to cfDNA from multiple patients for ctDNA detection seems likely to be required. It seems that SNVs are not the dominant drivers of MPM, with instead large scale deletions in *CDKN2A*, *BAP1* and *NF2* dominating MPM genomics. Bott and colleagues showed in their cohort of 53 patients, that 83% (44/53) had loss (heterozygous or homozygous) of at least one of *CDKN2A*,

BAP1 or *NF2* (Bott *et al.* 2011). Unfortunately, deletion events are extremely difficult to detect in cfDNA, due to the high proportion of healthy cfDNA found in the plasma of most patients. In this chapter, it is clear that even when MPM ctDNA is detectable in the blood, it is at low levels (2.96% – 0.16% VAF), and so the detection of gene deletions would be near-impossible using current methods.

For cfDNA analysis, SNVs are the easiest targets to interrogate. Bueno and colleagues found point mutations in either *BAP1* or *NF2* (or both) in a third (72/216) of their large MPM patient cohort (Bueno *et al.* 2016). Therefore, there may be value in the generation of a cfDNA sequencing panel providing exome-wide coverage of both of these genes. Their data may also warrant the inclusion of *SETD2*, *TP53*, *DDX3X* and *TRAF7* which would have extended this to 49% (105/216) of their patients. Although several have not yet been investigated with regards to ctDNA status, MEDUSA has now conducted multiregional-WES on a total of 20 MPM tumours (including the 11 studied in this chapter). In support of Bueno and colleagues data, when MEDUSA multiregional-WES was analysed, 10 / 20 (50%) patients would have had ≥ 1 nonsynonymous SNV(s) within at least one of these six genes, and 5 / 20 (25%) if only clonal SNVs were included (Figure 4.21). Therefore there may be value in the design of a targeted NGS panel for the routine detection of ctDNA in parallel targets, instead of the single variants interrogated in this chapter (that requires prior tumour WES). It would be interesting to see whether ctDNA detection rate is improved when several clonal variants are analysed; this was clearly the case in cfDNA from patients with NSCLC for example (see Chapter 3, bespoke method) (Abbosh *et al.* 2017).

Although only 1/86 tumour regions had a *TP53* SNV found, the inclusion of *TP53* on a NGS panel would probably be warranted considering these seem to be more associated with sarcomatoid histology; a subtype not studied by MEDUSA (Bueno *et al.* 2016). As epithelioid MPM is the most studied, the abundance of *TP53* mutations are likely to be underrepresented in MPM as a whole. Furthermore, the aforementioned study showed that patients with *TP53* mutations had shorter survival.

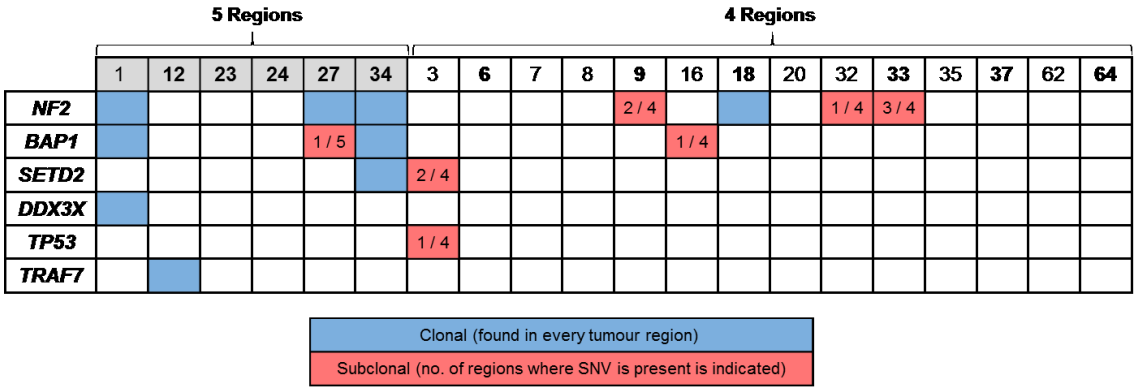


Figure 4.21 - Nonsynonymous SNVs within six genes present in 20 MEDUSA MPM tumours. SNVs that were found in every tumour region analysed (clonal; blue) are shown, as well as subclonal variants, with the number of regions involved indicated. The patients in bold represent those that have had ctDNA analysis to date.

4.5 Conclusions

This chapter has shown that ctDNA is detectable within the total cfDNA of patients with MPM. MEDUSA plans to extend cfDNA/ctDNA analysis to further patients, where it is hoped that ctDNA positivity proves to serve as a prognostic biomarker for this aggressive malignancy. Once this has been conducted, the association between ctDNA detection and clinicopathological variables could potentially uncover novel stratification methods, such as determining those patients that may not seek benefit from invasive surgery.

MEDUSA will endeavour to characterise the intratumour heterogeneity present in MPM to discover new targets that seed new clinical trials, potentially with cfDNA analysis serving as an ancillary test. Unlike in many solid tumours where tumour resection can be a curative treatment, in MPM long-term survivors are exceedingly rare, meaning should survival be improved it is likely to involve highly targeted, multimodal therapy. Henceforth, the use of minimally-invasive blood testing could provide an attractive method of monitoring MPM in patients receiving this therapy.

4.6 Future direction

Considering these preliminary data suggest that ctDNA detection could be informative of prognosis in patients with MPM, the extension of these analyses to larger sample numbers represents a natural progression for this project. Through power analysis of these data, it will be possible to determine the sample size required for this question to be addressed. Providing cfDNA-based analyses are informative of prognosis in these patients, an ultimate goal will be to design a custom SNV NGS panel for the routine detection of ctDNA in patients with MPM, therefore removing the requirement for patient-specific assay design; a process that relies on tumour WES data.

Finally, another avenue to explore could be to conduct genomic analyses on cells obtained from pleural fluid; an approach that could circumvent the low VAFs observed from plasma-based testing in this chapter. The molecular analysis of these cells could potentially provide detailed tumour information without the need for highly morbid techniques such as solid tissue biopsy and/or surgical resection. Depending on the tumour fraction amongst these cells, there is potential that unlike using plasma-based sources, analyses may be able to determine copy number loss and hence provide a more comprehensive genomic landscape for individual MPM tumours.

Chapter 5 Next-generation sequencing as a diagnostic tool: The PRIMINGS study

5.1 Introduction

5.1.1 A Diagnostic Problem

Tumours that arise following previous surgical resection for a primary malignancy may undergo biopsy to confirm the presence of metastasis. Often a confident diagnosis is straightforward *via* conventional light microscopy, particularly if the tumour is present in a lymph node draining the site of the primary tumour, or if the tumour displays morphological characteristics also seen in the primary disease, such as the production of melanin pigment in metastatic melanoma. For diagnoses where there is less certainty, immunohistochemistry is a valuable tool to decipher whether a second tumour represents metastatic spread, or an independent primary malignancy (Duraiyan *et al.* 2012). This is particularly true for biopsies of tumours arising in the lung; a common site of both primary and metastatic carcinoma (Budczies *et al.* 2015).

The differentiation between metastasis and independent primary malignancy is an important clinical distinction. For patients diagnosed with metastasis in a different organ system, this is indicative of an advanced stage of malignancy, which largely represents incurable disease. In these patients, palliative care is the predominant treatment decision. By contrast, patients diagnosed with an independent (second) primary tumour may be amenable to surgery and treatment with curative intent. Furthermore, optimal medical therapy is often alternate for a different tumour type; a key factor for improved patient outcomes. For instance, an independent lung adenocarcinoma diagnosed as stage IA would typically receive solely surgery (no adjuvant chemotherapy), whereas a colorectal metastases to a lung may receive improved outcomes with pulmonary metastastectomy with perioperative bevacizumab (Renaud *et al.* 2017).

Although immunohistochemistry can prove a useful diagnostic tool, no single immunohistochemical marker shows absolute sensitivity or specificity for individual tumour types. Therefore several immunohistochemical markers are often employed to reduce misdiagnosis. Nevertheless, immunohistochemistry panels can still be inconclusive, as some tumours occurring in different anatomical systems show similar or near-identical immunohistochemical profiles. Sometimes it is not possible to confidently differentiate between two independent

primary sites using immunohistochemistry alone, such as intestinal type ovarian and pancreatic mucinous tumours (Silverman *et al.* 2009), mucinous tumours arising from the lung and colon (Rossi *et al.* 2004), or squamous cell carcinomas arising from different organs (Pereira *et al.* 2011). In these cases, diagnosis is often made on the 'balance of probabilities' by a multidisciplinary team upon consideration of radiological, histological and clinical reports. Ultimately, this may lead to compromises in the selection of appropriate therapy, burden to the patient and potentially result in premature deaths.

Prior to the development of next-generation sequencing (NGS) technologies, alternative DNA-based molecular tests were shown to eliminate this ambiguity in certain cases (Brinkmann *et al.* 2004, Huang *et al.* 2001). However, complex methodologies including the need to perform tumour microdissection for enrichment of tumour cells represented a significant obstacle for translation into the clinic. A few alternative methods have been shown to be useful in the research setting, however these often exploit the presence or absence of molecular markers between tissues. For example, Nass and colleagues found that the expression of the microRNAs MiR-92b and miR-9/9* were specifically expressed by primary brain tumours, in contrast to metastatic brain lesions, using a combination of microRNA microarray and RT-qPCR (Nass *et al.* 2009). Unfortunately this serendipity is largely untrue for most solid tumours, meaning other avenues are required to be investigated.

5.1.2 Next-generation sequencing: A diagnostic tool?

The rapid rise of NGS methods in the last decade has proved valuable for tissue diagnosis. As discussed in previous chapters, tumours of a common origin share certain truncal driver mutations, meaning the identification of common mutations between two (or more) masses would be indicative of metastatic spread (Jamal-Hanjani *et al.* 2017). Alternatively, independent primary tumours are likely to possess discrete mutational profiles, based upon the statistical improbability of independently acquired, often single nucleotide changes (Figure 5.1). Identification of either a shared or unshared somatic lineage *via* NGS could not only aid in the accurate diagnosis of these tumours, but could also identify key targets for therapeutic intervention; actionable mutations are understandably a well-represented population found on commercially-available targeted

sequencing panels (tNGS). With the development of therapies that target specific mutated forms of proteins (e.g. G12C-specific KRAS inhibitors) becoming more prominent, there are additional potential benefits of tNGS for patient stratification moving forward (Janes *et al.* 2018). It also important to state that for the purposes of diagnosis, the identification of synonymous mutations, or any variants that may simply be "passengers", will serve the same utility to those of driver mutations.

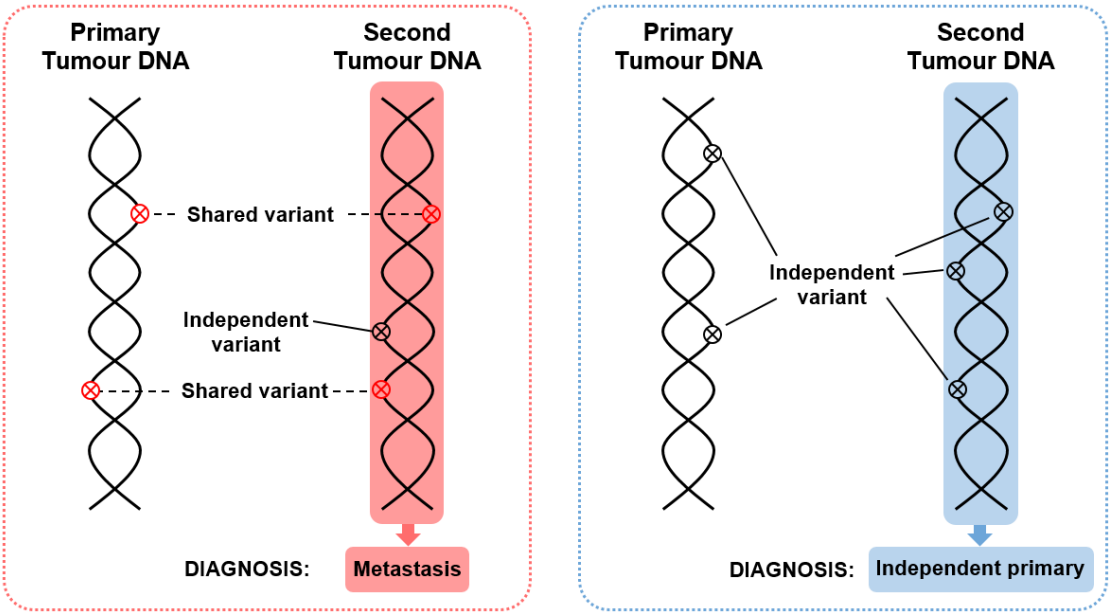


Figure 5.1 – Schematic of tNGS-based diagnosis of metastatic disease and independent primary cancers. The presence of ≥1 shared somatic variants would be indicative of shared clonal lineage, whereas tumours lacking shared variants are likely to be unrelated malignancies.

Primary and metastatic tumour sample pairs from a large series of patients have been shown to contain matching somatic variants using a commercial tNGS panel (Goswami *et al.* 2015). In this study, 85.7% (227/265) of the primary-metastatic pairs possessed identical mutation profiles, with the remaining 14.3% (38/265) featuring one or more discordant mutations. In all, 94% (249/265) of the pairs analysed had at least one shared mutation, providing proof of principle that this panel (Ion AmpliSeq™ Cancer Panel v1) was sufficiently comprehensive for the analysis of shared clonality across a broad spectrum of tumour types. It is hoped that newer, broader panels that cover more of the genome will be able to improve upon this percentage and ultimately prove to be a highly accurate tool in the clinical setting. This study also proposed that the genomic divergence observed between primary and metastatic tumours could be descriptive of clonal evolution, highlighting mutations in certain genes being more commonly associated with

metastasis (e.g. *TP53*, *PIK3CA* and *SMAD4*). They noted that their data displayed occasions where the absence of a hotspot mutation in *TP53* in a metastatic sample, could be attributed to a loss of heterozygosity in the gene; an event that must be considered when analysing such data.

Alike targeted methods, whole exome sequencing (WES) has also been employed to identify shared and unshared clonality between anatomically distinct tumours, including bilateral breast carcinomas (Bao *et al.* 2015). However, the cost of this technology and complexity of the analysis of data it generates, makes this method clinically inefficient at present. Admittedly, with the undoubted development of more user-friendly analysis pipelines, accompanied by a reduction in cost over time, WES will be a very attractive technology to utilise for diagnostic means in the future. WES has important advantages, as it would provide a complete protein-coding profile of tumours, meaning false negatives would be minimised and diagnostic confidence would be improved. Nevertheless, tNGS is currently the most attractive methodology for investigating clonality, as it is well suited to the heavily-fragmented DNA derived diagnostic formalin fixed paraffin embedded (FFPE) tissues, as well as producing less complex analysis (de Leng *et al.* 2016) . Although these previous case reports and studies suggest great promise for the clinical application of tNGS to help resolve this diagnostic uncertaintly, this approach has not yet been assessed using a suitable series of clinical cases to support its diagnostic value.

In this chapter, 20 pairs of anatomically separate tumours from patients with either unambiguous dual primary malignancies or unambiguous metastatic disease, where that distinction was made on the basis of tissue morphology and immunohistochemistry, were tested using a commercially available tNGS cancer panel; the Ion AmpliSeq™ Cancer Hotspot Panel v2 (CHPv2). The sequencing data was used to test whether somatic variants identified can be used to predict shared lineage and could therefore be beneficial in distinguishing metastatic disease from independent primary malignancies. Furthermore, five cases of histologically ambiguous tumours were also tested using this workflow, using the CHPv2, as well as with a recently launched diagnostically approved panel, the Oncomine™ Solid Tumour DNA Kit (OST) to assess the feasibility of clinical adoption.

5.2 Aims and objectives

The aims of this chapter were to:

i) Successfully inform the diagnosis of secondary tumours (defined as, either new primary malignancies or metastases), based on tNGS data alone, prior to receiving histopathological/immunohistochemical diagnosis (unambiguous cases);
ii) Determine the feasibility of tNGS for diagnosis in clinically relevant, ambiguous cases; and
iii) Determine an optimal tNGS method for the detection of shared and unshared clonality between tumours in a clinical setting.

The objectives were to:

i) Obtain FFPE tissue from archival cases including normal lymph node tissue that are unequivocally metastatic or new primary examples respectively;
ii) Following an assessment of DNA quality, conduct tNGS using the Ion AmpliSeq™ Cancer Hotspot Panel v2 (CHPv2) on the Ion PGM™ system; and
iii) Compare the variant detection between the research-orientated CHPv2 and the CE/IVD-marked Oncomine™ Solid Tumour DNA Kit (OST)* in clinically relevant, ambiguous cases.

*'IVD' (*in vitro* diagnostics) devices are regulated devices for the analysis of human tissues for diagnostic purposes. A 'CE' mark (Conformité Européene) denotes a declaration that the product complies with the essential requirements of the relevant European health, safety and environmental protection legislation (http://www.presafe.com/standard/ivdce-marking).

5.3 Results

5.3.1 Study overview

This study was separated into two chronological phases; the first assessed the accuracy of tNGS-based diagnosis (Phase A), and the second tested this technology in clinically applicable cases using two separate tNGS methods (Phase B). The study overview is summarised in Figure 5.2.

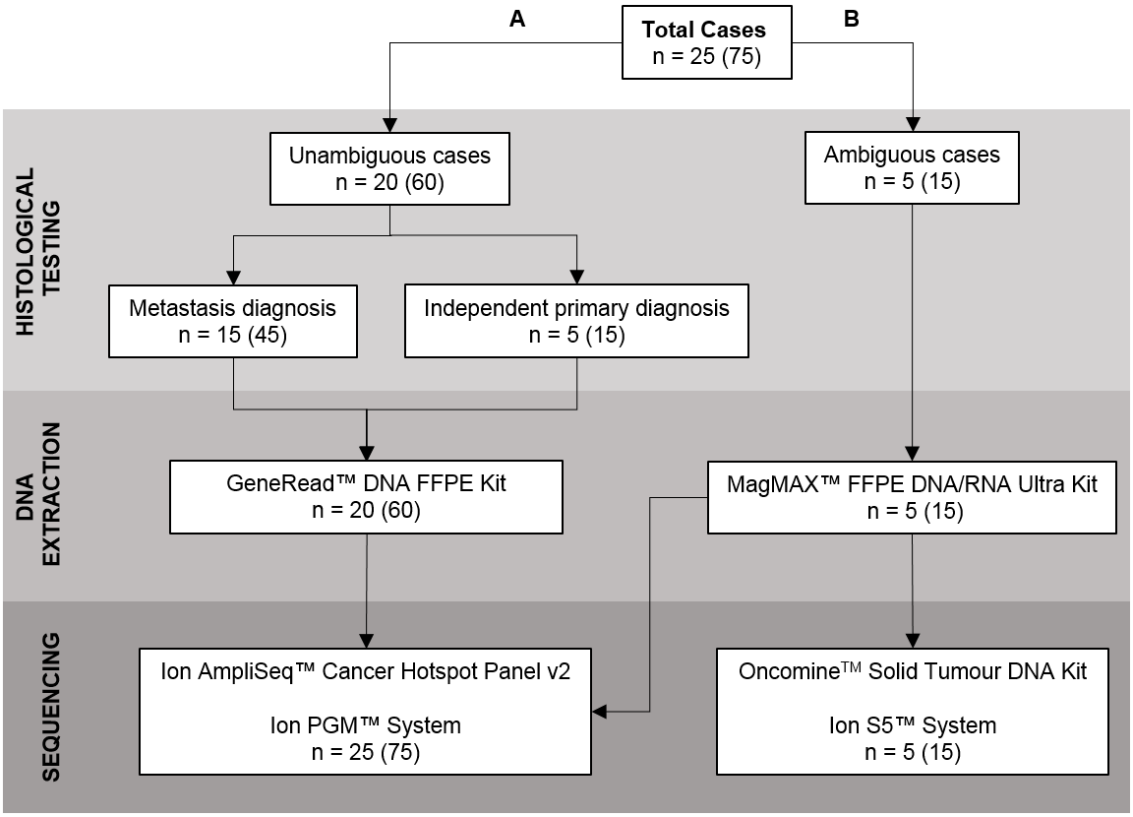


Figure 5.2 – Schematic summarising the PRIMINGS study. The number of cases are specified, with the number of DNA samples analysed in brackets (three samples per case). Phase A features cases that are unequivocally either metastatic or independent primary cases. By contrast Phase B are cases are clinically-relevant ambiguous cases that this study aims to resolve. The DNA extraction and sequencing methods are stated.

Each case was comprised of three samples; a primary tumour, a second tumour (either metastatic or independent primary) and a healthy tissue sample (either an uninvolved lymph node, or histologically confirmed normal tissue). Healthy tissue was a prerequisite for the study, as this acts as an effective filter for germline variants that may otherwise be misidentified as somatic variants, especially in samples that are high in tumour cellularity and coincidentally may have variants

called at ~50% variant allele frequency (VAF), alike a germline heterozygous SNP (Figure 5.3).

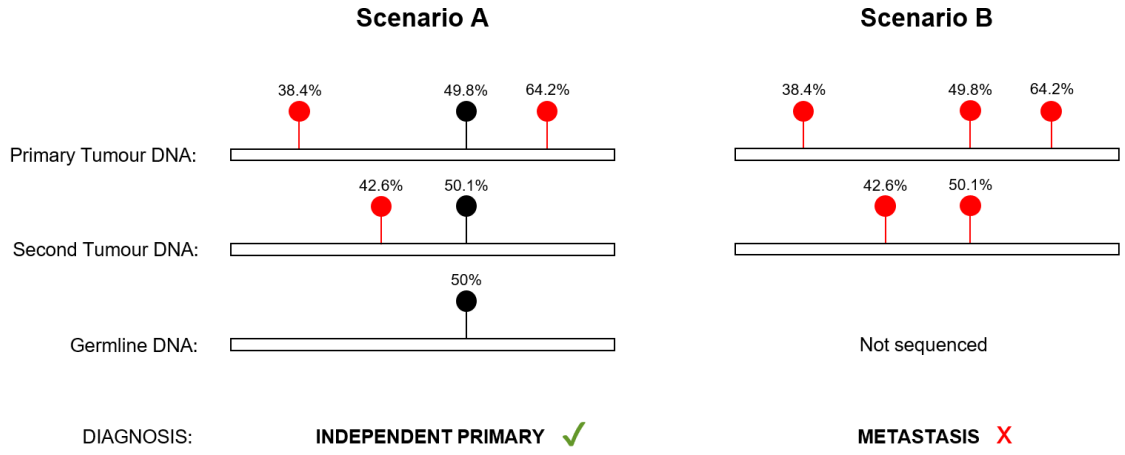


Figure 5.3 – Rationale for the inclusion of germline DNA for NGS. In scenario A, sequencing germline DNA enables identification of a heterozygous germline variant (black; ~50% VAF), in addition to unshared somatic variants in tumour samples (red), thus diagnosing two independent tumours. In scenario B, where germline DNA is not analysed, the germline variant is incorrectly deemed somatic and one that is present in both tumour samples, potentially misdiagnosing the second tumour as a metastasis. VAFs of variants are indicated above each lollipop.

5.3.2 Sample isolation and assessment of DNA quality

Following inspection by a consultant histopathologist (Dr David Moore), regions of tissue containing the highest tumour content were accessed using a 1mm corer and extracted using either the GeneRead™ DNA FFPE Kit (unambiguous cases) or an automated protocol using the MagMAX™ FFPE DNA/RNA Ultra Kit (ambiguous cases). A different extraction method was used for the ambiguous cases, as an automated process became available during this project; an aspect that is generally favoured for clinical implementation, as minimal human involvement is ideal for consistency.

DNA quality was assessed *via* a DNA integrity assay as formalin fixation introduces cross-links into DNA, which often results in a high degree of fragmentation; not unlike the extent observed in cfDNA. Highly fragmented DNA could potentially reduce the quality of tNGS. Therefore, using two real-time qPCR assays of differing amplicon lengths (95bp & 186bp) targeting a housekeeping gene *GAPDH*, the DNA samples were quantified and a DNA integrity index was calculated (for oligonucleotide sequences, see section 2.4.1.1). Sample concentrations and calculated DNA integrity index for each sample are found in the Appendix (Supplementary Table 1.3). Although these were different samples,

the two extraction methods had differing DNA recovery, with the GeneRead™ kit (60 samples) typically having higher recovery of DNA (95bp assay mean concentration, 113.2 ng/µL vs. 39.6 ng/µL) while the MagMAX™ kit (15 samples) had a higher mean DNA integrity index (mean, 0.31 vs. 0.25).

Results of the DNA integrity assay were used to prevent more fragmented samples from having poorer sequencing depth, and so in these cases, a higher amount of DNA was used during library preparation to minimise this occurrence (Table 5.1).

DNA integrity	DNA Integrity value range	DNA quantity used for library preparation (ng)
High	0.4 – 1	10
Medium	0.2 – 0.4	12.5
Low	0 – 0.2	15

Table 5.1 - DNA quantity used for library preparation based on integrity assay. The DNA quantity used for library preparation was based on the ng/µL values using the 95bp *GAPDH* assay (qPCR).

This strategy was considered effective as no correlation (R^2 =0.002591) was found between tNGS mean depth and the DNA integrity index of respective samples (Figure 5.4).

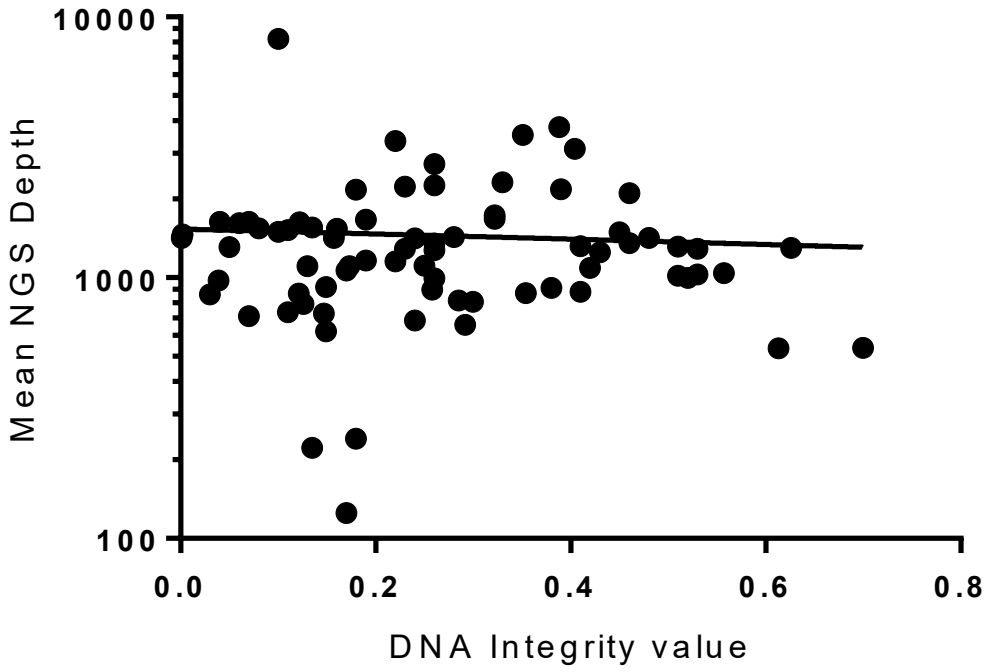


Figure 5.4 - Correlation between Mean NGS depth and DNA integrity value. The DNA quality of extracted FFPE DNA assessed by an in-house 95bp/186bp qPCR assay, was compared with observed sequencing depth of respective samples.

5.3.3 Targeted next-generation sequencing of diagnostically unambiguous cases (Phase A)

DNA samples from all unambiguous cases (n = 60) were sequenced using the CHPv2 on the Ion PGM™ system. This panel was selected as it covers ~2,800 COSMIC hotspot mutations in 50 commonly mutated oncogenes and tumour suppressor genes, thereby providing a relatively broad coverage of common mutations across multiple cancers. Sequencing data was blinded from the related clinical diagnostic information, to prevent bias. Diagnosis from the tNGS data alone was submitted to a consultant pathologist, Dr David Moore, who then released the pathological diagnosis, allowing the tabulation of findings.

Clinicopathological characteristics of the patients are shown in Table 5.2. Twelve different primary (primary & subsequent primary) tumours were represented, with colorectal adenocarcinomas being the most common, featuring in eight cases. Nine of the cases were anatomically involving the lung, either as a metastatic (n = 5) or as a primary tumour site (n = 4).

Case	Primary tumour sample	Metastatic site	Sex	Age at diagnosis of first tumour	Time lapse between tumours (months)
U1	colorectal adenocarcinoma	vulva	F	48	14
U2	endometrial adenocarcinoma	vulva	F	58	48
U3	colorectal adenocarcinoma	lung	M	40	11
U4	oesophageal adenocarcinoma	lung	M	63	5
U5	colorectal adenocarcinoma	ovary	F	34	0
U6	breast carcinoma	omentum	F	49	190
U7	duodenal adenocarcinoma	lung	F	71	21
U8	cervical carcinoma	vulva	F	29	23
U9	colorectal adenocarcinoma	bone	F	79	6
U10	ovarian serous adenocarcinoma	rectum	F	46	0
U11	breast carcinoma	bone	F	44	68
U12	colorectal adenocarcinoma	pancreas	F	53	9
U13	pancreatic adenocarcinoma	peritoneum	F	68	12
U14	oesophageal adenocarcinoma	lung	M	55	2
U15	urothelial carcinoma	lung	M	69	16
Case	**Primary tumour sample**	**Secondary tumour sample**	**Sex**	**Age at diagnosis of first tumour**	**Time lapse between tumours (months)**
U16	lung adenocarcinoma	colorectal adenocarcinoma	F	57	1
U17	colorectal adenocarcinoma	lung adenocarcinoma	F	78	4
U18	colorectal adenocarcinoma	hepatocellular carcinoma	M	66	22
U19	breast carcinoma	lung adenocarcinoma	F	64	184
U20	clear cell renal cell carcinoma	lung adenocarcinoma	M	80	16

Table 5.2 - Clinical details from the 20 unambiguous cases in the study. U1 - U15 represent patients with confirmed metastases (red). U16 - U20 are those with confirmed independent primary tumours (blue).

From the tNGS, the most commonly altered genes amongst the samples (two per case) were *TP53*, *KRAS* and *PIK3CA* with 22 (34%), 13 (20%) and 8 (13%) variant calls respectively. The majority of variants in this cohort were missense mutations (75%), with synonymous mutations being the rarest (5%) (Figure 5.5). As mentioned, for the purposes of this chapter, all somatic variants are included, not just those that are likely driver mutations.

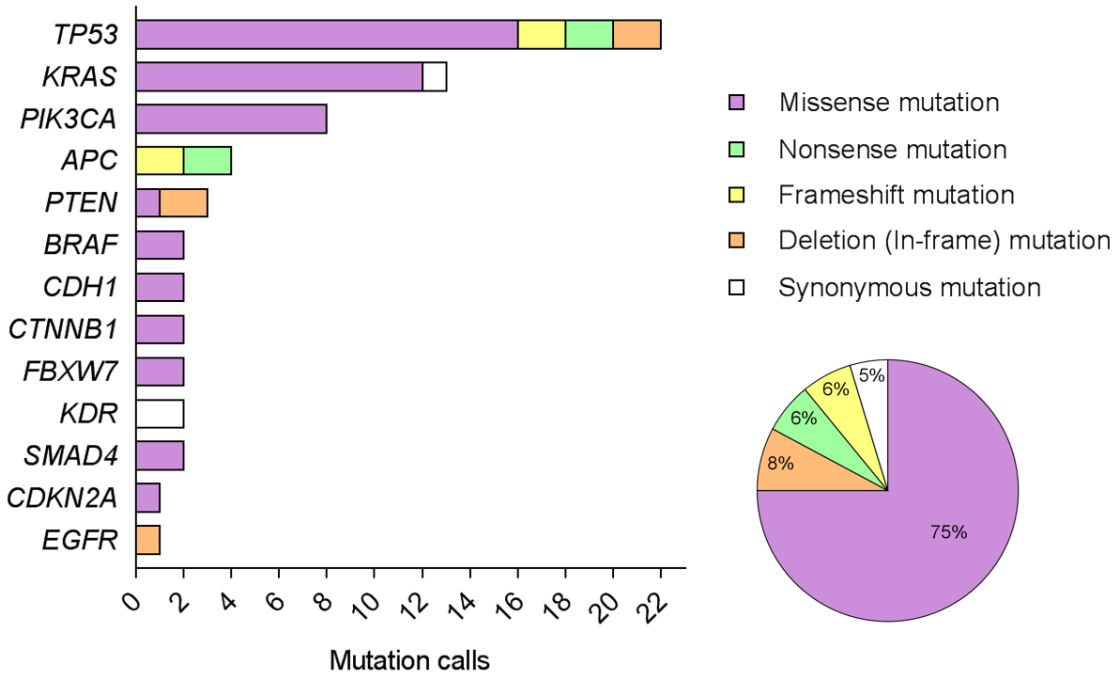


Figure 5.5 - Summary of the variants called in the diagnostically unambiguous cohort. Schematic showing in descending frequency, the variant calls respective to gene in the unambiguous cases (n = 20), with the type of mutation specified. The proportion of each mutation type is depicted in the pie chart.

Fourteen of the 15 pairs of matched primary and metastatic tumour pairs had one or more somatic variants identified (Table 5.3). One case had no somatic variants identified in either tumour sample and was therefore uninformative with this panel (Table 5.3, Case U15). In all, 5/14 metastatic cases shared two or more variants, with the other 9 positive cases sharing a single genomic variant. Three other variants amongst the 14 positive cases were discordant between tumour pairs, with two being found in only the primary tumour and the remaining variant within solely the metastatic sample (Table 5.3; U2, U3 and U6). Interestingly, a metastatic colorectal adenocarcinoma (U3) was shown to have an absence of an *APC* mutation in the metastatic sample; a mutation typically considered an early

(clonal) somatic event. A potential explanation for the loss of *APC* mutation in the metastasis, could be loss of heterozygosity (LOH) of the copy containing this allele. Of the 14 positive cases, Case U8 is one that potentially represents a false positive; the single shared variant in *KDR* was found at around 50% VAF in both tumours, however was not identified in the uninvolved lymph node sample and therefore was called. However the LN sample had poor sequencing depth (mean depth, 125X), meaning this variant could have been missed in this sample, leading to the false impression it was a somatic event.

In contrast, all of the independent primary cases (n = 5) exhibited discrete mutational profiles, with no shared variants identified between tumour pairs (Table 5.4). Four of these cases contained more variants in their subsequent primary sample, respective to their initial tumour. There were a few occasions where primary and subsequent primary DNA samples contained shared variants at VAFs resembling heterozygous (~50% VAF) or homozygous (~100% VAF) SNPs, yet were absent on the Single Nucleotide Polymorphism Database (dbSNP). However these were also found within their matching uninvolved lymph node DNA (with similar VAFs; 50% or 100%) and therefore were interpreted as a non-somatic (germline) event.

Position	Ref	Variant	Type	Gene	COSMIC ID	Primary Tumour VAF	Secondary Tumour VAF	CDS change	AA change
Case U1									
chr5:112174631	C	T	SNV	***APC***	COSM13125	**31.4%**	**78.3%**	c.3340C>T	p.R1114*
chr7:140453154	T	C	SNV	***BRAF***	COSM467	**19.3%**	**38.5%**	c.1781A>G	p.D594G
chr12:25398266	G	T	SNV	***KRAS***	COSM542	**24.3%**	**56.4%**	c.53C>A	p.A18D
chr17:7578263	G	A	SNV	***TP53***	COSM99668	**29.5%**	**79.0%**	c.586C>T	p.R196*
Case U2									
chr3:178936082	G	A	SNV	***PIK3CA***	COSM760	**27.7%**	**19.7%**	c.1624G>A	p.E542K
chr10:89624254	AGC	-	DEL	***PTEN***	COSM921053	**52.1%**	**37.2%**	c.28_30delAGC	p.S10delS
chr3:41266104	G	A	SNV	***CTNNB1***	COSM5671	**26.8%**	-	c.101G>A	p.G34E
Case U3									
chr5:112175676	AG	-	DEL	***APC***	COSM18873	**36.9%**	**-**	c.4385_4386delAG	p.S1465fs*3
chr12:25398284	C	G	SNV	***KRAS***	COSM522	**50.7%**	**50.6%**	c.35G>C	p.G12A
chr17:7579401	A	-	DEL	***TP53***	-	**68.1%**	**69.5%**	c.286delT	p.S96fs*27
Case U4									
chr17:7578384	GCAGCGCTC ATGGTGGGG	-	DEL	***TP53***	COSM43570	**54.3%**	**36.0%**	c.529_546del18	p.P177_C182delPHHERC
chr18:48591919	G	A	SNV	***SMAD4***	COSM14122	**60.3%**	**42.7%**	c.1082G>A	p.R361H
Case U5									
chr16:68847269	C	T	SNV	***TP53***	COSM6932	**36.5%**	**33.2%**	c.733G>A	p.G245S
chr17:7577548	A	C	SNV	***CDH1***	-	**46.1%**	**45.8%**	c.1191A>C	p.K397N
Case U6									
chr3:178936091	G	A	SNV	***PIK3CA***	COSM763	**21.9%**	**33.7%**	c.1633G>A	p.E545K
chr3:178952006	T	C	SNV	***PIK3CA***	COSM17444	-	**34.7%**	c.3061T>C	p.Y1021H
Case U7									
chr17:7578406	C	T	SNV	***TP53***	COSM10648	**8.4%**	**6.7%**	c.524G>A	p.R175H
Case U8									
chr4:55946171	G	A	SNV	***KDR***	COSM35855	**50.7%**	**53.2%**	c.4008C>T	p.T1336T
Case U9									
chr17:7577094	G	A	SNV	***TP53***	COSM10704	**20.1%**	**20.6%**	c.844C>T	p.R282W
Case U10									
chr17:7577120	C	T	SNV	***TP53***	COSM10660	**68.1%**	**80.7%**	c.818G>A	p.R273H
Case U11									
chr3:178936091	G	A	SNV	***PIK3CA***	COSM763	**41.8%**	**29.1%**	c.1633G>A	p.E545K
Case U12									
chr17:7577130	A	T	SNV	***TP53***	COSM43809	**61.7%**	**53.5%**	c.808T>A	p.F270I
Case U13									
chr12:25398284	C	A	SNV	***KRAS***	COSM520	**11.2%**	**18.7%**	c.35G>T	p.G12V
Case U14									
chr17:7577142	C	T	SNV	***TP53***	COSM10794	**70.4%**	**45.6%**	c.796G>A	p.G266R
Case U15									
No variants identified									

Table 5.3 - Summary of the variants identified in the fifteen primary-metastasis pairs. Details of each somatic variant identified within each respective case, including COSMIC ID (where applicable) and the variant allele frequency (VAF) within each tumour sample. Any variants identified in germline tissue (uninvolved lymph node) were excluded.

Position	Ref	Variant	Type	Gene	COSMIC ID	Primary Tumour VAF	Secondary Tumour VAF	CDS Mutation	AA Mutation
Case U16									
chr17:7577094	G	A	SNV	***TP53***	COSM10704	**27.0%**	-	c.844C>T	p.R282W
chr7:55242468	ATTAAGAG AAGCAAC	-	DEL	***EGFR***	COSM23571	**31.0%**	-	c.2238_2252del15	p.L747_T751delLREAT
chr17:7577124	C	T	SNV	***TP53***	COSM99950	-	**19.7%**	c.814G>A	p.V272M
chr4:153249384	C	T	SNV	***FBXW7***	COSM117308	-	**18.5%**	c.1154G>A	p.R385H
chr12:25378562	C	T	SNV	***KRAS***	COSM19404	-	**33.9%**	c.436G>A	p.A146T
Case U17									
chr5:112175212	AAAAG	-	DEL	***APC***	COSM18764	**68.1%**	-	c.3921_3925delAAAAG	p.E1309fs*4
chr12:25398284	C	G	SNV	***KRAS***	COSM522	**32.7%**	-	c.35G>C	p.G12A
chr9:21971199	C	T	SNV	***CDKN2A***	COSM3737053	-	**16.3%**	c.159G>A	p.M53I
chr12:25398285	C	A	SNV	***KRAS***	COSM516	-	**18.2%**	c.34G>T	p.G12C
chr17:7577118	C	A	SNV	***TP53***	COSM165075	-	**13.1%**	c.820G>T	p.V274F
Case U18									
chr12:25398281	C	T	SNV	***KRAS***	COSM532	**28.8%**	-	c.38G>A	p.G13D
chr10:89692883	C	T	SNV	***PTEN***	COSM5078	**60.4%**	-	c.367C>T	p.H123Y
chr17:7578403	C	T	SNV	***TP53***	COSM10687	**57.6%**	-	c.527G>A	p.C176Y
chr3:41266107	T	G	SNV	***CTNNB1***	COSM5674	-	**37.8%**	c.104T>G	p.I35S
Case U19									
chr3:178936082	G	A	SNV	***PIK3CA***	COSM760	**39.5%**	-	c.1624G>A	p.E542K
chr12:25398284	C	T	SNV	***KRAS***	COSM521	-	**34.5%**	c.35G>A	p.G12D
chr4:153247324	C	T	SNV	***FBXW7***	-	-	**12.7%**	c.1238G>A	p.C413Y
Case U20									
chr12:25398281	C	A	SNV	***KRAS***	COSM534	-	**31.7%**	c.38G>T	p.G13V
chr12:25398295	T	G	SNV	***KRAS***	COSM1360891	-	**32.2%**	c.24A>G	p.V8V

Table 5.4 - Summary of the variants identified in the five primary - primary pairs. Details of each somatic variant identified within each respective case, including COSMIC ID (where applicable) and the variant allele frequency (VAF) within each tumour sample. Any variants identified in germline tissue (uninvolved lymph, node) were excluded.

As validation of tNGS findings, a number of variants were also assessed using wet-lab validated Droplet Digital PCR (ddPCR) assays. A total of 14 variants encompassing 12/20 cases were studied, where both methods showed much reproducibility (R^2 = 0.9889; p < 0.0001) (Figure 5.6). Variants that were not validated was due to either a lack of DNA or unavailability of respective wet-lab validated ddPCR assays.

Although analyses need to be expanded to a larger cohort, the results of the unambiguous group allowed the CHPv2 to be assessed as a diagnostic test. If case U15 was considered as a 'false negative', then the CHPv2 had a sensitivity and specificity of 93.3% and 100% respectively (Table 5.5). The misdiagnosis of independent primary tumours as metastases are of particular concern in this study, as 'false positives' could lead to patients being administered solely palliative treatment in preference to potentially curative surgery. However encouragingly in this dataset there was a Positive Predictive Value of 100%. The Negative Predictive Value (NPV) was 83.3%.

		Histopathological diagnosis		
		Metastasis	**No Metastasis**	
CHPv2 diagnosis	**Metastasis (≥ 1 shared variants)**	**True Positive (TP)** **14**	**False Positive (FP)** **0**	**Positive Predictive Value (PPV)** **= TP / (TP + FP)** **= 14 / (14 + 0)** **= 100%**
	No metastasis (no shared variants)	**False Negative (FN)** **1**	**True Negative (TN)** **5**	**Negative Predictive Value (NPV)** **= TN / (FN + TN)** **= 5 / (1 + 5)** **≈ 83.3%**
		Sensitivity **= TP / (TP + FN)** **= 14 / (14 + 1)** **≈ 93.3%**	**Specificity** **= TN / (FP + TN)** **= 5 / (0 + 5)** **= 100 %**	

Table 5.5 - Assessment of the CHPv2 as a diagnostic test in the 20 unambiguous cases. Table summarising the performance of CHPv2-based diagnosis, compared to histopathological diagnosis, presuming the latter is correct. The sensitivity, specificity, and positive/negative predictive values are also shown.

A

		NGS VAF (%)		ddPCR VAF (%)	
Case	Mutation	Primary Tumour	Second Tumour	Primary Tumour	Second Tumour
U1	*TP53* p.R196*	29.5	79.0	27.8	78.5
U2	*PIK3CA* p.E542K	27.7	19.7	20.5	17.6
U5	*TP53* p.G245S	36.5	33.2	35.6	28.9
U6	*PIK3CA* p.E545K	21.9	33.7	20.6	31.4
U7	*TP53* p.R175H	8.4	6.7	5.4	3.9
U9	*TP53* p.R282W	20.1	20.6	17.0	17.6
U10	*TP53* p.R273H	68.1	80.7	66.3	77.0
U11	*PIK3CA* p.E545K	41.8	29.1	38.7	22.2
U16	*TP53* p.R282W	27.0	0	22.6	0
U17	*KRAS* p.G12A	32.7	0	32.5	0
	KRAS p.G12C	0	18.2	0	15.3
U18	*KRAS* p.G13D	28.8	0	27.5	0
U19	*PIK3CA* p.E542K	39.5	0	34.2	0.03
	KRAS p.G12D	0	34.5	0	34.0

B

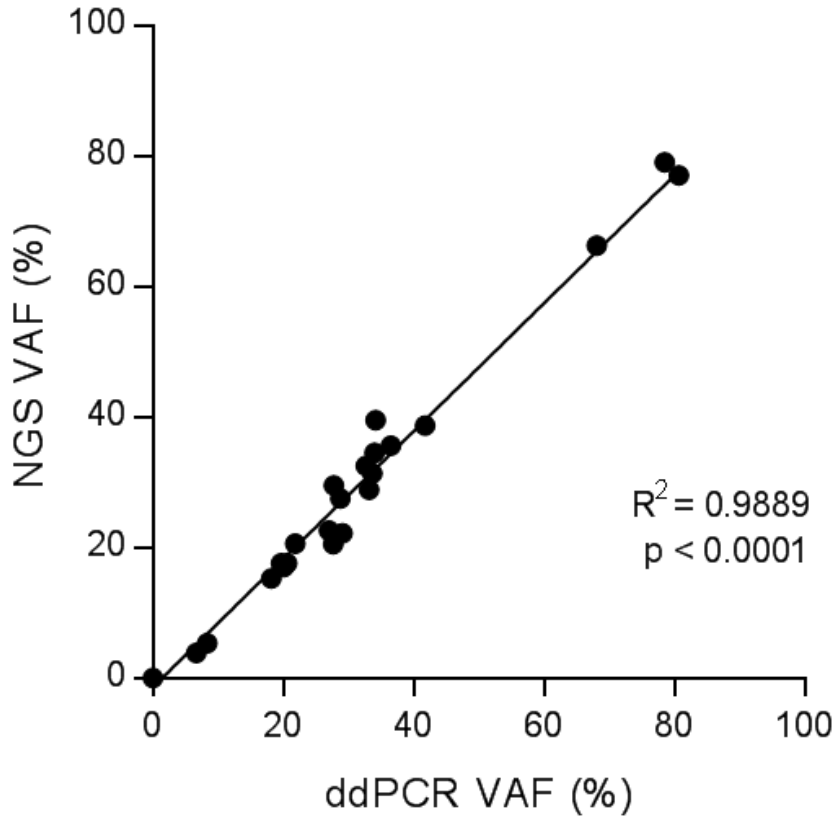


Figure 5.6 - Concordant variants found using NGS and ddPCR. (A) Table summarising the variant allele frequencies (VAFs) identified between NGS and ddPCR in both the primary tumour and second (either independent primary or metastasis) tumour samples. (B) Graph displaying the correlation between NGS and ddPCR.

5.3.4 Investigation of diagnostic utility of tNGS in ambiguous cases (Phase B)

As proof of principle had been established, the use of tNGS for the diagnosis of tumour pairs was subsequently investigated in five clinically relevant ambiguous case studies (A1-A5). The clinicopathological characteristics of these cases are outlined in Table 5.6. All of these samples were independently sequenced using both the CHPv2 and the Oncomine^{TM} Solid Tumour (OST) panel, as the latter represents a potential candidate for clinical implementation due to its CE/IVD certification. The research-orientated CHPv2 covers more genes than the OST, however 20 genes are shared between the two panels (Figure 5.7).

The CHPv2 identified at least one variant within each of the five cases, whereas the OST panel was unable to identify any variants within Case A4 (Table 5.7). For variants located within genes covered by both panels, the same specific variants were found by the CHPv2 and OST and at similar VAFs (Cases A1, A2, A3 and A5). Likewise, the only variants that were found exclusively by the CHPv2 were not covered by the OST panel, and therefore this is was an expected finding (*APC*, *CDKN2A* and *HRAS* variants). The OST had a higher sequencing depth and hence a lower limit of detection (LOD) than the CHPv2, with median depths of 5,105 (range 2,836 – 8,359) and 870 (range 535 – 1,302), however for these variants, a low LOD was not required considering their VAF (see Supplementary Table 1.4).

The focussed nature of the OST panel appears to have hindered its diagnostic utility when compared to the CHPv2, as it would provide misleading data in Case A3 and diagnostic doubt in Case A4 (more information regarding these to follow).

Case	Primary tumour diagnosis	Subsequent tumour site	Sex	Age at diagnosis of first tumour	Time lapse between tumours (months)
A1	Urothelial carcinoma (bladder)	Lung	M	75	48
A2	Colorectal adenocarcinoma	Lung	M	58	63
A3	Gastro-oesophageal adenocarcinoma	Lung	M	71	2
A4	Lower lobe lung squamous cell carcinoma	Upper Lung lobe/pleura	M	70	16
A5	Colorectal adenocarcinoma	Lung	M	74	2

Table 5.6 - Clinical information for ambiguous cases.

Gene	CHPv2	OST
ABL1		
AKT1		
ALK		
APC		
ATM		
BRAF		
CDH1		
CDKN2A		
CSF1R		
CTNNB1		
DDR2		
EGFR		
ERBB2		

Gene	CHPv2	OST
ERBB4		
EZH2		
FBXW7		
FGFR1		
FGFR2		
FGFR3		
FLT3		
GNA11		
GNAQ		
GNAS		
HNF1A		
HRAS		
IDH1		

Gene	CHPv2	OST
IDH2		
JAK2		
JAK3		
KDR		
KIT		
KRAS		
MAP2K1		
MET		
MLH1		
MPL		
NOTCH1		
NPM1		
NRAS		

Gene	CHPv2	OST
PDGFRA		
PIK3CA		
PTEN		
PTPN11		
RB1		
RET		
SMAD4		
SMARCB1		
SMO		
SRC		
STK11		
TP53		
VHL		
Total	**50**	**22**

Figure 5.7 - Comparison of genes covered by the CHPv2 and OST. Genes that are covered by the CHPv2 (yellow) & OST (green) respectively. Genes present on both panels are in bold. The total genes covered by each panel is also shown.

CHPv2									
Position	**Ref**	**Var**	**Type**	**Gene**	**COSMIC ID**	**Primary Tumour VAF**	**Second Tumour VAF**	**CDS Mutation**	**AA Mutation**
Case A1									
chr17:7578271	T	A	SNV	***TP53***	COSM11066	-	**19.2%**	c.578A>T	p.H193L
Case A2									
chr5:112175600	A	T	SNV	***APC***	COSM18818	**46.2%**	**22.4%**	c.4309A>T	p.K1437*
chr12:25398284	C	A	SNV	***KRAS***	COSM520	**60.4%**	**72.7%**	c.35G>T	p.G12V
chr17:7577580	T	C	SNV	***TP53***	COSM10725	**67.6%**	**69.0%**	c.701A>G	p.Y234C
Case A3									
chr17:7577094	G	A	SNV	***TP53***	COSM99925	-	**35.6%**	c.844C>T	p.R282W
chr9:21971120	G	A	SNV	***CDKN2A***	COSM12475	**36.7%**	**30.9%**	c.238C>T	p.R80*
Case A4									
chr11:534242	A	G	SNV	***HRAS***	COSM249860	**28.3%**	**11.0%**	c.81T>C	p.H27H
Case A5									
chr17:7577543	C	T	SNV	***TP53***	COSM44310	-	**44.9%**	c.738G>A	p.M246I
chr17:7577539	-	G	INS	***TP53***	-	**39.7%**	-	c.742_743insC	p.R248fs
OST									
Case A1									
chr17:7578271	T	A	SNV	***TP53***	COSM11066	-	**20.7%**	c.578A>T	p.H193L
Case A2									
chr12:25398284	C	A	SNV	***KRAS***	COSM520	**61.0%**	**72.9%**	c.35G>T	p.G12V
chr17:7577580	T	C	SNV	***TP53***	COSM10725	**69.1%**	**72.7%**	c.701A>G	p.Y234C
Case A3									
chr17:7577094	G	A	SNV	***TP53***	COSM99925	-	**39.7%**	c.844C>T	p.R282W
Case A4									
No variants identified									
Case A5									
chr17:7577543	C	T	SNV	***TP53***	COSM44310	-	**42.9%**	c.738G>A	p.M246I
chr17:7577539	-	G	INS	***TP53***	-	**35.7%**	-	c.742_743insC	p.R248fs

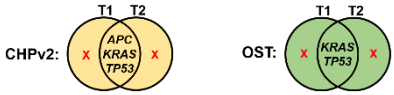
Table 5.7 - Comparison of NGS findings from the CHPv2 and OST in the 5 ambiguous cases. Details of each somatic variant identified within each respective case using the two different NGS assays; CHPv2 (yellow) and OST (green). All variants also found in germline DNA were excluded.

5.3.4.1 Ambiguous Case 1 (A1)

This patient was originally diagnosed and treated for urothelial carcinoma of the bladder, however subsequently presented with a lung tumour with mediastinal lymphadenopathy. Initially macro- and microscopic examination of this secondary lesion showed the appearance consistent with a metastatic non-small cell lung carcinoma (NSCLC), and was subsequently sent for immunohistochemical (IHC) analysis. IHC showed positive nuclear staining of GATA-3, and negative staining for CK20, Uroplakin, p40, CK5/6 and TTF-1, where the pathologist concluded that this would favour the diagnosis of a metastatic urothelial carcinoma, rather than a lung adenocarcinoma. Clear uncertainty was shown however, as they then stated that a metastatic adenocarcinoma of breast origin may also show this immunoprofile, questioning whether this could be a clinical possibility.

Both the CHPv2 and the OST panel found a single *TP53* p.H193L variant within solely the lung tumour tissue sample, at VAFs of 19.2% and 20.7% respectively. The primary bladder cancer sample was negative for a somatic variant using either panel. Based on the results of the unambiguous cohort, this would favour an independent primary diagnosis, however compared to cases U16 – U20, this would be have decreased confidence, having only one variant found in this case.

5.3.4.2 Ambiguous Case 2 (A2)



This case featured a primary colorectal adenocarcinoma at the peritoneal reflection which was treated by a R0 (curative) resection, but had noted ulceration into the peritoneal cavity. Over five years later, this same patient presented with a lung tumour. Initial morphological examination favoured the diagnosis of this tumour as a recurrent colorectal adenocarcinoma, however a pathological staging was noted, in the event that it was later deemed a primary lung tumour. IHC found strong expression of CDX2 and CK29, while being negative for CK7 and TTF-1 which when all pathological data was combined, lead to metastatic colorectal cancer as the most probable diagnosis.

Despite some ambiguity regarding histopathological diagnosis, the tNGS data on these samples was definitive. Three variants were shared between the colorectal adenocarcinoma and the lung tumour mass using the CHPv2; *APC* p.K1437*, *KRAS* p.G12V and *TP53* p.Y234C. The OST panel was also concordant, finding

the latter two variants at similar VAFs. A result of this kind would strongly support the diagnosis of colorectal to lung metastasis. In this specific case, the treatment modality may not have altered dramatically irrespective of diagnosis, as pulmonary metastasectomy from metastatic colorectal adenocarcinoma is a common practice i.e. if the second tumour as diagnosed as an independent lung adenocarcinoma, the tumour may have been resected also. Despite this, it is true that differential diagnosis of this case would have an impact prognostically.

5.3.4.3 Ambiguous Case 3 (A3)

This patient was also recruited into a ctDNA research study focussed on oesophageal adenocarcinoma. This patient had a primary tumour biopsy sequenced using a custom Ion AmpliSeq tNGS panel ("GOA" panel; see Mark Openshaw PhD thesis) with no variants being identified. Just one month subsequent to their tissue biopsy, the patient presented with a separate lung tumour mass. If confidently deemed a metastasis this lesion would not usually be resected, however there was diagnostic doubt as there were no other clear sites of metastasis upon CT imaging and so potentially this represented an independent primary malignancy. Therefore this patient had the lung tumour resected and was subsequently given chemoradiotherapy in an attempt to eradicate residual disease. The patient understood that in the setting of metastatic disease this would not be a curative treatment, but may prolong survival. Unlike the oesophageal adenocarcinoma, the GOA panel detected a *TP53* p.R282W mutation in the lung tumour sample at 38.1% VAF, which suggested independent primary diagnosis, supported by the fact that the patient was also performing well following surgery (no clinical progression).

Without this case being passed into the PRIMINGS study, there would be much doubt regarding the accurate diagnosis of this lung mass. However, using the CHPv2, a *CDKN2A* p.R80* mutation was found to be present in both the primary tumour (36.7% VAF) and second tumour sample (30.9% VAF); a hotspot that is not covered by the GOA or the OST panel. In accordance with the GOA panel, the CHPv2 and OST panel also identified the *TP53* p.R282W in solely the second tumour, implicating this variant as either a subclonal event enriched in the lung metastasis, or a variant that emerged post-seeding of metastasis. As this case was involved in a separate study, further verification was conducted. ddPCR

targeting the *TP53* p.R282W mutation was performed to ensure the primary tumour did not contain any traces of the variant. Although at a very low VAF, there was evidence that the mutation was present in the oesophageal adenocarcinoma (0.16%), hence providing a confident metastatic diagnosis (Figure 5.8). This example shows how ultrasensitive methods such as ddPCR can aid in building diagnostic confidence, but also how intratumour heterogeneity and clonal evolution can confound these studies.

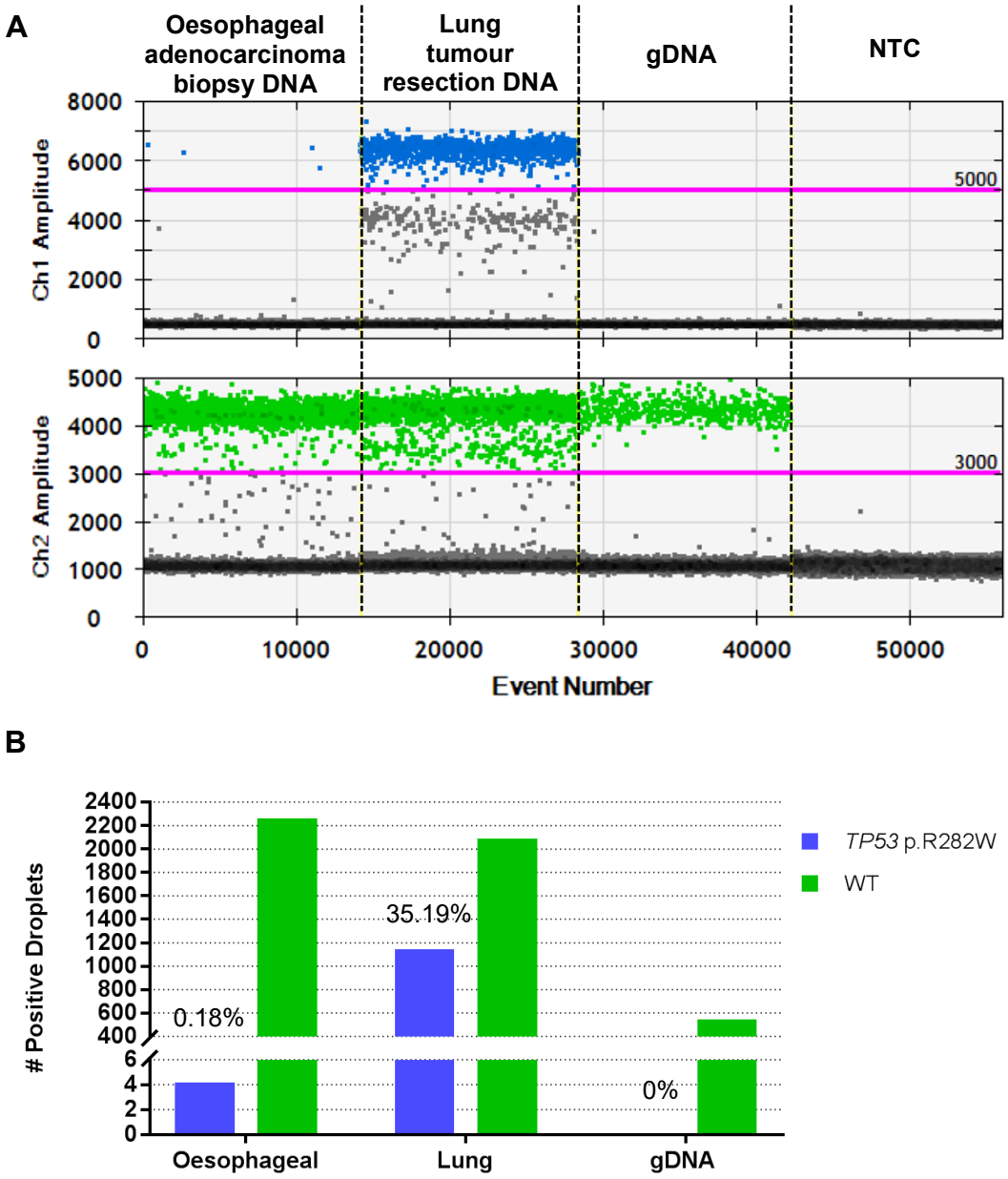


Figure 5.8 – ddPCR evidence of *TP53* p.R282W within both tumours (Case A3). A) 1D amplitudes of the oesophageal adenocarcinoma primary tumour DNA sample and the subsequent lung tumour DNA, showing the presence of four mutant droplets in the primary tumour, despite not being called by NGS. B) Graph showing the number of positive droplets present within each sample, with the VAF indicated.

5.3.4.4 Ambiguous Case 4 (A4)

This case involved two lung tumour masses that arose within 16 months of one another. The first was diagnosed confidently as a poorly differentiated squamous cell carcinoma, undergoing a left lower lobectomy. Subsequently, a tumour mass was found in the upper left lung lobe was identified which focally breached the pleural surface (PL2). Alike the first tumour, this had characteristics consistent with that of a squamous cell carcinoma, with IHC showing strong staining for P40 and CK56. Importantly however, it was noted in the pathology report that there was potential that this second lung tumour could represent a metachrononous primary as well as being a metastasis of the first malignancy.

The CHPv2 identified a single shared variant between the two tumours; a *HRAS* p.H27H at 28.3% and 11% VAFs respectively and therefore would point towards metastatic diagnosis. The OST panel did not identify any variants in this case and therefore would not be able to provide diagnostic information, much like in unambiguous case U15.

5.3.4.5 Ambiguous Case 5 (A5)

The final case involved the presentation of colorectal adenocarcinoma and a lung lesion, only two months apart. The lung tumour mass was removed by a right VATS lower lobectomy and had mixed adenocarcinoma architecture. Of particular doubt was the presence of signet ring carcinoma microscopically; a histology that is rare in both colorectal and lung adenocarcinomas. Nevertheless, on balance the pathological diagnosis leaned towards that of an independent lung adenocarcinoma.

The CHPv2 and OST panels found a *TP53* p.M246I variant in just the lung tumour at 44.9% and 42.9% respectively. Also present using both panels was the presence of a *TP53* insertion in only the colorectal adenocarcinoma sample (39.7 and 35.7% VAFs), interestingly just four bases (two codons) away from the p.M246I variant. Therefore, these findings support independent primary diagnosis for this case.

5.4 Discussion

5.4.1 Comparison between histopathological diagnosis and tNGS in unambiguous cases

For the majority of patients that present with a second tumour mass at a distant site, the classification of this is a routine process using a combination of histological analysis and immunohistochemistry (Duraiyan *et al.* 2012). Yet, for a minority of cases, doubt persists regarding diagnosis. Communication with Dr David Moore and Dr John Le Quesne estimate the frequency of this occurrence to be ~one patient per week at the University Hospitals of Leicester Lung MDT.

Through a combination of increased human longevity, expansion of cancer screening programmes, improved treatment and development of highly sensitive radiological diagnostic methods, the number of patients presenting with multiple tumours over their lifetime is due to rise in the future (Smittenaar *et al.* 2016). It is therefore apparent that our ability to correctly distinguish between primary and metastatic tumours is paramount for selecting the appropriate treatment modality, as well as accurate prognosis. Modern sequencing technologies and a better understanding of cancer genomics enable the high-throughput, rapid testing of a series of genes covering key driver events in most malignant tumours. This chapter aimed to test whether this technology could correctly predict tumour clonality in representative pairs of tumours.

From the pathologically unambiguous cases (U1 - U20), proof of principle has been provided that tNGS can diagnose second tumour masses as metastases or independent primary tumours, prior to the consultation of pathology reports. Once these data are combined, there is confidence that this method can be extrapolated to aid pathologists in the diagnosis of ambiguous tumours in real time clinical cases. Within the independent primary cohort, there were no cases shared variants were identified between tumours, although this was a small sample size (n = 5) and certainly would need extending to a larger group.

In this series there was a positive predictive value of 100% as a test of metastasis. If the single primary-metastasis case with no variants detected (U15) was interpreted as a ‘false negative’ for metastasis, then this would give a negative predictive value of 83%. It is worth considering that a result of this nature would

not indicate the absence of metastasis, but more a lack of evidence supporting metastasis and/or independent primary diagnosis. In practice, a test result of this kind, in which no variants were identified in either lesion would be non-contributory to the diagnostic process; there would be no false impression of independent primary tumours, meaning the finding in this sense would be 'safe'. Nonetheless, it would be much better to eliminate these events altogether and amendments to panel design to allow greater coverage of mutation hotspots is an important consideration (see section 5.4.3).

The greatest challenge that can arise from this study is that of 'false positives' in which both tumours share a common variant, but in reality are independent primary malignancies. This may occur with two lineage-independent tumours of the same type where there are common driver mutations such as *BRAF* p.V600E in melanoma, or *KRAS* p.G12C in lung adenocarcinoma. Although such pairs were not tested in this dataset, only a single shared variant was detected in nine primary/metastatic tumours (Table 5.3, U6 - U14). If taken forward into a prospective clinical setting, the use of a single shared variant as an indicator of shared lineage should also be accompanied by a clear report stating this may potentially represent independent mutational events, reflecting decreased confidence. In such cases, it may be useful to provide compiled data featuring very large sample sizes displaying the regularity of respective variants to help quantify this risk. Challenges of this type will occur with increasing regularity as molecular testing becomes more sophisticated and the role of multidisciplinary 'molecular tumour boards' to provide such nuanced interpretations is gradually being realised.

5.4.2 Inferring the diagnosis of ambiguous cases using tNGS

Providing tNGS could aid in the diagnosis of subsequent tumour masses, it was important to provide examples of how this could influence clinical decisions. Here, the CHPv2 has been shown to provide diagnostic confidence in Cases A2 and A5. For A2, the detection of three shared variants was clearly indicative of a common tumour lineage and hence metastatic diagnosis. While in A5, the two unshared variants reveal the likely presence of independent diagnoses. Both of these cases would be regarded as confident tNGS-directed diagnoses.

NGS in Case A1 was not as successful. Considering intratumour heterogeneity, the detection of a single variant in solely the second tissue sample is admittedly unconvincing. The accumulation of additional variants within common tumours has after all been described in various tumour types (Gerlinger *et al.* 2012, Jamal-Hanjani *et al.* 2017, Saito *et al.* 2018). Therefore this scenario would have trepidations of its own with regards to diagnosis. It seems likely that without sequencing coverage widening, situations like Case A1 are likely to recur in the future; where despite various tests, both histological and genomic, pathologists are no closer to arriving at confident diagnoses. Amendments to methodologies are discussed in the next section (see section 5.4.3).

For Case A3, tNGS with the CHPv2 was critical in deciphering diagnosis in this patient. As this patient was involved in another research project (Mark Openshaw Thesis), the CHPv2 results facilitated further analyses that cemented metastatic diagnosis, such as the ultra-low (0.16%) presence of a *TP53* p.R282W in the oesophageal primary tumour. Furthermore, additional findings have been conducted where high level somatic copy number amplification (SCNA) of *KRAS* was identified within both of these tumours (data not shown). Although tNGS cannot take credit for the detection of a shared *TP53* mutation and *KRAS* SCNA events, it certainly provided a crucial foundation to uncover confident diagnosis. With the development of increasingly deep NGS technologies, it seems likely that the lack of detection of this *TP53* p.R282W variant in the oesophageal primary would be circumvented with newer panels, due to incrementally lower LODs (< 0.1% VAF).

Case A4 had CHPv2 results resembling many of those seen in the unambiguous metastatic cohort (Cases U7 - U14), where only a single shared variant was identified between the first and second tumour samples. Arguably, this result would be less confident than those seen in Cases A2 and A5, but more confident than those in Case A1. For variants that are not particularly common in a number of tumour types, it seems statistically more probable for a single shared variant between tumours to be indicative of metastasis than the identification of a single variant in only one tumour (like in Case A1) being indicative of independent primary diagnosis. The development of statistical methods to quantify confidence in this setting would be a critical step to take if targeted NGS methods are to be used for diagnostic testing.

An important finding from this section, was that the OST panel is not fit for purpose with regards to differential diagnosis. Although this panel was not developed for this use, it was hoped to provide equivalent diagnostic strength to the CHPv2, while also being a certified test (CE-IVD) that could be clinically adopted in an efficient manner. The OST panel had identical findings to the CHPv2 in Cases A1 and A5 and would have been proficient in the metastatic diagnosis of A2 (2/3 shared variants detected), but would have potentially changed the diagnosis in A3 and provided no information for A4. Therefore, attention must turn to other methods, both panel-based and exome/genome-wide.

5.4.3 Future considerations for clinical implementation of tNGS for diagnosis

This study highlights how tNGS can infer the diagnosis of anatomically distinct tumours using a commercially available sequencing panel, however there are some methodological limitations which would need consideration prior to implementation.

The CHPv2 has proved successful for research purposes, however this is not an approved method in the diagnostic setting (no CE/IVD certification), nor is it likely to be so considering the rapidly advancing, newer technologies now becoming available in the field of genomics. This study also investigated a more emerging tNGS chemistry, the OST panel. However as previously mentioned, this is not likely to be the panel of choice for diagnostic means, due to its narrow coverage of mutation hotspots, irrespective of its CE/IVD certification. Therefore, if targeted NGS is to be the method of choice, other larger panels with rigorous standardisation are desired. Accreditation of such panels, workflows and operating laboratories by bodies such as ISO (International Organization for Standardization), CLIA (Clinical Laboratory Improvement Amendments) and CAP (College of American Pathologists) would be important.

Of course, small scale nucleotide changes are not the only method of detecting clonality between tumours, with large scale changes (somatic copy number alterations (SCNA), translocations etc.), not to mention epigenetic markers being key areas of interest. In the case of the latter, DNA methylation have been shown to be varied among tumour types (Witte *et al.* 2014) and therefore a multifaceted approach to tumour analysis could be key to minimising misdiagnosis.

Regarding starting sample material, although FFPE-derived uninvolved lymph node DNA worked well in filtering-out germline variants in this study, it is always worth considering using DNA derived from higher quality tissue, should the opportunity arise. As blood-based monitoring of solid tumours becomes increasingly adopted, the use of buffy coat DNA isolated from fresh whole blood would be a useful source of high molecular weight germline DNA. This would be less fragmented, reducing the incidence of sequencing failure compared to that of FFPE-derived sources (Einaga *et al.* 2017). This is not a trivial matter,

considering sequencing failure may delay (or prevent) diagnosis in real time cases. Of course, the use of FFPE-DNA from the tumour samples would be a more difficult hurdle to overcome, as the storage of fresh tumour tissue from surgery/biopsy is more challenging. Additionally, fresh-frozen tissue would not currently suit the structure of this study, as FFPE tissue is required for histopathological analysis. Only following this analysis would an ambiguous case be sent for "NGS diagnosis". Technologies to analyse highly fragmented DNA, such as FFPE-DNA are ever-increasing and so the use of this DNA is not likely to be a long-term issue.

As our understanding intratumour heterogeneity expands, it may be worth considering the prospect of taking multiple tissue cores (from FFPE blocks) for each respective tumour, with the aim of maximising the spectrum of variants that may be present spatially within tumours, thereby minimising the frequency of inconclusive results. Unlike studies investigating the extent of intratumour heterogeneity, a study for predominantly diagnostic means would not require the independent sequencing of multiple tumour regions, and instead could potentially pool DNA extracted from multiple regions of tumour prior to NGS, as a cost efficient alternative (maintaining three samples per patient case).

For clinical implementation, a key step will be to develop an analysis pipeline that integrates tNGS data with statistical analyses to assess the probability of false diagnosis. As mentioned in this chapter, not all genomic variants are equally common in tumours, with mutations such as *KRAS* p.G12C occurring in 11%–16% of lung adenocarcinomas and 2 – 4% of colorectal adenocarcinomas (Janes *et al.* 2018). Therefore, statistical algorithms are required to quantify the confidence of individual tNGS findings. Studies using random effects models have recently been published, which could be applied to such data (Mauguen *et al.* 2018). The reality is that as a diagnostic test, tNGS findings should be processed and ultimately presented to clinicians as a binary result; metastasis or independent primary.

5.4.4 Targeted NGS or Exome/Genome-wide methods?

Overall, targeted gene sequencing (tNGS) methods are currently the simplest approach for clinical implementation, as they simplify the analysis process, focusing on select hotspot regions of cancer consensus genes. Yet, other methods have advantages.

Broader sequencing methods such as whole exome sequencing (WES) and whole genome sequencing (WGS) are attractive avenues to explore, but clinical adoption is difficult, due to a combination of high cost and data convolution. Competitive prices for WES for a diagnostic case (two tumours plus germline DNA) would be ~£1,150 - £2,000, while WGS would likely cost ~£3,800 - £9,000 (Schwarze *et al.* 2018). Meanwhile, targeted methods would typically cost £500 - £800 per case (Marino *et al.* 2018) and turnaround time is likely to be quicker than with WES/WGS; an important factor for cancer diagnosis. After all, by the time pathologists have deemed such cases as ambiguous, time would have already elapsed. Increasingly, cost-benefit analyses are being conducted for the clinical use of WES/WGS methods, with the likely verdict being the additional cost of these technologies would not be a long-term limiting factor (Schwarze *et al.* 2018, Stark *et al.* 2017). To circumvent the complexity of WES/WGS data, working with clinical bioinformaticians in order to develop an analysis pipeline to streamline this output would be a key clinical consideration.

Regardless of price, WES/WGS would be diagnostically superior to targeted methods, as a number of variants present in tumours are located away from the quintessential cancer genes, yet would serve the same purpose as identifiers of shared/unshared clonality. Additionally, these could be highly variable genomic regions likely to possess lineage markers independent of natural selection (Lawrence *et al.* 2013). An advantage of targeted methods are that they allow a higher sequencing depth (>1,000 X) than WES/WGS. Such a characteristic is important for the detection of ctDNA for example where tumour-fractions are typically below 1% VAF. However, for a tumour tissue DNA-based study such as this, a low limit of detection is not a necessity. Instead, the comprehensive coverage of genes provided by WES/WGS will be vital for robust and accurate tissue diagnosis.

It seems the so called “genomics revolution” is ever approaching, with the NHS becoming the first healthcare system in the world to begin offering WGS systematically, where clinically appropriate (Turnbull *et al.* 2018). Brought about by the 100,000 Genomes Project, seven NHS Genomic Medicine Centres have been established by NHS England and it is not difficult to see the potential in utilising this setup to bring direct benefits to patients through accurate diagnosis. Additional long-term benefits of this would be through the contribution of further genomic data to the 100,000 Genomes knowledgebase that could accelerate us towards the era of precision medicine.

5.5 Conclusions

In this chapter, it has been shown that commercial tNGS panels can be utilised to determine clonality in tumour pairs from the same patient using FFPE tumour material. There have been examples of how this technology could resolve diagnostically ambiguous cases and ultimately aid clinicians in decision-making. With ongoing developments on the coverage of tNGS panels, in addition to broader methods such as WES/WGS, sequencing technology is likely to become increasing accurate, affordable and available for clinical implementation for resolving difficult diagnostic cases, enabling the most appropriate treatment to be offered to the patient.

5.6 Future direction

Although this chapter has described how targeted gene sequencing panels could be used to resolve ambiguous diagnostic cases, a key future direction will be to extend analyses using broader sequencing methods such as WES, as this will inevitably increase the number of shared variants between primary-metastatic tumour pairs and thereby increase diagnostic confidence. Conversely, for independent primary pairs this technology will provide information on the extent of genomic independence between samples, but also identify potential shared variants that need to be accounted for should such a test be clinically implemented.

Conducting this analysis in diagnostically unambiguous case studies will provide insight into the ideal thresholds to utilise for shared variants numbers and ultimately allow for the derivation of statistical algorithms to quantify diagnostic confidence. The meta-analysis of WES data from multiple studies focussed on primary-metastatic pairs will also be a valuable resource moving forward.

Finally, the integration of WES data within a bioinformatics analysis pipeline will be another significant area of focus, to allow for efficient, time-effective turnaround times to facilitate accurate and timely diagnoses.

Chapter 6 General Discussion

6.1 ctDNA: Are we at our limits?

From the results of this thesis and the current literature, it is clear that the detection of ctDNA is at the mercy of the technologies we have available. The majority of patients with early-stage NSCLC and MPM (various stages), were pre-surgically ctDNA-negative (35/52) with a ~0.1% LOD using ddPCR and/or NGS. Even when patient-tailored approaches are taken were several tumour SNVs are targeted, ctDNA is not detected in a large proportion of tumours, for example in 47/58 (81%) of LUADs (Abbosh *et al.* 2017). Considering the high cellular turnover within tumours, it is difficult to accept that ctDNA is absent within plasma of these patients, posing questions regarding the biological mechanisms governing ctDNA release, in addition to the sensitivity of the technologies we use to detect this DNA.

In this study, including all NSCLC subtypes, 46/96 (47.9%) of patients had ctDNA-detected, highlighting the large variation between cancer types. For example, other studies have shown ctDNA was detectable in > 60% of localised colorectal adenocarcinomas and > 90% of ovarian and hepatocellular carcinoma (Bettegowda *et al.* 2014, Cohen *et al.* 2018). Although light has been shed upon the potential release mechanisms pertaining ctDNA release, such as necrotic rate, proliferation index and lymphovascular invasion, it remains to be seen whether ctDNA detection rates will improve to routinely detect early-stage tumours, or whether ctDNA analysis should not make early detection a priority. When searching < 0.1% mutant VAF, the differentiation between somatic variants and stochastic noise (e.g. PCR-errors) becomes much less pronounced. The utilisation of unique molecular identifiers are addressing this issue somewhat, such as those used by the OLcfDNA (Newman *et al.* 2014, Schmitt *et al.* 2012).

During analytical validation of the bespoke approach (Abbosh *et al.* 2017), when using a background of 10,000 copies of cell line DNA, the sensitivity of the method to detect mutations at between 0.01% and 0.05% VAF was ~46%, and just 4.2% for those at 0.01% VAF (1/10,000 mutant copies). If extrapolated into a cfDNA context, 10,000 copies would equate to ~33 ng input. Therefore, if we are to consistently detect ctDNA at these allele fractions, higher input DNA amounts may be required (Table 6.1). cfDNA yield varies widely between extraction method, stage of disease and cancer type, but a typical estimate would be ~10

ng / mL plasma in early-stage disease. Presuming plasma equates to half of the volume of a blood draw, 20 mL of blood could potentially yield 100 ng cfDNA, however some of this would be consumed during quantification and/or fragment profiling.

cfDNA input (ng)	Haploid genome equivalents (copies)	Copies of a tumour SNV present at 0.01% VAF
10	~3,000	0.3
20	~6,000	0.6
30	~9,000	0.9
40	~12,000	1.2
50	~15,000	1.5
60	~18,000	1.8
70	~21,000	2.1
80	~24,000	2.4
90	~27,000	2.7
100	~30,000	3.0

Table 6.1 - ctDNA detection at 0.01% VAF based on cfDNA input. Red shading indicates where the detection of a heterozygous tumour SNV would not be possible (< 1 copy).

Other methods are being explored in order to potentially enrich for ctDNA, namely through the size-selection of smaller-sized cfDNA fragments (Hellwig *et al.* 2018, Mouliere *et al.* 2017). If such size-selection is successful in minimising healthy cfDNA background and if the sensitivity of NGS technologies continue to improve in LOD, there is hope that ctDNA detection could succeed in even these difficult to detect tumour types (e.g. LUAD, MPM) and smaller tumours.

Of course, an all too familiar problem with NGS, is that typically the more focussed the assay (e.g. less genes covered), the higher median depth you can achieve (usually lower LOD), however you decrease your probability of detecting multiple tumour SNVs and lower coverage of patients; the “Goldilocks principle” prevails. Nevertheless, one key finding from this thesis, was that the majority of patients analysed using a generic approach (OLcfDNA) had ≥ 1 clonal tumour SNVs targeted, emphasising that such methods are broadly applicable to ctDNA detection, and that the majority of these “hotspot” mutations are acquired early in tumour evolution and hence remain the ideal therapeutic targets (Jamal-Hanjani *et al.* 2017).

Appendix

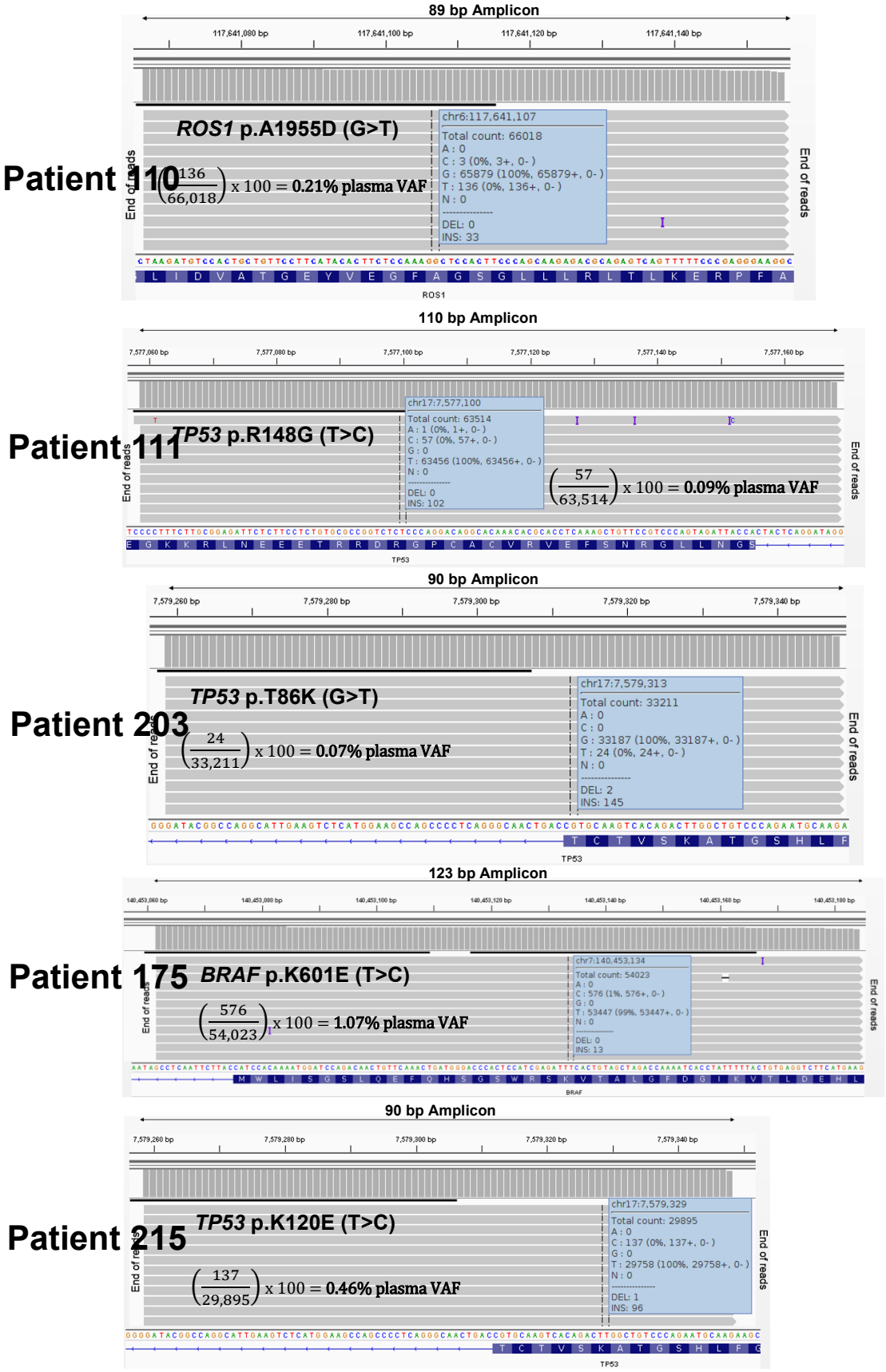
Patient ID	CRUK ID*	Clonal tumour variants searched in Pre-surgical cfDNA (covered by OLcfDNA)													Bespoke Method (Validation)			
		Chrom	Position	Ref	Var	Gene	OLcfDNA call?	VAF (%)	Allele Mol Cov	Total Mol Cov	Allele Read Cov	Total Read Cov	Hotspot/Novel?	Detected?	Bespoke VAF (%)	Allele Read Cov	Total Read Cov	Called/Detected?
1	CRUK0034	12	25398284	C	A	*KRAS*	N	ND	ND	ND	ND	57,948	Hotspot	N	ND	ND	73,340	N
22	CRUK0081	**17**	**7578190**	**T**	**C**	***TP53***	**Y**	**0.19**	**8**	**4,320**	**120**	**50,367**	**Hotspot**	**Y**	**0.42**	**127**	**29,984**	**Y**
33	CRUK0077	**17**	**7577082**	**C**	**T**	***TP53***	**N**	**4.89**	**n/a**	**n/a**	**1,710**	**34,985**	**Novel**	**Y**	**5.82**	**1,806**	**31,055**	**Y**
41	CRUK0051	12	25398285	C	A	*KRAS*	N	ND	ND	ND	ND	54,576	Hotspot	N	ND	ND	53,167	N
46	CRUK0004	7	55241708	G	C	*EGFR*	N	ND	ND	ND	ND	73,136	Hotspot	N	ND	ND	44,705	N
49	CRUK0057	12	25398284	C	A	*KRAS*	N	ND	ND	ND	ND	56,484	Hotspot	N	ND	ND	45,600	N
50	CRUK0003	3	178936094	C	A	*PIK3CA*	N	ND	ND	ND	ND	ND	Hotspot	N	n/a	n/a	n/a	n/a
51	CRUK0026	7	55259515	T	G	*EGFR*	N	ND	ND	ND	ND	94,754	Hotspot	N	ND	ND	38,709	N
		17	7577130	A	G	*TP53*	N	ND	ND	ND	ND	71,901	Hotspot	N	ND	ND	38,744	N
55	CRUK0039	12	25398285	C	A	*KRAS*	N	ND	ND	ND	ND	52,948	Hotspot	N	ND	ND	44,550	N
60	CRUK0016	17	7578455	C	G	*TP53*	N	ND	ND	ND	ND	53,210	Hotspot	N	ND	ND	18,349	N
62	CRUK0018	12	25398285	C	A	*KRAS*	N	ND	ND	ND	ND	68,702	Hotspot	N	ND	ND	56,951	N
64	CRUK0062	**17**	**7577121**	**G**	**A**	***TP53***	**Y**	**0.39**	**14**	**3,623**	**441**	**75,596**	**Hotspot**	**Y**	**1.07**	**304**	**28,355**	**Y**
75	CRUK0021	7	55259515	T	G	*EGFR*	N	ND	ND	ND	ND	70,130	Hotspot	N	ND	ND	76,660	N
		17	7578403	C	A	*TP53*	N	ND	ND	ND	ND	46,605	Hotspot	N	ND	ND	26,938	N
77	n/a	17	7577580	T	C	*TP53*	N	ND	ND	ND	ND	ND	Hotspot	N	n/a	n/a	n/a	n/a
84	CRUK0027	12	25398284	C	G	*KRAS*	N	ND	ND	ND	ND	68,799	Hotspot	N	ND	ND	61,185	N
97	CRUK0029	**1**	**115256529**	**T**	**C**	***NRAS***	**Y**	**2.05**	**153**	**7,460**	**1,284**	**61,235**	**Novel**	**Y**	**3.25**	**807**	**24,859**	**Y**
		17	**7577580**	**T**	**C**	***TP53***	**Y**	**1.14**	**69**	**6,047**	**635**	**56,609**	**Hotspot**	**Y**	**2.45**	**586**	**23,933**	**Y**
103	CRUK0035	17	7578406	C	T	*TP53*	N	ND	ND	ND	ND	56,427	Hotspot	N	ND	ND	15,018	N
110	CRUK0089	**3**	**178936091**	**G**	**A**	***PIK3CA***	**Y**	**0.28**	**12**	**4,360**	**259**	**73,235**	**Hotspot**	**Y**	n/a	n/a	n/a	n/a
		6	**117641107**	**G**	**T**	***ROS1***	**N**	**0.21**	**n/a**	**n/a**	**136**	**66,018**	**Novel**	**Y**	**0.11**	**54**	**48,901**	**Y**
111	CRUK0067	**17**	**7577100**	**T**	**C**	***TP53***	**N**	**0.09**	**n/a**	**n/a**	**57**	**63,514**	**Hotspot**	**Y**	**0.10**	**22**	**21,494**	**Y**
115	CRUK0030	12	25398285	C	A	*KRAS*	N	ND	ND	ND	ND	68,796	Hotspot	N	ND	ND	65,241	N
135	CRUK0052	**12**	**25398284**	**C**	**A**	***KRAS***	**N**	**ND**	**ND**	**ND**	**ND**	**55,796**	**Hotspot**	**N**	**0.1**	**47**	**45,945**	**Y**
		17	**7578190**	**T**	**C**	***TP53***	**Y**	**0.08**	**2**	**2,474**	**103**	**69,214**	**Hotspot**	**Y**	**ND**	**ND**	**28,487**	**N**
142	CRUK0058	17	7577099	C	T	*TP53*	N	ND	ND	ND	ND	59,336	Hotspot	N	ND	ND	28,160	N
		7	55259524	T	A	*EGFR*	N	ND	ND	ND	ND	71,767	Hotspot	N	ND	ND	50,322	N
163	CRUK0001	17	7577099	C	T	*TP53*	N	ND	ND	ND	ND	63,238	Hotspot	N	ND	ND	30,955	N
175	CRUK0009	**7**	**140453134**	**T**	**C**	***BRAF***	**N**	**1.07**	**n/a**	**n/a**	**576**	**54,023**	**Novel**	**Y**	**n/a**	**n/a**	**n/a**	**n/a**
		17	**7577570**	**C**	**T**	***TP53***	**Y**	**2.70**	**85**	**3,150**	**2,678**	**86,481**	**Hotspot**	**Y**	**1.63**	**534**	**32,812**	**Y**
180	CRUK0049	17	7578457	C	A	*TP53*	N	ND	ND	ND	ND	49,225	Hotspot	N	ND	ND	14,477	N
183	CRUK0025	12	25398285	C	A	*KRAS*	N	ND	ND	ND	ND	ND	Hotspot	N	n/a	n/a	n/a	n/a
		17	7578190	T	C	*TP53*	N	ND	ND	ND	ND	110,620	Hotspot	N	ND	ND	22,106	N
193	CRUK0007	3	178952085	A	G	*PIK3CA*	N	ND	ND	ND	ND	64,097	Hotspot	N	ND	ND	36,371	N
203	CRUK0086	**17**	**7579313**	**G**	**T**	***TP53***	**N**	**0.07**	**n/a**	**n/a**	**24**	**33,211**	**Novel**	**Y**	**0.20**	**30**	**15,335**	**Y**
206	CRUK0048	17	7578475	G	A	*TP53*	N	ND	ND	ND	ND	66,070	Hotspot	N	ND	ND	8,558	N
		17	7577548	C	A	*TP53*	N	ND	ND	ND	ND	88,860	Hotspot	N	ND	ND	18,129	N
		7	140453155	C	T	*BRAF*	N	ND	ND	ND	ND	95,562	Hotspot	N	ND	ND	40,345	N
215	n/a	**17**	**7579329**	**T**	**C**	***TP53***	**N**	**0.46**	**n/a**	**n/a**	**137**	**29,895**	**Novel**	**Y**	**n/a**	**n/a**	**n/a**	**n/a**
223	CRUK0037	12	25380276	T	A	*KRAS*	N	ND	ND	ND	ND	59,278	Hotspot	N	ND	ND	49,937	N
225	CRUK0014	12	25398284	C	A	*KRAS*	N	ND	ND	ND	ND	ND	Hotspot	N	n/a	n/a	n/a	n/a
		17	7578205	C	A	*TP53*	N	ND	ND	ND	ND	88,788	Hotspot	N	ND	ND	43,281	N
242	CRUK0064	17	7578461	C	A	*TP53*	N	ND	ND	ND	ND	54,944	Hotspot	N	ND	ND	11,588	N
251	CRUK0022	7	55259515	T	G	*EGFR*	N	ND	ND	ND	ND	88,363	Hotspot	N	0.03	9	28,574	Y
		17	7577082	C	A	*TP53*	N	ND	ND	ND	ND	65,489	Hotspot	N	ND	ND	23,796	N
252	CRUK0093	17	7578469	C	A	*TP53*	N	ND	ND	ND	ND	73,472	Hotspot	N	ND	ND	15,301	N
255	n/a	7	55259524	T	A	*EGFR*	N	ND	ND	ND	ND	ND	Hotspot	N	n/a	n/a	n/a	n/a
291	n/a	12	25398285	C	A	*KRAS*	N	ND	ND	ND	ND	ND	Hotspot	N	n/a	n/a	n/a	n/a
309	n/a	**12**	**25398284**	**C**	**A**	***KRAS***	**Y**	**0.15**	**3**	**2,022**	**122**	**61,950**	**Hotspot**	**Y**	**n/a**	**n/a**	**n/a**	**n/a**
310	n/a	3	178936082	G	A	*PIK3CA*	N	ND	ND	ND	ND	ND	Hotspot	N	n/a	n/a	n/a	n/a
428	n/a	**7**	**55248998**	**-**	**TGGCCAGCG**	***EGFR***	**Y**	**0.07**	**1**	**1,487**	**24**	**12,703**	**Hotspot**	**Y**	**n/a**	**n/a**	**n/a**	**n/a**
429	n/a	**12**	**25398285**	**C**	**A**	***KRAS***	**Y**	**0.10**	**3**	**3,016**	**32**	**47,280**	**Hotspot**	**Y**	**n/a**	**n/a**	**n/a**	**n/a**
		17	7577120	C	A	*TP53*	N	n/a	1	2,966	28	70,071	Hotspot	N	n/a	n/a	n/a	n/a

Supplementary Table 1.1 – Clonal tumour variants covered by the OLcfDNA assay with NGS findings. Table showing 54 variants covered by the OLcfDNA with NGS information, including the bespoke method (where applicable). When called by the OLcfDNA software, the molecular VAF (unique molecular barcoding) is used to calculate the VAF. When manually assessed, the VAF is calculated from the "Read" figures. CRUK ID (*) refers to the ID given to the patient in the publication (see Abbosh et al. 2017). Variants in bold denote those that were detected by the OLcfDNA assay. ND, Not detected; Y, yes; N, no.

Patient	CRUK ID*	cfDNA Time point	Chrom	Position	Ref	Var	Gene	AA change	VAF (%)	Allele Mol Cov	Total Mol Cov	Allele Read Cov	Total Read Cov	Hotspot/ Novel?
64	CRUK0062	Pre-surgical	17	7577121	G	A	*TP53*	p.R273C	0.39	14	3,623	441	75,596	Hotspot
		30 days (Pre-chemo)	no variants											
		181 days	no variants											
		258 days	17	7577121	G	A	*TP53*	p.R273C	0.23	7	3,004	197	80,716	Hotspot
		420 days (First Recurrence)	17	7577121	G	A	*TP53*	p.R273C	0.22	1	466	82	31,693	Hotspot
77	n/a	Pre-surgical	no variants											
		601 days	no variants											
		853 days	17	7577580	T	C	*TP53*	p.Y234C	0.22	5	2,309	179	76,936	Hotspot
		957 days	17	7577580	T	C	*TP53*	p.Y234C	2.65	129	4,875	2,883	98,408	Hotspot
110	CRUK0089	Pre-surgical	3	178936091	G	A	*PIK3CA*	p.E545K	0.28	12	4,360	259	73,235	Hotspot
			6	117641107	G	T	*ROS1*	p.A1955D	0.21	n/a	n/a	136	66,018	Novel
			17	7577121	G	A	*TP53*	p.R273C	0.06	2	3,479	121	59,849	Hotspot
		65 days	17	7577121	G	A	*TP53*	p.R273C	0.12	4	3,391	41	33,309	Hotspot
		209 days	17	7577121	G	A	*TP53*	p.R273C	0.25	6	2,409	140	61,405	Hotspot
		286 days	17	7577121	G	A	*TP53*	p.R273C	0.57	4	702	111	33,623	Hotspot
		734 days	17	7577121	G	A	*TP53*	p.R273C	0.57	4	702	111	33,623	Hotspot
		916 days	17	7577121	G	A	*TP53*	p.R273C	0.55	7	1,267	383	48,939	Hotspot
		1098 days	17	7577121	G	A	*TP53*	p.R273C	0.14	3	2,211	107	70,184	Hotspot
215	n/a	Pre-surgical	17	7579329	T	C	*TP53*	p.K120E	0.46	n/a	n/a	137	29,895	Novel
		76 days	17	7579329	T	C	*TP53*	p.K120E	0.67	n/a	n/a	393	58,284	Novel
309	n/a	Pre-surgical	12	25398284	C	A	*KRAS*	p.G12V	0.15	3	2022	122	61950	Hotspot
		469 days	no variants											
		638 days	no variants											
		916 days	no variants											
428	n/a	Pre-surgical	7	55249010	-	GCCAGCGTG	*EGFR*	p.M766delinsMASV	0.07	1	1487	24	12703	Hotspot
		92 days	7	55249010	-	GCCAGCGTG	*EGFR*	p.M766delinsMASV	0.06	2	3374	16	26882	Hotspot
		218 days	7	55249010	-	GCCAGCGTG	*EGFR*	p.M766delinsMASV	0.36	8	2255	167	24505	Hotspot
		505 days	7	55249010	-	GCCAGCGTG	*EGFR*	p.M766delinsMASV	3.60	140	3891	1968	44854	Hotspot
429	n/a	Pre-surgical	12	25398285	C	A	*KRAS*	p.G12C	0.10	3	3016	32	47280	Hotspot
		114 days	no variants											
		217 days	no variants											
		318 days	no variants											
		435 days	no variants											

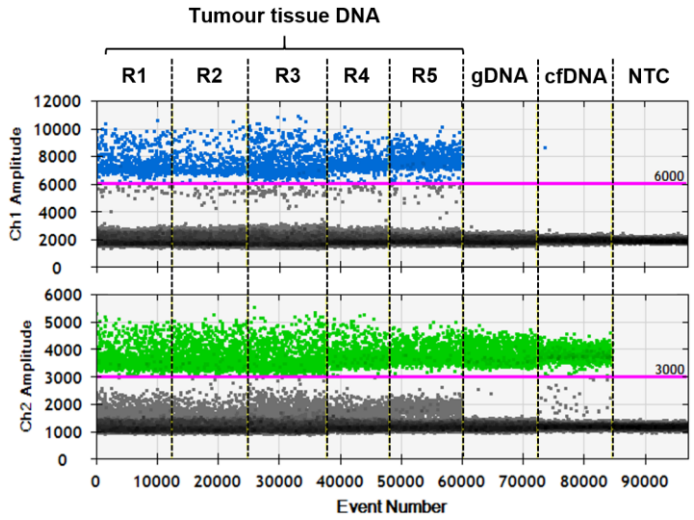
Supplementary Table 1.2 – Longitudinal ctDNA analysis in TRACERx Patients. Raw data for the variants detected in the seven Leicester patients in serial samples, with the time point indicated.

.



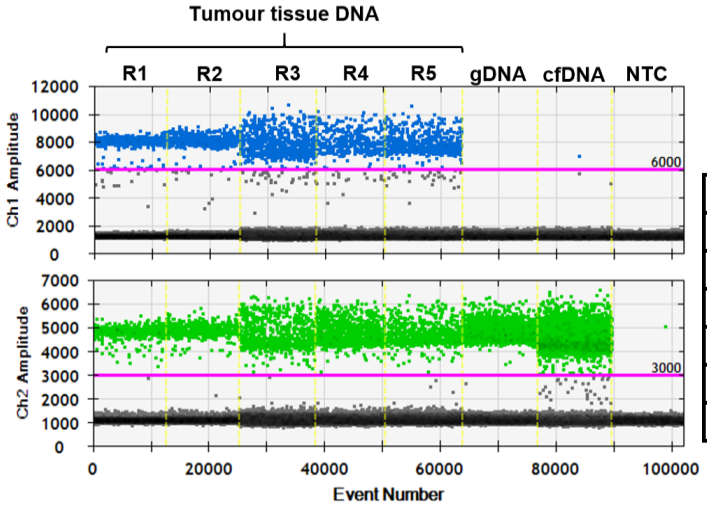
Supplementary Figure 1.1 - Manually detected tumour-confirmed clonal variants in pre-surgical TRACERx cfDNA. IGV images of variants not called by the OLcfDNA software, with read counts shown and amplicon size shown. Each respective mutation and how manually the VAF is calculated.

A Patient 24 – *LRP6* p.I1185L



	VAF (%)		# positive droplets	
	WES	ddPCR	Mut	WT
R1	40.88	37.40	628	1,051
R2	25.00	22.79	389	1,318
R3	46.26	42.68	881	1,183
R4	30.00	29.80	464	1,093
R5	35.71	42.81	827	1,105
gDNA	0.00	0.00	0	1,678
cfDNA	n/a	0.04	1	2,320

B Patient 34 – *BAP1* p.L635Sfs*2



	VAF (%)		# positive droplets	
	WES	ddPCR	Mut	WT
R1	45.21	43.73	499	642
R2	53.41	52.38	573	521
R3	47.43	51.28	781	742
R4	27.98	28.80	366	905
R5	44.97	45.04	540	659
gDNA	0.00	0.00	0	1,735
cfDNA	n/a	0.02	1	4,387

Supplementary Figure 1.2 - Repeat cfDNA analysis in Patients 24 (A) and 34 (B). 1D amplitudes of mutant (blue) and WT (green) channels showing positive droplets. Also present are tables providing the ddPCR VAF and number of positive droplets for each respective run. For samples that had undergone WES, the VAFs were also shown.

Case	Sample	*GAPDH* 95bp Assay Conc. (ng/µL)	*GAPDH*-186bp Assay Conc. (ng/µL)	DNA Integrity Index (*GAPDH186*/*GAPDH95*)
U1	Primary	66.00	16.40	0.250
	Metastasis	136.80	71.90	0.530
	LN	68.00	13.00	0.190
U2	Primary	117.90	4.90	0.040
	Metastasis	50.50	8.30	0.160
	LN	85.20	8.10	0.100
U3	Primary	42.10	11.00	0.260
	Metastasis	24.40	2.30	0.100
	LN	41.60	9.90	0.240
U4	Primary	42.40	13.90	0.330
	Metastasis	69.10	12.40	0.180
	LN	36.70	8.20	0.220
U5	Primary	33.40	2.80	0.080
	Metastasis	49.90	9.30	0.190
	LN	65.50	4.80	0.070
U6	Primary	17.30	0.02	0.001
	Metastasis	23.63	2.90	0.122
	LN	63.00	3.70	0.060
U7	Primary	59.30	0.13	0.002
	Metastasis	39.19	6.16	0.157
	LN	63.70	6.93	0.110
U8	Primary	93.30	4.80	0.050
	Metastasis	14.23	6.50	0.460
	LN	5.80	1.00	0.170
U9	Primary	142.42	60.49	0.420
	Metastasis	117.60	46.14	0.390
	LN	328.80	85.73	0.260
U10	Primary	165.70	42.82	0.260
	Metastasis	307.81	87.03	0.280
	LN	452.24	119.08	0.260
U11	Primary	49.34	11.24	0.230
	Metastasis	17.52	7.97	0.450
	LN	202.39	13.59	0.070
U12	Primary	67.28	20.15	0.300
	Metastasis	134.19	35.47	0.260
	LN	250.49	28.08	0.110
U13	Primary	42.60	17.30	0.410
	Metastasis	21.40	0.50	0.460
	LN	455.10	82.90	0.180
U14	Primary	227.23	88.21	0.388
	Metastasis	185.67	65.19	0.351
	LN	247.64	100.03	0.404
U15	Primary	58.51	13.28	0.230
	Metastasis	75.37	16.47	0.220
	LN	120.68	20.12	0.170
U16	Primary	178.90	68.10	0.380
	Second Primary	26.80	13.80	0.510
	LN	179.50	74.00	0.410
U17	Primary	125.40	64.00	0.510
	Second Primary	158.70	81.80	0.520
	LN	201.60	106.80	0.530
U18	Primary	85.96	11.65	0.135
	Second Primary	66.30	21.32	0.322
	LN	141.00	33.60	0.240
U19	Primary	85.96	11.65	0.135
	Second Primary	66.30	21.32	0.322
	LN	80.50	10.50	0.130
U20	Primary	9.60	0.30	0.030
	Second Primary	235.70	100.70	0.430
	LN	168.00	81.00	0.480
A1	Primary	20.25	3.51	0.173
	Second Tumour	51.94	36.38	0.700
	NT	31.28	4.67	0.149
A2	Primary	60.43	7.61	0.126
	Second Tumour	133.61	83.69	0.626
	LN	10.56	0.41	0.039
A3	Primary	32.78	11.60	0.354
	Second Tumour	5.95	0.88	0.149
	NT	14.96	8.34	0.557
A4	Primary	13.38	3.91	0.292
	Second Tumour	18.81	11.53	0.613
	LN	31.05	8.86	0.285
A5	Primary	94.95	13.92	0.147
	Second Tumour	42.84	11.05	0.258
	LN	30.51	3.68	0.121

Supplementary Table 1.3 - DNA Integrity Index of study samples. DNA samples were quantified using two assays of differing amplicon lengths (95bp & 186bp) targeting *GAPDH*, to provide insight into the extent of DNA fragmentation. Unambiguous (U) cases (red & blue) & Ambiguous (A) cases (grey) are indicated. LN=lymph node; NT=Normal tissue

Case	Sample	Mapped reads	On Target	Mean Depth	Uniformity
U1	Primary	250,627	98.69%	1,112	97.32%
	Metastasis	225,199	98.83%	1,028	97.42%
	LN	265,716	98.33%	1,162	92.60%
U2	Primary	385,568	96.60%	1,640	94.71%
	Metastasis	348,479	98.50%	1,537	95.89%
	LN	353,802	97.13%	1,498	93.31%
U3	Primary	518,862	99.24%	2,263	81.85%
	Metastasis	1,894,452	98.88%	8,253	80.55%
	LN	157,296	99.36%	686	85.16%
U4	Primary	537,367	97.68%	2,319	92.65%
	Metastasis	511,861	96.71%	2,177	92.36%
	LN	779,532	97.94%	3,349	94.34%
U5	Primary	346,713	98.81%	1,546	90.39%
	Metastasis	374,565	98.96%	1,670	88.77%
	LN	376,158	97.81%	1,639	91.24%
U6	Primary	342,403	97.82%	1,419	83.32%
	Metastasis	369,553	98.34%	1,629	91.84%
	LN	387,886	97.29%	1,617	90.03%
U7	Primary	335,835	98.16%	1,467	83.74%
	Metastasis	321,560	98.26%	1,421	86.81%
	LN	353,199	97.32%	1,520	91.79%
U8	Primary	298,438	98.80%	1,312	87.98%
	Metastasis	301,210	99.25%	1,356	92.13%
	LN	81,921	39.20%	125	79.61%
U9	Primary	240,438	99.04%	1,090	94.38%
	Metastasis	484,597	99.22%	2,185	94.32%
	LN	218,276	98.98%	996	95.04%
U10	Primary	615,416	98.98%	2,729	94.19%
	Metastasis	318,592	98.75%	1,430	91.30%
	LN	299,403	99.06%	1,349	94.73%
U11	Primary	587,720	93.89%	2,237	87.42%
	Metastasis	383,129	96.23%	1,495	80.21%
	LN	182,914	95.79%	712	82.53%
U12	Primary	206,337	95.77%	809	79.48%
	Metastasis	326,329	95.62%	1,275	80.26%
	LN	188,392	96.61%	736	77.66%
U13	Primary	204,776	97.69%	883	91.24%
	Metastasis	534,773	98.35%	2,107	75.92%
	LN	56,903	97.05%	241	91.84%
U14	Primary	849,892	99.45%	3,784	76.77%
	Metastasis	804,226	99.56%	3,529	71.63%
	LN	706,708	99.27%	3,121	83.82%
U15	Primary	362,594	81.46%	1,292	84.44%
	Metastasis	261,099	98.55%	1,156	93.11%
	LN	239,607	98.75%	1,064	92.52%

Case	Sample	Mapped reads	On Target	Mean Depth	Uniformity
U16	Primary	203,650	98.86%	914	96.12%
	Secondary	292,870	99.05%	1,314	93.60%
	LN	298,289	96.95%	1,319	96.22%
U17	Primary	222,491	98.82%	1,019	96.94%
	Secondary	222,639	98.93%	997	98.50%
	LN	285,616	98.59%	1,289	95.33%
U18	Primary	353,628	98.98%	1,558	88.26%
	Secondary	372,308	98.86%	1,679	90.20%
	LN	319,936	98.52%	1,414	92.86%
U19	Primary	221,626	26.06%	223	71.49%
	Secondary	396,024	98.90%	1,735	81.81%
	LN	256,552	97.51%	1,108	94.47%
U20	Primary	207,667	97.25%	861	80.17%
	Secondary	275,058	98.80%	1,250	99.00%
	LN	315,872	98.51%	1,422	95.81%

Case	Sample	Mapped reads	On Target	Mean Depth	Uniformity
A1	Primary	277,872	98.22%	1,108	83.93%
	Second tumour	148,161	93.55%	537	67.17%
	NT	152,638	97.22%	622	89.19%
A2	Primary	179,081	98.57%	794	92.64%
	Second tumour	295,178	98.60%	1,302	92.78%
	LN	221,262	98.46%	974	93.63%
A3	Primary	204,408	99.14%	920	92.12%
	Second tumour	197,535	98.94%	872	85.16%
	NT	232,729	99.05%	1,043	94.16%
A4	Primary	195,411	88.27%	660	73.72%
	Second tumour	230,266	79.99%	535	42.77%
	LN	223,487	92.36%	818	80.79%
A5	Primary	163,119	98.25%	728	96.94%
	Second tumour	199,085	98.94%	899	95.78%
	LN	197,972	98.58%	870	95.04%

Case	Sample	Mapped reads	On Target	Mean Depth	Uniformity
A1	Primary	384,184	98.95%	3,878	93.39%
	Second tumour	583,895	98.81%	5,984	87.92%
	NT	868,622	96.65%	8,359	97.93%
A2	Primary	351,774	98.65%	3,608	100.00%
	Second tumour	494,453	98.37%	5,105	92.36%
	LN	310,391	97.03%	2,836	86.95%
A3	Primary	878,514	98.49%	8,995	92.48%
	Second tumour	632,133	97.26%	6,235	81.93%
	NT	966,650	79.04%	7,397	92.19%
A4	Primary	349,145	98.27%	3,519	74.16%
	Second tumour	676,485	98.33%	7,086	75.90%
	LN	441,385	98.84%	4,491	83.96%
A5	Primary	385,997	98.53%	3,915	100.00%
	Second tumour	463,982	99.08%	4,770	97.89%
	LN	696,232	82.96%	5,232	90.18%

Supplementary Table 1.4: General tNGS metrics for all cases in the PRIMINGS study (Chapter 5). Tables showing the general sequencing metrics of all chip runs, including Mapped reads, the percentage of On Target reads, Mean Depth and Uniformity. Unambiguous metastatic (red) and independent primary (blue) cases are shown (all CHPv2). Additionally shown are those of the ambiguous cohort sequenced using the CHPv2 (yellow) and OST panel (green).

Bibliography

Abbosh, C., Birkbak, N. J., Wilson, G. A., Jamal-Hanjani, M., Constantin, T., Salari, R., Le Quesne, J., Moore, D. A., Veeriah, S., Rosenthal, R., Marafioti, T., Kirkizlar, E., Watkins, T. B. K., McGranahan, N., Ward, S., Martinson, L., Riley, J., Fraioli, F., Al Bakir, M., Gronroos, E., Zambrana, F., Endozo, R., Bi, W. L., Fennessy, F. M., Sponer, N., Johnson, D., Laycock, J., Shafi, S., Czyzewska-Khan, J., Rowan, A., Chambers, T., Matthews, N., Turajlic, S., Hiley, C., Lee, S. M., Forster, M. D., Ahmad, T., Falzon, M., Borg, E., Lawrence, D., Hayward, M., Kolvekar, S., Panagiotopoulos, N., Janes, S. M., Thakrar, R., Ahmed, A., Blackhall, F., Summers, Y., Hafez, D., Naik, A., Ganguly, A., Kareht, S., Shah, R., Joseph, L., Marie Quinn, A., Crosbie, P. A., Naidu, B., Middleton, G., Langman, G., Trotter, S., Nicolson, M., Remmen, H., Kerr, K., Chetty, M., Gomersall, L., Fennell, D. A., Nakas, A., Rathinam, S., Anand, G., Khan, S., Russell, P., Ezhil, V., Ismail, B., Irvin-Sellers, M., Prakash, V., Lester, J. F., Kornaszewska, M., Attanoos, R., Adams, H., Davies, H., Oukrif, D., Akarca, A. U., Hartley, J. A., Lowe, H. L., Lock, S., Iles, N., Bell, H., Ngai, Y., Elgar, G., Szallasi, Z., Schwarz, R. F., Herrero, J., Stewart, A., Quezada, S. A., Peggs, K. S., Van Loo, P., Dive, C., Lin, C. J., Rabinowitz, M., Aerts, H., Hackshaw, A., Shaw, J. A., Zimmermann, B. G. and Swanton, C. (2017) Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature, 545*(7655), pp. 446-451.

Aberle, D. R., Adams, A. M., Berg, C. D., Black, W. C., Clapp, J. D., Fagerstrom, R. M., Gareen, I. F., Gatsonis, C., Marcus, P. M. and Sicks, J. D. (2011) Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med, 365*(5), pp. 395-409.

Acuna-Hidalgo, R., Sengul, H., Steehouwer, M., van de Vorst, M., Vermeulen, S. H., Kiemeney, L., Veltman, J. A., Gilissen, C. and Hoischen, A. (2017) Ultra-sensitive Sequencing Identifies High Prevalence of Clonal Hematopoiesis-Associated Mutations throughout Adult Life. *Am J Hum Genet, 101*(1), pp. 50-64.

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Borresen-Dale, A. L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjord, J. E., Foekens, J. A., Greaves, M., Hosoda, F., Hutter, B., Ilicic, T., Imbeaud, S., Imielinski, M., Jager, N., Jones, D. T., Jones, D., Knappskog, S., Kool, M., Lakhani, S. R., Lopez-Otin, C., Martin, S., Munshi, N. C., Nakamura, H., Northcott, P. A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J. V., Puente, X. S., Raine, K., Ramakrishna, M., Richardson, A. L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T. N., Span, P. N., Teague, J. W., Totoki, Y., Tutt, A. N., Valdes-Mas, R., van Buuren, M. M., van 't Veer, L., Vincent-Salomon, A., Waddell, N., Yates, L. R., Zucman-Rossi, J., Futreal, P. A.,

McDermott, U., Lichter, P., Meyerson, M., Grimmond, S. M., Siebert, R., Campo, E., Shibata, T., Pfister, S. M., Campbell, P. J. and Stratton, M. R. (2013) Signatures of mutational processes in human cancer. *Nature, 500*(7463), pp. 415-21.

Ali, G., Bruno, R. and Fontanini, G. (2018) The pathological and molecular diagnosis of malignant pleural mesothelioma: a literature review. *J Thorac Dis, 10*(Suppl 2), pp. S276-s284.

Alley, E. W., Lopez, J., Santoro, A., Morosky, A., Saraf, S., Piperdi, B. and van Brummelen, E. (2017) Clinical safety and activity of pembrolizumab in patients with malignant pleural mesothelioma (KEYNOTE-028): preliminary results from a non-randomised, open-label, phase 1b trial. *Lancet Oncol, 18*(5), pp. 623-630.

Anderson, K., Lutz, C., van Delft, F. W., Bateman, C. M., Guo, Y., Colman, S. M., Kempski, H., Moorman, A. V., Titley, I., Swansbury, J., Kearney, L., Enver, T. and Greaves, M. (2011) Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature, 469*(7330), pp. 356-61.

Andujar, P., Wang, J., Descatha, A., Galateau-Salle, F., Abd-Alsamad, I., Billon-Galland, M. A., Blons, H., Clin, B., Danel, C., Housset, B., Laurent-Puig, P., Le Pimpec-Barthes, F., Letourneux, M., Monnet, I., Regnard, J. F., Renier, A., Zucman-Rossi, J., Pairon, J. C. and Jaurand, M. C. (2010) p16INK4A inactivation mechanisms in non-small-cell lung cancer patients occupationally exposed to asbestos. *Lung Cancer, 67*(1), pp. 23-30.

Artal Cortes, A., Calera Urquizu, L. and Hernando Cubero, J. (2015) Adjuvant chemotherapy in non-small cell lung cancer: state-of-the-art. *Transl Lung Cancer Res, 4*(2), pp. 191-7.

Balduyck, B., Trousse, D., Nakas, A., Martin-Ucar, A. E., Edwards, J. and Waller, D. A. (2010) Therapeutic surgery for nonepithelioid malignant pleural mesothelioma: is it really worthwhile? *Ann Thorac Surg, 89*(3), pp. 907-11.

Bao, L., Messer, K., Schwab, R., Harismendy, O., Pu, M., Crain, B., Yost, S., Frazer, K. A., Rana, B., Hasteh, F., Wallace, A. and Parker, B. A. (2015) Mutational Profiling Can Establish Clonal or Independent Origin in Synchronous Bilateral Breast and Other Tumors. *PLoS One, 10*(11), pp. e0142487.

Bean, J., Brennan, C., Shih, J. Y., Riely, G., Viale, A., Wang, L., Chitale, D., Motoi, N., Szoke, J., Broderick, S., Balak, M., Chang, W. C., Yu, C. J., Gazdar, A., Pass, H., Rusch, V., Gerald, W., Huang, S. F., Yang, P. C., Miller, V., Ladanyi, M., Yang,

C. H. and Pao, W. (2007) MET amplification occurs with or without T790M mutations in EGFR mutant lung tumors with acquired resistance to gefitinib or erlotinib. *Proc Natl Acad Sci U S A, 104*(52), pp. 20932-7.

Bedard, P. L., Hansen, A. R., Ratain, M. J. and Siu, L. L. (2013) Tumour heterogeneity in the clinic. *Nature, 501*(7467), pp. 355-64.

Bettegowda, C., Sausen, M., Leary, R. J., Kinde, I., Wang, Y., Agrawal, N., Bartlett, B. R., Wang, H., Luber, B., Alani, R. M., Antonarakis, E. S., Azad, N. S., Bardelli, A., Brem, H., Cameron, J. L., Lee, C. C., Fecher, L. A., Gallia, G. L., Gibbs, P., Le, D., Giuntoli, R. L., Goggins, M., Hogarty, M. D., Holdhoff, M., Hong, S. M., Jiao, Y., Juhl, H. H., Kim, J. J., Siravegna, G., Laheru, D. A., Lauricella, C., Lim, M., Lipson, E. J., Marie, S. K., Netto, G. J., Oliner, K. S., Olivi, A., Olsson, L., Riggins, G. J., Sartore-Bianchi, A., Schmidt, K., Shih I, M., Oba-Shinjo, S. M., Siena, S., Theodorescu, D., Tie, J., Harkins, T. T., Veronese, S., Wang, T. L., Weingart, J. D., Wolfgang, C. L., Wood, L. D., Xing, D., Hruban, R. H., Wu, J., Allen, P. J., Schmidt, C. M., Choti, M. A., Velculescu, V. E., Kinzler, K. W., Vogelstein, B., Papadopoulos, N. and Diaz, L. A., Jr. (2014) Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med, 6*(224), pp. 224ra24.

Blakely, C. M. and Bivona, T. G. (2012) Resiliency of lung cancers to EGFR inhibitor treatment unveiled, offering opportunities to divide and conquer EGFR inhibitor resistance. *Cancer Discov, 2*(10), pp. 872-5.

Bloom, H. J. and Richardson, W. W. (1957) Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *Br J Cancer, 11*(3), pp. 359-77.

Boch, C., Kollmeier, J., Roth, A., Stephan-Falkenau, S., Misch, D., Gruning, W., Bauer, T. T. and Mairinger, T. (2013) The frequency of EGFR and KRAS mutations in non-small cell lung cancer (NSCLC): routine screening data for central Europe from a cohort study. *BMJ Open, 3*(4).

Bonelli, M. A., Digiacomo, G., Fumarola, C., Alfieri, R., Quaini, F., Falco, A., Madeddu, D., La Monica, S., Cretella, D., Ravelli, A., Ulivi, P., Tebaldi, M., Calistri, D., Delmonte, A., Ampollini, L., Carbognani, P., Tiseo, M., Cavazzoni, A. and Petronini, P. G. (2017) Combined Inhibition of CDK4/6 and PI3K/AKT/mTOR Pathways Induces a Synergistic Anti-Tumor Effect in Malignant Pleural Mesothelioma Cells. *Neoplasia, 19*(8), pp. 637-648.

Bononi, A., Yang, H., Giorgi, C., Patergnani, S., Pellegrini, L., Su, M., Xie, G., Signorato, V., Pastorino, S., Morris, P., Sakamoto, G., Kuchay, S. and Gaudino,

G. (2017) Germline BAP1 mutations induce a Warburg effect. *24*(10), pp. 1694-1704.

Bott, M., Brevet, M., Taylor, B. S., Shimizu, S., Ito, T., Wang, L., Creaney, J., Lake, R. A., Zakowski, M. F., Reva, B., Sander, C., Delsite, R., Powell, S., Zhou, Q., Shen, R., Olshen, A., Rusch, V. and Ladanyi, M. (2011) The nuclear deubiquitinase BAP1 is commonly inactivated by somatic mutations and 3p21.1 losses in malignant pleural mesothelioma. *Nat Genet, 43*(7), pp. 668-72.

Bradley, J. D., Paulus, R., Komaki, R., Masters, G., Blumenschein, G., Schild, S., Bogart, J., Hu, C., Forster, K., Magliocco, A., Kavadi, V., Garces, Y. I., Narayan, S., Iyengar, P., Robinson, C., Wynn, R. B., Koprowski, C., Meng, J., Beitler, J., Gaur, R., Curran, W., Jr. and Choy, H. (2015) Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patients with stage IIIA or IIIB non-small-cell lung cancer (RTOG 0617): a randomised, two-by-two factorial phase 3 study. *Lancet Oncol, 16*(2), pp. 187-99.

Brinkmann, D., Ryan, A., Ayhan, A., McCluggage, W. G., Feakins, R., Santibanez-Koref, M. F., Mein, C. A., Gayther, S. A. and Jacobs, I. J. (2004) A molecular genetic and statistical approach for the diagnosis of dual-site cancers. *J Natl Cancer Inst, 96*(19), pp. 1441-6.

Brock, A., Chang, H. and Huang, S. (2009) Non-genetic heterogeneity--a mutation-independent driving force for the somatic evolution of tumours. *Nat Rev Genet, 10*(5), pp. 336-42.

Bronte, G., Incorvaia, L., Rizzo, S., Passiglia, F., Galvano, A., Rizzo, F., Rolfo, C., Fanale, D., Listi, A., Natoli, C., Bazan, V. and Russo, A. (2016) The resistance related to targeted therapy in malignant pleural mesothelioma: Why has not the target been hit yet? *Crit Rev Oncol Hematol, 107*, pp. 20-32.

Brose, M. S., Volpe, P., Feldman, M., Kumar, M., Rishi, I., Gerrero, R., Einhorn, E., Herlyn, M., Minna, J., Nicholson, A., Roth, J. A., Albelda, S. M., Davies, H., Cox, C., Brignell, G., Stephens, P., Futreal, P. A., Wooster, R., Stratton, M. R. and Weber, B. L. (2002) BRAF and RAS mutations in human lung cancer and melanoma. *Cancer Res, 62*(23), pp. 6997-7000.

Budczies, J., von Winterfeld, M., Klauschen, F., Bockmayr, M., Lennerz, J. K., Denkert, C., Wolf, T., Warth, A., Dietel, M., Anagnostopoulos, I., Weichert, W., Wittschieber, D. and Stenzinger, A. (2015) The landscape of metastatic progression patterns across major human cancers. *Oncotarget, 6*(1), pp. 570-83.

Bueno, R., Stawiski, E. W., Goldstein, L. D., Durinck, S., De Rienzo, A., Modrusan, Z., Gnad, F., Nguyen, T. T., Jaiswal, B. S., Chirieac, L. R., Sciaranghella, D., Dao, N., Gustafson, C. E., Munir, K. J., Hackney, J. A., Chaudhuri, A., Gupta, R., Guillory, J., Toy, K., Ha, C., Chen, Y. J., Stinson, J., Chaudhuri, S., Zhang, N., Wu, T. D., Sugarbaker, D. J., de Sauvage, F. J., Richards, W. G. and Seshagiri, S. (2016) Comprehensive genomic analysis of malignant pleural mesothelioma identifies recurrent mutations, gene fusions and splicing alterations. *Nat Genet, 48*(4), pp. 407-16.

Burdett, S., Rydzewska, L., Tierney, J., Fisher, D., Parmar, M. K., Arriagada, R., Pignon, J. P. and Le Pechoux, C. (2016) Postoperative radiotherapy for non-small cell lung cancer. *Cochrane Database Syst Rev, 10*, pp. Cd002142.

Burdett, S., Stephens, R., Stewart, L., Tierney, J., Auperin, A., Le Chevalier, T., Le Pechoux, C., Pignon, J. P., Arriagada, R., Higgins, J., Johnson, D., van Meerbeeck, J., Parmar, M., Souhami, R., Bell, D., Cartei, G., Cormier, Y., Cullen, M., Ganz, P., Gridelli, C., Kaasa, S., Quoix, E., Rapp, E., Seymour, L., Spiro, S., Thatcher, N., Tummarello, D., Williams, C. and Williamson, I. (2008) Chemotherapy in addition to supportive care improves survival in advanced non-small-cell lung cancer: a systematic review and meta-analysis of individual patient data from 16 randomized controlled trials. *J Clin Oncol, 26*(28), pp. 4617-25.

Caldas, C. (2012) Cancer sequencing unravels clonal evolution. *Nat Biotechnol, 30*(5), pp. 408-10.

Camidge, D. R., Bang, Y. J., Kwak, E. L., Iafrate, A. J., Varella-Garcia, M., Fox, S. B., Riely, G. J., Solomon, B., Ou, S. H., Kim, D. W., Salgia, R., Fidias, P., Engelman, J. A., Gandhi, L., Janne, P. A., Costa, D. B., Shapiro, G. I., Lorusso, P., Ruffner, K., Stephenson, P., Tang, Y., Wilner, K., Clark, J. W. and Shaw, A. T. (2012) Activity and safety of crizotinib in patients with ALK-positive non-small-cell lung cancer: updated results from a phase 1 study. *Lancet Oncol, 13*(10), pp. 1011-9.

Cao, C., Tian, D., Park, J., Allan, J., Pataky, K. A. and Yan, T. D. (2014) A systematic review and meta-analysis of surgical treatments for malignant pleural mesothelioma. *Lung Cancer, 83*(2), pp. 240-5.

Cappuzzo, F., Marchetti, A., Skokan, M., Rossi, E., Gajapathy, S., Felicioni, L., Del Grammastro, M., Sciarrotta, M. G., Buttitta, F., Incarbone, M., Toschi, L., Finocchiaro, G., Destro, A., Terracciano, L., Roncalli, M., Alloisio, M., Santoro, A. and Varella-Garcia, M. (2009) Increased MET gene copy number negatively

affects survival of surgically resected non-small-cell lung cancer patients. *J Clin Oncol, 27*(10), pp. 1667-74.

Carbone, M., Yang, H., Pass, H. I., Krausz, T., Testa, J. R. and Gaudino, G. (2013) BAP1 and cancer. *Nat Rev Cancer, 13*(3), pp. 153-9.

Cardarella, S., Ogino, A., Nishino, M., Butaney, M., Shen, J., Lydon, C., Yeap, B. Y., Sholl, L. M., Johnson, B. E. and Janne, P. A. (2013) Clinical, pathologic, and biologic features associated with BRAF mutations in non-small cell lung cancer. *Clin Cancer Res, 19*(16), pp. 4532-40.

Caruso, R., Parisi, A., Bonanno, A., Paparo, D., Quattrocchi, E., Branca, G., Scardigno, M. and Fedele, F. (2012) Histologic coagulative tumour necrosis as a prognostic indicator of aggressiveness in renal, lung, thyroid and colorectal carcinomas: A brief review. *Oncol Lett, 3*(1), pp. 16-18.

Chaudhuri, A. A., Chabon, J. J., Lovejoy, A. F., Newman, A. M., Stehr, H., Azad, T. D., Khodadoust, M. S., Esfahani, M. S., Liu, C. L., Zhou, L., Scherer, F., Kurtz, D. M., Say, C., Carter, J. N., Merriott, D. J., Dudley, J. C., Binkley, M. S., Modlin, L., Padda, S. K., Gensheimer, M. F., West, R. B., Shrager, J. B., Neal, J. W., Wakelee, H. A., Loo, B. W., Jr., Alizadeh, A. A. and Diehn, M. (2017) Early Detection of Molecular Residual Disease in Localized Lung Cancer by Circulating Tumor DNA Profiling. *Cancer Discov, 7*(12), pp. 1394-1403.

Chernova, T., Murphy, F. A., Galavotti, S., Sun, X. M., Powley, I. R., Grosso, S., Schinwald, A., Zacarias-Cabeza, J., Dudek, K. M., Dinsdale, D., Le Quesne, J., Bennett, J., Nakas, A., Greaves, P., Poland, C. A., Donaldson, K., Bushell, M., Willis, A. E. and MacFarlane, M. (2017) Long-Fiber Carbon Nanotubes Replicate Asbestos-Induced Mesothelioma with Disruption of the Tumor Suppressor Gene Cdkn2a (Ink4a/Arf). *Curr Biol, 27*(21), pp. 3302-3314.e6.

Christoph, D. C. and Eberhardt, W. E. (2014) Systemic treatment of malignant pleural mesothelioma: new agents in clinical trials raise hope of relevant improvements. *Curr Opin Oncol, 26*(2), pp. 171-81.

Ciuleanu, T., Brodowicz, T., Zielinski, C., Kim, J. H., Krzakowski, M., Laack, E., Wu, Y. L., Bover, I., Begbie, S., Tzekova, V., Cucevic, B., Pereira, J. R., Yang, S. H., Madhavan, J., Sugarman, K. P., Peterson, P., John, W. J., Krejcy, K. and Belani, C. P. (2009) Maintenance pemetrexed plus best supportive care versus placebo plus best supportive care for non-small-cell lung cancer: a randomised, double-blind, phase 3 study. *Lancet, 374*(9699), pp. 1432-40.

Cohen, J. D., Li, L., Wang, Y., Thoburn, C., Afsari, B., Danilova, L., Douville, C., Javed, A. A., Wong, F., Mattox, A., Hruban, R. H., Wolfgang, C. L., Goggins, M. G., Dal Molin, M., Wang, T.-L., Roden, R., Klein, A. P., Ptak, J., Dobbyn, L., Schaefer, J., Silliman, N., Popoli, M., Vogelstein, J. T., Browne, J. D., Schoen, R. E., Brand, R. E., Tie, J., Gibbs, P., Wong, H.-L., Mansfield, A. S., Jen, J., Hanash, S. M., Falconi, M., Allen, P. J., Zhou, S., Bettegowda, C., Diaz, L. A., Tomasetti, C., Kinzler, K. W., Vogelstein, B., Lennon, A. M. and Papadopoulos, N. (2018) Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science, 359*(6378), pp. 926-930.

Cooper, J. and Giancotti, F. G. (2014) Molecular insights into NF2/Merlin tumor suppressor function. *FEBS Lett, 588*(16), pp. 2743-52.

Couraud, S., Vaca-Paniagua, F., Villar, S., Oliver, J., Schuster, T., Blanche, H., Girard, N., Tredaniel, J., Guilleminault, L., Gervais, R., Prim, N., Vincent, M., Margery, J., Larive, S., Foucher, P., Duvert, B., Vallee, M., Le Calvez-Kelm, F., McKay, J., Missy, P., Morin, F., Zalcman, G., Olivier, M. and Souquet, P. J. (2014) Noninvasive diagnosis of actionable mutations by deep sequencing of circulating free DNA in lung cancer from never-smokers: a proof-of-concept study from BioCAST/IFCT-1002. *Clin Cancer Res, 20*(17), pp. 4613-24.

Creaney, J., Dick, I. M. and Robinson, B. W. (2015) Discovery of new biomarkers for malignant mesothelioma. *Curr Pulmonol Rep, 4*(1), pp. 15-21.

Davidson, M. R., Gazdar, A. F. and Clarke, B. E. (2013) The pivotal role of pathology in the management of lung cancer. *J Thorac Dis, 5 Suppl 5*, pp. S463-78.

Dawson, S. J., Tsui, D. W., Murtaza, M., Biggs, H., Rueda, O. M., Chin, S. F., Dunning, M. J., Gale, D., Forshew, T., Mahler-Araujo, B., Rajan, S., Humphray, S., Becq, J., Halsall, D., Wallis, M., Bentley, D., Caldas, C. and Rosenfeld, N. (2013) Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med, 368*(13), pp. 1199-209.

de Bruin, E. C., McGranahan, N., Mitter, R., Salm, M., Wedge, D. C., Yates, L., Jamal-Hanjani, M., Shafi, S., Murugaesu, N., Rowan, A. J., Gronroos, E., Muhammad, M. A., Horswell, S., Gerlinger, M., Varela, I., Jones, D., Marshall, J., Voet, T., Van Loo, P., Rassl, D. M., Rintoul, R. C., Janes, S. M., Lee, S. M., Forster, M., Ahmad, T., Lawrence, D., Falzon, M., Capitanio, A., Harkins, T. T., Lee, C. C., Tom, W., Teefe, E., Chen, S. C., Begum, S., Rabinowitz, A., Phillimore, B., Spencer-Dene, B., Stamp, G., Szallasi, Z., Matthews, N., Stewart, A., Campbell, P. and Swanton, C. (2014) Spatial and temporal diversity in

genomic instability processes defines lung cancer evolution. *Science, 346*(6206), pp. 251-6.

de Leng, W. W., Gadellaa-van Hooijdonk, C. G., Barendregt-Smouter, F. A., Koudijs, M. J., Nijman, I., Hinrichs, J. W., Cuppen, E., van Lieshout, S., Loberg, R. D., de Jonge, M., Voest, E. E., de Weger, R. A., Steeghs, N., Langenberg, M. H., Sleijfer, S., Willems, S. M. and Lolkema, M. P. (2016) Targeted Next Generation Sequencing as a Reliable Diagnostic Assay for the Detection of Somatic Mutations in Tumours Using Minimal DNA Amounts from Formalin Fixed Paraffin Embedded Material. *PLoS One, 11*(2), pp. e0149405.

Dewhurst, S. M., McGranahan, N., Burrell, R. A., Rowan, A. J., Gronroos, E., Endesfelder, D., Joshi, T., Mouradov, D., Gibbs, P., Ward, R. L., Hawkins, N. J., Szallasi, Z., Sieber, O. M. and Swanton, C. (2014) Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution. *Cancer Discov, 4*(2), pp. 175-85.

Diaz, L. A., Jr. and Bardelli, A. (2014) Liquid biopsies: genotyping circulating tumor DNA. *J Clin Oncol, 32*(6), pp. 579-86.

Ding, L., Getz, G., Wheeler, D. A., Mardis, E. R., McLellan, M. D., Cibulskis, K., Sougnez, C., Greulich, H., Muzny, D. M., Morgan, M. B., Fulton, L., Fulton, R. S., Zhang, Q., Wendl, M. C., Lawrence, M. S., Larson, D. E., Chen, K., Dooling, D. J., Sabo, A., Hawes, A. C., Shen, H., Jhangiani, S. N., Lewis, L. R., Hall, O., Zhu, Y., Mathew, T., Ren, Y., Yao, J., Scherer, S. E., Clerc, K., Metcalf, G. A., Ng, B., Milosavljevic, A., Gonzalez-Garay, M. L., Osborne, J. R., Meyer, R., Shi, X., Tang, Y., Koboldt, D. C., Lin, L., Abbott, R., Miner, T. L., Pohl, C., Fewell, G., Haipek, C., Schmidt, H., Dunford-Shore, B. H., Kraja, A., Crosby, S. D., Sawyer, C. S., Vickery, T., Sander, S., Robinson, J., Winckler, W., Baldwin, J., Chirieac, L. R., Dutt, A., Fennell, T., Hanna, M., Johnson, B. E., Onofrio, R. C., Thomas, R. K., Tonon, G., Weir, B. A., Zhao, X., Ziaugra, L., Zody, M. C., Giordano, T., Orringer, M. B., Roth, J. A., Spitz, M. R., Wistuba, II, Ozenberger, B., Good, P. J., Chang, A. C., Beer, D. G., Watson, M. A., Ladanyi, M., Broderick, S., Yoshizawa, A., Travis, W. D., Pao, W., Province, M. A., Weinstock, G. M., Varmus, H. E., Gabriel, S. B., Lander, E. S., Gibbs, R. A., Meyerson, M. and Wilson, R. K. (2008) Somatic mutations affect key pathways in lung adenocarcinoma. *Nature, 455*(7216), pp. 1069-75.

Duraiyan, J., Govindarajan, R., Kaliyappan, K. and Palanisamy, M. (2012) Applications of immunohistochemistry. *J Pharm Bioallied Sci, 4*(Suppl 2), pp. S307-9.

Dutt, A., Ramos, A. H., Hammerman, P. S., Mermel, C., Cho, J., Sharifnia, T., Chande, A., Tanaka, K. E., Stransky, N., Greulich, H., Gray, N. S. and Meyerson, M. (2011) Inhibitor-sensitive FGFR1 amplification in human non-small cell lung cancer. *PLoS One, 6*(6), pp. e20351.

Edwards, J. G., Stewart, D. J., Martin-Ucar, A., Muller, S., Richards, C. and Waller, D. A. (2006) The pattern of lymph node involvement influences outcome after extrapleural pneumonectomy for malignant mesothelioma. *J Thorac Cardiovasc Surg, 131*(5), pp. 981-7.

Einaga, N., Yoshida, A., Noda, H., Suemitsu, M., Nakayama, Y., Sakurada, A., Kawaji, Y., Yamaguchi, H., Sasaki, Y., Tokino, T. and Esumi, M. (2017) Assessment of the quality of DNA from various formalin-fixed paraffin-embedded (FFPE) tissues and the use of this DNA for next-generation sequencing (NGS) with no artifactual mutation. *PLoS One, 12*(5), pp. e0176280.

Ellis, P. M. and Vandermeer, R. (2011) Delays in the diagnosis of lung cancer. *J Thorac Dis, 3*(3), pp. 183-8.

Fennell, D. A., Kirkpatrick, E., Cozens, K., Nye, M., Lester, J., Hanna, G., Steele, N., Szlosarek, P., Danson, S., Lord, J., Ottensmeier, C., Barnes, D., Hill, S., Kalevras, M., Maishman, T. and Griffiths, G. (2018) CONFIRM: a double-blind, placebo-controlled phase III clinical trial investigating the effect of nivolumab in patients with relapsed mesothelioma: study protocol for a randomised controlled trial. *Trials, 19*(1), pp. 233.

Fernando, M. R., Jiang, C., Krzyzanowski, G. D. and Ryan, W. L. (2017) New evidence that a large proportion of human blood plasma cell-free DNA is localized in exosomes. *PLoS One, 12*(8), pp. e0183915.

Finn, R. S., Brims, F. J. H., Gandhi, A., Olsen, N., Musk, A. W., Maskell, N. A. and Lee, Y. C. G. (2012) Postmortem findings of malignant pleural mesothelioma: a two-center study of 318 patients. *Chest, 142*(5), pp. 1267-1273.

Finn, R. S., Crown, J. P., Lang, I., Boer, K., Bondarenko, I. M., Kulyk, S. O., Ettl, J., Patel, R., Pinter, T., Schmidt, M., Shparyk, Y., Thummala, A. R., Voytko, N. L., Fowst, C., Huang, X., Kim, S. T., Randolph, S. and Slamon, D. J. (2015) The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): a randomised phase 2 study. *Lancet Oncol, 16*(1), pp. 25-35.

Fribbens, C., O'Leary, B., Kilburn, L., Hrebien, S., Garcia-Murillas, I., Beaney, M., Cristofanilli, M., Andre, F., Loi, S., Loibl, S., Jiang, J., Bartlett, C. H., Koehler, M., Dowsett, M., Bliss, J. M., Johnston, S. R. and Turner, N. C. (2016) Plasma ESR1 Mutations and the Treatment of Estrogen Receptor-Positive Advanced Breast Cancer. *J Clin Oncol, 34*(25), pp. 2961-8.

Friedberg, J. S., Simone, C. B., 2nd, Culligan, M. J., Barsky, A. R., Doucette, A., McNulty, S., Hahn, S. M., Alley, E., Sterman, D. H., Glatstein, E. and Cengel, K. A. (2017) Extended Pleurectomy-Decortication-Based Treatment for Advanced Stage Epithelial Mesothelioma Yielding a Median Survival of Nearly Three Years. *Ann Thorac Surg, 103*(3), pp. 912-919.

Frost, G. (2013) The latency period of mesothelioma among a cohort of British asbestos workers (1978-2005). *Br J Cancer, 109*(7), pp. 1965-73.

Garrido-Castro, A. C. and Felip, E. (2013) HER2 driven non-small cell lung cancer (NSCLC): potential therapeutic approaches. *Transl Lung Cancer Res, 2*(2), pp. 122-7.

Gazdar, A. F. (2009) Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. *Oncogene, 28 Suppl 1*, pp. S24-31.

Genovese, G., Jaiswal, S., Ebert, B. L. and McCarroll, S. A. (2015) Clonal hematopoiesis and blood-cancer risk. *N Engl J Med, 372*(11), pp. 1071-2.

Gerlinger, M., Rowan, A. J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., Varela, I., Phillimore, B., Begum, S., McDonald, N. Q., Butler, A., Jones, D., Raine, K., Latimer, C., Santos, C. R., Nohadani, M., Eklund, A. C., Spencer-Dene, B., Clark, G., Pickering, L., Stamp, G., Gore, M., Szallasi, Z., Downward, J., Futreal, P. A. and Swanton, C. (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med, 366*(10), pp. 883-92.

Gibbons, D. L., Byers, L. A. and Kurie, J. M. (2014) Smoking, p53 mutation, and lung cancer. *Mol Cancer Res, 12*(1), pp. 3-13.

Go, H., Jeon, Y. K., Park, H. J., Sung, S. W., Seo, J. W. and Chung, D. H. (2010) High MET gene copy number leads to shorter survival in patients with non-small cell lung cancer. *J Thorac Oncol, 5*(3), pp. 305-13.

Goldstraw, P., Chansky, K., Crowley, J., Rami-Porta, R., Asamura, H., Eberhardt, W. E., Nicholson, A. G., Groome, P., Mitchell, A. and Bolejack, V. (2016) The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol, 11*(1), pp. 39-51.

Goswami, R. S., Patel, K. P., Singh, R. R., Meric-Bernstam, F., Kopetz, E. S., Subbiah, V., Alvarez, R. H., Davies, M. A., Jabbar, K. J., Roy-Chowdhuri, S., Lazar, A. J., Medeiros, L. J., Broaddus, R. R., Luthra, R. and Routbort, M. J. (2015) Hotspot mutation panel testing reveals clonal evolution in a study of 265 paired primary and metastatic tumors. *Clin Cancer Res, 21*(11), pp. 2644-51.

Govindan, R., Ding, L., Griffith, M., Subramanian, J., Dees, N. D., Kanchi, K. L., Maher, C. A., Fulton, R., Fulton, L., Wallis, J., Chen, K., Walker, J., McDonald, S., Bose, R., Ornitz, D., Xiong, D., You, M., Dooling, D. J., Watson, M., Mardis, E. R. and Wilson, R. K. (2012) Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell, 150*(6), pp. 1121-34.

Greaves, M. and Maley, C. C. (2012) Clonal evolution in cancer. *Nature, 481*(7381), pp. 306-13.

Guo, G., Chmielecki, J., Goparaju, C., Heguy, A., Dolgalev, I., Carbone, M., Seepo, S., Meyerson, M. and Pass, H. I. (2015) Whole-exome sequencing reveals frequent genetic alterations in BAP1, NF2, CDKN2A, and CUL1 in malignant pleural mesothelioma. *Cancer Res, 75*(2), pp. 264-9.

Guo, N., Lou, F., Ma, Y., Li, J., Yang, B., Chen, W., Ye, H., Zhang, J. B., Zhao, M. Y., Wu, W. J., Shi, R., Jones, L., Chen, K. S., Huang, X. F., Chen, S. Y. and Liu, Y. (2016) Circulating tumor DNA detection in lung cancer patients before and after surgery. *Sci Rep, 6*, pp. 33519.

Hadano, N., Murakami, Y., Uemura, K., Hashimoto, Y., Kondo, N., Nakagawa, N., Sueda, T. and Hiyama, E. (2016) Prognostic value of circulating tumour DNA in patients undergoing curative resection for pancreatic cancer. *Br J Cancer, 115*(1), pp. 59-65.

Hasako, S., Terasaka, M., Abe, N., Uno, T., Ohsawa, H., Hashimoto, A., Fujita, R., Tanaka, K., Okayama, T., Wadhwa, R., Miyadera, K., Aoyagi, Y., Yonekura, K. and Matsuo, K. (2018) TAS6417, A Novel EGFR Inhibitor Targeting Exon 20 Insertion Mutations. *Mol Cancer Ther, 17*(8), pp. 1648-1658.

Heintz, N. H., Janssen-Heininger, Y. M. and Mossman, B. T. (2010) Asbestos, lung cancers, and mesotheliomas: from molecular approaches to targeting tumor survival pathways. *Am J Respir Cell Mol Biol, 42*(2), pp. 133-9.

Heitzer, E. and Perakis, S. (2017) The potential of liquid biopsies for the early detection of cancer. *1*(1), pp. 36.

Hellwig, S., Nix, D. A., Gligorich, K. M., O'Shea, J. M., Thomas, A. and Fuertes, C. L. (2018) Automated size selection for short cell-free DNA fragments enriches for circulating tumor DNA and improves error correction during next generation sequencing. *13*(7), pp. e0197333.

Heppner, G. H. and Miller, B. E. (1983) Tumor heterogeneity: biological implications and therapeutic consequences. *Cancer Metastasis Rev, 2*(1), pp. 5-23.

Hida, T., Nokihara, H., Kondo, M., Kim, Y. H., Azuma, K., Seto, T., Takiguchi, Y., Nishio, M., Yoshioka, H., Imamura, F., Hotta, K., Watanabe, S., Goto, K., Satouchi, M., Kozuki, T., Shukuya, T., Nakagawa, K., Mitsudomi, T., Yamamoto, N., Asakawa, T., Asabe, R., Tanaka, T. and Tamura, T. (2017) Alectinib versus crizotinib in patients with ALK-positive non-small-cell lung cancer (J-ALEX): an open-label, randomised phase 3 trial. *Lancet, 390*(10089), pp. 29-39.

Hirsch, F. R., Varella-Garcia, M., Bunn, P. A., Jr., Di Maria, M. V., Veve, R., Bremmes, R. M., Baron, A. E., Zeng, C. and Franklin, W. A. (2003) Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *J Clin Oncol, 21*(20), pp. 3798-807.

Hu, Y., Ulrich, B. C., Supplee, J., Kuang, Y. and Lizotte, P. H. (2018) False-Positive Plasma Genotyping Due to Clonal Hematopoiesis.

Huang, J., Behrens, C., Wistuba, I., Gazdar, A. F. and Jagirdar, J. (2001) Molecular analysis of synchronous and metachronous tumors of the lung: impact on management and prognosis. *Ann Diagn Pathol, 5*(6), pp. 321-9.

Huh, D. and Paulsson, J. (2011) Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat Genet, 43*(2), pp. 95-100.

Hylebos, M., Op de Beeck, K., Pauwels, P., Zwaenepoel, K., van Meerbeeck, J. P. and Van Camp, G. (2018) Tumor-specific genetic variants can be detected in

circulating cell-free DNA of malignant pleural mesothelioma patients. *Lung Cancer, 124*, pp. 19-22.

Hylebos, M., Van Camp, G., Vandeweyer, G., Fransen, E., Beyens, M., Cornelissen, R., Suls, A., Pauwels, P., van Meerbeeck, J. P. and Op de Beeck, K. (2017) Large-scale copy number analysis reveals variations in genes not previously associated with malignant pleural mesothelioma. *Oncotarget, 8*(69), pp. 113673-113686.

Illei, P. B., Rusch, V. W., Zakowski, M. F. and Ladanyi, M. (2003) Homozygous deletion of CDKN2A and codeletion of the methylthioadenosine phosphorylase gene in the majority of pleural mesotheliomas. *Clin Cancer Res, 9*(6), pp. 2108-13.

Inai, K. (2008) Pathology of mesothelioma. *Environ Health Prev Med, 13*(2), pp. 60-4.

Iwakawa, R., Kohno, T., Anami, Y., Noguchi, M., Suzuki, K., Matsuno, Y., Mishima, K., Nishikawa, R., Tashiro, F. and Yokota, J. (2008) Association of p16 homozygous deletions with clinicopathologic characteristics and EGFR/KRAS/p53 mutations in lung adenocarcinoma. *Clin Cancer Res, 14*(12), pp. 3746-53.

Jahr, S., Hentze, H., Englisch, S., Hardt, D., Fackelmayer, F. O., Hesch, R. D. and Knippers, R. (2001) DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res, 61*(4), pp. 1659-65.

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P. V., Mar, B. G., Lindsley, R. C., Mermel, C. H., Burtt, N., Chavez, A., Higgins, J. M., Moltchanov, V., Kuo, F. C., Kluk, M. J., Henderson, B., Kinnunen, L., Koistinen, H. A., Ladenvall, C., Getz, G., Correa, A., Banahan, B. F., Gabriel, S., Kathiresan, S., Stringham, H. M., McCarthy, M. I., Boehnke, M., Tuomilehto, J., Haiman, C., Groop, L., Atzmon, G., Wilson, J. G., Neuberg, D., Altshuler, D. and Ebert, B. L. (2014) Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med, 371*(26), pp. 2488-98.

Jaiswal, S., Natarajan, P., Silver, A. J., Gibson, C. J., Bick, A. G., Shvartz, E., McConkey, M., Gupta, N., Gabriel, S., Ardissino, D., Baber, U., Mehran, R., Fuster, V., Danesh, J., Frossard, P., Saleheen, D., Melander, O., Sukhova, G. K., Neuberg, D., Libby, P., Kathiresan, S. and Ebert, B. L. (2017) Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med, 377*(2), pp. 111-121.

Jamal-Hanjani, M., Hackshaw, A., Ngai, Y., Shaw, J., Dive, C., Quezada, S., Middleton, G., de Bruin, E., Le Quesne, J., Shafi, S., Falzon, M., Horswell, S., Blackhall, F., Khan, I., Janes, S., Nicolson, M., Lawrence, D., Forster, M., Fennell, D., Lee, S. M., Lester, J., Kerr, K., Muller, S., Iles, N., Smith, S., Murugaesu, N., Mitter, R., Salm, M., Stuart, A., Matthews, N., Adams, H., Ahmad, T., Attanoos, R., Bennett, J., Birkbak, N. J., Booton, R., Brady, G., Buchan, K., Capitano, A., Chetty, M., Cobbold, M., Crosbie, P., Davies, H., Denison, A., Djearman, M., Goldman, J., Haswell, T., Joseph, L., Kornaszewska, M., Krebs, M., Langman, G., MacKenzie, M., Millar, J., Morgan, B., Naidu, B., Nonaka, D., Peggs, K., Pritchard, C., Remmen, H., Rowan, A., Shah, R., Smith, E., Summers, Y., Taylor, M., Veeriah, S., Waller, D., Wilcox, B., Wilcox, M., Woolhouse, I., McGranahan, N. and Swanton, C. (2014) Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. *PLoS Biol, 12*(7), pp. e1001906.

Jamal-Hanjani, M., Wilson, G. A., Horswell, S., Mitter, R., Sakarya, O., Constantin, T., Salari, R., Kirkizlar, E., Sigurjonsson, S., Pelham, R., Kareht, S., Zimmermann, B. and Swanton, C. (2016) Detection of ubiquitous and heterogeneous mutations in cell-free DNA from patients with early-stage non-small-cell lung cancer. *Ann Oncol, 27*(5), pp. 862-7.

Jamal-Hanjani, M., Wilson, G. A., McGranahan, N., Birkbak, N. J., Watkins, T. B. K., Veeriah, S., Shafi, S., Johnson, D. H., Mitter, R., Rosenthal, R., Salm, M., Horswell, S., Escudero, M., Matthews, N., Rowan, A., Chambers, T., Moore, D. A., Turajlic, S., Xu, H., Lee, S. M., Forster, M. D., Ahmad, T., Hiley, C. T., Abbosh, C., Falzon, M., Borg, E., Marafioti, T., Lawrence, D., Hayward, M., Kolvekar, S., Panagiotopoulos, N., Janes, S. M., Thakrar, R., Ahmed, A., Blackhall, F., Summers, Y., Shah, R., Joseph, L., Quinn, A. M., Crosbie, P. A., Naidu, B., Middleton, G., Langman, G., Trotter, S., Nicolson, M., Remmen, H., Kerr, K., Chetty, M., Gomersall, L., Fennell, D. A., Nakas, A., Rathinam, S., Anand, G., Khan, S., Russell, P., Ezhil, V., Ismail, B., Irvin-Sellers, M., Prakash, V., Lester, J. F., Kornaszewska, M., Attanoos, R., Adams, H., Davies, H., Dentro, S., Taniere, P., O'Sullivan, B., Lowe, H. L., Hartley, J. A., Iles, N., Bell, H., Ngai, Y., Shaw, J. A., Herrero, J., Szallasi, Z., Schwarz, R. F., Stewart, A., Quezada, S. A., Le Quesne, J., Van Loo, P., Dive, C., Hackshaw, A. and Swanton, C. (2017) Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl J Med, 376*(22), pp. 2109-2121.

Janes, M. R., Zhang, J., Li, L. S., Hansen, R., Peters, U., Guo, X., Chen, Y., Babbar, A., Firdaus, S. J., Darjania, L., Feng, J., Chen, J. H., Li, S., Li, S., Long, Y. O., Thach, C., Liu, Y., Zarieh, A., Ely, T., Kucharski, J. M., Kessler, L. V., Wu, T., Yu, K., Wang, Y., Yao, Y., Deng, X., Zarrinkar, P. P., Brehmer, D., Dhanak, D., Lorenzi, M. V., Hu-Lowe, D., Patricelli, M. P., Ren, P. and Liu, Y. (2018) Targeting KRAS Mutant Cancers with a Covalent G12C-Specific Inhibitor. *Cell, 172*(3), pp. 578-589.e17.

Janne, P. A., Yang, J. C., Kim, D. W., Planchard, D., Ohe, Y., Ramalingam, S. S., Ahn, M. J., Kim, S. W., Su, W. C., Horn, L., Haggstrom, D., Felip, E., Kim, J. H., Frewer, P., Cantarini, M., Brown, K. H., Dickinson, P. A., Ghiorghiu, S. and Ranson, M. (2015) AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. *N Engl J Med, 372*(18), pp. 1689-99.

Jin, G., Kim, M. J., Jeon, H. S., Choi, J. E., Kim, D. S., Lee, E. B., Cha, S. I., Yoon, G. S., Kim, C. H., Jung, T. H. and Park, J. Y. (2010) PTEN mutations and relationship to EGFR, ERBB2, KRAS, and TP53 mutations in non-small cell lung cancers. *Lung Cancer, 69*(3), pp. 279-83.

Johnson, J. L., Pillai, S. and Chellappan, S. P. (2012) Genetic and biochemical alterations in non-small cell lung cancer. *Biochem Res Int, 2012*, pp. 940405.

Jung, K., Fleischhacker, M. and Rabien, A. (2010) Cell-free DNA in the blood as a solid tumor biomarker--a critical appraisal of the literature. *Clin Chim Acta, 411*(21-22), pp. 1611-24.

Kadota, K., Suzuki, K., Sima, C. S., Rusch, V. W., Adusumilli, P. S. and Travis, W. D. (2011) Pleomorphic epithelioid diffuse malignant pleural mesothelioma: a clinicopathological review and conceptual proposal to reclassify as biphasic or sarcomatoid mesothelioma. *J Thorac Oncol, 6*(5), pp. 896-904.

Karlovich, C., Goldman, J. W., Sun, J. M., Mann, E., Sequist, L. V., Konopa, K., Wen, W., Angenendt, P., Horn, L., Spigel, D., Soria, J. C., Solomon, B., Camidge, D. R., Gadgeel, S., Paweletz, C., Wu, L., Chien, S., O'Donnell, P., Matheny, S., Despain, D., Rolfe, L., Raponi, M., Allen, A. R., Park, K. and Wakelee, H. (2016) Assessment of EGFR Mutation Status in Matched Plasma and Tumor Tissue of NSCLC Patients from a Phase I Study of Rociletinib (CO-1686). *Clin Cancer Res, 22*(10), pp. 2386-95.

Kastenhuber, E. R. and Lowe, S. W. (2017) Putting p53 in Context. *Cell, 170*(6), pp. 1062-1078.

Kaufman, A. J. and Flores, R. M. (2011) Surgical treatment of malignant pleural mesothelioma. *Curr Treat Options Oncol, 12*(2), pp. 201-16.

Kidess, E. and Jeffrey, S. S. (2013) Circulating tumor cells versus tumor-derived cell-free DNA: rivals or partners in cancer care in the era of single-cell analysis? *Genome Med, 5*(8), pp. 70.

Kohno, T., Nakaoku, T., Tsuta, K., Tsuchihara, K., Matsumoto, S., Yoh, K. and Goto, K. (2015) Beyond ALK-RET, ROS1 and other oncogene fusions in lung cancer. *Transl Lung Cancer Res, 4*(2), pp. 156-64.

Krebs, M. G., Metcalf, R. L., Carter, L., Brady, G., Blackhall, F. H. and Dive, C. (2014) Molecular analysis of circulating tumour cells-biology and biomarkers. *Nat Rev Clin Oncol, 11*(3), pp. 129-44.

Ladanyi, M. and Pao, W. (2008) Lung adenocarcinoma: guiding EGFR-targeted therapy and beyond. *Mod Pathol, 21 Suppl 2*, pp. S16-22.

Ladanyi, M., Zauderer, M. G., Krug, L. M., Ito, T., McMillan, R., Bott, M. and Giancotti, F. (2012) New strategies in pleural mesothelioma: BAP1 and NF2 as novel targets for therapeutic development and risk assessment. *Clin Cancer Res, 18*(17), pp. 4485-90.

Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., Carter, S. L., Stewart, C., Mermel, C. H., Roberts, S. A., Kiezun, A., Hammerman, P. S., McKenna, A., Drier, Y., Zou, L., Ramos, A. H., Pugh, T. J., Stransky, N., Helman, E., Kim, J., Sougnez, C., Ambrogio, L., Nickerson, E., Shefler, E., Cortes, M. L., Auclair, D., Saksena, G., Voet, D., Noble, M., DiCara, D., Lin, P., Lichtenstein, L., Heiman, D. I., Fennell, T., Imielinski, M., Hernandez, B., Hodis, E., Baca, S., Dulak, A. M., Lohr, J., Landau, D. A., Wu, C. J., Melendez-Zajgla, J., Hidalgo-Miranda, A., Koren, A., McCarroll, S. A., Mora, J., Crompton, B., Onofrio, R., Parkin, M., Winckler, W., Ardlie, K., Gabriel, S. B., Roberts, C. W. M., Biegel, J. A., Stegmaier, K., Bass, A. J., Garraway, L. A., Meyerson, M., Golub, T. R., Gordenin, D. A., Sunyaev, S., Lander, E. S. and Getz, G. (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature, 499*(7457), pp. 214-218.

Lee, S. Y., Kim, M. J., Jin, G., Yoo, S. S., Park, J. Y., Choi, J. E., Jeon, H. S., Cho, S., Lee, E. B., Cha, S. I., Park, T. I., Kim, C. H., Jung, T. H. and Park, J. Y. (2010) Somatic mutations in epidermal growth factor receptor signaling pathway genes in non-small cell lung cancers. *J Thorac Oncol, 5*(11), pp. 1734-40.

Leon, S. A., Shapiro, B., Sklaroff, D. M. and Yaros, M. J. (1977) Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res, 37*(3), pp. 646-50.

Li, W., You, L., Cooper, J., Schiavon, G., Pepe-Caprio, A., Zhou, L., Ishii, R., Giovannini, M., Hanemann, C. O., Long, S. B., Erdjument-Bromage, H., Zhou, P., Tempst, P. and Giancotti, F. G. (2010) Merlin/NF2 suppresses tumorigenesis by

inhibiting the E3 ubiquitin ligase CRL4(DCAF1) in the nucleus. *Cell, 140*(4), pp. 477-90.

Liang, W., Zhao, Y., Huang, W., Liang, H., Zeng, H. and He, J. (2018) Liquid biopsy for early stage lung cancer. *J Thorac Dis, 10*(Suppl 7), pp. S876-s881.

Lindberg, K., Nyman, J., Riesenfeld Kallskog, V., Hoyer, M., Lund, J. A., Lax, I., Wersall, P., Karlsson, K., Friesland, S. and Lewensohn, R. (2015) Long-term results of a prospective phase II trial of medically inoperable stage I NSCLC treated with SBRT - the Nordic experience. *Acta Oncol, 54*(8), pp. 1096-104.

Liu, W., Ernst, J. D. and Broaddus, V. C. (2000) Phagocytosis of crocidolite asbestos induces oxidative stress, DNA damage, and apoptosis in mesothelial cells. *Am J Respir Cell Mol Biol, 23*(3), pp. 371-8.

Mandel, P. and Metais, P. (1948) Les acides nucléiques du plasma sanguin chez l'homme. *C R Seances Soc Biol Fil, 142*(3-4), pp. 241-3.

Marino, P., Touzani, R., Perrier, L., Rouleau, E., Kossi, D. S., Zhaomin, Z., Charrier, N., Goardon, N., Preudhomme, C., Durand-Zaleski, I., Borget, I. and Baffert, S. (2018) Cost of cancer diagnosis using next-generation sequencing targeted gene panels in routine practice: a nationwide French study. *Eur J Hum Genet, 26*(3), pp. 314-323.

Marusyk, A., Almendro, V. and Polyak, K. (2012) Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer, 12*(5), pp. 323-34.

Mauguen, A., Seshan, V. E., Ostrovnaya, I. and Begg, C. B. (2018) Estimating the probability of clonal relatedness of pairs of tumors in cancer patients. *Biometrics, 74*(1), pp. 321-330.

McGranahan, N. and Swanton, C. (2017) Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell, 168*(4), pp. 613-628.

Meyerhoff, R. R., Yang, C. F., Speicher, P. J., Gulack, B. C., Hartwig, M. G., D'Amico, T. A., Harpole, D. H. and Berry, M. F. (2015) Impact of mesothelioma histologic subtype on outcomes in the Surveillance, Epidemiology, and End Results database. *J Surg Res, 196*(1), pp. 23-32.

Mirsadraee, S., Oswal, D., Alizadeh, Y., Caulo, A. and van Beek, E., Jr. (2012) The 7th lung cancer TNM classification and staging system: Review of the changes and implications. *World J Radiol, 4*(4), pp. 128-34.

Mok, T. S., Wu, Y. L., Thongprasert, S., Yang, C. H., Chu, D. T., Saijo, N., Sunpaweravong, P., Han, B., Margono, B., Ichinose, Y., Nishiwaki, Y., Ohe, Y., Yang, J. J., Chewaskulyong, B., Jiang, H., Duffield, E. L., Watkins, C. L., Armour, A. A. and Fukuoka, M. (2009) Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med, 361*(10), pp. 947-57.

Mouliere, F., Piskorz, A. M., Chandrananda, D., Moore, E., Morris, J., Smith, C. G., Goranova, T., Heider, K., Mair, R., Supernat, A., Gounaris, I., Ros, S., Wan, J. C. M., Jimenez-Linan, M., Gale, D., Brindle, K., Massie, C. E., Parkinson, C. A., Brenton, J. D. and Rosenfeld, N. (2017) Selecting Short DNA Fragments In Plasma Improves Detection Of Circulating Tumour DNA. *bioRxiv*.

Mouliere, F. and Rosenfeld, N. (2015) Circulating tumor-derived DNA is shorter than somatic DNA in plasma. *Proc Natl Acad Sci U S A, 112*(11), pp. 3178-9.

Murakami, H., Mizuno, T., Taniguchi, T., Fujii, M., Ishiguro, F., Fukui, T., Akatsuka, S., Horio, Y., Hida, T., Kondo, Y., Toyokuni, S., Osada, H. and Sekido, Y. (2011) LATS2 is a tumor suppressor gene of malignant mesothelioma. *Cancer Res, 71*(3), pp. 873-83.

Murtaza, M., Dawson, S. J., Tsui, D. W., Gale, D., Forshew, T., Piskorz, A. M., Parkinson, C., Chin, S. F., Kingsbury, Z., Wong, A. S., Marass, F., Humphray, S., Hadfield, J., Bentley, D., Chin, T. M., Brenton, J. D., Caldas, C. and Rosenfeld, N. (2013) Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature, 497*(7447), pp. 108-12.

Nass, D., Rosenwald, S., Meiri, E., Gilad, S., Tabibian-Keissar, H., Schlosberg, A., Kuker, H., Sion-Vardy, N., Tobar, A., Kharenko, O., Sitbon, E., Lithwick Yanai, G., Elyakim, E., Cholakh, H., Gibori, H., Spector, Y., Bentwich, Z., Barshack, I. and Rosenfeld, N. (2009) MiR-92b and miR-9/9* are specifically expressed in brain primary tumors and can be used to differentiate primary from metastatic brain tumors. *Brain Pathol, 19*(3), pp. 375-83.

Nesbitt, J. C., Putnam, J. B., Jr., Walsh, G. L., Roth, J. A. and Mountain, C. F. (1995) Survival in early-stage non-small cell lung cancer. *Ann Thorac Surg, 60*(2), pp. 466-72.

Newman, A. M., Bratman, S. V., To, J., Wynne, J. F., Eclov, N. C., Modlin, L. A., Liu, C. L., Neal, J. W., Wakelee, H. A., Merritt, R. E., Shrager, J. B. and Loo, B. W., Jr. (2014) An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *20*(5), pp. 548-54.

Nowak, A. K. (2012) Chemotherapy for malignant pleural mesothelioma: a review of current management and a look to the future. *Ann Cardiothorac Surg, 1*(4), pp. 508-15.

Nowak, A. K., Chansky, K., Rice, D. C., Pass, H. I., Kindler, H. L., Shemanski, L., Bille, A., Rintoul, R. C., Batirel, H. F., Thomas, C. F., Friedberg, J., Cedres, S., de Perrot, M. and Rusch, V. W. (2016) The IASLC Mesothelioma Staging Project: Proposals for Revisions of the T Descriptors in the Forthcoming Eighth Edition of the TNM Classification for Pleural Mesothelioma. *J Thorac Oncol, 11*(12), pp. 2089-2099.

Nowell, P. C. (1976) The clonal evolution of tumor cell populations. *Science, 194*(4260), pp. 23-8.

Nygaard, A. D., Garm Spindler, K. L., Pallisgaard, N., Andersen, R. F. and Jakobsen, A. (2013) The prognostic value of KRAS mutated plasma DNA in advanced non-small cell lung cancer. *Lung Cancer, 79*(3), pp. 312-7.

O'Dowd, E. L. and Baldwin, D. R. (2018) Lung cancer screening-low dose CT for lung cancer screening: recent trial results and next steps. *Br J Radiol, 91*(1090), pp. 20170460.

Ordonez, N. G. (2012) Pleomorphic mesothelioma: report of 10 cases. *Mod Pathol, 25*(7), pp. 1011-22.

Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A. and Shokat, K. M. (2013) K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature, 503*(7477), pp. 548-51.

Oxnard, G. R., Binder, A. and Janne, P. A. (2013) New targetable oncogenes in non-small-cell lung cancer. *J Clin Oncol, 31*(8), pp. 1097-104.

Page, K., Guttery, D. S., Fernandez-Garcia, D., Hills, A., Hastings, R. K., Luo, J., Goddard, K., Shahin, V., Woodley-Barker, L., Rosales, B. M., Coombes, R. C., Stebbing, J. and Shaw, J. A. (2017) Next Generation Sequencing of Circulating

Cell-Free DNA for Evaluating Mutations and Gene Amplification in Metastatic Breast Cancer. *Clin Chem, 63*(2), pp. 532-541.

Pereira, T. C., Share, S. M., Magalhaes, A. V. and Silverman, J. F. (2011) Can we tell the site of origin of metastatic squamous cell carcinoma? An immunohistochemical tissue microarray study of 194 cases. *Appl Immunohistochem Mol Morphol, 19*(1), pp. 10-4.

Perez-Moreno, P., Brambilla, E., Thomas, R. and Soria, J. C. (2012) Squamous cell carcinoma of the lung: molecular subtypes and therapeutic opportunities. *Clin Cancer Res, 18*(9), pp. 2443-51.

Peters, S., Camidge, D. R., Shaw, A. T., Gadgeel, S., Ahn, J. S., Kim, D. W., Ou, S. I., Perol, M., Dziadziuszko, R., Rosell, R., Zeaiter, A., Mitry, E., Golding, S., Balas, B., Noe, J., Morcos, P. N. and Mok, T. (2017) Alectinib versus Crizotinib in Untreated ALK-Positive Non-Small-Cell Lung Cancer. *N Engl J Med, 377*(9), pp. 829-838.

Petrilli, A. M. and Fernandez-Valle, C. (2015) Role of Merlin/NF2 inactivation in tumor biology. *Oncogene*.

Provencio, M., Torrente, M., Calvo, V., Perez-Callejo, D., Gutierrez, L., Franco, F., Perez-Barrios, C., Barquin, M., Royuela, A., Garcia-Garcia, F., Bueno, C., Garcia-Grande, A., Camps, C., Massuti, B., Sotomayor, E. and Romero, A. (2018) Prognostic value of quantitative ctDNA levels in non small cell lung cancer patients. *Oncotarget, 9*(1), pp. 488-494.

Redig, A. J., Capelletti, M., Dahlberg, S. E., Sholl, L. M., Mach, S., Fontes, C., Shi, Y., Chalasani, P. and Janne, P. A. (2016) Clinical and Molecular Characteristics of NF1-Mutant Lung Cancer. *Clin Cancer Res, 22*(13), pp. 3148-56.

Renaud, S., Schaeffer, M., Falcoz, P. E., Seitlinger, J., Romain, B., Voegeli, A. C., Legrain, M., Reeb, J., Santelmo, N., Rohr, S., Brigand, C., Olland, A., Guenot, D. and Massard, G. (2017) Perioperative bevacizumab improves survival following lung metastasectomy for colorectal cancer in patients harbouring v-Ki-ras2 Kirsten rat sarcoma viral oncogene homologue exon 2 codon 12 mutationsdagger. *Eur J Cardiothorac Surg, 51*(2), pp. 255-262.

Ricciardi, S., Cardillo, G., Zirafa, C. C., Carleo, F., Facciolo, F., Fontanini, G., Mutti, L. and Melfi, F. (2018) Surgery for malignant pleural mesothelioma: an international guidelines review. *J Thorac Dis, 10*(Suppl 2), pp. S285-s292.

Rice, D. (2012) Standardizing surgical treatment in malignant pleural mesothelioma. *Ann Cardiothorac Surg, 1*(4), pp. 497-501.

Rios Velazquez, E., Parmar, C., Liu, Y., Coroller, T. P., Cruz, G., Stringfield, O., Ye, Z., Makrigiorgos, M., Fennessy, F., Mak, R. H., Gillies, R., Quackenbush, J. and Aerts, H. (2017) Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. *Cancer Res, 77*(14), pp. 3922-3930.

Robinson, B. M. (2012) Malignant pleural mesothelioma: an epidemiological perspective. *Ann Cardiothorac Surg, 1*(4), pp. 491-6.

Robinson, B. W., Creaney, J., Lake, R., Nowak, A., Musk, A. W., de Klerk, N., Winzell, P., Hellstrom, K. E. and Hellstrom, I. (2005) Soluble mesothelin-related protein--a blood test for mesothelioma. *Lung Cancer, 49 Suppl 1*, pp. S109-11.

Rosell, R., Carcereny, E., Gervais, R., Vergnenegre, A., Massuti, B., Felip, E., Palmero, R., Garcia-Gomez, R., Pallares, C., Sanchez, J. M., Porta, R., Cobo, M., Garrido, P., Longo, F., Moran, T., Insa, A., De Marinis, F., Corre, R., Bover, I., Illiano, A., Dansin, E., de Castro, J., Milella, M., Reguart, N., Altavilla, G., Jimenez, U., Provencio, M., Moreno, M. A., Terrasa, J., Munoz-Langa, J., Valdivia, J., Isla, D., Domine, M., Molinier, O., Mazieres, J., Baize, N., Garcia-Campelo, R., Robinet, G., Rodriguez-Abreu, D., Lopez-Vivanco, G., Gebbia, V., Ferrera-Delgado, L., Bombaron, P., Bernabe, R., Bearz, A., Artal, A., Cortesi, E., Rolfo, C., Sanchez-Ronco, M., Drozdowskyj, A., Queralt, C., de Aguirre, I., Ramirez, J. L., Sanchez, J. J., Molina, M. A., Taron, M. and Paz-Ares, L. (2012) Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol, 13*(3), pp. 239-46.

Rossi, G., Murer, B., Cavazza, A., Losi, L., Natali, P., Marchioni, A., Migaldi, M., Capitanio, G. and Brambilla, E. (2004) Primary mucinous (so-called colloid) carcinomas of the lung: a clinicopathologic and immunohistochemical study with special reference to CDX-2 homeobox gene and MUC2 expression. *Am J Surg Pathol, 28*(4), pp. 442-52.

Rothe, F., Laes, J. F., Lambrechts, D., Smeets, D., Vincent, D., Maetens, M., Fumagalli, D., Michiels, S., Drisis, S., Moerman, C., Detiffe, J. P., Larsimont, D., Awada, A., Piccart, M., Sotiriou, C. and Ignatiadis, M. (2014) Plasma circulating tumor DNA as an alternative to metastatic biopsies for mutational analysis in breast cancer. *Ann Oncol, 25*(10), pp. 1959-65.

Rusch, V. W., Chansky, K., Kindler, H. L., Nowak, A. K., Pass, H. I., Rice, D. C., Shemanski, L., Galateau-Salle, F., McCaughan, B. C., Nakano, T., Ruffini, E., van Meerbeeck, J. P. and Yoshimura, M. (2016) The IASLC Mesothelioma Staging Project: Proposals for the M Descriptors and for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Mesothelioma. *J Thorac Oncol, 11*(12), pp. 2112-2119.

Russell, P. A., Wainer, Z., Wright, G. M., Daniels, M., Conron, M. and Williams, R. A. (2011) Does lung adenocarcinoma subtype predict patient survival?: A clinicopathologic study based on the new International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary lung adenocarcinoma classification. *J Thorac Oncol, 6*(9), pp. 1496-504.

Saito, T., Niida, A., Uchi, R., Hirata, H., Komatsu, H., Sakimura, S. and Hayashi, S. (2018) A temporal shift of the evolutionary principle shaping intratumor heterogeneity in colorectal cancer. *9*(1), pp. 2884.

Sato, T. and Sekido, Y. (2018) NF2/Merlin Inactivation and Potential Therapeutic Targets in Mesothelioma. *Int J Mol Sci, 19*(4).

Schmitt, M. W., Kennedy, S. R., Salk, J. J., Fox, E. J., Hiatt, J. B. and Loeb, L. A. (2012) Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci U S A, 109*(36), pp. 14508-13.

Schreuer, M., Meersseman, G., Van Den Herrewegen, S., Jansen, Y., Chevolet, I., Bott, A., Wilgenhof, S., Seremet, T., Jacobs, B., Buyl, R., Maertens, G. and Neyns, B. (2016) Quantitative assessment of BRAF V600 mutant circulating cell-free tumor DNA as a tool for therapeutic monitoring in metastatic melanoma patients treated with BRAF/MEK inhibitors. *J Transl Med, 14*, pp. 95.

Schuz, J., Schonfeld, S. J., Kromhout, H., Straif, K., Kashanskiy, S. V., Kovalevskiy, E. V., Bukhtiyarov, I. V. and McCormack, V. (2013) A retrospective cohort study of cancer mortality in employees of a Russian chrysotile asbestos mine and mills: study rationale and key features. *Cancer Epidemiol, 37*(4), pp. 440-5.

Schwarze, K., Buchanan, J., Taylor, J. C. and Wordsworth, S. (2018) Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. *Genet Med*.

Scott, W. J., Howington, J., Feigenberg, S., Movsas, B. and Pisters, K. (2007) Treatment of non-small cell lung cancer stage I and stage II: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest, 132*(3 Suppl), pp. 234s-242s.

Sekido, Y. (2013) Molecular pathogenesis of malignant mesothelioma. *Carcinogenesis, 34*(7), pp. 1413-9.

Servais, E. L., Colovos, C., Rodriguez, L., Bograd, A. J., Nitadori, J., Sima, C., Rusch, V. W., Sadelain, M. and Adusumilli, P. S. (2012) Mesothelin overexpression promotes mesothelioma cell invasion and MMP-9 secretion in an orthotopic mouse model and in epithelioid pleural mesothelioma patients. *Clin Cancer Res, 18*(9), pp. 2478-89.

Shapiro, I. M., Kolev, V. N., Vidal, C. M., Kadariya, Y., Ring, J. E., Wright, Q., Weaver, D. T., Menges, C., Padval, M., McClatchey, A. I., Xu, Q., Testa, J. R. and Pachter, J. A. (2014) Merlin deficiency predicts FAK inhibitor sensitivity: a synthetic lethal relationship. *Sci Transl Med, 6*(237), pp. 237ra68.

Sharkey, A. J., Bilancia, R., Tenconi, S., Nakas, A. and Waller, D. A. (2017) Extended pleurectomy decortication for malignant pleural mesothelioma in the elderly: the need for an inclusive yet selective approach. *Interact Cardiovasc Thorac Surg, 25*(5), pp. 696-702.

Shaw, J. A., Guttery, D. S., Hills, A., Fernandez-Garcia, D., Page, K., Rosales, B. M., Goddard, K. S., Hastings, R. K., Luo, J., Ogle, O., Woodley, L., Ali, S., Stebbing, J. and Coombes, R. C. (2017) Mutation Analysis of Cell-Free DNA and Single Circulating Tumor Cells in Metastatic Breast Cancer Patients with High Circulating Tumor Cell Counts. *Clin Cancer Res, 23*(1), pp. 88-96.

Shaw, J. A., Page, K., Blighe, K., Hava, N., Guttery, D., Ward, B., Brown, J., Ruangpratheep, C., Stebbing, J., Payne, R., Palmieri, C., Cleator, S., Walker, R. A. and Coombes, R. C. (2012) Genomic analysis of circulating cell-free DNA infers breast cancer dormancy. *Genome Res, 22*(2), pp. 220-31.

Shi, H., Hugo, W., Kong, X., Hong, A., Koya, R. C., Moriceau, G., Chodon, T., Guo, R., Johnson, D. B., Dahlman, K. B., Kelley, M. C., Kefford, R. F., Chmielowski, B., Glaspy, J. A., Sosman, J. A., van Baren, N., Long, G. V., Ribas, A. and Lo, R. S. (2014) Acquired resistance and clonal evolution in melanoma during BRAF inhibitor therapy. *Cancer Discov, 4*(1), pp. 80-93.

Shigematsu, H., Lin, L., Takahashi, T., Nomura, M., Suzuki, M., Wistuba, II, Fong, K. M., Lee, H., Toyooka, S., Shimizu, N., Fujisawa, T., Feng, Z., Roth, J. A., Herz, J., Minna, J. D. and Gazdar, A. F. (2005) Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst, 97*(5), pp. 339-46.

Sholl, L. M., Yeap, B. Y., Iafrate, A. J., Holmes-Tisch, A. J., Chou, Y. P., Wu, M. T., Goan, Y. G., Su, L., Benedettini, E., Yu, J., Loda, M., Janne, P. A., Christiani, D. C. and Chirieac, L. R. (2009) Lung adenocarcinoma with EGFR amplification has distinct clinicopathologic and molecular features in never-smokers. *Cancer Res, 69*(21), pp. 8341-8.

Silverman, J. F., Zhu, B., Liu, Y. and Lin, X. (2009) Distinctive immunohistochemical profile of mucinous cystic neoplasms of pancreas, ovary and lung. *Histol Histopathol, 24*(1), pp. 77-82.

Siravegna, G., Mussolin, B., Buscarino, M. and Corti, G. (2015) Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *21*(7), pp. 795-801.

Smittenaar, C. R., Petersen, K. A., Stewart, K. and Moitt, N. (2016) Cancer incidence and mortality projections in the UK until 2035. *Br J Cancer, 115*(9), pp. 1147-1155.

Sneddon, S., Dick, I., Lee, Y. C. G., Musk, A. W. B., Patch, A. M., Pearson, J. V., Waddell, N., Allcock, R. J. N., Holt, R. A., Robinson, B. W. S. and Creaney, J. (2018) Malignant cells from pleural fluids in malignant mesothelioma patients reveal novel mutations. *Lung Cancer, 119*, pp. 64-70.

Soda, M., Choi, Y. L., Enomoto, M., Takada, S., Yamashita, Y., Ishikawa, S., Fujiwara, S., Watanabe, H., Kurashina, K., Hatanaka, H., Bando, M., Ohno, S., Ishikawa, Y., Aburatani, H., Niki, T., Sohara, Y., Sugiyama, Y. and Mano, H. (2007) Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature, 448*(7153), pp. 561-6.

Sorenson, G. D., Pribish, D. M., Valone, F. H., Memoli, V. A., Bzik, D. J. and Yao, S. L. (1994) Soluble normal and mutated DNA sequences from single-copy genes in human blood. *Cancer Epidemiol Biomarkers Prev, 3*(1), pp. 67-71.

Sottoriva, A., Kang, H., Ma, Z. and Graham, T. A. (2015) A Big Bang model of human colorectal tumor growth. *47*(3), pp. 209-16.

Stahel, R. A., Weder, W., Lievens, Y. and Felip, E. (2010) Malignant pleural mesothelioma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol, 21 Suppl 5*, pp. v126-8.

Stark, Z., Schofield, D., Alam, K., Wilson, W., Mupfeki, N. and Macciocca, I. (2017) Prospective comparison of the cost-effectiveness of clinical whole-exome sequencing with that of usual care overwhelmingly supports early use and reimbursement. *19*(8), pp. 867-874.

Sugarbaker, D. J., Strauss, G. M., Lynch, T. J., Richards, W., Mentzer, S. J., Lee, T. H., Corson, J. M. and Antman, K. H. (1993) Node status has prognostic significance in the multimodality therapy of diffuse, malignant mesothelioma. *J Clin Oncol, 11*(6), pp. 1172-8.

Sulzmaier, F. J., Jean, C. and Schlaepfer, D. D. (2014) FAK in cancer: mechanistic findings and clinical applications. *Nat Rev Cancer, 14*(9), pp. 598-610.

Sun, S., Schiller, J. H. and Gazdar, A. F. (2007) Lung cancer in never smokers--a different disease. *Nat Rev Cancer, 7*(10), pp. 778-90.

Takezawa, K., Pirazzoli, V., Arcila, M. E., Nebhan, C. A., Song, X., de Stanchina, E., Ohashi, K., Janjigian, Y. Y., Spitzler, P. J., Melnick, M. A., Riely, G. J., Kris, M. G., Miller, V. A., Ladanyi, M., Politi, K. and Pao, W. (2012) HER2 amplification: a potential mechanism of acquired resistance to EGFR inhibition in EGFR-mutant lung cancers that lack the second-site EGFRT790M mutation. *Cancer Discov, 2*(10), pp. 922-33.

Tallet, A., Nault, J. C., Renier, A., Hysi, I., Galateau-Salle, F., Cazes, A., Copin, M. C., Hofman, P., Andujar, P., Le Pimpec-Barthes, F., Zucman-Rossi, J., Jaurand, M. C. and Jean, D. (2014) Overexpression and promoter mutation of the TERT gene in malignant pleural mesothelioma. *Oncogene, 33*(28), pp. 3748-52.

Tamborero, D., Rubio-Perez, C., Deu-Pons, J., Schroeder, M. P., Vivancos, A., Rovira, A., Tusquets, I., Albanell, J., Rodon, J., Tabernero, J., de Torres, C., Dienstmann, R., Gonzalez-Perez, A. and Lopez-Bigas, N. (2018) Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med, 10*(1), pp. 25.

Tammemagi, M. C., McLaughlin, J. R. and Bull, S. B. (1999) Meta-analyses of p53 tumor suppressor gene alterations and clinicopathological features in resected lung cancers. *Cancer Epidemiol Biomarkers Prev, 8*(7), pp. 625-34.

Taniguchi, T., Karnan, S., Fukui, T., Yokoyama, T., Tagawa, H., Yokoi, K., Ueda, Y., Mitsudomi, T., Horio, Y., Hida, T., Yatabe, Y., Seto, M. and Sekido, Y. (2007) Genomic profiling of malignant pleural mesothelioma with array-based comparative genomic hybridization shows frequent non-random chromosomal alteration regions including JUN amplification on 1p32. *Cancer Sci, 98*(3), pp. 438-46.

Testa, J. R., Cheung, M., Pei, J., Below, J. E., Tan, Y., Sementino, E., Cox, N. J., Dogan, A. U., Pass, H. I., Trusa, S., Hesdorffer, M., Nasu, M., Powers, A., Rivera, Z., Comertpay, S., Tanji, M., Gaudino, G., Yang, H. and Carbone, M. (2011) Germline BAP1 mutations predispose to malignant mesothelioma. *Nat Genet, 43*(10), pp. 1022-5.

Thress, K. S., Paweletz, C. P., Felip, E., Cho, B. C., Stetson, D., Dougherty, B., Lai, Z., Markovets, A., Vivancos, A., Kuang, Y., Ercan, D., Matthews, S. E., Cantarini, M., Barrett, J. C., Janne, P. A. and Oxnard, G. R. (2015) Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. *Nat Med, 21*(6), pp. 560-2.

Thunnissen, E., Bubendorf, L., Dietel, M., Elmberger, G., Kerr, K., Lopez-Rios, F., Moch, H., Olszewski, W., Pauwels, P., Penault-Llorca, F. and Rossi, G. (2012) EML4-ALK testing in non-small cell carcinomas of the lung: a review with recommendations. *Virchows Arch, 461*(3), pp. 245-57.

Thurneysen, C., Opitz, I., Kurtz, S., Weder, W., Stahel, R. A. and Felley-Bosco, E. (2009) Functional inactivation of NF2/merlin in human mesothelioma. *Lung Cancer, 64*(2), pp. 140-7.

Tie, J., Wang, Y., Tomasetti, C., Li, L., Springer, S., Kinde, I., Silliman, N., Tacey, M., Wong, H. L., Christie, M., Kosmider, S., Skinner, I., Wong, R., Steel, M., Tran, B., Desai, J., Jones, I., Haydon, A., Hayes, T., Price, T. J., Strausberg, R. L., Diaz, L. A., Jr., Papadopoulos, N., Kinzler, K. W., Vogelstein, B. and Gibbs, P. (2016) Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci Transl Med, 8*(346), pp. 346ra92.

Tokumo, M., Toyooka, S., Kiura, K., Shigematsu, H., Tomii, K., Aoe, M., Ichimura, K., Tsuda, T., Yano, M., Tsukuda, K., Tabata, M., Ueoka, H., Tanimoto, M., Date, H., Gazdar, A. F. and Shimizu, N. (2005) The relationship between epidermal

growth factor receptor mutations and clinicopathologic features in non-small cell lung cancers. *Clin Cancer Res, 11*(3), pp. 1167-73.

Travis, W. D., Brambilla, E., Burke, A. P., Marx, A. and Nicholson, A. G. (2015) Introduction to The 2015 World Health Organization Classification of Tumors of the Lung, Pleura, Thymus, and Heart. *J Thorac Oncol, 10*(9), pp. 1240-1242.

Travis, W. D., Brambilla, E., Noguchi, M., Nicholson, A. G., Geisinger, K., Yatabe, Y., Ishikawa, Y., Wistuba, I., Flieder, D. B., Franklin, W., Gazdar, A., Hasleton, P. S., Henderson, D. W., Kerr, K. M., Nakatani, Y., Petersen, I., Roggli, V., Thunnissen, E. and Tsao, M. (2013a) Diagnosis of lung adenocarcinoma in resected specimens: implications of the 2011 International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society classification. *Arch Pathol Lab Med, 137*(5), pp. 685-705.

Travis, W. D., Brambilla, E. and Riely, G. J. (2013b) New pathologic classification of lung cancer: relevance for clinical practice and clinical trials. *J Clin Oncol, 31*(8), pp. 992-1001.

Treasure, T., Lang-Lazdunski, L., Waller, D., Bliss, J. M., Tan, C., Entwisle, J., Snee, M., O'Brien, M., Thomas, G., Senan, S., O'Byrne, K., Kilburn, L. S., Spicer, J., Landau, D., Edwards, J., Coombes, G., Darlison, L. and Peto, J. (2011) Extra-pleural pneumonectomy versus no extra-pleural pneumonectomy for patients with malignant pleural mesothelioma: clinical outcomes of the Mesothelioma and Radical Surgery (MARS) randomised feasibility study. *Lancet Oncol, 12*(8), pp. 763-72.

Turnbull, C., Scott, R. H., Thomas, E., Jones, L., Murugaesu, N., Pretty, F. B., Halai, D., Baple, E., Craig, C., Hamblin, A., Henderson, S., Patch, C., O'Neill, A., Devereaux, A., Smith, K., Martin, A. R., Sosinsky, A., McDonagh, E. M., Sultana, R., Mueller, M., Smedley, D., Toms, A., Dinh, L., Fowler, T., Bale, M., Hubbard, T., Rendon, A., Hill, S. and Caulfield, M. J. (2018) The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *Bmj, 361*, pp. k1687.

Ulz, P., Belic, J., Graf, R., Auer, M., Lafer, I., Fischereder, K., Webersinke, G., Pummer, K., Augustin, H., Pichler, M., Hoefler, G., Bauernhofer, T., Geigl, J. B., Heitzer, E. and Speicher, M. R. (2016) Whole-genome plasma sequencing reveals focal amplifications as a driving force in metastatic prostate cancer. *7*, pp. 12008.

UyBico, S. J., Wu, C. C., Suh, R. D., Le, N. H., Brown, K. and Krishnam, M. S. (2010) Lung cancer staging essentials: the new TNM staging system and potential imaging pitfalls. *Radiographics, 30*(5), pp. 1163-81.

van Zandwijk, N., Clarke, C., Henderson, D., Musk, A. W., Fong, K., Nowak, A., Loneragan, R., McCaughan, B., Boyer, M., Feigen, M., Currow, D., Schofield, P., Nick Pavlakis, B. I., McLean, J., Marshall, H., Leong, S., Keena, V. and Penman, A. (2013) Guidelines for the diagnosis and treatment of malignant pleural mesothelioma. *J Thorac Dis, 5*(6), pp. E254-307.

Ventii, K. H., Devi, N. S., Friedrich, K. L., Chernova, T. A., Tighiouart, M., Van Meir, E. G. and Wilkinson, K. D. (2008) BRCA1-associated protein-1 is a tumor suppressor that requires deubiquitinating activity and nuclear localization. *Cancer Res, 68*(17), pp. 6953-62.

Verstegen, N. E., Lagerwaard, F. J., Hashemi, S. M., Dahele, M., Slotman, B. J. and Senan, S. (2015) Patterns of Disease Recurrence after SABR for Early Stage Non-Small-Cell Lung Cancer: Optimizing Follow-Up Schedules for Salvage Therapy. *J Thorac Oncol, 10*(8), pp. 1195-200.

Vogelzang, N. J., Rusthoven, J. J., Symanowski, J., Denham, C., Kaukel, E., Ruffie, P., Gatzemeier, U., Boyer, M., Emri, S., Manegold, C., Niyikiza, C. and Paoletti, P. (2003) Phase III study of pemetrexed in combination with cisplatin versus cisplatin alone in patients with malignant pleural mesothelioma. *J Clin Oncol, 21*(14), pp. 2636-44.

Wagner, P. L., Stiedl, A. C., Wilbertz, T., Petersen, K., Scheble, V., Menon, R., Reischl, M., Mikut, R., Rubin, M. A., Fend, F., Moch, H., Soltermann, A., Weder, W., Altorki, N. K. and Perner, S. (2011) Frequency and clinicopathologic correlates of KRAS amplification in non-small cell lung carcinoma. *Lung Cancer, 74*(1), pp. 118-23.

Wan, J. C. M., Massie, C., Garcia-Corbacho, J., Mouliere, F., Brenton, J. D., Caldas, C., Pacey, S., Baird, R. and Rosenfeld, N. (2017) Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer, 17*(4), pp. 223-238.

Wang, R., Li, X., Zhang, H., Wang, K. and He, J. (2017) Cell-free circulating tumor DNA analysis for breast cancer and its clinical utilization as a biomarker. *Oncotarget, 8*(43), pp. 75742-75755.

Wang, S., An, T., Wang, J., Zhao, J., Wang, Z., Zhuo, M., Bai, H., Yang, L., Zhang, Y., Wang, X., Duan, J., Wang, Y., Guo, Q. and Wu, M. (2010) Potential clinical significance of a plasma-based KRAS mutation analysis in patients with advanced non-small cell lung cancer. *Clin Cancer Res, 16*(4), pp. 1324-30.

Wang, W., Kong, P., Ma, G., Li, L., Zhu, J., Xia, T., Xie, H., Zhou, W. and Wang, S. (2017) Characterization of the release and biological significance of cell-free DNA from breast cancer cell lines. *Oncotarget, 8*(26), pp. 43180-43191.

Weiss, J., Sos, M. L., Seidel, D., Peifer, M., Zander, T., Heuckmann, J. M., Ullrich, R. T., Menon, R., Maier, S., Soltermann, A., Moch, H., Wagener, P., Fischer, F., Heynck, S., Koker, M., Schottle, J., Leenders, F., Gabler, F., Dabow, I., Querings, S., Heukamp, L. C., Balke-Want, H., Ansen, S., Rauh, D., Baessmann, I., Altmuller, J., Wainer, Z., Conron, M., Wright, G., Russell, P., Solomon, B., Brambilla, E., Brambilla, C., Lorimier, P., Sollberg, S., Brustugun, O. T., Engel-Riedel, W., Ludwig, C., Petersen, I., Sanger, J., Clement, J., Groen, H., Timens, W., Sietsma, H., Thunnissen, E., Smit, E., Heideman, D., Cappuzzo, F., Ligorio, C., Damiani, S., Hallek, M., Beroukhim, R., Pao, W., Klebl, B., Baumann, M., Buettner, R., Ernestus, K., Stoelben, E., Wolf, J., Nurnberg, P., Perner, S. and Thomas, R. K. (2010) Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. *Sci Transl Med, 2*(62), pp. 62ra93.

Williams, T., Duraid, H., Watson, S., Durkin, A., Todd, K., Kindler, H. L. and Vigneswaran, W. T. (2015) Extended Pleurectomy and Decortication for Malignant Pleural Mesothelioma Is an Effective and Safe Cytoreductive Surgery in the Elderly. *Ann Thorac Surg, 100*(5), pp. 1868-74.

Witte, T., Plass, C. and Gerhauser, C. (2014) Pan-cancer patterns of DNA methylation. *Genome Med, 6*(8), pp. 66.

Wu, J. Y., Yu, C. J., Chang, Y. C., Yang, C. H., Shih, J. Y. and Yang, P. C. (2011) Effectiveness of tyrosine kinase inhibitors on "uncommon" epidermal growth factor receptor mutations of unknown clinical significance in non-small cell lung cancer. *Clin Cancer Res, 17*(11), pp. 3812-21.

Wu, S., Liu, S., Liu, Z., Huang, J., Pu, X., Li, J., Yang, D., Deng, H., Yang, N. and Xu, J. (2015) Classification of circulating tumor cells by epithelial-mesenchymal transition markers. *PLoS One, 10*(4), pp. e0123976.

Xu, J., Alexander, D. B., Futakuchi, M., Numano, T., Fukamachi, K., Suzui, M., Omori, T., Kanno, J., Hirose, A. and Tsuda, H. (2014) Size- and shape-dependent pleural translocation, deposition, fibrogenesis, and mesothelial proliferation by multiwalled carbon nanotubes. *Cancer Sci, 105*(7), pp. 763-9.

Yang, H., Bocchetta, M., Kroczynska, B., Elmishad, A. G., Chen, Y., Liu, Z., Bubici, C., Mossman, B. T., Pass, H. I., Testa, J. R., Franzoso, G. and Carbone,

M. (2006) TNF-alpha inhibits asbestos-induced cytotoxicity via a NF-kappaB-dependent pathway, a possible mechanism for asbestos-induced oncogenesis. *Proc Natl Acad Sci U S A, 103*(27), pp. 10397-10402.

Yang, P. (2009) Epidemiology of lung cancer prognosis: quantity and quality of life. *Methods Mol Biol, 471*, pp. 469-86.

Yasuda, H., Kobayashi, S. and Costa, D. B. (2012) EGFR exon 20 insertion mutations in non-small-cell lung cancer: preclinical data and clinical implications. *Lancet Oncol, 13*(1), pp. e23-31.

Yasuda, H., Park, E., Yun, C. H., Sng, N. J., Lucena-Araujo, A. R., Yeo, W. L., Huberman, M. S., Cohen, D. W., Nakayama, S., Ishioka, K., Yamaguchi, N., Hanna, M., Oxnard, G. R., Lathan, C. S., Moran, T., Sequist, L. V., Chaft, J. E., Riely, G. J., Arcila, M. E., Soo, R. A., Meyerson, M., Eck, M. J., Kobayashi, S. S. and Costa, D. B. (2013) Structural, biochemical, and clinical characterization of epidermal growth factor receptor (EGFR) exon 20 insertion mutations in lung cancer. *Sci Transl Med, 5*(216), pp. 216ra177.

Yates, L. R. and Campbell, P. J. (2012) Evolution of the cancer genome. *Nat Rev Genet, 13*(11), pp. 795-806.

Yates, L. R., Gerstung, M., Knappskog, S., Desmedt, C., Gundem, G., Van Loo, P., Aas, T., Alexandrov, L. B., Larsimont, D., Davies, H., Li, Y., Ju, Y. S., Ramakrishna, M., Haugland, H. K., Lilleng, P. K., Nik-Zainal, S., McLaren, S., Butler, A., Martin, S., Glodzik, D., Menzies, A., Raine, K., Hinton, J., Jones, D., Mudie, L. J., Jiang, B., Vincent, D., Greene-Colozzi, A., Adnet, P.-Y., Fatima, A., Maetens, M., Ignatiadis, M., Stratton, M. R., Sotiriou, C., Richardson, A. L., Lønning, P. E., Wedge, D. C. and Campbell, P. J. (2015) Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med, 21*, pp. 751.

Yoshikawa, Y., Sato, A., Tsujimura, T., Emi, M., Morinaga, T., Fukuoka, K., Yamada, S., Murakami, A., Kondo, N., Matsumoto, S., Okumura, Y., Tanaka, F., Hasegawa, S., Nakano, T. and Hashimoto-Tamaoki, T. (2012) Frequent inactivation of the BAP1 gene in epithelioid-type malignant mesothelioma. *Cancer Sci, 103*(5), pp. 868-74.

Yu, H. A., Arcila, M. E., Rekhtman, N., Sima, C. S., Zakowski, M. F., Pao, W., Kris, M. G., Miller, V. A., Ladanyi, M. and Riely, G. J. (2013) Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. *Clin Cancer Res, 19*(8), pp. 2240-7.

Zappa, C. and Mousa, S. A. (2016) Non-small cell lung cancer: current treatment and future advances. *Transl Lung Cancer Res, 5*(3), pp. 288-300.

Zauderer, M. G., Kass, S. L., Woo, K., Sima, C. S., Ginsberg, M. S. and Krug, L. M. (2014) Vinorelbine and gemcitabine as second- or third-line therapy for malignant pleural mesothelioma. *Lung Cancer, 84*(3), pp. 271-4.

Zhang, J., Fujimoto, J., Zhang, J., Wedge, D. C., Song, X., Zhang, J., Seth, S., Chow, C. W., Cao, Y., Gumbs, C., Gold, K. A., Kalhor, N., Little, L., Mahadeshwar, H., Moran, C., Protopopov, A., Sun, H., Tang, J., Wu, X., Ye, Y., William, W. N., Lee, J. J., Heymach, J. V., Hong, W. K., Swisher, S., Wistuba, II and Futreal, P. A. (2014) Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science, 346*(6206), pp. 256-9.

Zhang, Y. L., Yuan, J. Q., Wang, K. F., Fu, X. H., Han, X. R., Threapleton, D., Yang, Z. Y., Mao, C. and Tang, J. L. (2016) The prevalence of EGFR mutation in patients with non-small cell lung cancer: a systematic review and meta-analysis. *Oncotarget, 7*(48), pp. 78985-78993.

Zhou, L., Lyons-Rimmer, J., Ammoun, S., Muller, J., Lasonder, E., Sharma, V., Ercolano, E., Hilton, D., Taiwo, I., Barczyk, M. and Hanemann, C. O. (2016) The scaffold protein KSR1, a novel therapeutic target for the treatment of Merlin-deficient tumors. *Oncogene, 35*(26), pp. 3443-53.

Zink, F., Stacey, S. N., Norddahl, G. L., Frigge, M. L., Magnusson, O. T., Jonsdottir, I., Thorgeirsson, T. E., Sigurdsson, A., Gudjonsson, S. A., Gudmundsson, J., Jonasson, J. G., Tryggvadottir, L., Jonsson, T., Helgason, A., Gylfason, A., Sulem, P., Rafnar, T., Thorsteinsdottir, U., Gudbjartsson, D. F., Masson, G., Kong, A. and Stefansson, K. (2017) Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood, 130*(6), pp. 742-752.