Genetic Biomarkers of Chemotherapy Response and Resistance in Lung Cancer Patients

Daniel John Nelmes
MRCP

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

Cardiff University

School of Medicine

March 2018

Summary

In advanced lung cancer, careful selection of systemic anticancer therapy (SACT) is of vital importance. Companion biomarkers can optimise treatment selection, such as with the use of EGFR tyrosine kinase inhibitors (EGFR TKi) in patients with *EGFR*$^{mut+ve}$ adenocarcinoma of the lung. There is increasing interest in mutation detection and monitoring, in circulating cell free tumour DNA (ctDNA).

This thesis reports that Next Generation Sequencing (NGS) with software VarScan with Annovar, can detect mutations at a 10-fold lower alternate allele frequency compared to alternative software available through Ion Torrent, but with a greater number of low level 'false positive' genetic variants. Droplet digital PCR (ddPCR) is more sensitive than NGS, successfully detecting mutations as low as 0.1% alternate allele frequency.

Lung cancer mutations were successfully detected in small, formalin-fixed, paraffin embedded (FFPE) tumour tissue samples, and ctDNA, from lung cancer patients, using the same NGS technique, with a commercially available, targeted 50 gene cancer hotspot panel. Results are compared to a custom 22-gene panel.

The kinetics of mutation levels in serial ctDNA samples is reported in a case series. In *EGFR*$^{mut+ve}$ lung adenocarcinoma patients treated with EGFR TKi, decreases in levels of mutant *EGFR* in ctDNA were observed. Levels remained undetectable during periods of disease control/stability, and increases in mutant *EGFR* in ctDNA were seen several weeks before the diagnosis of clinical or radiological disease progression. NGS of ctDNA during disease progression revealed novel genetic mutations that were not detected in the original tumour biopsy, and may inform subsequent treatment options.

Similar ctDNA kinetics was seen in advanced SCLC patients treated with SACT. The level of mutated ctDNA, at diagnosis may be an independent prognostic biomarker, using a cut-off of 44.3% alternate allele frequency. SCLC patients who experienced a greater absolute decrease in mutant ctDNA had a poorer prognosis.

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed (candidate) Date

STATEMENT 1

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

Signed (candidate) Date

STATEMENT 2

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

Signed (candidate) Date

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed (candidate) Date

Acknowledgments

I would like to thank the funders of this research project, namely Cancer Research UK (CRUK), and Stepping Stones Charity, part of Velindre Charitable Funds.

I am grateful for the help of colleagues within the lab who helped to train me in the various laboratory techniques, and who I worked with, especially Kirsty Hambridge, Angharad Williams and Ian Williams. My fellow researchers also helped me with getting lab work done and keeping morale up, namely Alex Georgiades, Matt Summers and Magda Meissner.

I thank Dr Mick Button and Dr Jason Lester for allowing me to attend lung cancer clinics, and for facilitating patient consent and recruitment into this research project.

I thank my supervisor Rachel Butler, the head of the All Wales Medical Genetics Laboratory, and her staff, and my supervisor Professor John Chester for his input with generating this thesis. I also thank Dr Mark Davies for his guidance, support and advice.

The sheer volume of time dedicated to research and writing, especially after returning to commuting to full time clinical practice, with an out-of-hours on-call rota, after my time in the lab, meant I made several sacrifices. I am very grateful for the patience and understanding of my family and friends.

Table of Contents

List of Figures

List of Tables

Commonly Used Abbreviations

AAF	Alternate Allele Frequency
COLD-PCR	Co-amplification at lower denaturation temperature-PCR
CT	Computed Tomography
CTC	Circulating Tumour Cell
cfDNA	Cell Free Deoxyribose Nucleic Acid
ctDNA	Circulating Tumour Deoxyribose Nucleic Acid
ddPCR	Droplet Digital Polymerase Chain Reaction
DNA	Deoxyribose nucleic acid
EBUS	Endobronchial Ultrasound Scan
EDTA	Ethylenediaminetetraacetic Acid
EGFR	Epidermal growth factor receptor
FFPE	Formalin Fixed Paraffin Embedded
H&E	Hematoxylin & Eosin
ITVC	Ion Torrent Variant Caller
KRAS	Kirsten Rat Sarcoma Virus [Gene]
LN	Lymph Node
NGS	Next generation sequencing
NSCLC	Non-Small Cell Lung Cancer
PCI	Prophylactic Cranial Irradiation
PCR	Polymerase chain reaction
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha [Gene]
PTEN	Phosphatase and Tensin Homolog [Gene]
SACT	Systemic Anti-Cancer Therapy
SCLC	Small Cell Lung Cancer
TNM	Tumour Nodes Metastases
TP53	Tumour Protein p53 [Gene]
WCB	Wales Cancer Bank

Abstract

Lung cancer is the leading cause of cancer related death. Most patients present with advanced, or metastatic, small cell lung cancer (SCLC) or non-small cell lung cancer (NSCLC). Careful selection of optimal anticancer therapy is therefore key in minimising the adverse consequences of palliative treatment. Systemic anti-cancer therapy (SACT) remains the mainstay of management of advanced lung cancer. The presence of specific mutations in lung tumours can help to predict response to some SACTs, for example, the use of targeted EGFR tyrosine kinase inhibitors (EGFR TKi), in patients with *EGFR*$^{mut+ve}$ adenocarcinoma of the lung. These same mutations can often be detected in peripheral blood, as circulating cell free tumour DNA (ctDNA). However, mutations may occur at very low level, against a background of genomic or 'wild-type' DNA (expressed as an alternate allele frequency), requiring sensitive technologies, such as Next Generation Sequencing (NGS) and droplet digital PCR (ddPCR), for reliable detection.

For downstream analysis of ctDNA in samples from lung cancer patients in this thesis, limits of detection of NGS and ddPCR,, are explored using reference standard DNA. The Ion Proton NGS platform with VarScan and Annovar bioinformatics tool detected mutations at a 10-fold lower alternate allele frequency, but with a greater number of low level 'false positive' genetic variants, compared to the Ion Torrent Variant Caller software. Droplet digital PCR was more sensitive than NGS, successfully detecting mutations as low as 0.1% alternate allele frequency. Mutation detection by ddPCR was limited, however, by the amount of available DNA for the assay.

Using the same NGS approach, it is possible to detect mutations in small FFPE tumour tissue samples from patients with adenocarcinoma of the lung, squamous cell carcinoma of the lung and SCLC. The commercially available 50 gene cancer hotspot panel was more robust in detecting *EGFR* mutations in lung cancer tumour samples compared to a custom 22-gene panel which includes extra genes of clinical relevance to NSCLC. A novel *EGFR* exon 19 deletion sequence variant has been detected.

EGFR mutations were detected in ctDNA by NGS and ddPCR, even at low level, in patients with *EGFR*$^{mut+ve}$ adenocarcinoma of the lung. Quantitative changes in mutant *EGFR* in ctDNA, were monitored in the same patients, in real-time, during treatment with oral EGFR TKi. Decreases were observed and related to radiological responses to treatment. Levels remained undetectable during periods of disease control, or stability, but increases in mutant *EGFR* in ctDNA were seen several weeks before the diagnosis of clinical or radiological disease progression. Mutations detected by NGS during disease progression included detecting genetic mutations that were not detected in the original tumour biopsy. Detection of acquired mutations in ctDNA, that convey therapeutic resistance (such as *EGFR* c.2369C>T p.T790M resistance to EGFR TKi), provide a basis for decision-making in the use of newer SACT, such as osimertinib.

Similar ctDNA kinetics were seen in advanced SCLC patients treated with SACT. SCLC patients who experienced a greater absolute decrease in mutant ctDNA had a poorer overall survival. It is feasible that the level of mutated ctDNA, at diagnosis may be an independent prognostic biomarker, using a cut-off of 44.3% alternate allele frequency. This needs validation in a larger, prospective cohort of patients. Using ctDNA as a prognostic biomarker, may be particularly clinically useful for advanced SCLC patients, to help stratify them to different therapeutic options, to maximize therapeutic benefits, and to avoid potentially unnecessary treatment toxicities.

Data presented in this thesis demonstrate that NGS and ddPCR can successfully detect mutations in small FFPE tumour tissue samples, and ctDNA samples, and these can act as biomarkers in lung cancer. Changes in ctDNA level are dynamic, and can be used as a monitoring biomarker in ongoing response to therapy, as well as detecting disease relapse. Earlier signals of disease progression may facilitate an earlier change in the patient's clinical management plan, which could offer potential benefits for the patient's quality of life, during the late stages of their cancer journey. This thesis demonstrates that it is feasible to apply ctDNA mutation detection, and ctDNA kinetics, to specific clinical situations. This may help clinicians and patients make better-informed choices, about optimising the management of lung cancer, and improving patient outcomes. This should form the basis of future clinical studies.

1 Introduction

1.1 Incidence

Lung cancer is currently the second most common cancer diagnosed in the UK. In 2010, there were 42,000 new cases of lung cancer diagnosed in the UK. Lung cancer remains the most common cause of cancer death in the UK, with 34,859 deaths in 2010, at a rate of 38.6 per 100,000 population(1). Currently, there is less than a 10% overall 5-year survival associated with lung cancer(2). Up to 9 out of 10 cases of lung cancer are significantly associated with a history of tobacco smoking(3).

1.2 Histological Sub-types of Lung Cancer

Lung cancers have traditionally been classified according to their histological subtype. The latest World Health Organisation (WHO) update (2015) for histological classification of lung tumours has been widely accepted(4). The main histological subtypes of lung cancer include small cell lung cancer (SCLC) (approximately 25% of cases) and non-small cell lung cancer (NSCLC) (approximately 75% of cases). NSCLC is further divided into squamous and non-squamous cell carcinoma sub-types, including adenocarcinoma and large cell carcinoma. Traditionally, classifying lung cancers into these histological sub-types has been important in prognostication, and to direct appropriate therapeutic approaches to their treatment. For example, small cell lung cancer, usually associated with a history of heavy smoking, is usually a much more aggressive form of lung cancer, and is associated with a worse prognosis.

1.3 Presentation and Survival

The peak incidence of lung cancer is in the 65 – 89 years age group, and the majority of cases of lung cancer are diagnosed with stage IIIB (locally advanced)

or IV (metastatic) disease(2). These stages of lung cancer are associated with worse survival than earlier stages of disease(5).

1.4 Treatment of Lung Cancer

The aim of most aspects of the clinical management of cancer is to detect and diagnose cancer early enough (ideally stages I and II) to enable radical or curative therapy, most often involving surgery and/or chemo-radiotherapy. This is provided the patient is fit enough to tolerate these treatments. However, for patients with lung cancer who present with stage III (locally advanced) or stage IV (metastatic) disease, the currently available anticancer treatments can only offer a period of disease control, rather than cure. Complicating issues are that many patients with lung cancer are either too unwell to tolerate standard palliative systemic anti-cancer therapies, or have other medical problems which preclude them(6). Despite improvements in chemotherapy regimens and radiotherapy techniques for patients with lung cancer, these treatments can still be associated with significant toxicities that have an adverse impact upon patients' quality of life. This is especially so if the patients are elderly, or have concurrent medical co-morbidities. Therefore, one very important therapeutic goal in cancer therapy is to optimise efficacy whilst minimising the toxicities that may detract from the patient's quality of life.

1.5 Radical Treatment for Lung Cancer

Curative treatment for lung cancer includes surgical resection of part of a lobe of a lung (wedge resection), a whole lobe (lobectomy), or whole lung (pneumonectomy). If surgical resection is not possible, then radical radiotherapy can be considered(7),(8). There have been some clinical studies examining the role of adjuvant chemotherapy, and radiotherapy, following surgical resection of lung cancer. Adjuvant chemotherapy may offer an absolute reduction in 5 year mortality in the order of 5%(9).

1.6 Treatment of advanced NSCLC

1.6.1 First Line Chemotherapy for NSCLC

In 2008, the NSCLC Collaborative Group reported an improvement in 1 year survival from 5% to 15%, for lung cancer patients with locally advanced, or metastatic disease, treated with chemotherapy and best supportive care, compared to best supportive care alone(10).

Small-molecule cytotoxic chemotherapy agents commonly used in routine clinical practice include platinum agents that disrupt DNA synthesis and repair (cisplatin and carboplatin), thymidylate synthase inhibitors that impair DNA synthesis (pemetrexed), tubulin inhibitors such as vinca alkaloids (vinorelbine), and nucleoside analogues such as gemcitabine. In general, they work by interfering with DNA synthesis and cellular proliferation. In treating a cancer patient with these chemotherapy drugs, cancer cells are preferentially affected, owing to their higher rate of cellular proliferation, thus rendering them more susceptible to the effects of these drugs. These agents can affect non-cancer cells of the body, causing many of the toxicities associated with these therapies. Common toxicities of these agents include fatigue, nausea and vomiting, diarrhoea or constipation, neutropenia that can result in severe infection/sepsis, hospitalisation and death.

It has been reported (Schiller *et al,* 2002) that platinum-based doublet chemotherapy regimens offer the best disease control and improvements in overall survival in patients with incurable NSCLC and SCLC. There is no significant difference between the different agents combined with cisplatin or carboplatin chemotherapies, and response rates were reported in the order of 19%, with median survival of 7.9 months (7.3 – 8.5 months)(11). Survival rates with cytotoxic chemotherapy in NSCLC have not improved significantly since 1995(10). Scagliotti et al (2008) reported an improvement in survival from 10.9 months to 12.6 months, in patients with advanced stage adenocarcinoma of the lung, when treated with cisplatin/pemetrexed chemotherapy, compared to

cisplatin/gemcitabine chemotherapy. A similar survival benefit was also observed in patients with large cell carcinoma of the lung(12). Due to treatment toxicities, lung cancer patients can usually tolerate no more than 4-6 cycles of chemotherapy. After this, the potential risks of continuing chemotherapy can outweigh the clinical benefits, and so a course of first line chemotherapy is usually discontinued at this point.

1.6.2 Maintenance Chemotherapy

Maintenance chemotherapy is the term given to chemotherapy used as an ongoing treatment strategy, immediately after a course of first line chemotherapy has brought about a clinical response or stable disease. The PARAMOUNT clinical trial examined the role of maintenance pemetrexed chemotherapy, following initial therapy with cisplatin and pemetrexed chemotherapy. Maintenance pemetrexed was associated with a 22% reduction in the risk of death at median follow-up of 24.3 months(13). This has been adopted in routine clinical practice, provided the patient can tolerate on-going chemotherapy.

1.6.3 Second Line Chemotherapy

For patients who tolerate first line palliative chemotherapy, but whose cancer then progresses, second line chemotherapy may be appropriate. An example is the use of second-line, single-agent docetaxel, in previously treated, advanced stage, NSCLC. However, only modest improvements in survival, in the order of a few months, have been observed in clinical studies(14).

1.6.4 Immunotherapy in advanced non-small cell lung cancer

A mechanism by which lung cancer cells evade the immune system is by the expression of cell surface ligand, called programmed death ligand 1 (PD-L1), which inhibits the anti-tumour function of programmed death receptor 1 (PD-1) positive cytotoxic T-cells. Monoclonal antibodies that target PD-1 or PD-L1, and promote immune related tumour cell death, have been developed as systemic

anti-cancer therapy (SACT) in patients with advanced lung cancer. Examples include pembrolizumab (Keytruda®, Merck), nivolumab (Opdivo®, Bristol Myers Squibb) and atezolizumab (Tecentriq®, Roche). Levels of PD-L1 expression can be determined by immunohistochemistry of FFPE tumour tissue, and this is expressed as a percentage of nucleated tumour cells that are 'positive' for the PD-L1 stain, compared to nucleated tumour cells that are 'negative'. There is debate as to the value of PD-L1 expression levels as a predictive biomarker of response to therapy(15).

The KEYNOTE-010 clinical study revealed that previously treated, advanced NSCLC patients, with PD-L1 expression on >1% of nucleated tumour cells, have a significantly superior median overall survival (12.7 months, compared with 10.4 months for patients treated with docetaxel chemotherapy)(16). In the first line setting clinical study KEYNOTE-024, patients with PD-L1 expressed on >50% of nucleated tumour cells, who were treated with pembrolizumab, experienced a significantly greater median progression free survival (10.3 months) and overall survival at 6 months (80.2%), compared to standard first line palliative chemotherapy (median progression free survival of 6.0 months and overall survival of 72.9% at 6 months)(17). Both KEYNOTE-010 and KEYNOTE-024 report fewer significant toxicities associated with pembrolizumab, compared to chemotherapy.

Pembrolizumab is now in routine clinical use in the first line palliative setting, for patients with advanced NSCLC expressing PD-L1 >50%, or as a second or greater line therapy, for patients with advanced NSCLC expressing PD-L1 >1%(18),(19).

1.7 Small Cell Lung Cancer

Small Cell Lung Cancer (SCLC) is a particularly aggressive form of lung cancer, significantly associated with a history of heavy smoking. There were approximately 11300 new cases of SCLC in the UK in 2013. This equates to approximately 30 new diagnoses a day.(20)

Prognosis associated with SCLC is poor, and is linked to the stage at diagnosis. Historically, SCLC was described as limited stage i.e. disease contained within the thorax, and extensive stage i.e. disease that has spread outside the thorax. However, as therapeutic strategies become more refined, the TNM classification and staging system is increasingly used for SCLC. Only about 30% of patients are diagnosed with limited, or earlier stage of SCLC, and, even with treatment of less advanced disease, 2 year survival rate is only 25%, due to high recurrence rates. The majority of SCLC patients (up to 70%) present with metastatic disease (stage IV), and here treatment is even less successful, with 5-year survival less than 5%.(21) Median overall survival rates for metastatic SCLC have been estimated at 9 months from diagnosis.(22)

1.7.1 Radical treatment for Small Cell Lung Cancer

Nearly one-third of patients will be diagnosed with early stage SCLC, and radical treatments are in the form of surgical lung resection or concurrent radical chemo-radiotherapy(7). However, there is a high relapse rate, and for the majority of patients with relapsed or metastatic disease, treatment aims are disease control and prolonged survival.

1.7.2 Treatment for Advanced Small Cell Lung Cancer

As with advanced stage NSCLC, the mainstay of treatment for patients with advanced stage SCLC is palliative SACT. This most commonly entails use of a platinum compound, usually cisplatin or carboplatin, often in combination with the topoisomerase inhibitor, etoposide.

50 – 85% of patients with incurable SCLC experience quite dramatic initial responses to chemotherapy, but time to disease relapse is often only of the order of a few months, and relapsed disease is often much more chemotherapy-resistant(22). Survival in relapsed SCLC is poor, with estimated median overall survival of 4 months.(23) Furthermore, patients with relapsed SCLC tend to have a poor performance status. It has been estimated that second-line palliative

chemotherapy is a clinically appropriate option for only 42% of these patients, while the remainder receive best supportive care.(24, 25) This highlights the importance of trying to select patients who are going to receive a clinical benefit from second-line therapies, when clinically appropriate. Therefore, there is currently an unmet clinical need to develop novel therapeutic approaches, with companion biomarkers, for patients with advanced stage SCLC.

1.7.3 Brain Metastases in Small Cell Lung Cancer

Up to 30% of SCLC patients, will develop brain metastases, during progression of their cancer, resulting in marked deterioration in quality of life.(26)

Patients with early or intermediate stage disease, for example stage III, who respond to palliative chemotherapy, can subsequently receive prophylactic cranial irradiation (PCI), to reduce the incidence of developing cerebral metastatic disease, but this is associated with its own toxicities. In the study by Slotman, *et al* (2007), most eligible patients with stage III and IV SCLC, were offered PCI. The authors report that from the time of randomization, patients treated with PCI had a lower risk of developing symptomatic brain metastases, and a lower cumulative risk of developing brain metastases at one year (14.6% v 40.4%, treated v untreated groups). Also, for patients treated with PCI, the median disease free survival increased from 12 to 14.7 weeks, median overall survival increased from 5.4 to 6.7 months, and the one-year survival increased from 13.3% to 27.1%(27).

Potential toxicities of PCI include hair loss, headache, nausea and vomiting and profound fatigue. Patients treated with PCI can also experience hyper-somnolence syndrome, 6-8 weeks following completion of radiotherapy. Patients who were well enough to complete the questionnaire reported treatment toxicity, though it was not thought to be clinically significant, compared to their global health status(27). Despite these concerns over toxicity, PCI is now offered routinely to patients with metastatic, or extensive stage SCLC who respond to chemotherapy.

1.8 Pathogenesis of Lung Cancer

The pathogenesis of a cancer is thought to be due to the accumulation of cellular damage, at the genetic level, which conveys the characteristics of a tumour cell, namely uncontrolled growth, proliferation and metastatic potential(28). Tobacco smoking is the most common cause of acquired cellular genetic damage in the aetiology of lung cancer(29). Different types of genetic abnormality include point mutations, small and large deletions and/or insertions, gene translocations, gene copy number gain, loss of heterozygosity and gross chromosomal abnormalities, including translocations.

In principle, drugs that selectively target an aberrant, oncogenic molecular pathway, will selectively kill cancer cells, with reduced toxicity to other, healthy cells. These are called targeted therapies(30).

Since the 1990's, abnormalities in the Epidermal Growth Factor Receptor (EGFR) intra-cellular signaling pathway has been associated with lung cancer. In particular gain-of-function mutations have been identified in the tyrosine kinase genes in that pathway, including *EGFR, KRAS, NRAS, BRAF, MEK1(MAP2K1)(31)*. A more recent on-line database, my-cancer-genome, identifies additional oncogenic driver mutations in lung cancer, namely *AKT1, ALK, DDR2, FGFR, HER2, MET, PIK3CA*, *RET* and *ROS1.* Mutations in these genes can be found in all histological sub-types of NSCLC, thought lung cancer patients who have never smoked and are of adenocarcinoma sub-type have higher incidences of *EGFR, HER2, ALK, RET* and *ROS1* mutations(32).

Figure 1-1shows the frequency of major driver mutations found in adenocarcinoma of the lung. Novel anti-cancer agents, targeted against several of the gene products of these driver mutations, are now in routine clinical use, or in various phases of clinical development.

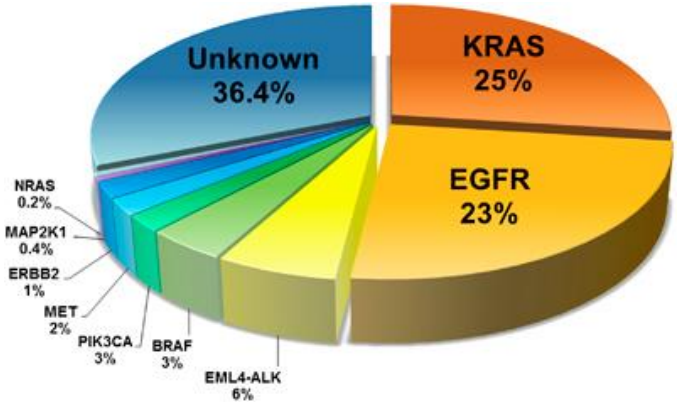

Figure 1-1 Frequency of Major Driver Mutations in Signalling Molecules in Lung Adenocarcinomas. Adapted from Cheng et al (2012)(30)

1.8.1 Clinically Relevant Genetic Changes in Lung Cancer

Table 1-1 summaries the up-to-date frequency, and availability of targeted therapies in NSCLC. The more commonly occurring genetic abberations are delt with individually below.

Table 1-1 Clinically relavent genetic abberations in NSCLC. Adapted from my-cancer-genome(32), last updated 28 March 2016.

Gene	Genetic Alteration	Frequency in NSCLC (%)	Availability of Targeted Therapy
AKT1	Mutation	1	In clinical development
ALK	Rearrangement	3 – 7	Drugs approved in NSCLC
BRAF	Mutation	1 – 3	Drugs approved in other cancer
DDR2	Mutation	~ 4	Drugs approved in other cancer
EGFR	Mutation	10 – 35	Drugs approved in NSCLC
FGFR1	Amplification	20	In clinical development
HER2	Mutation	2 – 4	Drugs approved in other cancer
KRAS	Mutation	15 – 25	In clinical development
MEK1	Mutation	1	Drugs approved in other cancer
MET	Amplification	2 – 4	Drug approved in NSCLC, but another molecular sub-type
NRAS	Mutation	1	In clinical development
PIK3CA	Mutation	1 – 3	In clinical development
PTEN	Mutation	4 – 8	In clinical development
RET	Rearrangement	1	Drugs approved in other cancer
ROS1	Rearrangement	1	Drugs approved in NSCLC

1.8.1.1 Epidermal Growth Factor Receptor (EFGR) Tyrosine Kinase Inhibitors (TKi)

Small molecule tyrosine kinase inhibitors have been developed which specifically target the tyrosine kinase domain of the cell surface receptor protein, Epidermal Growth Factor Receptor (EGFR). They work by blocking aberrant catalytic activity of the intracellular tyrosine kinase domain of EGFR, and subsequent downstream cellular growth and proliferation signals, including the

MAPkinase pathway. The first EFGR TKi developed was gefitinib (Iressa®, AstraZeneca), approved for clinical use in 2004. However, initial clinical trials involving EGFR tyrosine kinase inhibitors proved inconsistent. The ISEL study compared gefitinib with active supportive care, which demonstrated no statistically significant improvement in survival. However, a sub-group analysis revealed a significant survival benefit with gefitinib, in patients who have never smoked and were of Asian ethnicity(33).

A seminal publication, in 2004, noted that approximately 10% of the patients in these sub-groups had a rapid and often dramatic clinical response to gefitinib(34). Several mutations in the Epidermal Growth Factor Receptor (*EGFR)* gene have since been shown to be associated with the sensitivity of the tumours to EGFR tyrosine kinase inhibitors (EGFR TKi), gefitinib and erlotinib. It is estimated that between 7 -15% of patients with non-small cell lung cancer harbour a somatic 'driver' mutation in the *EGFR* gene(35). *EGFR* mutations are more strongly associated with lung cancers of the adenocarcinoma histological sub-type, and in patients who were never-smokers, female, and of Asian ethnicity(33). The clinical response to oral EGFR TKi also depends on the type of *EGFR* mutation detected, whereby some mutations act as a positive predictive biomarker of response to certain oral EGFR TKi (so called 'sensitising mutations') most commonly EGFR c.2573T>G, p.L858R, whilst others have proved a predictive biomarker for resistance (so called 'resistance mutations'), most commonly *EGFR* c.2369C>T, p.T790M (34). These are illustrated in figure 1-2.

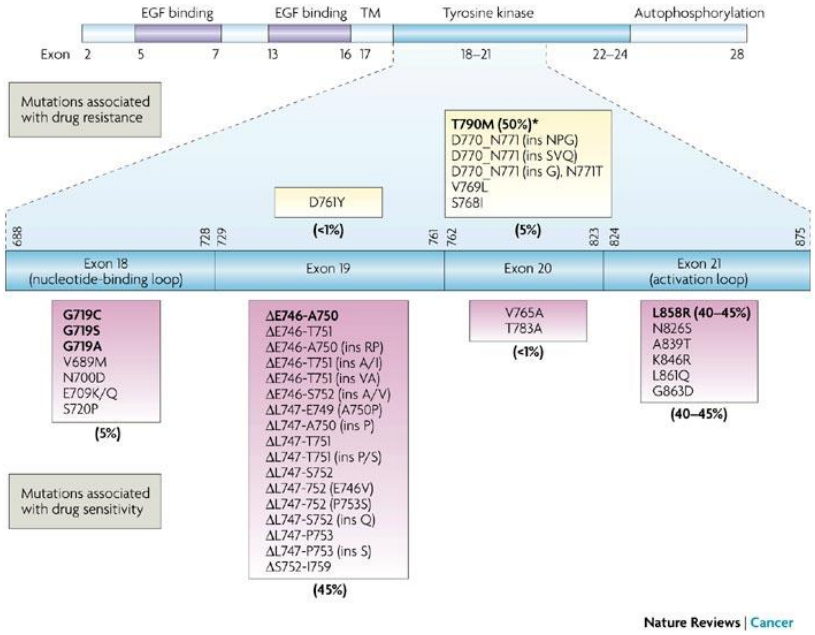

Figure 1-2 Gefitinib- and erlotinib- sensitizing mutations of EGFR in NSCLC. Adapted from Sharma et al (2007)(36)

1.8.2 Clinical Data for EGFR TKi use

The EURTAC clinical trial has demonstrated that the oral EGFR TKi, erlotinib can increase progression free survival, compared with cytotoxic chemotherapy (9.7 versus 5.2 months), in patients with locally advanced and metastatic adenocarcinoma that harbour sensitising *EGFR* mutations. Conversely, patients with *EGFR* 'wild-type' tumours do not gain a clinical benefit from EGFR TKi (37, 38). Consequently, erlotinib is approved in the UK and elsewhere, either as first line therapy for patients with locally advanced or metastatic NSCLC harbouring *EGFR* mutations(39), or second line as an alternative to docetaxel chemotherapy(40).

In the IPASS clinical trial, patients with adenocarcinoma of the lung were allocated to the EGFR tyrosine kinase inhibitor, gefitinib, or chemotherapy (carboplatin and paclitaxel), and analysis was performed on patients with lung tumours that harboured *EGFR* mutations. This demonstrated a 3-month improvement in median progression free survival (from 7 months to 10 months) in $EGFR^{mut+ve}$ patients treated with gefitinib(37). In the UK, gefitinib has been approved since 2010, for the first line treatment in patients with locally advanced or metastatic NSCLC whose tumours harbour *EGFR* mutations(41).

The BR21 clinical study utilized EGFR TKi, erlotinib (Tarceva®, Roche), as a second line palliative therapy for patients with NSCLC(42). This demonstrated a modest improvement in overall survival for erlotinib, compared with placebo. In this study, patients were not selected to receive erlotinib based on their tumour *EGFR* mutation status.

Compared to erlotinib and gefitinib, afatinib is a second generation, irreversible EGFR TKi. It has been used in the LUX-Lung 3 clinical trial, which randomised *EGFR* mutant-positive lung cancer patients (with *EGFR* sensitizing mutations L858R or exon 19 deletions) to afatinib or chemotherapy (cisplatin and pemetrexed). Median progression free survival for the afatinib group was 13.6 months, and 6.9 months in the chemotherapy group (HR 0.47 (0.34-0.65) $p<0.0001$)(43).

1.8.3 EGFR Testing

Based on this clinical data, it is therefore extremely important to be able to offer a routine test for *EGFR* mutational analysis of tumour DNA from lung cancer patients, at diagnosis. This is to ensure distinguish between patients with $EGFR^{mut+ve}$ adenocarcinoma of the lung, who are most likely to benefit from EGFR TKi therapy, and patients with $EGFR^{wild\text{-}type}$ adenocarcinoma of the lung, who are unlikely to obtain a clinical benefit, but may experience unnecessary toxicity. This is an example of patient 'stratification'.

1.9 Therapeutic Resistance to EGFR tyrosine Kinase inhibitors

Almost invariably, patients with locally advanced or metastatic, $EGFR^{mut+ve}$ adenocarcinoma of the lung will eventually develop resistance to continuous EGFR TKi therapy, and the cancer will progress. Molecular mechanisms of resistance to EGFR TKi have been characterised. Approximately 50% of molecular resistance mechanisms involve the *EGFR* gene, specifically via *EGFR* gene amplification, or via a secondary mutation within the *EGFR* tyrosine kinase domain, that may convey resistance to EGFR TKi. The most common type of

resistance-conferring mutation of this nature is the *EGFR* c.2369C>T, p.T790M. The remaining 50% of resistance mechanisms involve other genes such as mutations within *PIK3CA*, and gene amplifications including *HER2* and *MET(44)*.

Precise identification of molecular mechanism of resistance to therapy are becoming increasingly important for patients with metastatic *EGFR*$^{mut+ve}$ adenocarcinoma of the lung, as subsequent therapeutic options become available for these patients. For example, there has been the development of a 'third generation' of EGFR TKi (osimertinib, Tagrisso® - previously AZD9291, AstraZeneca), that has shown to be effective against both the original sensitising *EGFR* mutation and the secondary *EGFR* c.2369C>T, p.T790M resistance mutation(45).

1.9.1 ALK and ROS1 Translocated Lung Cancer

1.9.1.1 EML4-ALK Translocation

In 2007, an additional oncogenic mechanism was discovered in NSCLC, in the form of echinoderm microtubule associate protein like 4 – anaplastic lymphoma kinase (*EML4-ALK*) genetic translocation(46). NSCLC patients harbouring these *ALK* translocations have been reported in up to 5% of lung cancer cases, and tend to be most prevalent in younger patients, with adenocarcinoma histological sub-type, with a light- or never-smoking history(47).

1.9.2 ALK Inhibitors

Clinical trials have randomised patients with ALK translocation positive lung cancer, between the ALK inhibitor, crizotinib (Pfizer), and chemotherapy, in the first line setting (Profile1014(48)), and in the second line setting, between crizotinib and docetaxel or pemetrexed chemotherapy (Profile1007(49)). Profile1014 observed an increased median progression free survival from 7.0 months (chemotherapy group) to 10.9 months for ALK positive patients treated with crizotinib. However, an overall survival advantage with crizotinib was not demonstrated, possibly due to cross-over of patients between arms, whereby

patients who developed progression of their cancer after chemotherapy, were then treated with crizotinib, and vice versa.

Like patients with *EGFR*$^{mut+ve}$ adenocarcinoma cancer, patients who have an *ALK*-translocation positive lung cancer may develop resistance to ALK inhibition therapy, due the development of a secondary mutation within the tyrosine kinase domain of *ALK*, in 50% of cases(50).

1.10 Other Clinically-Relevant Genetic Changes in NSCLC

KRAS

Lung cancer with mutated *KRAS* has been known to be a poor prognostic factor since 1990. They tend to occur in Caucasian patients with a history of smoking. (51). Mutations in *KRAS* and *EGFR* tend to be mutually exclusive as they exist in the same intracellular signaling pathway(52, 53). It is generally thought that EGFR TKi are not effective in *KRAS* mutant lung tumours, but Miller et al demonstrated some mild tumour shrinkage with erlotinib (not significant by RECIST criteria). More targeted therapies are in clinical development (Table 1-2).

HER2

The most commonly occurring mutation in *HER2* in lung cancer are Exon 20 insertions, and tend to occur in female, non-smoker patients(54). These tumours are sensitive to HER2 targeted therapies in-vivo and in-vitro(55, 56). Generally, *HER2* mutated lung cancers are resistant to EGFR TKi therapy, but increases in *HER2* gene copy number may convey sensitivity to gefitinib therapy(57). Targeted HER2 therapies are approved for the clinical use in *HER2* amplified breast cancers and gastro-oesophageal adenocarcinomas(58, 59). Further clinical studies in lung cancer are underway (Table 1-2).

BRAF

It has been proven in a mouse model that lung cancers containing mutant *BRAF* c.1799T>A, p.V600E, may be sensitive to treatment with a down-stream MEK inhibitors, such as CI-1040(60). A clinical benefit of BRAF and MEK inhibitors

has been observed in clinical studies involving *BRAF* mutated lung cancer. These are summarized in Table 1-2.

PIK3CA

Mutations in this gene can be observed in up to a third of cases of colon, liver and breast cancer, but only 1-4% of lung cancer cases. So far, mutations in *PIK3CA* have not been shown to be mutually exclusive with other mutations, and further research is required to determine whether *PIK3CA* mutational status predicts tumour response to EGFR TKi therapy(61).

FGFR

Fibroblast Growth Factor Receptor (FGFR) is a cell surface tyrosine kinase receptor, and mutations in lung cancer have been found in upto 2% of cases of NSCLC, and include activating point mutations and gene fusions(62). Pan-FGFR inhibitors and FGFR antibodies are being developed in pre-clinical research, and there has been some evidence that these agents demonstrate anticancer activity in a few patients with mutant *FGFR* bladder cancer (63), but clinical data for the efficacy of anti-FGFR therapy in FGFR mutant NSCLC is currently lacking.

MET

Another cell surface tyrosine kinase receptor – MNNG-HOS transforming gene (MET), can also be amplified, or mutated in NSCLC, specifically via an exon 14 skipping mutation, leading to the enhanced signaling through the MET receptor pathway(64). The exon 14 skipping mutation has been reported in 3-4% of cases of adenocarcinoma of the lung(65). There are numerous case reports in the literature that suggest that patients with MET mutant adenocarcinoma of the lung may get a clinical benefit from MET inhibition therapy, for example Crizotinib (66).

RET gene fusionsFusions of this gene are thought to occur in approximately 1% of cases of adenocarcinoma of the lung (67). A phase 2 study using a multi-targeted kinase inhibitor – cabozantinib, only reported 3 cases of RET fusion

positive patients (68), so further clinical data is required to determine whether this is an effective treatment strategy for these patients.

Table 1-2 Clinically Relevant Genetic Changes in NSCLC and Potential Therapeutic Agents (other than *EGFR* mutations and *ALK* translocations). PFS = Progression Free Survival.

Genetic abberation	Clinical Implication	Agent	Clinical Study	Clinical Outcome	Reference
KRAS mutation	Mutant KRAS conveys poorer prognosis and relative resistance to chemotherapy	Selumetinib (MEK inhibitor) – downstream of KRAS	Phase II & Phase III (SELECT-1) for Docetaxel +/- selumetinib for KRAS mutant NSCLC	Addition of selumetinib to docetaxel did not improve PFS	(69)
MET gene amplification	Target for MET Tyrosine Kinase Inhibitors	Tivantinib (ARQ197)(Arqule Inc)	Marquee Clinical Trial – Erlotinib +/- Tivantinib as 2nd or 3rd line therapy in NSCLC	Improved PFS and OS in patients with increased MET expression (confirmed by IHC)	(70)
BRAF mutation	Target for BRAF inhibitors	Vemurafenib and debrafenib	Phase II Basket Clinical Trial – treating BRAF mutant NSCLC with vemurafenib	RR 42% Median PFS 7.3 months One year PFS 23%	(71)
			EURAF – retrospective study of BRAF mut+ve NSCLC patients treated with BRAF inhibitors	RR – 50% Median PFS 5 months	(72)
			BRAFmut+ve NSCLC treated with combination therapy dabrafenib and trametinib (MEK inhibitor)	RR >60%	(72)
DDR2	Target for DDR2 inhibition therapy	DDR2 inhibitors	Phase II clinical study dasatinib v chemotherapy	Lower RR with dasatinib compared with chemotherapy and increase in pleural effusion rates	(73)
RET gene fusions	Multi-targeted kinase inhibitor	Cabozantinib	Phase II	3 patients experience partial response	(68)

1.10.1 Next Generation Sequencing in Lung Cancer

A key contributor to our understanding of the molecular abnormalities of lung and other cancers has been next generation sequencing (NGS), also known as massively parallel genetic sequencing. This allows the analysis of the genetic sequence of the whole genome, whole exome, or targeted panels of genes, from a single sample of tumour DNA. This significantly reduces the cost, time and DNA requirement, to achieve the same degree of genetic sequencing, compared with more conventional techniques such as Sanger sequencing(74). The Cancer Genome Atlas (TCGA) pan-cancer analysis project, uses NGS to profile and analyse large numbers of human tumours, generating vast databases of genetic information which help inform cancer pathogenesis, and identify molecular targets for anticancer therapy or causes of resistance to anticancer therapies(75).

1.10.2 Tumour Clonal Evolution

NGS has made a major contribution to the understanding of the theory of tumour clonal evolution(76). This theory proposes that, as tumours evolve, they accumulate acquired genetic mutations, and different sub-clones of tumour cells develop. It may be that different tumour sub-clones convey different tumour characteristics, such as the ability to metastasise, or convey resistance to anti-cancer therapies.

1.10.3 Intra-patient Heterogeneity

Within each patient, different tumour sub-clones, characterised by the presence of different somatic mutations, may be observed both within the primary tumour, and between metastatic tumours(77). Some genetic aberrations are common to all sub-clones in an individual patient's disease. These are thought to have originated earlier in tumourigenesis, and are common to all sub-clones, are called 'trunk' mutations. Mutations that occur later in tumour evolution, and that may be unique to particular tumour sub-clones, are so called 'branch mutations'(78).

1.10.4 Inter-patient Tumour Heterogeneity

Patients with seemingly the same histological sub-type of lung cancer, may exhibit very different outcomes, both in terms of their response to treatment and their prognosis. This may reflect observations, from tumour sequencing data, that different tumours, with identical microscopic appearances, harbour different oncogenic somatic mutations(79),(80).

A landmark study, TRACERx, published in 2017 (by the Swanton group) aimed to prospectively characterize intra-tumour heterogeneity and determine the clonal nature of key driver events and the evolutionary process in early NSCLC. This was achieved by whole-exome-sequencing of 100 early stage, resected NSCLC tumours, from a total of 327 different tumoral regions. Driver mutations involving *EGFR, MET, BRAF* and *TP53*, almost always occur early in tumour evolution, and are common to all tumour sub-clones. At least 75% of driver mutations that occur later in tumour evolution, that are unique to certain tumour sub-clones, involve *PIK3CA*, *NF1*, genes involved in chromatin modification and DNA damage response and repair. Ongoing dynamic chromosome instability and genomic doubling were features associated with intra-tumour heterogeneity and parallel evolution of driver somatic increases in gene copy number including *CDK4*, *FOXA1* and *BCL11A*. Particularly, intra-tumour heterogeneity mediated via chromosomal instability was associated with an increased risk of recurrence of resected NSLC and death(81).

1.10.5 Molecular Resistance

Unfortunately, even lung cancer patients who experience a dramatic initial clinical response to systemic treatment, will eventually develop resistance to treatment and/or progression of their disease. It is thought that clonal evolution in the presence of an anti-cancer agent, leads to the development of one or more tumour sub-clonal populations that are resistant to that particular anti-cancer therapy. These resistant sub-clones may be present prior to therapy (innate

resistance), or may emerge during the course of treatment and/or clinical follow-up (acquired resistance).

1.10.6 Molecular landscape of Small Cell Lung Cancer

As in NSCLC, a large number of acquired somatic mutations are observed in small cell lung cancers, often due to the mutagenic effects of tobacco smoke, in patients with a heavy smoking history. One study of NGS in a small cell lung cancer cell line (NCI-H209) identified 22,910 somatic substitutions, including 134 contained within coding exons.(82)

Inactivation of TP53 or RB1 occurs in over 90% of cases of SCLC.(83) This is confirmed in a recent comprehensive genomic profiling of 110 cases of small cell lung cancer.(84) George et al (2015) describe bi-allelic inactivation of TP53 and Rb1, sometimes by complex rearrangements. Unlike in NSCLC, few cases of SCLC have been associated with mutations in kinase genes e.g. *KRAS*, *EGFR* and *PTEN*. Inactivating mutations in the *NOTCH* family of genes were seen in 25% of cases of SCLC.(84)

1.11 Epigenetic Alteration in Small Cell Lung Cancer

Other molecular mechanisms can indirectly affect cancer gene expression (epigenetic changes). These are a hallmark of SCLC. These can be secondary to genetic alterations in genes that are involved in transcriptional regulation (e.g. abnormalities in the transcriptional regulator *CHD7* and chromatin modification(82). A massive genetic insult, followed by remodelling of chromosome structure (chromothripsis) is alternate mechanism of Rb1 deregulation (when Rb1 was wild-type). This results in an overexpression of cyclin D1.(84) The *SOX* transcription factor family of genes has also been implicated in SCLC, with amplification of SOX2 occurring in approximately 27% of cases. Supporting evidence comes from the observation that proliferation of SOX2-amplified SCLC cell lines could be achieved, using anti-SOX2 shRNA. MYCL1 is also a transcription factor, and *RLF-MYCL1* fusion transcripts have also

been observed in SCLC cell lines, that have demonstrated reduced proliferation when MYCL1 is subsequently silenced.(85)

Similarly to NSCLC, there are newer therapies, in various stages of clinical development, which are more specifically targeted to the aberrant, mutational processes implicated in SCLC pathogenesis. These are summarised in table 1-2.

Table 1-3 Summary of targeted therapeutic strategies in Small Cell Lung Cancer (reviewed in Santarpia et al 2016)(86)

Therapeutic Target	Potential Predictive Biomarker	Potential Therapeutic Strategy
Receptor Tyrosine Kinases	Pre-treatment phosphor-ERK level	Low levels may convey sensitivity to IGF-R1 inhibitor
	FGFR1 amplification	Sensitivity to FGFR inhibitors
	FGF/FGFR1 member molecular aberrations	Sensitivity to FGFR inhibitors
	Activating mutations in c-MET	Sensitivity to small molecule MET inhibitors and MET monoclonal antibodies, or synergy of these with Top1 inhibitor
PI3K/AKT/mTOR pathway	PI3K/AKT pathway activation and low BCL-2 expression	Sensitivity to mTOR inhibitors. Overcome chemo-resistance with mTOR inhibitors
	MYC and eIF4E overexpression	MYC-eIF4E inhibitors may convey sensitivity to mTOR inhibitors
	PIK3CA mutation or amplification, *AKT* or *PTEN* mutation	Phase II trial using AKT inhibitor
Targeting apoptosis	Overexpression BCL-2	BCL-2 antisense oligonucleotide enhance efficacy of cis/etop chemo
	Overexpression BIM	Synergy of mTOR and BCL-2 inhibitors
Hh pathway	To be established	Small Molecule Inhibitors of Hh signalling pathway or anti-IGF-1R monoclonal antibodies. Potential synergy with chemotherapy
DNA damage/repair	Overexpression of PARP1. Elevated baseline expression levels of multiple DNA repair proteins ("DNA repair score")	PARP inhibitors. Potential of PARP inhibitors to enhance chemotherapy and radiotherapy efficacy. Potential to combine PARP inhibitors/ WEE1 inhibitors.
MYC	Increased expression of enhancer associated transcription factor genes eg. MYC or SOX2	Sensitivity to CDK7 inhibitors
Aurora kinase	MYC overexpression	Sensitivity to AURKA inhibitors
HDAC	To be established	Sensitivitiy to HDAC inhibitors and potentiate effect of chemotherapy
Hsp90	Increased expression of RIP1	Sensitivity to Hsp90 inhibitors and potential use in combination with chemotherapies that up-regualte RIP1
Anti-angiogenesis	To be established	Sensitivity to mono-clonal antibodies against VEGF and other anti-VEGF/VEGFRs agents
PD-L1	PD-L1 expression, Mutational load and higher burden of candidate neoantigens, molecular smoking marker and DNA repair pathway mutations	Immune checkpoint inhibitors against CTLA-4 and PD-1/PD-L1
	TP53 Mutation, overexpression of GD3	Vaccines
Neuro-endocrine marker inhibitors	Increased expression of NCAM1, CD56	Anti-body drug conjugate - anti CD56-DM1

1.12 Biomarkers

It is important to be able to monitor the success of lung cancer treatment and, with an increasing list of effective drugs from which to choose, prospectively predict response to each, so that we achieve the primary aim of 'precision medicine' - "the right treatment for the right patient at the right time". It is therefore imperative that new and better predictive biomarkers of response are developed, comparable to *EGFR* mutations in *EGFR*$^{mut+ve}$ adenocarcinoma, for each of the new targeted agents, as they are developed.

A biomarker can be defined as a "*Characteristic, objectively measured and evaluated, indicator of normal biological processes, pathogenic processes or pharmacologic responses to a therapeutic intervention*." (US National Institute of Health Biomarkers Definition Working Group). Biomarkers can be anything that signal the presence of a biological process, including cancer. There are several potential uses of cancer biomarkers, for example to help in cancer screening, cancer diagnosis, inform prognosis, predict response to treatment, determine minimal residual disease, or detect disease relapse or progression. Although 'biomarker' is a popular, recent terminology, in fact such molecules have been used for decades, for example, serum Alpha-fetoprotein (AFP), human Chorionic Gonadotrophin (hCG) and Lactate dehydrogenase (LDH) in germ cell tumours. Further examples of blood-borne biomarkers, in routine clinical practice include Prostate Specific Antigen (PSA) in prostate cancer, Chorionic Embryonic Antigen (CEA) in colorectal cancers and Ca125 in Ovarian cancer(87).

1.12.1 Biomarkers in lung cancer

Until recently, the main biomarkers used in routine clinical practice for patients with lung cancer are imaging, rather than molecular, biomarkers - abnormal radiological appearances, detected by chest radiograph or CT scan. Monitoring of therapeutic response to anticancer therapies includes clinical assessment, and interval radiology comparing serial radiological appearances to base-line scans.

There are currently no validated biomarkers for predicting response to palliative chemotherapy, monitoring periods of disease control, or detecting disease relapse and prognosis in patients with lung cancer.

1.12.2 Predictive Biomarkers to Targeted therapies in lung cancer

Specific mutations within certain genes can act as 'predictive biomarkers' of response of lung tumours to targeted therapies, such as *EGFR* mutations for EGFR TKi therapy, as above. The USA's National Comprehensive Cancer Network (NCCN) currently recommends molecular mutational analysis of the genes *EGFR, BRAF, ERBB2*, and *MET*, and for gene rearrangement analysis for *ALK*, *ROS1* and *RET*, and *MET* gene amplification analysis, in patients with metastatic non-small cell lung cancer, to guide the use or avoidance of potential targeted therapies(88). The most commonly accepted method of detecting these mutations is by genetic analysis of tumour DNA, extracted from tumour tissue, usually obtained as a result of an invasive biopsy, or surgical resection. Recently, the value of mutational analysis of alternative sources of tumour DNA, such as circulating cell free tumour DNA (ctDNA), has also started to become clearer. Clinically relevant analysis of these sources of DNA for genetic abnormalities has been dependent on major technological advances, particularly in nucleic acid sequencing.

1.13 Relevant Genetic Technologies for Cancer Mutation Detection

1.13.1 Sanger Sequencing

Sanger sequencing allows in-vitro DNA replication of a target DNA sequence, using short DNA primer sequences and selective incorporation of fluorescently-labelled, chain-terminating dideoxynucleotides (ddNTP), by DNA-directed DNA polymerase. A different fluorophore is used to represent each of the 4 ddNTPs. The resulting DNA strands of different lengths and fluorophore labels, are separated by size using the principle of chromatography, and genetic sequence data is generated by fluorescence detectors. This technique can be used to assess the genetic sequence of single genes from any given DNA sample. It requires a relatively large amount of input DNA, which is not always available from small

lung cancer tissue samples. Hence, while this technique can be used for single gene analysis in tumour tissue samples, it is not able to detect low-level genetic mutations, such as those seen in ctDNA samples.

1.13.2 Pyrosequencing

Pyrosequencing is a quantitative, real-time DNA sequence analysis performed on biotinylated PCR products. It involves four key enzymes; DNA polymerase, ATP sulfurase, firefly luciferase and apyrase. Iterative incorporation of nucleotides by DNA polymerase causes an equivalent release of inorganic pyrophosphate. This by-product of DNA elongation is utilised by luciferase, to produce equimolar quantities of ATP and light. The release of light is detected and recorded. Apyrase degrades any ATP and unincorporated nucleotides before the cycle of incorporating the next nucleotide starts again. Resulting sequence data files are visualized and analysed using software, to detect the presence of a mutation.

1.13.3 COLD-PCR

Co-amplification at Lower Denaturation temperature PCR (COLD-PCR) is a modified PCR technology that reduces the signal to 'wild-type' ratio. For samples containing a mix of mutant and wild-type DNA strands, a critical denaturation temperature is used, allowing a mutant: wild-type heteroduplex to form. This heteroduplex preferentially denatures at lower denaturation temperatures, relative to a wild-type: wild-type duplex. This allows the preferential amplification of DNA strands containing a mutant sequence. When used prior to pyrosequencing, the limit of detection is lowered 100 fold, which is particularly more useful for detecting a low level of mutant alleles in a background of wild-type DNA(89).

1.13.4 Next Generation Sequencing (NGS)

NGS generates genetic sequence data for a whole genome, whole exome, multiple genes, or genetic regions simultaneously, from a single tumour DNA sample, rather than performing multiple, sequential, single-gene analyses. This is particularly advantageous when there is a limited amount of DNA from small tumour tissue samples, as is commonly the case with lung cancer samples. Figure

1-3 gives an overview of the principle of amplification based, next generation sequencing.

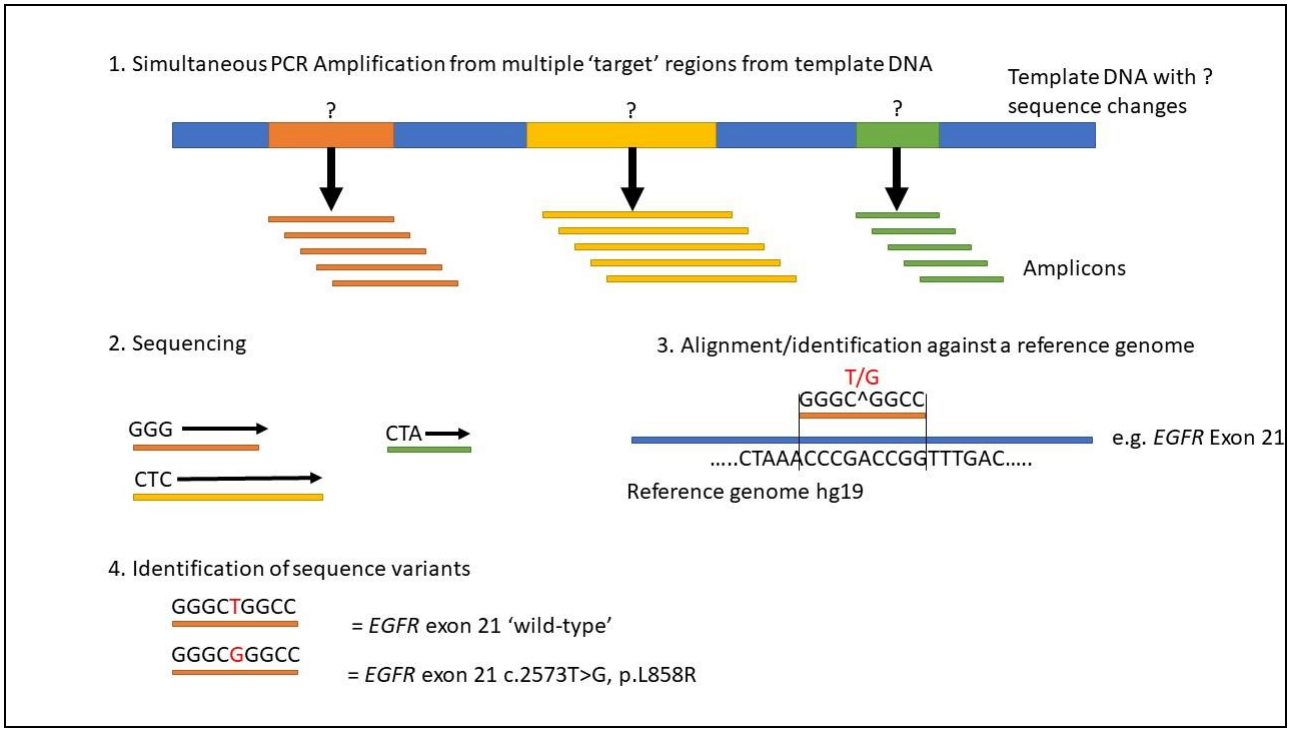

Figure 1-3 Principles of amplification based next generation sequencing (NGS).

NGS works by utilizing a 'pool' of DNA primers that amplify multiple, specific genetic regions, in a single PCR reaction. The resulting 'amplified' genetic regions are called amplicons. Multiple 'overlapping' amplicons are 'sequenced', and analysed using bioinformatics software tools, to generate sequence data for whole genes. Subsequently, multiple genes from a single DNA sample can be screened for genetic sequence changes, including pathogenic mutations.

1.13.5 Droplet Digital PCR

The droplet digital PCR process is summarized in Figure 1-4 below.

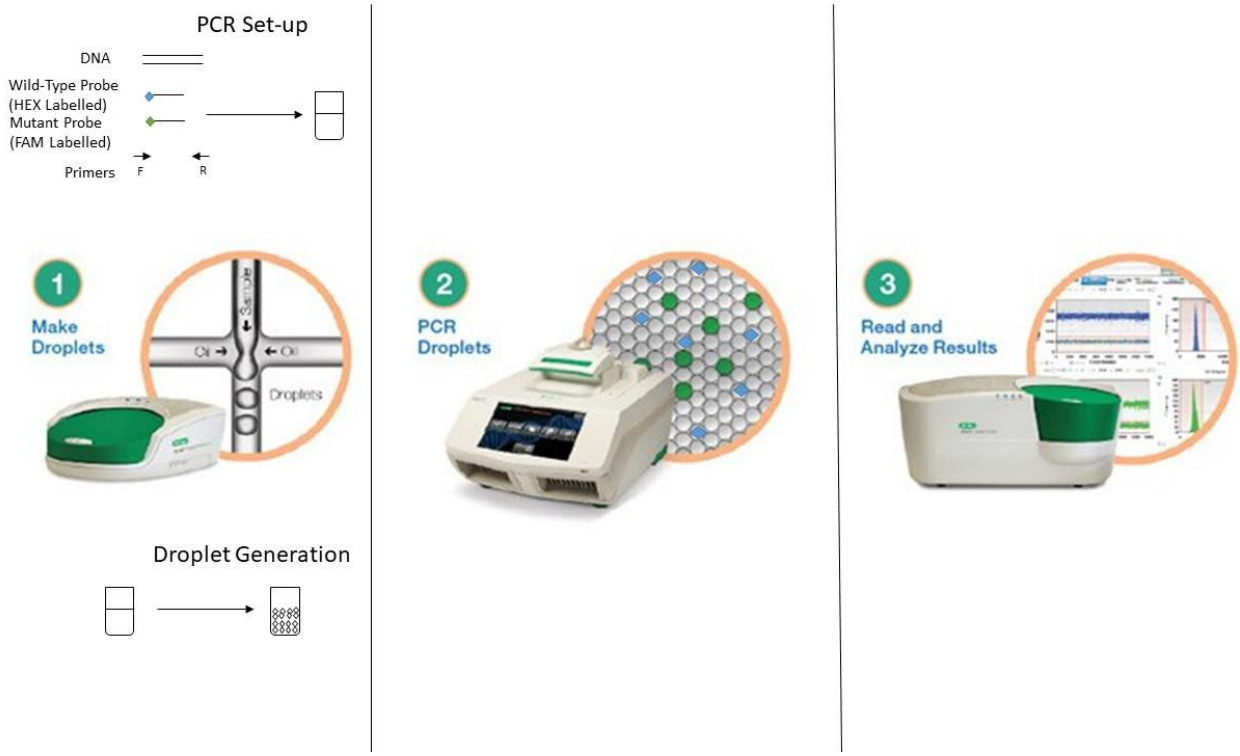

Figure 1-4 Droplet Digital PCR (BioRad technology). 1. Initial emulsion PCR mixture is dispersed into approx.. 20000 droplets, each containing an average one copy of DNA (mutant or wild-type), primers, mutant probe and wild-type probe. 2. PCR of droplet. 3. Each droplet is 'read' individually and classified as positive or negative. Positive droplets further divided into mutant or wild-type. Adapted from http://www.bioradiations.com/digital-pcr-at-the-next-level-users-find-many-applications-for-the-qx100-droplet-digital-pcr-system/ (accessed 11 Feb 2018).

Droplet digital polymerase chain reaction (ddPCR) is a quantitative PCR method. DNA primers and fluorescently labelled DNA probes are designed against a specific DNA mutation sequence and their corresponding wild-type DNA sequence. The PCR reaction is dispersed into 20,000 individual reaction droplets, and utilising Poisson distribution statistics, the emulsion PCR of the droplets occurs with an average of 1.2 copies of DNA per droplet. Depending on whether the copy of DNA in each droplet is mutant or wild-type, the corresponding fluorescent probe will be incorporated, so that each droplet can then be individually detected as 'mutant' or 'wild-type' on a droplet reading machine. This gives a digital, and subsequently very accurate DNA quantitation, and ratio of mutant allele compared to a background of wild-type alleles (alternate allele frequency (AAF)), with Poisson error statistics. Droplet digital PCR has a much lower limit of mutant allele detection than NGS - the sensitivity of this platform has been reported to detect alleles as rare as 0.01% frequency(90). However, unlike NGS, which can provide sequence data of multiple genes from one DNA

sample, ddPCR is based around hybridization of sequence-specific probes, and thus can only detect the presence of specific genetic sequences, depending on which probes are used.

1.14 Detecting EGFR Mutations and ALK translocations in lung cancer samples

There are many different accepted methods, in widespread clinical-diagnostic use, to detect somatic *EGFR* mutations. These include Sanger sequencing and pyrosequencing of DNA extracted from formalin fixed, paraffin embedded (FFPE) lung tumour tissue(91). The current accepted standard to detect ALK translocation is by fluorescence in-site hybridization (FISH) analysis of the FFPE tumour tissue, using a break-away probe(92)

1.15 NGS in Lung Cancer

1.15.1 Targeted NGS Gene Panel for lung cancer

There exist different platforms on which to perform NGS. Therefore, when choosing an appropriate NGS strategy for lung cancer DNA analysis, important factors to be considered include the quantitative DNA requirement of the NGS reaction and the choice of genes to be included by any targeted panel.

1.15.2 DNA Requirement

The DNA requirement for NGS can range, from as little as 10ng DNA per sample (with platforms such as Ion Torrent, Ion Proton NGS platform (LifeTech)) to 30-50ng per sample (with NGS platforms such as Illumina).

1.15.3 Gene Panels

Despite the NCCN current recommendation (section 1.12.2), there is no consensus NGS platform, or gene panel, for the analysis of lung cancer. Commercial gene panels are available. For example, a targeted 50 gene cancer hotspot panel (ThermoFischer Scientific), focuses on specific regions within 50 oncogenic or tumour suppressor genes that are known to be hotspots for pathogenic mutations, many of which are commonly observed in lung cancer.

However, to investigate genetic regions of interest to lung cancer which are not contained within commercially available panels, there is the option to design 'custom' gene panels for use in NGS.

1.16 Sources of Lung Cancer Tissue

The most common source of tumour tissue from patients with advanced or metastatic lung cancer is tumour tissue obtained from an invasive, percutaneous or trans-bronchial biopsy. Sometimes, tumour cells can be obtained from cytology of malignant pleural fluid, or from an endoscopic bronchial ultrasound (EBUS) guided fine needle aspiration of a malignant lymph node. All of these invasive procedures are associated with complications, and patients often find them uncomfortable and sometimes intolerable.(93)

A diagnostic difficulty is presented by the fact that these biopsy procedures are only possible in 80% of patients with advanced lung cancer. Furthermore, even where they are possible, in some cases the biopsies only contain very small amounts of tumour tissue. Multiple analyses need to be performed on these small samples, including diagnostic histopathological examination and immunohistochemistry, which often exhaust the tissue sample, prior to any genetic molecular analyses, such as NGS[64]. As a result, and as exemplified by *EGFR* mutational analysis, the ability to detect the presence, or confirm the absence, of cancer mutations, using more conventional genetic technologies, can depend on the tumour tissue sampling method. For examples, image-guided percutaneous transthoracic core-needle biopsies have a 31.8% mutation test failure rate, and bone samples from core biopsies, especially after decalcification have up to 40% mutation test failure rate(94). Therefore, methods of molecular analysis that require only small amounts of DNA are desirable. Alternatively, alternative sources of tumour cells and tumour DNA for mutational analysis need to be explored, such as circulating tumour cells (CTCs) and circulating cell free tumour DNA (ctDNA).

1.17 Circulating Tumour Cells (CTCs)

Individual tumour cells can be identified, isolated and enriched from the peripheral circulation in patients with cancer. Many technologies, including the CELLSEARCH® platform, and Parsortix (Angle Plc.) platform, have been developed to identify, isolate and enrich CTCs, on the basis of cell surface protein expression.

There is a relative abundance of CTCs in small cell lung cancer, and these decrease in number when a patient is exposed to carboplatin/etoposide chemotherapy(95). In non-small cell lung cancer, 32% of stage IV patients are positive for CTCs (defined as >2 absolute count) at baseline(96).

It is possible to extract DNA for molecular genetic analysis from CTCs. However, these technologies are expensive and not readily available, certainly compared to technologies used for ctDNA extraction and analysis, below.

Advantages of CTC analysis include the discovery of concurrent genetic mutations that co-exist within the same cell, rather than ctDNA, which represents a pool of DNA from all tumour cells that shed DNA into the circulation. It is also possible to extract DNA or mRNA from the same cell. Current limitations of CTC research are that it appears less cost effective than ctDNA research, and is more technically challenging.

1.18 Circulating Cell Free Tumour DNA (ctDNA)

The idea of circulating cell-free DNA (cfDNA) has been developed since it became known that free foetal DNA can be identified within the maternal circulation(97). Similarly, it is now known that cancers also shed DNA into the blood circulation. This sub-fraction of cfDNA is known as circulating tumour DNA (ctDNA). There are several hypotheses about the mechanism of DNA shedding. One is that ctDNA is shed from dead or dying tumour cells(98). The amount of ctDNA in the circulation is proportional to tumour burden(99), for example in patients with lymphoma(100).

There are between 1,800 and 13,000 fragments of cfDNA per millilitre of plasma, of which it is estimated that between 0.4 and 11% can be ctDNA. The fraction, or

percentage, of a genetic mutation within ctDNA, compared to the background of 'wild-type' or genomic DNA, is known as the alternate allele frequency (AAF), or variant allele frequency (VAF).

Total cell free DNA, and hence ctDNA, yields from plasma tend to be low, in the order of 0.5 – 1.5 ng/µl in elution volumes of 20 – 50 µl. Despite the low cfDNA and ctDNA concentrations, however, it is possible to perform molecular characterisation, following amplification with polymerase chain reaction (PCR) based technologies(101).

1.19 Clinical Uses of ctDNA

It is relatively simple to obtain ctDNA samples for analysis, via a minimally invasive, peripheral circulating blood draw, requiring only a few millilitres of blood. Serial samples can be obtained, sequentially and in 'real time', as the patient undergoes various anti-cancer therapies, enabling monitoring of disease relapse, or progression. These sequential 'liquid biopsies' are a more acceptable alternative to repeat, invasive tissue biopsy for patients, especially if the patient is unwell.

CtDNA analysis has been used as a biomarker to help monitor response to treatment, detect minimal residual disease (MRD), detect relapse and predict prognosis(102). Dawson et al (2013) monitored ctDNA in 30 patients with metastatic breast cancer, and compared it with the biomarker Ca15-3, a commonly used biomarker in breast cancer. The authors concluded that monitoring somatic mutations in ctDNA is an informative, sensitive and specific biomarker(103). Brenton and Rosenfield (2016) identified somatic *TP53* mutations in FFPE tumour tissue samples from patients with ovarian cancer. These *TP53* mutations were then monitored quantitatively in serial plasma samples using ddPCR, and comparisons made with circulating Ca-125 biomarker levels. Copy number of *TP53* mutations in cell free DNA correlated with changes in Ca-125. There could potentially be a greater prognostic value compared with Ca-125(104).

Analysis of ctDNA can provide insight into the genetic evolution of cancer as it progresses, as well as understanding molecular mechanisms of resistance to various anti-cancer therapies(105). This is exemplified by Diaz et al (2012), who detected emergent *KRAS* mutations in ctDNA in 28 '*KRAS* wild-type' colorectal cancer patients during the development of clinical resistance to anti-EGFR therapy (e.g. panitumumab or cetuximab)(106).

1.20 CtDNA in Lung Cancer

1.20.1 CtDNA Analysis of *EGFR* at Diagnosis

Clinicians are now able to prescribe EGFR TKi, such as gefitinib, based on detection of the *EGFR* mutations described above, in ctDNA from a lung cancer patient, rather than from samples taken from the tumour itself. It has been reported that there is a high concordance, in detecting *EGFR* mutations in ctDNA, compared with FFPE tumour tissue analysis, in patients with locally advanced, or metastatic *EGFR*$^{mut+ve}$ adenocarcinoma of the lung[16]. Therefore, *EGFR* mutational analysis of ctDNA is an acceptable alternative for *EGFR* mutational analysis of tumour tissue, especially for patients who do not have sufficient tumour tissue for molecular analysis, or where the DNA from tumour issue is not of sufficient quality or quantity.

1.21 CtDNA Analysis of EGFR mutations during Disease Progression/Resistance to Therapy

One study used a hybridization capture, NGS technology called Cancer Personalized Profiling by deep Sequencing (CAPP-Seq) to characterise therapeutic resistance mechanisms in ctDNA in 43 NSCLC patients treated with a 3rd generation EGFR TKi, rociletinib. Acquired mutations were simultaneously detected in *MET, EGFR* (including a novel p.L798I mutation), *PIK3CA, ERBB2*, *KRAS* and *RB1* genes (107).

The most commonly-observed mechanism of resistance to therapy in *EGFR*$^{mut+ve}$ adenocarcinoma of the lung is the T790M resistance mutation. This is detectable

in ctDNA, in patients who develop acquired resistance to first line EGFR TKi(108). For such patients, the EGFR TKi osimertinib, mentioned previously, can be given, with response rates as high as 67% reported in early phase clinical trials. This is where the T790M mutation is detected by analysis of tumour tissue obtained by invasive biopsy(45).

Previously, *EGFR* T790M detection has been described in a few case studies. Two case studies investigated mutational abundance in serial ctDNA samples in osimertinib-treated patients, using ddPCR. In both patients, circulating mutational load became undetectable within 1 month of treatment, and the original sensitising mutations became detectable again when the patients developed therapeutic resistance to osimertinib, as confirmed by standard radiology. However, levels of circulating *EGFR* c.2369C>T p.T790M were only detectable in one of these osimertinib-treated patients on radiological progression, and the therapeutic resistance mechanism was confirmed to be *MET* amplification on repeat tumour biopsy. The authors suggested that monitoring levels of the original EGFR TKi sensitizing mutation, rather than T790M, may facilitate more reliable detection of disease progression, in this disease setting(109).

1.22 Recommended ctDNA Sample Processing

The validity of genetic testing techniques will be limited by the quantity and quality of ctDNA employed, and so it is crucial that great care is taken in various aspects of sample handling and processing. Specialised blood collection tubes have been developed, CellSave (Janssen Diagnostics) and Streck tubes (Streck), which contain a preservative that binds to cells, and prevents shedding of DNA into the plasma, thereby reducing degradation of cfDNA, including ctDNA. These blood collection tubes have been shown to reduce genomic DNA shedding into the plasma, up to 4 days after collection, and thus reducing the relative amount of tumour-derived DNA(110).

The ECMC Cell-free DNA Consensus Meeting (2014) recommended a double-spin protocol, whereby two more gentle centrifugation steps are employed, the first to remove the majority of cellular material without excessive cell lysis, and the second, of the plasma, to remove any remaining cellular material(111).

1.23 Technical Limitations for ctDNA analysis

A key current issue in assessing ctDNA in 'liquid biopsies' is the ability of various molecular techniques to identify somatic mutations at low AAF. Given that cancer mutations in ctDNA can exist at low alternate AAF (previously published range 0.1 – 5%)(112), it is important to utilise molecular testing techniques that increase the changes of detecting a mutation (true-positive), while minimizing the chances of a false-negative and false-positive test results. It is also important to remember that the amount of cfDNA is limited, and whether this affects the ability to detect mutations at low AAF.

With regard to NGS, many genetic sequence changes are observed in NGS data, at low level frequency (1-2% alternate allele frequencies)(113). It is recognised that many of these changes are sequencing artefacts, resulting from the NGS sequencing process. This gives rise to a potential false-positive test result for the presence of cancer mutations. Different NGS techniques and bioinformatics software tools have been developed in an attempt to enhance true-positive results, and reduce false negative results.

In order to test the lower limit of variant detection using NGS, Lanman et al (2015) 'spiked' cfDNA with a panel of 10 known mutations at a level of 0.1% variant allele frequency, i.e. one copy of DNA containing a mutation, per 1000 copies of wild-type DNA. They then performed NGS on the Illumina NGS platform. They discovered a high level of background false positives, with alternate allele frequencies 0.05 – 5%(114).

Incorporating unique short DNA sequences, or 'barcodes', to ctDNA molecules, prior to NGS, can increase the true positivity rate, as is the case with tagged amplicon deep sequencing (TamSeq), an NGS method that can detect mutant

alleles as low as 2% with sensitivity and specificity of >97% (115). Another NGS method to improve low-level mutation detection, is Cancer Personalised Profiling by deep Sequencing (CAPP-Seq), which utilises DNA probes for specific cancer mutations, followed by a hybridisation-capture NGS technique. It can detect one mutant DNA strand in a background of 10,000 wild-type DNA molecules. Using this technique, the authors report an ultrasensitive ability to detect ctDNA in 100% of patients with stage II-IV NSCLC, and in 50% of patients with stage I. They report a 96% specificity for mutant allele fractions down to 0.02% alternate allele frequency(116). However, within this publication there was still the issue of false positive, or 'background' sequence noise, from which it is hard to differentiate the true positive genetic sequence changes. Therefore this group developed integrated digital error suppression (iDES). This is an in-silico elimination of highly stereotypical background artefacts and a molecular barcoding strategy to identify ctDNA molecules. The authors report that CAPP-Seq, in combination with iDES, can further increase the ability to detect low level ctDNA. In particular, when applied to patients with NSCLC, they achieved 90% sensitivity and 96% specificity in detecting *EGFR* mutations(117). TamSeq and CAPP-Seq are not yet widely available for routine use in the clinic setting.

1.24 Thesis Overview

This thesis will focus on the use of ctDNA as a biomarker in advanced lung cancer (NSCLC and SCLC). This thesis will specifically address ctDNA as a circulating biomarker to inform prognosis, and to predict and monitor response to therapy, as well as predict therapeutic resistance. The ultimate aim is to contribute to the improvement of the ability to select lung cancer patients most likely to respond to both conventional therapies, as well as to enhance the management of patients treated with a more molecularly targeted approach.

1.25 Thesis Aims

The aims of this thesis are:

- To assess whether ddPCR and/or NGS are sufficiently sensitive to detect mutations in tumour DNA, occurring at low alternate allele frequency, and therefore their suitability for use with ctDNA analyses.
- To determine whether carefully-selected NGS platform and gene panels are able to detect somatic mutations in FFPE tumour tissue and ctDNA samples, from patients with locally advanced or metastatic lung cancer.
- To assess the potential utility of *EGFR* mutation detection in ctDNA as a surrogate predictive biomarker of response/resistance to EGFR TKi treatment in *EGFR*$^{mut+ve}$ NSCLC cancer patients.
- To assess the utility of sequential ctDNA sampling as a surrogate biomarker of early disease relapse in patients with *EGFR*$^{mut+ve}$ lung cancer.
- To assess the potential clinical utility of ctDNA as a potential predictive, and prognostic surrogate biomarker in SCLC.
- To investigate the potential of ctDNA to supplement/supplant the use of existing methods of mutation detection, in routine lung cancer clinical practice, thereby facilitating earlier clinical decision-making.

2 Methods and Materials

2.1 Project governance, patients and clinical data collection

A research protocol, patient information sheet and patient consent sheets were written to specifically recruit patients to this research project, which is sponsored by Cardiff University (Section 7 Appendix – Patient Information Sheet, Section 8 – Patient Consent Sheet). Patients were invited to give written, informed consent to participate in this research, and for access to FFPE tumour blocks from their original diagnostic biopsy, access to the clinical case notes, and to provide serial blood samples during various anti-cancer therapies.

National ethics approval was gained (Edgbaston, IRAS 149571), and local research and development approvals provided by Cardiff and Vale University Health Board, Velindre NHS Trust, and Aneurin Bevan University Health Board, to allow patient identification, recruitment and research related procedures across these sites. Patients could also consent to this research project via the Wales Cancer Bank (WCB).

At total of 29 patients were recruited: 11 patients had locally advanced or metastatic, *EGFR* mutation-positive ($EGFR^{mut+ve}$) adenocarcinoma of the lung; 10 patients had locally advanced or metastatic small cell lung cancer; and 8 patients had NSCLC - 5 patients with adenocarcinoma of unknown *EGFR* mutation status and 3 patients with squamous cell carcinoma.

Consented patients were given a unique, anonymous research identifier, against which clinical samples and data were recorded, so that no patient identifiable information was included in any of the research sample processing, data analysis or presentation.

2.2 DNA Extraction

2.2.1 FFPE Tumour DNA extraction

For each consented patient, FFPE tumour tissue slides were obtained from the patient's diagnostic biopsy, with a matched H&E slide, with the tumour area ringed and the percentage of nucleated tumour cells denoted on the slide. This was performed by the local histopathology department. Tumour tissue was subsequently macro-dissected and DNA extracted using the EZ1 (Qiagen) according to the manufacturer's protocol, and stored at 4°C.

2.2.2 DNA Quantification

Final concentrations of extracted DNA samples were determined by fluorometric quantitation. Using the Qubit® dsDNA HS (High Sensitivity) Assay Kit (ThermoFisher Scientific, UK), 2µl of sample was mixed with a fluorescent dye and buffer mix, that fluoresces once intercalated into double stranded DNA molecules. Fluorescence of each sample is measured using Qubit® Fluorometer 2.0, and the concentration of DNA is directly proportional to the amount of fluorescence signal. The concentration detection range for this assay is 0.2 – 100 ng of DNA of double stranded DNA.

2.3 Collection and processing of ctDNA samples

Up to 2x 10mls peripheral venous blood, were collected in appropriate clinical environments, such as hospital out-patients department, into CellSave blood collection tubes (Janssen Diagnostics), at selected time-points during the patient's anti-cancer therapy. Where possible, CellSave blood samples were taken at the same time as a routine clinical blood test, in order to minimize the number of blood draws.

Blood samples in CellSave tubes were centrifuged at 1000g for 10 mins at 4°C using a double spin protocol, to minimize cell lysis and reducing any extra genomic DNA entering the plasma compartment, according to CRUK protocol, and in accordance with the ECMC cell free consensus opinion(111). The

supernatant was used for ctDNA analysis and the 'buffy coat' used for genomic DNA extraction, as below.

The supernatant (plasma) was transferred into 2 ml Eppendorf tubes and centrifuged at 2000g for a further 10 mins. The second supernatant was then stored at -80°C in 1 ml aliquots.

2.3.1 Cell Free DNA extraction

Frozen plasma aliquots were thawed to room temperature, and extracted using the QiaAmp circulating nucleic acid extraction kit (Qiagen) according to the manufacturer's protocol. This involves a proteinase-K digestion step, to liberate any plasma protein-bound nucleic acids. Samples are then mixed with the Buffer ACL (a Qiagen proprietary buffer containing guanidine thiocyanate and trometamil), to optimize DNA binding to a silica membrane, contained within a column attached to a vacuum manifold. The bound DNA was washed using ethanol. Bound cell free DNA was eluted into a final volume of 40µl of AVE elution buffer (a Qiagen proprietary buffer of 0.04% sodium azide in RNase free water). Eluted DNA samples are placed on ice, ready to be quantified, using fluorometic analysis with the Qubit system described above.

2.3.2 Genomic DNA extraction

After the first centrifugation, the 'Buffy coat' from each blood sample was transferred to fresh Eppendorf tubes and stored at -80°C. Genomic DNA was extracted from 300µl of thawed buffy coat. This was performed by lysing white blood cells contained within the buffy coat, liberating nucleic DNA Lysed samples were placed into a Maxwell® RSC Instrument (Promega UK) where nucleic acids are captured using paramagnetic MagneSil® particles which are then washed in a series of ethanol washes, and nucleic acids are then eluted according to the Maxwell protocol. Resulting DNA concentrations were determined using fluorometric quantitation with the Qubit 2.0, as above.

2.4 Next Generation Sequencing – Principles

The principles in amplification based next generation sequencing are illustrated in Figure 1-3 (see Introduction chapter). The amplification based, Ion Torrent - Ion Proton platform for NGS (Ion Ampliseq™) was chosen for this project (LifeTechnologies part of ThermoFisher Scientfic, UK). Steps involved in this process are library preparation, template preparation, DNA sequencing and data analysis.

2.4.1 Library Preparation

This is the process whereby the sample DNA is mixed with a combination of short, single stranded DNA primers, and DNA polymerase, in a polymerase chain reaction, to amplify the DNA sequences of interest. This polymerase chain reaction step can be multiplexed, so that one DNA sample can be mixed with a 'pool' of primers, to amplify several 'target' genetic regions, simultaneously. Depending on the number of 'target' regions that need to be amplified, the initial PCR of the sample may need to be divided across more than one primer pool, which are later combined. Amplified segments of DNA are called amplicons, and further short strands of unique DNA sequence can be further added to the amplicons, to provide a unique 'barcode-identifier' to denote the sample of origin. This is particularly useful when combining amplicons derived from different DNA samples. This project utilized the Ion Amplieq™ library preparation method (Figure 2-1).

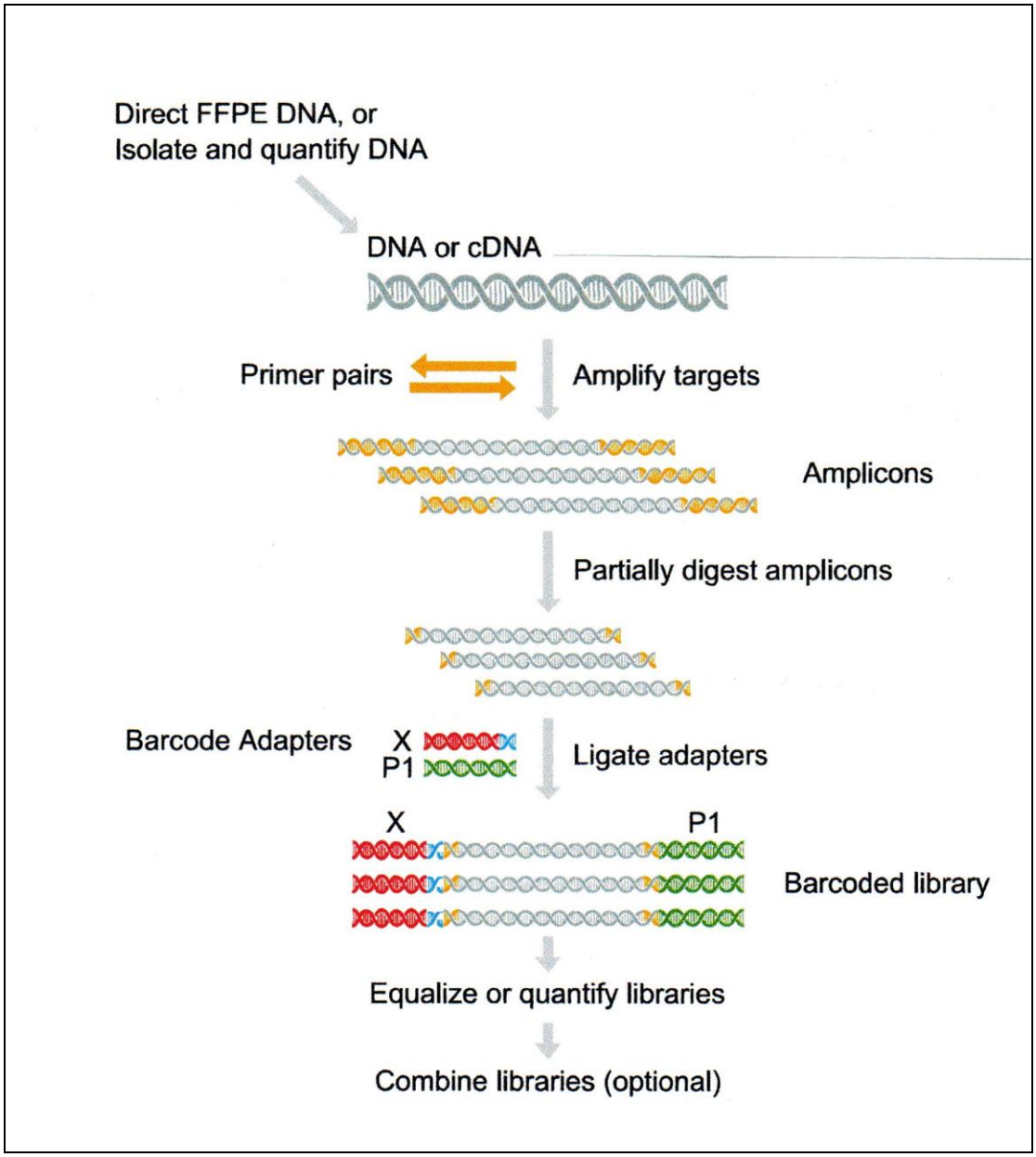

Figure 2-1 Ion Ampliseq™ library preparation workflow. Adapted from the Ion Ampliseq™ Library Kit 2.0 User Guide (ThermoFisher Scientific, Publication No. MAN0006735, Revision E.0 24 May 2017)

2.4.2 Template Preparation

Each amplicon needs to be joined to a small bead, or ion sphere particle, on which the DNA sequencing reaction takes place. Beads containing barcoded amplicons undergo a further emulsification polymerase chain reaction, to populate each bead with multiple monoclonal copies of the original amplicon. The monoclonal ion sphere particles are loaded onto a microchip, which contains 70 million individual wells. In principle, one monoclonal ion sphere particle will occupy an individual well on the microchip.

2.4.3 DNA sequencing

The prepared microchips are loaded onto the initialised Ion Proton sequencing machine. Single nucleotides (A, T, C or G) are individually passed over the microchip sequentially, and when a nucleotide is incorporated into the growing DNA template, there is a release of a hydrogen ion, resulting in a change in pH. This change in pH, within each individual well, is detected by the micro-electronic circuitry contained within the microchip. These signals are collated to generate raw DNA sequencing data, which is uploaded to an internet based sever, in real-time.

2.4.4 Data Analysis

Raw DNA sequence data is converted to data files. Genetic sequence data can then be separated according to each sample unique barcode, and compared with reference genome sequence databases (alignment) to determine which genetic areas of the sample DNA have been sequenced. Raw DNA sequence data files and aligned DNA sequence files, can be analysed using different bioinformatics software tools, to determine any genetic sequence changes, and these can be further characterised with in-silico analysis.

2.4.5 Sequencing Coverage and Depth

The degree of target DNA sequencing, ranging from single gene to whole genome, is known as sequencing coverage. The number of amplicons that cover a specific genetic region will determine sequence depth. As a general principle, the capacity of DNA sequencing, per DNA sample, per primer pool, per microchip, is limited. This means that NGS panels that cover a wide genomic region eg. whole genome, will have reduced sequence depth potential. Conversely, for NGS panels that cover a smaller, or more targeted genetic region, eg a panel of genes, a greater sequence depth will be possible. In relation to cancer, DNA samples contain a mixture of tumour DNA and germline DNA. Therefore the NGS sequence depth has to be sufficient, so that amplicons generated in the library

preparation step, will represent both tumour and germline DNA contained within the sample. Histopathological examination can provide an estimate of the number of nucleated tumour cells contained within FFPE tumour tissue sample. This can influence the degree of sequencing depth required. For example, a sample containing 10% tumour cells, may require at least 1000x sequencing depth, so that at least 100 amplicons will be generated from the tumour DNA. Samples containing 100% tumour cells, may only require at least 100x sequence depth, as all of the amplicons may be generated from tumour DNA.

Circulating cell free tumour DNA (ctDNA), may only account for a very small proportion of total circulating cell free DNA (as little as 0.01%), therefore NGS sequencing depth would need to be significantly deeper than with DNA derived from FFPE tumour tissue, to increase the chances of amplicons being generated from tumour DNA.

2.5 Next Generation Sequencing - Process

2.5.1 NGS Platform and Gene Panel Choice

The Ion Torrent, Ion Proton (LifeTechnologies, part of ThermoFisher Scientific, UK) platform was selected for somatic variant discovery via Next Generation Sequencing, since the AmpliSeq™ [118] method of library preparation only requires 10ng of input DNA per sample, per primer pool. The amplicon size range between 75 and 150 base pairs, which is most appropriate for FFPE tumour DNA and ctDNA, where tumour DNA fragments are of the same size range.

2.5.2 Custom Panel for NSCLC

A custom, 22 gene panel for non-small cell lung cancer, was designed for genes that are mutated at >1% frequency, that may also convey prognostic information, or they provide predictive biomarkers for response to targeted therapies. These genes are listed in Table 2-1. Specifically, this panel includes genes implicated in the pathogenesis of lung cancer. Genes encoding cell surface receptors, mutated at >1% in lung cancer, may provide a target for mono-clonal antibodies, or tyrosine kinase inhibitors, include *DDR2*, *MET*, *FGFR2*, *ERBB2*,

FLT3, KIT, PDGFRA and *EGFR*. Activated RAS/RAF/MEK/ERK and PIC3-K/AKT cellular signaling pathways are commonly seen in lung cancer. Many anticancer agents that target these aberrant cellular pathways are in development. Therefore, genes involved in these pathways, mutated at >1% in lung cancer were included in the custom gene panel. There are namely, *NRAS*, *KRAS*, *BRAF, MAP2K1 (MEK1)*, *PIK3CA, MTOR* and *AKT1*. Loss of NF1 (mutated >1% NSCLC) also activates the RAS pathway(119). DNA mis-match repair genes *BRCA1* and *BRCA2*, may convey sensitivity to PARP-inhibition therapy (120, 121). The tumour suppressor genes *TP53* and *STK11* are also mutated in >1% NSCLC. *CDKN2A* is mutated in >1% of NSCLC. This acts as a tumour suppressor by regulating the cell cycle. A new class of anti-cancer drugs, cyclin-dependent kinase inhibitors are being explored in various cancer types, eg breast cancer(122), and deletion/inactivation of *CDKN2A* may convey sensitivity to these agents(123).

The custom, whole exon, 22-gene panel was designed using the on-line Ion Ampliseq Designer tool v3.0 (http://www.ampliseq.com) specifying high specificity, for FFPE DNA, with amplicon range of 125 – 175 bp. Three primer pools were needed to achieve >90% NGS coverage of the intended target regions. Therefore 30ng of input DNA is required for the library preparation step for this gene panel.

Table 2-1 Genes chosen for the 22-gene custom targeted next generation sequencing for NSCLC samples, and the achievable gene sequence coverage with the current design (exonic regions). Gene mutational frequencies compiled from Kandoth, et al (2013)(80).

Genes Included in this Panel	Mutational Frequency in Adenocarcinoma of Lung (%)	Mutational Frequency in Squamous Cell Carcinoma of Lung (%)	Coverage with current design (%)
TP53	51.8	79.3	100
NRAS	1.8	0.6	100
PGDFRA	6.6	4.0	100
MTOR	7.5	4.6	99.36
DDR2	4.0	2.5 – 3.8	100
KIT	1.8	3.5	100
EGFR	11.4	2.9	99.93
MET	25	25	99.87
BRAF	6.6	4.6	98.59
FGFR2	3.1	2.3	99.81
BRCA2	5.7	5.8	98.68
AKT1	0.0	0.6	95.1
MAP2K1	1	1	93.11
BRCA1	5.2	3.5	99.86
CDKN2A	6.6	14.9	44.47
PTEN	2.2	8.1	96.77
KRAS	26.3	1.2	100
NF1	11.8	10.3	98.15
STK11	8.8	1.7	72.77
FLT3	4.0	4.0	97.7
ERBB2	2-4	2-4	91.97
PIK3CA	4.4	14.9	97.8

2.5.3 Cancer Hotspot Panel

In addition to the custom 22-gene panel, above, an off-the-shelf, Cancer Hotspot Panel (v2) (LifeTechnologies) was also used. This specifically covered 50 cancer mutation "hotspot" regions in contained within known oncogenes and tumour suppressor genes, and only required one primer pool per DNA sample (a total of 10ng of input DNA).

2.5.4 Input DNA requirements

At least 10% of the nucleated cells in the FFPE tumour tissue samples had to be tumour cells. There was a minimum requirement for a least 10ng of DNA for NGS using the Ion Ampliseq library preparation method.

2.5.5 Library Preparation

NGS library preparation was performed using the Ion AmpliSeq™ targeted panel method, using the manufacturer's Ion AmpliSeq™ Library Kit 2.0 User Guide (LifeTechnologies part of ThermoFisher Scientific, UK)(118). The initial polymerase chain reaction (PCR) amplification of the sample DNA was performed with the primer pool[a], with 20-25 cycles of the initial PCR, depending on the amount of input DNA. The flanking DNA sequences of the resulting amplicons were treated with the restriction endonuclease FuPa, to enable the addition of DNA 'barcode' sequences and adapter DNA sequences. After this, amplicons are bound to Agencourt® AMPure® XP magnetic beads (Beckman Coulter) and washed in ethanol, and eluted into a nested PCR for library enrichment. The enriched library is again bound to Agencourt® AMPure® XP magnetic beads (Beckman Coulter) and washed with ethanol. The resulting libraries were eluted in LowTE buffer, and concentrations were determined using Qubit (High Sensitivity kit) (Agilent Technologies), as above, with target concentrations 300-1500 ng/ml.

Libraries were further diluted in LowTE to obtain a final concentration of ~100 pM (equivalent to 15ng/ml for amplicons up to 225bp). Diluted, bar-coded libraries were then combined together, ready for template preparation and sequencing.

The following equation was used to determine the number of DNA samples that can be sequenced per microchip, as a function of the number primer pools per panel and the desired average sequence depth (per amplicon). To calculate an approximate minimum sequencing depth, then the average sequencing depth is divided by a factor of 5, as advised by the field application scientist (LifeTechnologies part of ThermoFisher Scientific).

[a] Either the 50-gene cancer hotspot panel primer pool, or the 3 primer pools that make up the custom 22-gene NSCLC panel.

$$\text{No. of Samples} = \frac{70{,}000{,}000^{b}}{\text{No. of primer pools per library} \times \text{No. of amplicons per primer pool} \times \text{average sequencing depth per amplicon}}$$

For this project, minimum sequencing depths were set at 100x for germline DNA, 1000x FFPE tumour DNA, and an average sequence depth was set at 25000x for ctDNA.

2.5.6 Planning the Run

Each sequencing run was planned on the Ion Torrent Server (https://www.thermofisher.com/uk/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-data-analysis-workflow/ion-torrent-suite-software.html), whereby each barcode used in the library was attributed to the research ID of the DNA sample.

2.5.7 Template Preparation

Template preparation and Ion Proton microchip loading was performed using the Ion Chef, according to manufacturer's user guide (Ion PI™ Hi-Q™ Chef Kit – User Guide (ThermoFisher Scientific), Catalogue No. A27198, Publication No. MAN0010967, Revision A.0 27 March 2015). Calibration of chip loading densities obtained, demonstrated that, in our hands, greater chip loading densities were obtained using a combined loading sample of 75 pM, rather than the 50 pM recommended in the user guide, without compromising loading quality.

[b] 70 Million is the number of individual sequencing wells on the IonProton Sequencing Chip.

2.5.8 Sequencing

Sequencing runs were performed on the Ion Proton machine, over 150 minutes in line with manufacturer's recommendations. Resulting data were then downloaded from the Ion Torrent Server.

2.5.9 NGS In-Silico Variant analysis

For each sample, resulting raw sequence data files (BAM and BAI) files, and data files containing sequence depth per amplicon per sample, were downloaded from the Ion Torrent Server.

The Ion Torrent Server also provided an Ion Torrent Variant Caller, for each sample, and resulting variant called data files (variant caller files (VCF)) were downloaded as Microsoft Excel spreadsheets.

2.6 Bioinformatics

2.6.1 VarScan

The open access VarScan version 2.4.0 (August 2015)(124), bioinformatics tool was used to analyse the raw sequencing data BAM files, to optimise low level sequence variants.

To facilitate somatic variant discovery, VarScan was also utilised to compare variants detected in genomic DNA, FFPE-embedded tumour DNA and ctDNA samples. Resulting variants were then annotated as being germline, somatic or loss of heterozygosity, based on comparisons between the 3 DNA populations.

2.6.2 Annovar

Novel genetic variants were identified within VCF files from the Ion Torrent Variant Caller and VarScan. The open access bioinformatics tool ANNOVAR(125) was used to characterize genetic sequence variants by comparison of genetic changes contained within databases of variants (COSMIC, dbSNP, 1000Genomes and Sift).

2.6.3 Integrative Genomics Viewer (IGV)

Genetic sequence changes contained within an amplicon was visually inspected, by loading BAM files into the software Integrative Genomics Viewer (IGV)(126).

2.6.4 Alamut

In order to classify genetic sequence changes into pathogenic mutations, single nucleotide polymorphisms (SNPs) or sequence changes of unknown significance, in silico-analysis was performed using Alamut® Visual software (Interactive Biosoftware). This software predicts the nature of the genetic change, reporting whether the change is intronic or exonic, mis-sennse, non-sense, and whether it effects gene splicing. For any given genetic sequence change, this software also searches the on-line databases including COSMIC and dbSNP, to determine whether the genetic mutation has been previously described.

2.6.5 Limit of Detection

Tru-Q Reference Standard DNA (HorizonDiscovery), is commercially available combined DNA from 10 engineered human cancer cell lines. This contains a panel of known pathogenic mutations, at known allelic frequencies determined by droplet digital PCR, for example Table 2-1. AKT1 wild type reference standard (HorizonDiscovery), is genomic DNA derived from human cell line MCF10A.

AKT1 wild-type human genomic DNA was 'spiked' with reference standard DNA (TruQ1), containing known pathogenic mutations at known frequencies (obtained from HorizonDiscovery), to mimic the presence of ctDNA mutations against a background of total circulating cell free DNA. NGS was then performed on these samples (in triplicate), using the 50 gene cancer hotspot panel.

Table 2-2 Panel of genetic mutations contained within reference standard DNA TruQ1 (HorizonDiscovery), commercially available at 50 ng/µl. Adapted from https://www.horizondiscovery.com/tru-q-1-5-tier-reference-standard-hd728 (accessed 4.2.2018)

Gene	Variant	Expected Allelic Frequency, %
BRAF	V600E	8.0%
BRAF	V600K	4.0%
EGFR	G719S	16.7%
EGFR	T790M	4.2%
FLT3	ΔI836	5.0%
IDH1	R132C	5.0%
JAK2	V617F	5.0%
KRAS	G12A	5.0%
KRAS	G12R	5.0%
KRAS	G13D	25.0%
MEK1	P124L	5.0%
NOTCH1	L1600P	4.8%
NRAS	Q61K	5.0%
PIK3CA	H1047R	30.0%

2.7 Droplet Digital Polymerase Chain Reaction (ddPCR)

2.7.1 ddPCR Reaction

Droplet Digital PCR reaction mixes were set up in a total volume of 25ul in ddPCR supermix containing template DNA at 50fg to 100ng plus, primers, and fluorescently labelled probes (FAM (carboxyfluorescein) for mutant probe and HEX (6-carboxy-2 ,4,4 ,5 ,7,7 -hexachlorofluorescein succinimidyl ester) for wild-type probes) at final concentrations of 450mM and 250nM, respectively (Table 2-2).

Table 2-3 Droplet Digital PCR Set-up Protocol.

Component	Volume Per Reaction (µl)	Final Concentration
2x ddPCR supermix for probes (no dUTP)	12.5	1x
20x target primers/probe (FAM)	1.25	450 nM primers/250 nM Probe
20x wild type primers/probe (HEX)	1.25	450 nM primers/250 nM Probe
DNA sample	γ	*
Water	10 - γ	
Final Volume (µl)	25	

*γ = 50 fg to 100 ng

22 µl of each reaction then dispersed into approximately 20,000 droplets, using the QX200™ Droplet Generator (BioRad), aiming for an average of 1 copy of DNA per droplet, given a range of input DNA (50fg – 100ng), based on a Poisson distribution. The reaction mix containing droplets are transferred to a clean 96-well Eppendorph plate for use in the T100 (BioRad) thermal cycler and later the QX200™ Droplet reader (BioRad).

Table 2-4 Droplet Digital PCR Cycling Conditions

Cycling Step	Temperature, °C	Time	Ramp Rate	No. of cycles
Enzyme Activation	95	10 min	~ 2°C/sec	1
Denaturation	94	30 sec	~ 2°C/sec	40
Annealing/Extension	*	1 min	~ 2°C/sec	
Enzyme Deactivation	98	10 min	~ 2°C/sec	1
Hold	4	infinite	~ 1°C/sec	1

Heated lid at 105°C and sample volume set to 45µl.
* Temperature gradient 65°C to 50°C, or optimal annealing temperature for a specific primer/probe set.

The water-oil emulsion based PCR was performed on a T100 (BioRad) thermal cycler. PCR was performed as per the protocol in Table 2-3. After PCR, the 96-well plates were briefly centrifuged, to bring the samples to the bottom of the wells, and placed on the QX200™ Droplet reader (BioRad). The template was set-up using the QuantaSoft software, and resulting data was collected and analysed using the same software.

2.8 Defining the Rain

Droplet digital technology aims to classify individual droplets as positive or negative based on the degree of fluorescence signal for each droplet, for particular fluorophore (either FAM (mutant probe) or HEX (wild-type probe). However, there are still some droplets with intermediate fluorescence values which fall between those that are positive and those that are negative, and these droplets are known as 'noise' or 'rain'. In order to overcome potential problems of data interpretation resulting from 'rain', minimum fluoresence amplitude thresholds are calculated for each droplet digital PCR data file, using a free, open access bioinformatics pipeline called 'definetherain' (http://www.definetherain.org.uk). After defining fluorescence apliture thresholds, Quantasoft was used to determine number and ratio of positive and negative droplets per PCR reaction.

2.9 Annealing Temperature Optimisation

For each ddPCR primer/probe set, and 10ng of a reference standard DNA[c], a temperature gradient was performed during the annealing phase, to determine the optimum annealing temperature, according to the PCR protocol in table 2-3. Fluorescent cut off values were determined for each temperature point using 'definetherain', and the no-template controls (NTCs) were analysed using the same cut offs, to ensure that no false positive droplets were detected. For all primer and probe sets, the optimum annealing temperatures were between 53.0˚C and 55.7˚C.

[c] Tru-Q tier 1 reference standard DNA (Horizon Discovery) contains a panel of known mutations at know alternate allele frequencies.

2.10 Other Molecular DNA Techniques

2.11 Pyrosequencing

An initial PCR mix is made with 12.5µl of 10x MegaMix Gold[d], 0.5 µl (at 10pmol/µl) F&R biotinylated primers for pyrosequencing (Table 2-4), 10µl water and 20ng of DNA, in a final reaction volume of 25µl.

[d] Proprietary reagent (Cambio Ltd. Cambridge, UK) provides a final reaction concentration of 2.5 mM $MgCl_2$, 1 unit of Taq polymerase, high purity dNTPs, reaction buffer and enzyme stabilizer.

Table 2-5 Pyrosequencing PCR Primers. Btn = Biotin

Gene	Exon	Primer Name	Primer Sequence
EGFR	18	EGFR_18_F	ATGGTGAGGGCTGAGGTGAC
		EGFR_18_Btn_R	[Btn]TATACACCGTGCCGAACGC
	T790M	EGFR_T790M_F	GTCTTTGTGTTCCCGGACAT
		EGFR_T790M_Btn_R	[Btn]GCATCTGCCTCACCTCCAC
	20	EGFR_20v2_F	CGAAGCCACACTGACGTG
		EGFR_20v2_Btn_R	[Btn]GGTGGAGGTGAGGCAGATG
	21	EGFR_21_F	GAAAACACCGCAGCATGTCAA
		EGFR_21_Btn_R	[Btn]CCTCCTTCTGCATGGTATTCTTTC
KRAS	2 (codons 12&13)	KRAS_c12&13 _F	GACTGAATATAAACTTGTGGTAGTTGG
		KRAS_c12&13_Btn_R	[Btn]TTGGATCATATTCGTCCACAA
	3 (codon 61)	KRAS_c61 _F	CAGGAAGCAAGTAGTAATTGATGG
		KRAS_61_Btn_R	[Btn]AAGAAAGCCCTCCCCAGTC
	4 (codon 146)	KRAS_c146_F	AGGCTCAGGACTTAGCAAGAAGTT
		KRAS_c146_Btn_R	[Btn]GATTAAGAAGCAATGCCCTCTC
PIK3CA	9v2	PIK3CA_9v2_F	GCACTTACYTGTGACTCCATAGAA
		PIK3CA_9v2_R	[Btn]GAAAATGACRAAGAACAGCTCAAA
	9v4	PIK3CA_9v2_F	GCACTTACYTGTGACTCCATAGAA
		PIK3CA_9v2_R	[Btn]GAAAATGACRAAGAACAGCTCAAA
	20	PIK3CA_20B_F	GTTCAATGCATGCTGTTTAATTGT
		PIK3CA_20B_R	[Btn]CGAAAGACYCTAGCCTTAGATAAA

The PCR cycle conditions were then as follows:

Table 2-6 Initial PCR for *EGFR/KRAS/PIK3CA* pyrosequencing of FFPE tumour DNA

Cycle Phase	Temp. (°C)	Time	Cycles
Initial Denaturation	95	10 min	1
Denaturation	95	30 Seconds	35
Annealing	60	30 Seconds	
Synthesis	72	30 Seconds	
Final Synthesis	72	10 min	1
Hold	4	∞	

40µl of primer/annealing buffer mix is added to each well in a separate clean low-well PyroMark Q96 plate. This consists of a final concentration of 0.4µM sequencing primer (Table 2-6) in 1x PyroMark Annealing Buffer (Qiagen).

Table 2-7 Pyrosequencing Primers

Gene	Exon	Primer Name	Primer Sequence
EGFR	18	EGFR_18_pyro	GAATTCAAAAAGATCAAAGT
	T790M	EGFR_T790M_pyro	CGAAGGGCATGAGC
	20	EGFR_20Av2_pyro	CCCTCCCTCCAGGAA
		EGFR_20Bv2_pyro	CAGGAAGCCTACGTGA
	21	EGFR_21A_pyro	AAGATCACAGATTTTGG
		EGFR_21B_pyro	ATTTTGGGCTGGCCAAAC
KRAS	2	KRAS_c12&13_pyro	CTTGTGGTAGTTGGAG
	2*	KRAS12&13_seqA3	TGTGGTAGTTGGAGCTG
	3	K-ras_c61_pyro	GGATATTCTCGACACAGC
	4	KRAS_c146v2_pyro	GGAATTCCTTTTATTGAAAC
PIK3CA	9v2	PIK3CA_9v2_pyro	CCTGTGACTCCATAGAAAA
	9v4	PIK3CA_9v4_pyro	CCATAGAAAATCTTTCTCC
	20	PIK3CA_20B_pyro	TTTGTTGTCCAGCCA

PCR products from the initial PCR were immobilized onto streptavidin Sepharose™ beads, by adding 55µl buffer/beads master mix (2µl streptavidin beads, 40µl 1x PyroMark Binding Buffer, 13µl sterile water) to each PCR product, and mixed for 10 minutes at room temperature. Using a vacuum pump manifold (Qiagen), the bead-bound DNA is washed in 70% ethanol for 5 seconds, followed by the denaturation solution (0.2M NaOH) for 5 seconds, and a further wash solution (PyroMark Wash Buffer (Qiagen)) for 10 seconds. The vacuum is switched off to release the cleaned/denatured PCR products into the low-well PyroMark Q96 plate containing sequencing primer and annealing buffer.

To anneal the sequencing primer, this plate was placed onto a thermocycler and heated to 80 °C for 2 mins, removed, and allowed to cool to room temperature for 10 minutes.

The low-well PyroMark Q96 plate was loaded into a PyroMark Q96 machine along with a reagent cartridge containing loaded with dNTPs and the key enzymes; DNA polymerase, ATP sulfurase, firefly luciferase and apyrase. The sequencing is run using the PyroMark Q96 v2.5.8 software (Qiagen). Raw sequencing data files were visualized as pyrograms using the same software.

2.12 EGFR Analysis – Fragment Sizing

Fluorescently labeled PCR primers (Table 2-7) amplify exon 19 or exon 20 of *EGFR*. PCR product fragment sizes are analysed with the ABI 3730 genetic analyser. A PCR product larger or smaller than the normal allele size indicates the presence of an insertion or deletion respectively. Samples are analysed alongside a negative control (H2O), a normal control for each exon and positive control for each exon, namely an exon 19 deletion and an exon 20 insertion.

The initial PCR reaction mix contains 5µl 10x Megamix Gold, 0.5µl (final concentration 10pm/µl) fluorescently labelled forward and reverse primers (Table 2-7), 3.5µl dH_2O and 10ng DNA, in a final reaction volume of 10µl. The PCR thermo-cycle conditions are given in Table 2-8.

Table 2-8 Primer Sequences for the fragment sizing assay [EGFR exon 19 and exon 20]

Exon	Primer Name	Primer Sequence
19	EGFR_19v2_FAM_F	[6FAM]ACTCTGGATCCCAGAAGGTGA
	EGFR_19v2_GS_R	ACACAGCAAAGCAGAAACTCA
20	EGFR_20v4_HEX_F	[HEX]CGAAGCCACACTGACGTG
	EGFR_20v4_GS_R	GGTGGAGGTGAGGCAGATG

Table 2-9 Thermo-cycling condition for EGFR fragment length analysis.

Cycle Phase	Temp. (°C)	Time	Cycles
Initial Denaturation	95	5 min	1
Denaturation	95	30 s	27
Annealing	60	30 s	
Synthesis	72	45 s	
Final Synthesis	72	5 min	1
Hold	4	∞	

2.12.1 COLD-PCR

COLD-PCR is a nested PCR. The initial PCR with external primers is set-up according to Table 2-9, with cycling conditions Table 2-10. The secondary PCR is set-up according to Table 2-12 or 2-13 depending on the downstream application, with cycle conditions in Table 2-11.

Table 2-10 Primary COLD-PCR set-up

Reagent	Final Conc.	1x (µl)
Megamix Gold	1x	12.5
F + R External Primer Mix*	10 pmol/µl	1.0
dH_2O		9.5
DNA (10ng/µl)		2.0
Total		25.0

* See Tables 2-10 for primer sequences.

Table 2-11 External Primer Sequences for COLD-PCR

Gene	Exon	Primer Name	Primer Sequence
EGFR	18	EGFR_18_eF	GGCACTGCTTTCCAGCAT
		EGFR_18_eR	CCCCACCAGACCATGAGA
	19	EGFR_19v2_eF	GCTGGTAACATCCACCCAGA
		EGFR_19v2_eR	GGCCAGTGCTGTCTCTAAGG
	20*	EGFR_20_eF	CATTCATGCGTCTTCACCTG
		EGFR_20v2_eR	CCGTATCTCCCTTCCCTGAT
	21	EGFR_21_eF	GCAGAGCTTCTTCCCATGAT
		EGFR_21_eR	AGGAAAATGCTGGCTGACCT
KRAS	2	KRAS_12&13_eF	AAAAGGTACTGGTGGAGTATTTGA
		KRAS_12&13_eR	TCATGAAAATGGTCAGAGAAACC
	3	KRAS_61_eF	TTTTTGAAGTAAAAGGTGCACTG
		KRAS_61_eR	TTTAAACCCACCTATAATGGTGAA
PIK3CA	9	PI3KCA_9_eF	TGAGATCAGCCAAATTCAGTT
		PI3KCA_9_eR	GGGAAAAATATGACAAAGAAAGC
	20	PI3KCA_20_eF	GGAATCCAGAGTGAGCTTTCA
		PI3KCA_20_eR	TTCTCAATGATGCTTGGCTCT

* Exon 20 external primers are used as the primary primers for T790M, 20A, 20B pyro and exon 20 fragment analysis assays.

Table 2-12 Primary COLD-PCR Cycle Conditions

Cycle Phase	Temp. (°C)	Time	Cycles
Initial Denaturation	95	5 mins	1
Denaturation	95	30 s	20
Annealing	60*	30 s	
Synthesis	72	45 s	
Final Synthesis	72	5 mins	1

*For PIK3CA Exon 9 use 61 °C, and PIK3CA Exon 20 use 63 °C.

Secondary COLD-PCR Reaction (for use with subsequent pyrosequencing)

Table 2-13 Secondary COLD-PCR Reaction (replaces initial PCR reaction for Pyrosequencing)

Reagent	Final Conc.	1x (µl)
Megamix Gold	1x	12.5
F + R Primer Mix*	10 pmol/µl	0.5
dH_2O		11.0
PCR Products from Primary COLD-PCR	Neat	1.0
Total		25.0

*Primer sequences as per Table 2-13, unless for EGFR Pyrosequencing (use primers in Table 2-6) or EGFR fragment length analysis (use primers in Table 2-7)

Table 2-14 Pyrosequencing COLD-PCR Internal primers

Exon	Primer Name	Primer Sequence
18	EGFR_18_F	ATGGTGAGGGCTGAGGTGAC
	EGFR_18_Btn_R	[Btn]TCCCTCCAGGAAGCCTACG
T790M	EGFR_T790M_F	TTGCGATCTGCACACACC
	EGFR_T790M_Btn_R	[Btn]GCATCTGCCTCACCTCCAC
20	EGFR_20v2_F	TGCGAAGCCACACTGACG
	EGFR_20v2_Btn_R	[Btn] GCACACGTGGGGGTTGTC
21	EGFR_21_F	GAAAACACCGCAGCATGTCAA
	EGFR_21_Btn_R	[Btn]CCTCCTTCTGCATGGTATTCTTTC

Table 2-15 Fragment sizing COLD-PCR primers [EGFR exon 19 and exon 20]

Exon	Primer Name	Primer Sequence
19	EGFR_19_FAM_F	[FAM]CGTCTTCCTTCTCTCTCTGTCA
	EGFR_19_GS_R	CCACACAGCAAAGCAGAAAC
20	EGFR_20_HEX_F	[HEX]CTCCAGGAAGCCTACGTGAT
	EGFR_20_GS_R	GTCTTTGTGTTCCCGGACAT

Table 2-16 Secondary COLD-PCR Cycle Conditions

Cycle Phase	Temp. (°C)	Time	Cycles
Initial Denaturation	95	2 mins	1
Denaturation	95	15 s	10
Annealing	60	30 s	
Synthesis	72	1 min	
Denaturation	95	15 s	25
Hybridisation	70	8 mins	
Critical Denaturation	(see Table 2-16)	3 s	
Annealing	60	30 s	
Synthesis	72	1 min	

Table 2-17 Critical Denaturation Temperatures for Secondary COLD-PCR reaction

Gene	Assay	Pyrosequencing Critical Temp. (°C)	Fragment Size Critical Temp. (°C)
EGFR	Exon 18	85.4	N/A
	Exon 19	N/A	82.0
	Exon 20	85.4	88.3
	Exon 21	81.4	N/A
	T790M	88.3	N/A
KRAS	Codon 12/13	80.6	N/A
	Codon 61	81.3	N/A
PIK3CA	Exon 9	78.3	N/A
	Exon 20	79.7	N/A

2.12.2 Sanger Sequencing

DNA extracted from FFPE tumour tissue was first amplified using a PCR and M13 tagged primer sequences as per table 2-17. This initial PCR has a final reaction volume of 25 µl, and a final concentration of 1x Megamix Gold, 25 pmol/µl Forward and Reverse primer mix and 20ng of DNA. In the thermocycler, samples are preheated to 95˚C for 5 minutes, followed by 30 cycles of denaturation (95˚C for 1 minute), annealing (60˚C for 1 minute) and extension (72˚C for 1 minute). There is a final extension at 72˚C for 7 minutes, then hold at 4˚C.

Table 2-18 Dideoxy-sequencing primer sequences. Underlined sequences = M13 tag sequences.

Gene	Exon	Primer Name	Primer Sequence
EGFR	18	EGFR_18v2_MF	tgtaaaacgacggccagtGGCACTGCTTTCCAGCAT
		EGFR_18v2_MR	caggaaacagctatgaccCCCCACCAGACCATGAGA
	19	EGFR_19v2_MR	caggaaacagctatgaccCAGGGTCTAGAGCAGAGCAG
		EGFR_19v2_MF	tgtaaaacgacggccagtAGCATGTGGCACCATCTC
	20	EGFR_20v2_MF	tgtaaaacgacggccagtCATTCATGCGTCTTCACCTG
		EGFR_20v2_MR	caggaaacagctatgaccCATATCCCCATGGCAAACTC
	21	EGFR_21v2_MF	tgtaaaacgacggccagtATTCGGATGCAGAGCTTCTT
		EGFR_21v2_MR	caggaaacagctatgaccTGGCTCACACTACCAGGAGA
KRAS	2 (codon 12&13)	KRAS_c12&13_MF	tgtaaaacgacggccagtCTTAAGCGTCGATGGAGGAG
		KRAS_c12&13_MR	caggaaacagctatgaccAGAATGGTCCTGCACCAGTAA
	3 (codon 61)	KRAS_61v2_MF	tgtaaaacgacggccagtTGTTTCTCCCTTCTCAGGATTCCTA
		KRAS_61v2_MR	caggaaacagctatgaccAGAAAGCCCTCCCCAGTCC
PIK3CA	1A	PIK3CA_1A_MF	tgtaaaacgacggccagtAGCCTAATTTAGAGCAACAGTCTAGAT
		PIK3CA_1A_MR	caggaaacagctatgaccAGGGGGTATTTTCTTGCTTCTTTAAATA
	1B	PIK3CA_1B_MF	tgtaaaacgacggccagtTCCACGACCATCATCAGGTGAACT
		PIK3CA_1B_MR	caggaaacagctatgaccAGATTACGAAGGTATTGGTTTAGACAGAAA
	9	PIK3CA_9_MF	tgtaaaacgacggccagtAGAGACAATGAATTAAGGGAAAATGACAA
		PIK3CA_9_MR	caggaaacagctatgaccAATCTCCATTTTAGCACTTACCTGTGAC
	20	PIK3CA_20_MF	tgtaaaacgacggccagtTGGTAAGAGAAGTGAGAGAGGAATGCT
		PIK3CA_20_MR	caggaaacagctatgaccACAGTGCAGTGTGGAATCCAGAGTG

The PCR products were cleaned to remove excess dNTPs and primers. To do this AMPure beads[e] bound to PCR products were washed in 70% ethanol. DNA was then eluted into sterile water, and transferred into a clean PCR plate.

Sanger sequencing of the PCR-derived fragments was performed using the BigDye Terminator v1.1 system (ThermoFisher Scientific), setting up forward

[e] AMPure (Agencourt) Beads (Beckman)

and reverse reactions separately. The sequencing reaction is set-up using final concentrations of 1x BigDye v1.1[f], 100 pmol/µl Forward OR Reverse M13 primer (Eurogentec), 1X Sequencing Buffer[g] and neat, cleaned PCR product.

The samples are placed in the thermocycler at 94˚C for 2 minutes, followed by 25 cycles of denaturation (94˚C for 10 seconds), annealing (50˚C for 5 seconds) and elongation (60˚C for 4 minutes). A subsequent clean-up of sequencing products was performed to remove excess dye terminators using the Agencourt's paramagnetic bead technology - CleanSEQ[h] method, and washed in a higher concentration of ethanol (85%), to prevent loss of the smaller DNA fragments. The final cleaned DNA products are eluted into 0.04 mM EDTA solution.

Capillary electrophoresis, using POP-7 polymer (Applied Bios. (LifeTechnologies)), in a 3730 DNA analyser machine (ThermoFisher Scientific), is performed on the cleaned sequencing products to generate electropherograms, according to manufacturers protocol. Raw sequence data in the form of electropherograms are then visually inspected using Mutation Surveyor® software v4.0.8, a DNA variant analysis software from SoftGenetics®.

[f] Commercially available proprietary reaction mix (ThermoFisher Scientific) contains DNA polymerase, dNTPs, fluorescently labelled ddNTPs.
[g] Proprietary Buffer - 5X Sequencing Buffer (For use with BigDye® Terminator v1.1 & v1.3) (Applied Bios, LifeTechnologies)
[h] CleanSEQ (Agencourt) (Beckman)

3 Detection of ctDNA in Lung Cancer Patients using NGS and ddPCR

3.1 Background

The amount of total cell free DNA available from a blood sample is generally very low, for example, previously reported as 24.3 ng/ml in patients with lung cancer(127). Circulating tumour DNA (ctDNA) represents only a fraction of total circulating cell free DNA. The fraction of ctDNA present in the circulation is dependent on tumour type, tumour stage and whether the patient has received any anti-cancer therapies.

It is essential to clearly establish the feasibility of obtaining, and analysing ctDNA in patients with advanced lung cancer, so that conclusions relating to the clinical context and clinical impact can subsequently be made. To do this, genetic analysis techniques must be able to detect a low frequency ctDNA mutation "signal", against background wild-type genomic cell free DNA "noise". The ratio of mutant DNA to total 'wild-type' DNA is known as alternate, or variant allele frequency.

It is important to explore the limits of detection in ctDNA analysis, as mutations usually occur at low alternate allele frequency, for example between 0.1 – 5%(112). The clinical consequences of false negative results could be significant. There is currently a lack of data in the literature regarding lower limits of mutation detection in ctDNA in terms of the clinical consequences for lung cancer.

Next generation sequencing (NGS), and droplet digital PCR (ddPCR) are two major technologies available for DNA sequence analysis. They can be used for DNA derived from tumour tissue and ctDNA. Droplet digital PCR is widely considered to be a more sensitive and specific approach to analysing mutations contained within ctDNA. However, unlike NGS, which can interrogate multiple mutations simultaneously, ddPCR can only detect the presence of very specific mutations. However, the amplification phase of NGS can generate low-level sequence artifacts. When applied to ctDNA

analysis, it is therefore important to be able to differentiate between low level 'true positive' mutations, and low level 'false positive' mutations.

Different software bio-informatics tools exist for the analysis of NGS genetic sequence data, and can detect different rare alleles, with different degrees of sensitivity and specificity. Technical exploration is needed into better defining the lower limits of detection of the Ion Torrent NGS platform, used in this thesis. Specifically, NGS data is compared, using a customised open access software analysis and annotation tool (VarScan(128) with Annovar(125)), with the software tool available with the Ion Torrent NGS platform, namely the Ion Torrent Variant Caller (ITVC).

We have used NGS and ddPCR technologies to investigate mutations in ctDNA from lung cancer patients, to see whether we should use either, or both, in subsequent research in this thesis.

3.2 Chapter Aims

This chapter aims to define some of the biological and technical factors that influence the ability to detect low-level mutations using NGS and ddPCR. Specific chapter aims are to:

- Using DNA derived from cancer cell lines:
    - Assess the limit of mutation detection of NGS, using the Ion Proton NGS platform, and comparing VarScan(128) with Annovar(125) and ITVC.
    - Assess the level of background sequence changes generated by Ion Torrent NGS platform.
    - Assess the limit of mutation detection of ddPCR, as a function of the amount of input DNA.
- Using DNA derived from patient samples, and experience gained from cell line DNA:
    - Assess the influence of clinical features of lung cancer, including stage and histological subtype of disease, on total cell free DNA levels.

3.3 Chapter Results

3.3.1 Limit of mutation detection using NGS

Comparative sequence data, analysed using two different bioinformatics software, revealed a difference in the ability to detect all of the expected mutations, and both software tools had different limits of detection (Figure 3-1). The VarScan(128) with Annovar(125) bioinformatics software tools were able to detect all of the expected mutations. ITVC invariably failed to detect the *FLT3* p.836_837delI mutation, regardless of alternate allele frequency of the mutation (Figure 3-1). This *FLT3* mutation is important in acute myeloid leukaemia(129), so the inability to detect this mutation may have important implications for certain clinical applications in this disease(130).

VarScan(128) with Annovar(125) was able to consistently detect mutations at lower alternate allele frequency (consistently as low as 0.5% alternate allele frequency), ITVC was only able to detect mutations confidently, at a minimum of 5% mutant allele frequency - a 10 fold higher limit of detection than VarScan with Annovar. The majority of mutations in ctDNA usually exist in the 0.1 – 5% alternate allele frequency range(112). Of the two software tools compared here, VarScan with Annovar is preferred to ITVC for clinical ctDNA mutations detection, since it offers a lower limit of mutation detection. VarScan and Annovar has subsequently been the preferred choice for all analysis of other NGS data generated in this thesis.

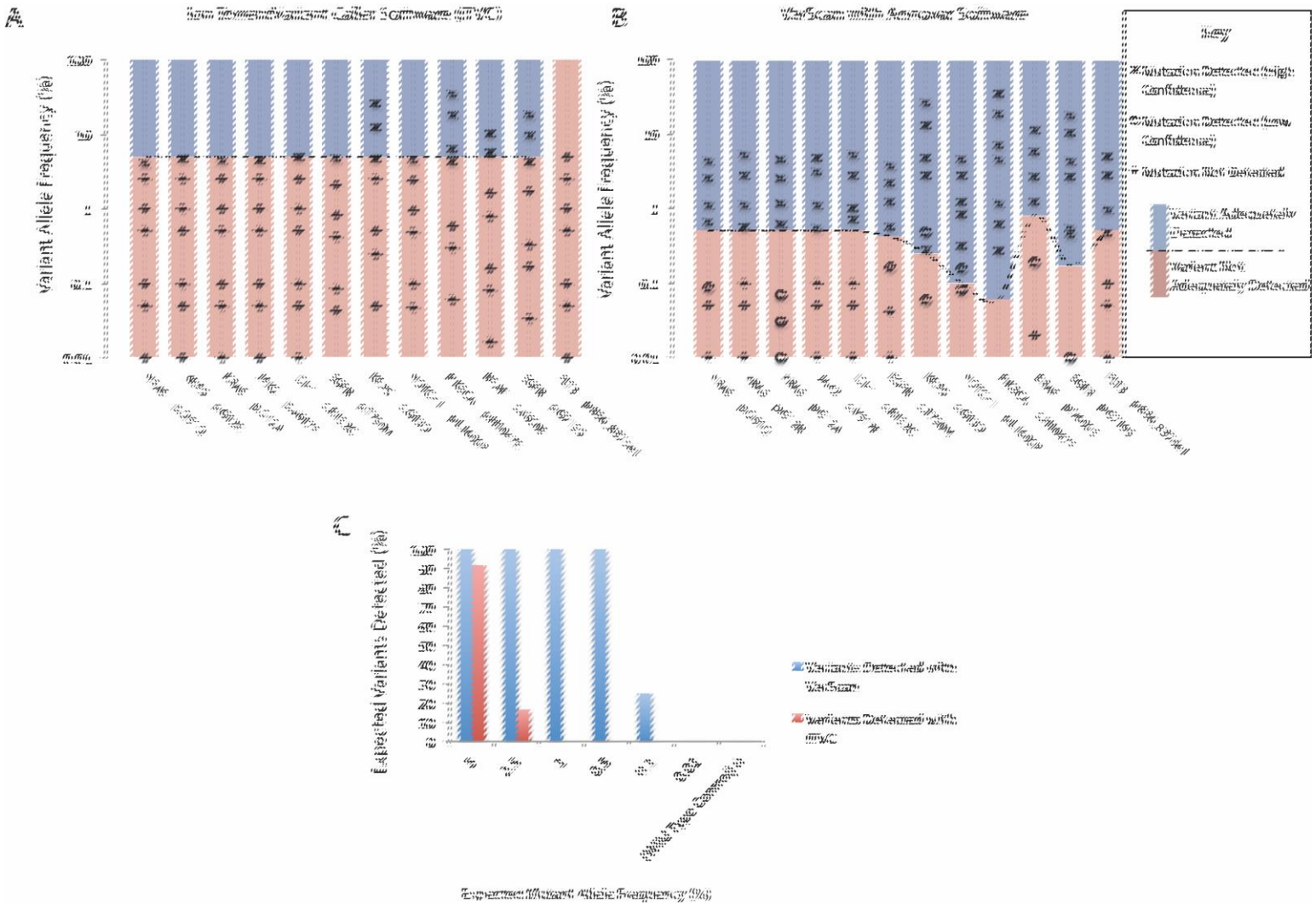

Figure 3-3 Limits of detection of spiked, mutant DNA against a background of wild-type DNA, using NGS and a 50-gene hotspot panel and 2 different software analysis tools (ITVC and VarScan), prior to application to ctDNA analysis.

3.4 NGS Background levels of noise

50 gene cancer hotspot panel NGS analysis of all of the DNA samples in the previous section revealed the presence of many other genetic sequence changes, in addition to the mutations known to be present in the DNA samples. The total number of sequence changes differed depending on the software used to detect them (Figure 3-2, panel C). The total number of variants detected by ITVC ranged from 11 to 42. This reflects that this software only identifies sequence changes that are present with an alternate allele frequency of >5%. The total number of variants detected using VarScan with Annovar were filtered so that only variants with a COSMIC ID were included. Using this method, the average number of variants detected ranged from 527 to 645 for SNPs (approximately 100 fold greater than ITVC), and 66 to 82 for Insertions/Deletions (approximately 10 fold greater). Data generated from VarScan with Annovar revealed that the bulk (95%) of the genetic variations, occur at the level of 0.1% alternate allele frequency or less. Less than 5% of the total variants occur at an alternate allele frequency of greater than 1% (Figure 3-2, panels A & B). It is thought that the vast majority of these 'low level' sequence changes are due to 'sequence artefacts' as a result of amplification errors during the NGS process. In terms of using NGS for ctDNA analysis, care must be taken when deciding whether a very low level genetic sequence variation (for example, an alternate allele frequency of less than 0.1%), represents a true somatic mutation contained within ctDNA, or whether it is sequencing artefact. In keeping with the literature, it is extremely difficult to differentiate between 'true-positive' changes and 'false-positive' changes, though there have been attempts to increase confidence in identifying 'true-positive' mutations by setting alternate allele frequency 'thresholds'.

To summarise, VarScan with Annovar appears satisfactory in detecting the presence of a known genetic mutation, with a 10-fold greater sensitivity of detection, compared to ITVC. This appears to be at the expense of a significantly increased number of total variants detected, many of which are 'sequence artifact'. Therefore, for the purposes of this thesis, to increase the changes of detecting clinically relevant genetic mutations, the NGS data is filtered according to changes occurring in exons, that are likely to result in a non-synonymous protein change, and that may have already been previously

catalogued in a database such as COSMIC. When a suspected 'true-postive' low-level mutation is detected by NGS, a second genetic analysis technique will be used to confirm the presence of that mutation, mostly in the form of ddPCR.

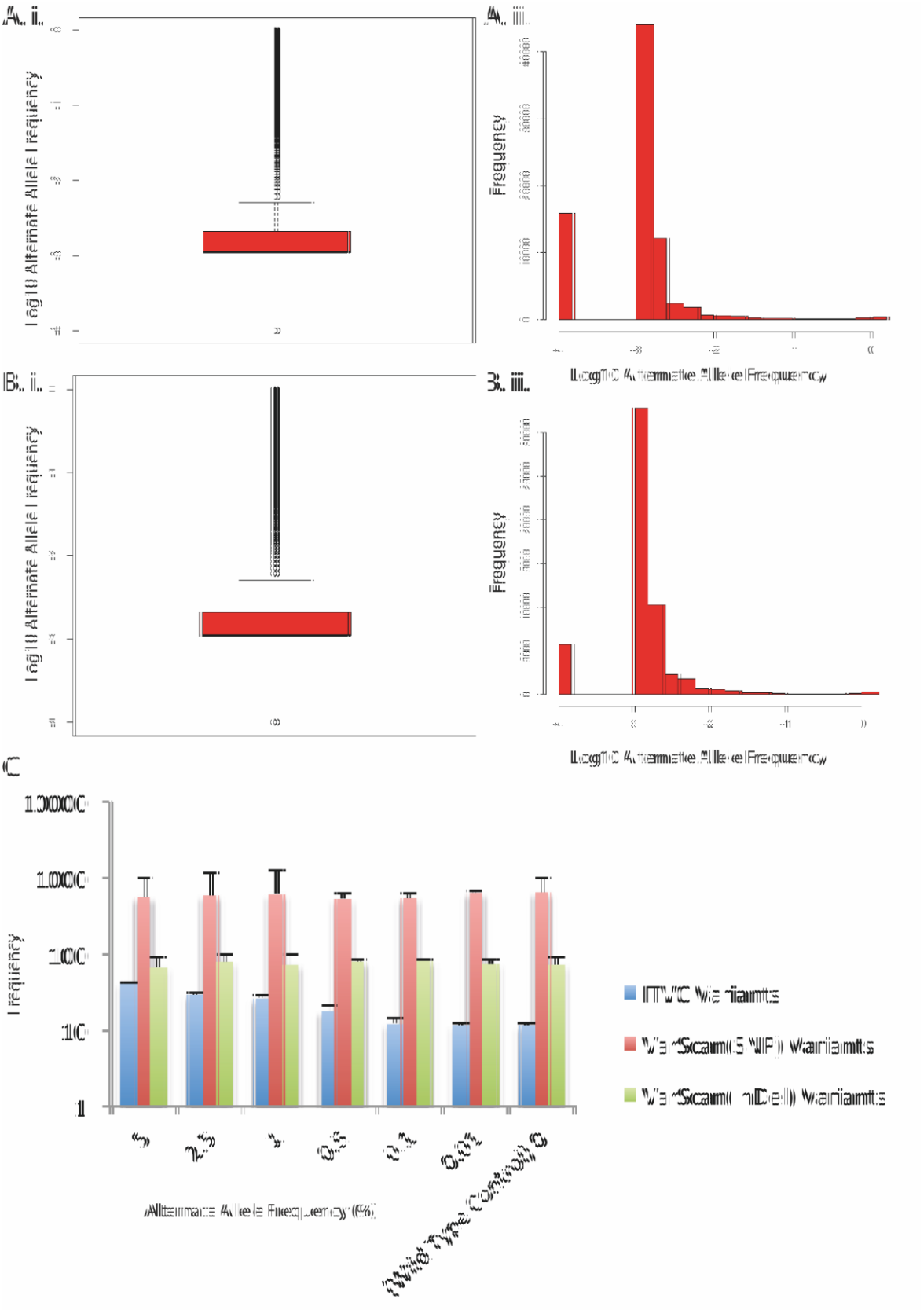

Figure 3-1 Total numbers of genetic sequence changes detected by NGS. Box-and-whisker plots (i) and histograms (ii) showing alternate allele frequencies for total variants, for samples containing known mutations at an alternate allele frequency of 0.01% (A) and 0.5% (B). The average total number of variants detected, at decreasing allele frequencies, with each software, is shown (C). Error Bars shown represent 95% CI.

3.5 Limit of Detection for Droplet Digital PCR

Droplet digital PCR is able to more confidently detect a specific mutation, against a background of wild type DNA, compared with NGS. Mutations occurring as low as 0.01% alternate allele frequency can be detected. Increasing, or decreasing the amount of input DNA in ddPCR, increases or decreases respectively, the confidence in detecting a 'true-positive' result, compared with a 'false-positive' result (Figure 3-3).

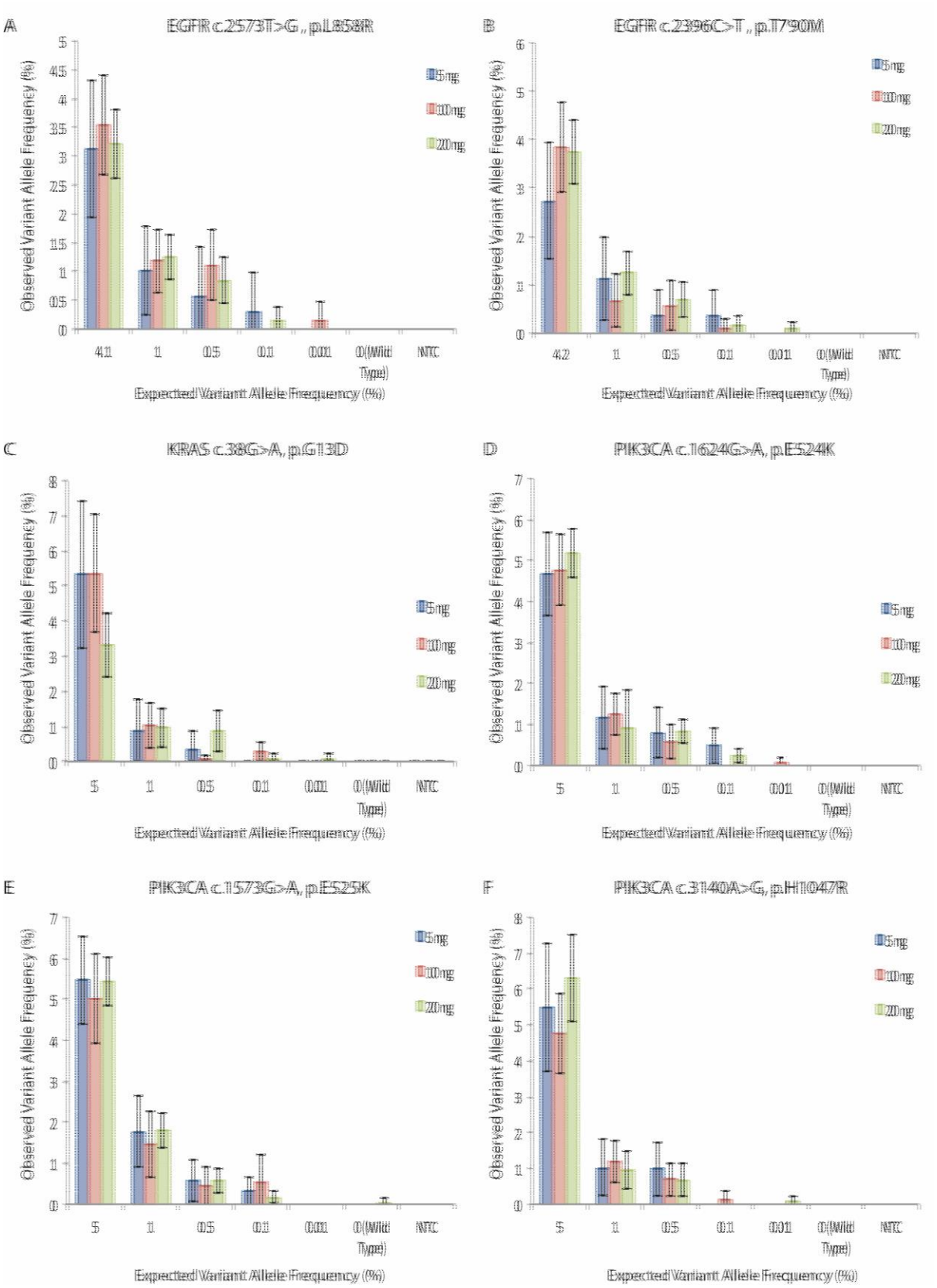

Figure 3-2 Limit of detection with droplet digital PCR for 6 different tumour-associated mutations. NTC = No Template Control. 95% Poisson Error Bars shown.

For each of the mutations tested, the resulting minimum level of detection is tabulated (Table 3-1), as a function of the amount of input DNA, reflective of the amount of circulating cell free DNA that is available from a blood test.

Table 3-1 Minimum level of detection (alternate allele frequency %) for each mutation, using ddPCR

		Amount of Input DNA		
		5ng	10ng	20ng
Mutation	*EGFR* c.2573T>G p.L858R	1%	0.5%	0.5%
	EGFR c.2369C>T p.T790M	1%	1%	0.5%
	KRAS c.38G>A p.G13D	1%	1%	0.5%
	PIK3CA c.1573G>A p.E525K	0.5%	0.5%	0.1%
	PIK3CA c.1624G>A p.E524K	0.5%	0.5%	0.1%
	PIK3CA c.3140A>G p.H1047R	0.5%	0.5%	0.5%

Comparing data from figure 3-2 with figure 3-3, it is possible to detect mutations at lower alternate allele frequencies with ddPCR (0.1% alternate allele frequency for 2 of 6 tested mutations), compared with NGS using VarScan with Annovar (0.5% alternate allele frequency). However, these data confirm that knowledge of the amount of input DNA available for ddPCR, crucially relevant when dealing with ctDNA, will impact on the sensitivity and specificity of the test, especially when mutations occur at very low alternate allele frequencies. It must also be remembered that ddPCR is very specific for any given genetic mutation, which is satisfactory if analysing samples for a specific mutation (e.g. the EGFR TKi resistance mutation - *EGFR* c.2369C>T p.T790M) but much less satisfactory when screening for a panel of potential mutations, as the amount ctDNA is a limiting factor.

3.6 Blood sample collection and cell free DNA extraction

In this research project, there were a total of 110 research blood samples taken from recruited patients, for total cell free DNA quantitation, and subsequent molecular analysis for mutations contained within ctDNA. There are no published data regarding factors that influence the total concentration of cell free DNA in blood samples from lung cancer patients. Therefore total cell free DNA (cfDNA) concentration was compared to tumour stage (III versus IV), treatment status (pre-treatment or on-treatment), and duration of blood storage, for both SCLC and NSCLC cancer sub-types (Figures 3-4 and 3-5). Concentrations of total extracted cell free DNA (ng/ml) was quantified using the Qubit (Agilent technologies). Non-parametric comparisons of median values were performed using Mann-Whitney-U test.

Statistically significant differences in total cell free DNA concentrations were seen in only 2 situations – a comparison of Stage IV SCLC blood samples with Stage III SCLC and with Stage IV adenocarcinoma blood samples. In both cases, this may reflect the differing biology of SCLC and NSCLC, with SCLC known to be a more aggressive disease, with a higher disease burden.

More specifically, there were 43 blood samples taken from patients with SCLC. A statistically significantly higher mean concentration of total cell free DNA was seen in patients with stage IV SCLC (n=32, median 20.8 ng/ml, interquartile range 12.8 – 37.3 ng/ml)), compared to patients with stage III SCLC (n=11, median 11.40 ng/ml, interquartile range 7.56 – 15.34 ng/ml) ($p<0.05$). This has not been previously detected in reports in the literature. This may be a reflection of a higher disease volume in patients with stage IV SCLC, compared to stage III.

There was a statistically significantly higher mean concentration of total circulating cell free DNA seen, in patients with stage IV SCLC compared to stage IV adenocarcinoma of the lung (n=40, median 8.45 ng/ml (interquartile range 5.96 – 10.13) ($p < 0.05$) (Figure 5). This may reflect a difference in the biology of the two cancer sub-types, or differences in disease volume between the two

cancer types, stage-for-stage. For example, there may be more cell free DNA in patients with stage IV SCLC, potentially due to a higher disease volume, and/or because it is a more aggressive form of lung cancer associated with a poorer prognosis.

There were a total of 47 blood samples from *EGFR*$^{mut+ve}$ lung adenocarcinoma patients. There was no apparent difference observed between the concentrations of total circulating cell free DNA between the blood samples from patients with stage III disease (n=9, median 7.58 ng/ml, (interquartile range 7.02 – 11.88 ng/ml) and stage IV disease (n=38, median 9.48 ng/ml, (interquartile range 7.47 – 13.82 ng/ml).

Equally, no apparent difference was observed between mean concentrations of total cell free DNA between blood samples from *EGFR*$^{mut+ve}$ lung adenocarcinoma patients that were EGFR TKi treatment naïve (n=12, median 8.09 ng/ml, (interquartile range 7.46 – 11.68 ng/ml) or receiving EGFR TKi therapy (n=35, median 8.02 ng/ml, (interquartile range 7.09 – 12.31 ng/ml).

There is no apparent effect of duration of storage (0 - 4 days) of blood samples collected in CellSave blood collection tubes, on the concentration of total cell free DNA (Figure 4, panels A iii. & B iii). This is consistent with, and extends previous findings for SCLC by Rothwell et al(110). These data are meaningful when considering use of ctDNA analyses as biomarkers in routine clinical practice, as they suggest that CellSave blood collection tubes represent a feasible method for blood collection from lung cancer patients, without reducing the relative mutational abundance of ctDNA, compared to 'wild-type' genomic cell free DNA. This is especially important when there may be a delay in getting the sample to the lab, for example the time taken to transport the blood sample from the clinic.

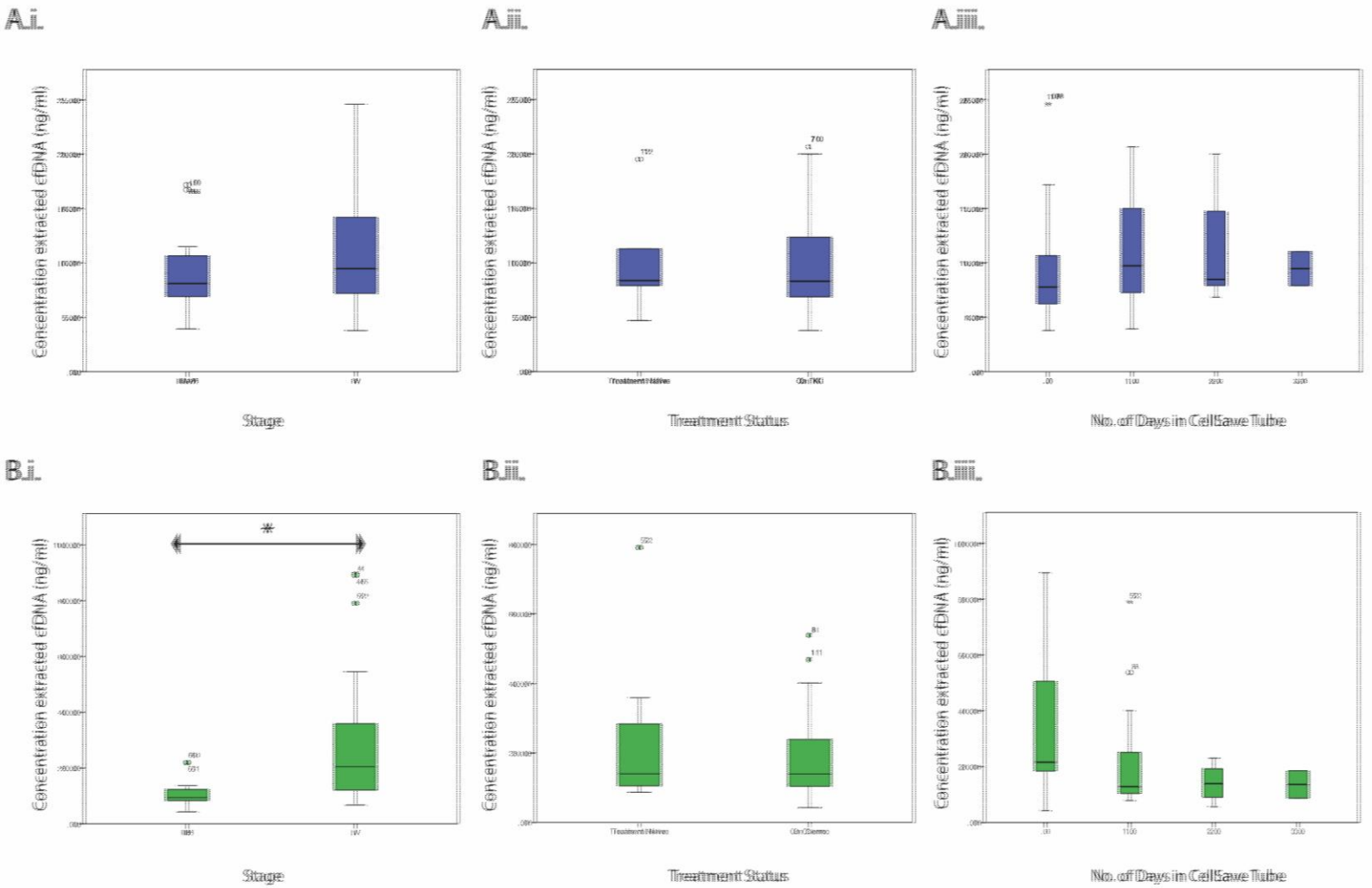

Figure 3-3 Potential Factors that may Influence Concentration of total cell free DNA. Box-and-whisker plots shown for patients with Adenocarcinoma of Lung (A) and Small Cell Lung Cancer (B), based on stage of disease (i), whether the sample is treatment naïve or not (ii) and the number of days between sample collection and sample processing (iii). * denotes that the difference between groups is statistically significant ($p<0.05$).

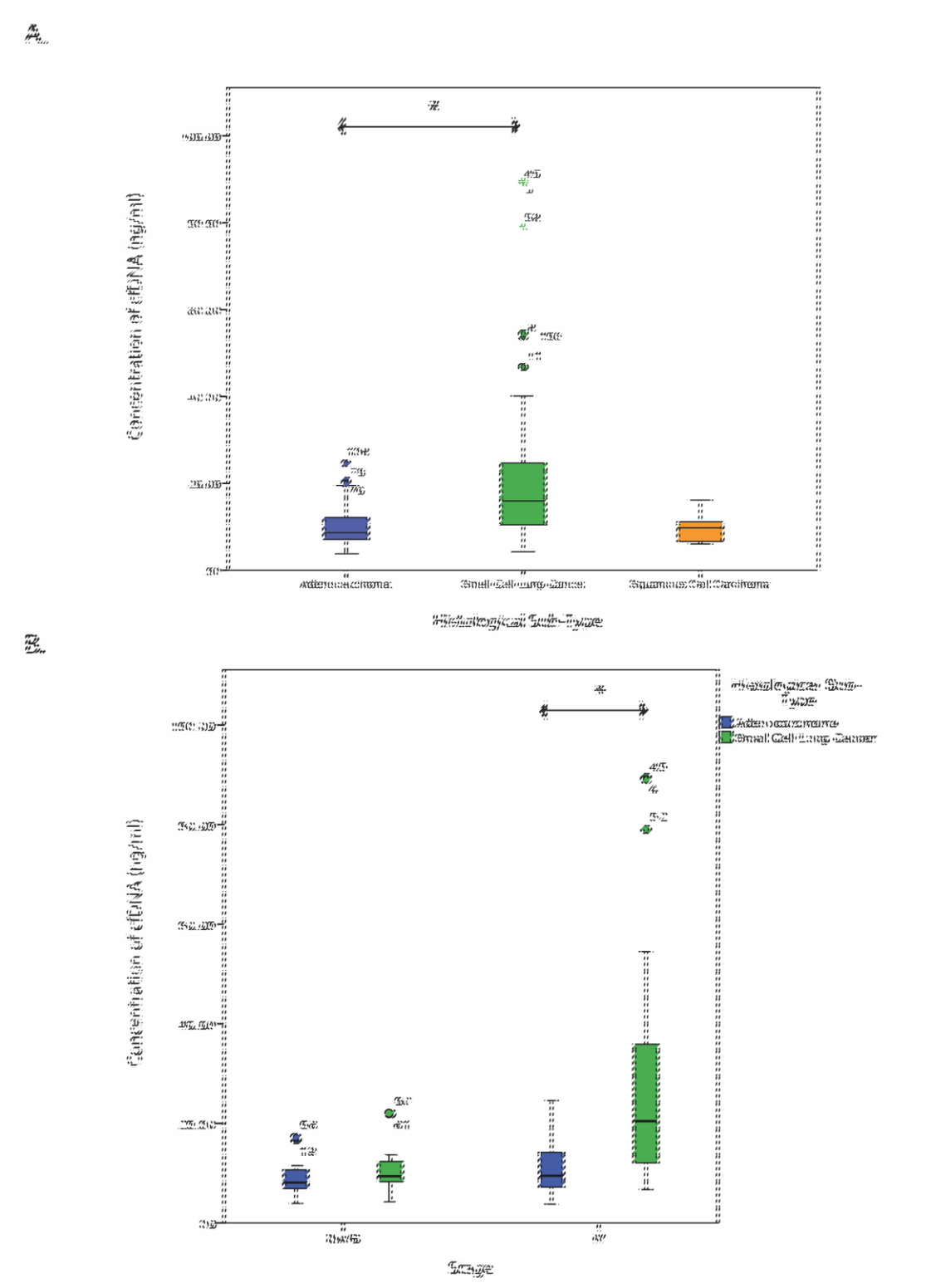

Figure 3-4 Differences in total cell free DNA (cfDNA) concentrations, between different histological subtypes of lung cancer (A). Taking into consideration the effect of stage of disease and the difference between adenocarcinoma and small cell lung cancer (B). * denotes that the difference between groups is statistically significant (p<0.05).

3.7 Clinical Implications

Detection and molecular characterisation of cancer mutations within ctDNA, using NGS or ddPCR, offers the prospect of facilitating personalised cancer therapy, using blood samples ("liquid biopsies") alone, where tumour tissue biopsies are unavailable or inadequate for confident diagnosis. However, it is important to define the ability of current technologies to confidently and reliably detect low levels of mutations against a background of 'wild-type' or genomic DNA, since (a) there is often only a small quantity of total circulating cell free DNA, and (b) mutations in ctDNA form only a fraction of total cell free DNA, and occur mostly between 0.1 – 5% alternate allele frequencies(112).

An advantage of NGS analysis is the ability to detect sequence changes contained within a panel of genetic regions, simultaneously, from one sample of DNA. The choice of NGS software analysis tool is important in determining the ability to detect the presence of mutations, especially at low mutational frequency. However, a commonly observed problem with NGS analysis, here and in the literature, is the ability to distinguish low level mutations from background 'NGS sequence artefact', especially where there is a large volume of 'sequence artefact' at low alternate allele frequencies.

A newer technology, such as third generation sequencing, is where raw DNA is sequenced directly, without the need for amplification. This may reduce the amount of 'sequencing artefact' generated, and this could offer a more sensitive technique for confidently detecting low level sequence variants(131). However, it is likely to be several years before this technology can be used for assessing ctDNA as a biomarker in routine clinical practice, mainly due to accessibility and cost issues.

Droplet digital PCR (ddPCR) is a more sensitive molecular technique, compared to NGS, both in the literature and in the data presented here. It is currently proposed by this thesis, that that the level of mutation detection using ddPCR, is

determined by the amount of input DNA, such that, more input DNA is required to confidently detect decreasing levels of mutational abundance. This is of direct relevance to using ddPCR for ctDNA analysis since the average concentration of total circulating cell free DNA in samples from patients with locally advanced/metastatic lung cancer is low, in the order of 1 ng/ul. This is consistent with the literature for other cancer types(132). A low amount of input DNA, may result in 'false-negative' test result, with important clinical implications. For example, patients with *EGFR*$^{mut+ve}$ adenocarcinoma of the lung may not be treated with an EGFR TKi, if they have a false negative *EGFR* mutation assay result on testing ctDNA. Conversely, the 'absence' of a mutation may also be important to guide a treatment decision, for example the use of cetuximab for *KRAS* 'wild-type' colorectal cancer patients. Caution needs to be taken if the presence of a 'wild-type' gene is an important predictive biomarker. In this case, it is important to reduce the 'false-negative' test result, since a false negative *KRAS* mutation test may result in the patient potentially being given a therapy, which would then be ineffectual, and associated with increased toxicity for the patient. Using ddPCR it would be difficult know whether a 'negative' test is due to there being not enough input DNA, the level of mutation is lower than the limit of detection, or the mutation is truly absent. In conclusion, for patients with advanced lung cancer, ddPCR is better suited for clinical applications where a 'true-positive' result is required, i.e. detecting the presence of a mutation, rather than a clinical application where a 'true-negative' result is required, i.e. to confirm the absence of a mutation.

For *EGFR*$^{mut+ve}$ adenocarcinoma of the lung, there are over 20 recognised activating mutations within the *EGFR* tyrosine kinase domain(36). Given that ddPCR is effective for detecting an individual mutation, per reaction, the amount of input ctDNA will also limit the number of mutations that can be tested for, per sample. Work is ongoing to multi-plex ddPCR reactions to overcome this problem.

For successful introduction of NGS/ddPCR mutation analysis into routine clinical practice, cancer mutations must be detected reproducibly, even if/where there

are significant delays between sample collection and processing/analysis. Data presented in this results chapter, demonstrate that the use a specialised blood collection tubes, such as CellSave, provide sufficient stability in the quantities of total cell free DNA concentration, to allow delays of up to 4 days in transport of samples, from any peripheral hospital to a centralised analytical laboratory for processing and testing. Therefore, CellSave tubes can facilitate the use of ctDNA analyses as biomarkers in routine, individualised clinical management of lung cancer patients, many of whom are managed in peripheral clinics without on-site genetic facilities.

No statistically significant difference could be detected in cfDNA concentration between samples from patients with small cell lung cancer, before or during intravenous chemotherapy, nor for adenocarcinoma patients before or during oral, EGFR TKi therapy.

Patients with stage IV small cell lung cancer have statistically significant more total circulating cell free DNA in their blood compared with patients with stage III SCLC and patients with stage IV adenocarcinoma of the lung. Further research can address whether this is a phenomenon observed as a result of a larger disease volume, as a consequence of more “aggressive” disease than NSCLC by correlating total cell free DNA concentration with radiological volume of disease, and disease activity, for example with CT-PET scans.

Despite the current limitations, analysis of ctDNA offers the prospect of a useful, quantitative and qualitative biomarker for real-time monitoring of disease status, prior to or on treatment, via serial blood samples from patients with locally advanced or metastatic lung cancer. Using the methods in this thesis, and understanding their limitations, it will be interesting to explore in subsequent results chapters, the potential utility of ctDNA as biomarker in specific clinical settings for patients with locally advanced/metastatic lung cancer.

4 Mutation Discovery in Tumour Tissue by Next Generation Sequencing

4.1 Chapter Background

Next generation sequencing (NGS) offers the ability to determine simultaneously multiple, target genetic regions, from a single DNA sample. There are many genetic changes of potential clinical interest and relevance to lung cancer e.g. mutations in *EGFR*, *BRAF*, *PIK3CA*, *MET*, *HER2* and *ALK*. These can be assessed by any one of multiple different NGS platforms, using various gene panels. Some gene panels are commercially available, but, for most NGS platforms, there is also the option of creating a customised gene panel.

It is possible to utilise NGS analysis for tumour DNA, whether this is DNA extracted from FFPE tumour tissue, or circulating cell-free tumour DNA (ctDNA). However, the quantity and quality of lung cancer tumour DNA is often a limiting factor in NGS analysis. In order to detect clinically relevant genetic changes in small quantities of tumour DNA from patients with lung cancer, it is important to select an appropriate gene panel and NGS platform combination.

4.2 Chapter Aims

The aim of this chapter is to determine whether it is possible to detect somatic genetic variants, in DNA samples obtained from FFPE tumour tissue samples, obtained from patients with NSCLC and SCLC, using NGS. Given the small amounts of tumour DNA often available, it is of interest whether sufficient NGS sequence 'depth' (see section on NGS methodology) can be achieved to determine the presence or absence of a somatic mutation.

4.3 Chapter Objectives

This chapter has the following objectives:

- To use NGS to assess the mutation status of DNA from tumour tissue samples from patients with adenocarcinoma of the lung (including *EGFR* mutation positive ($EGFR^{mut+ve}$) adenocarcinoma), squamous cell carcinoma of the lung, and small cell lung cancer.
- To assess the correlation between mutations detected via NGS and sequences obtained via complementary molecular techniques (Sanger gene sequencing, pyrosequencing and droplet digital PCR) confirm the presence of additional somatic mutations revealed by NGS.
- To compare the ability of a novel, custom-designed, in-house 22-gene non-small cell lung cancer panel with a commercially-available 50 cancer gene hotspot panel to detect the presence of known *EGFR* mutations, previously detected by routine diagnostic gene sequencing.
- To assess the ability of the 22-gene panel to detect mutations, which cannot be assessed, using the 50-gene panel.

4.4 Chapter Results

4.4.1 Mutation Detection in Lung Tumour Samples

A total of 29 patients with lung cancer were recruited to this research project. These included 11 patients with *EGFR*$^{mut+ve}$ adenocarcinoma of the lung, 5 patients with adenocarcinoma of lung (*EGFR* mutational status previously unknown), 3 patients with squamous cell carcinoma of the lung, and 10 patients with small cell lung cancer. Tumour tissue samples, or cytological 'cell-block' samples, were only available for DNA extraction and NGS analysis in 26 of the 29 patients (Table 4-1). For 2 patient samples (EGFR_2 and EGFR_5), there was insufficient tumour tissue remaining after routine histopathological examination. For 1 patient (SCLC_7), NGS was not performed since the available tissue sample did not contain any nucleated tumour cells.

Table 4-1 Summary of recruited lung cancer patients. NSCLC = Non Small Cell Lung Cancer, SCC = Squamous Cell Carcinoma, SCLC = Small Cell Lung Carcinoma.

Tumour type	Histology	Mutation status	Number of patients		Tumour DNA available for NGS analysis
NSCLC	Adenocarcinoma	Known *EGFR*$^{mut+ve}$	11	16	9
		Previously unknown mutation status	5		4
	SCC			3	3
SCLC				10	9
Total				29	26

The result of NGS analysis of the patient tumour DNA samples, using the 50 gene cancer hotspot panel, is shown in the heat map in Figure 4-1. In tumour samples from 23 of the 26 patients (88%), there was sufficient tumour DNA quantity and quality to enable a good sequence 'depth' (defined as 1000x sequence depth for this research project – see Methods and Materials – Next Generation Sequencing – Process – Library Preparation), to enable genetic characterisation. This is where mutations were confidently detected, or the genetic regions were confidently labelled as wild-type. Adequate sequence 'depth' was not achieved, in more than half of the genetic 'hotspot' regions contained

within the NGS panel, for tumour samples from 3 patients (EGFR_3, RT_2 and RT_4). This was because there was insufficient tumour DNA available for patient RT_2 (the sample contained only scanty cellular material). For the other two patients (EGFR_3 and RT_4), there was enough tumour DNA (>10ng), but the quality of the DNA may have been too poor. For example, it is known that DNA can fragment or otherwise degrade, as part of the formalin fixation process.

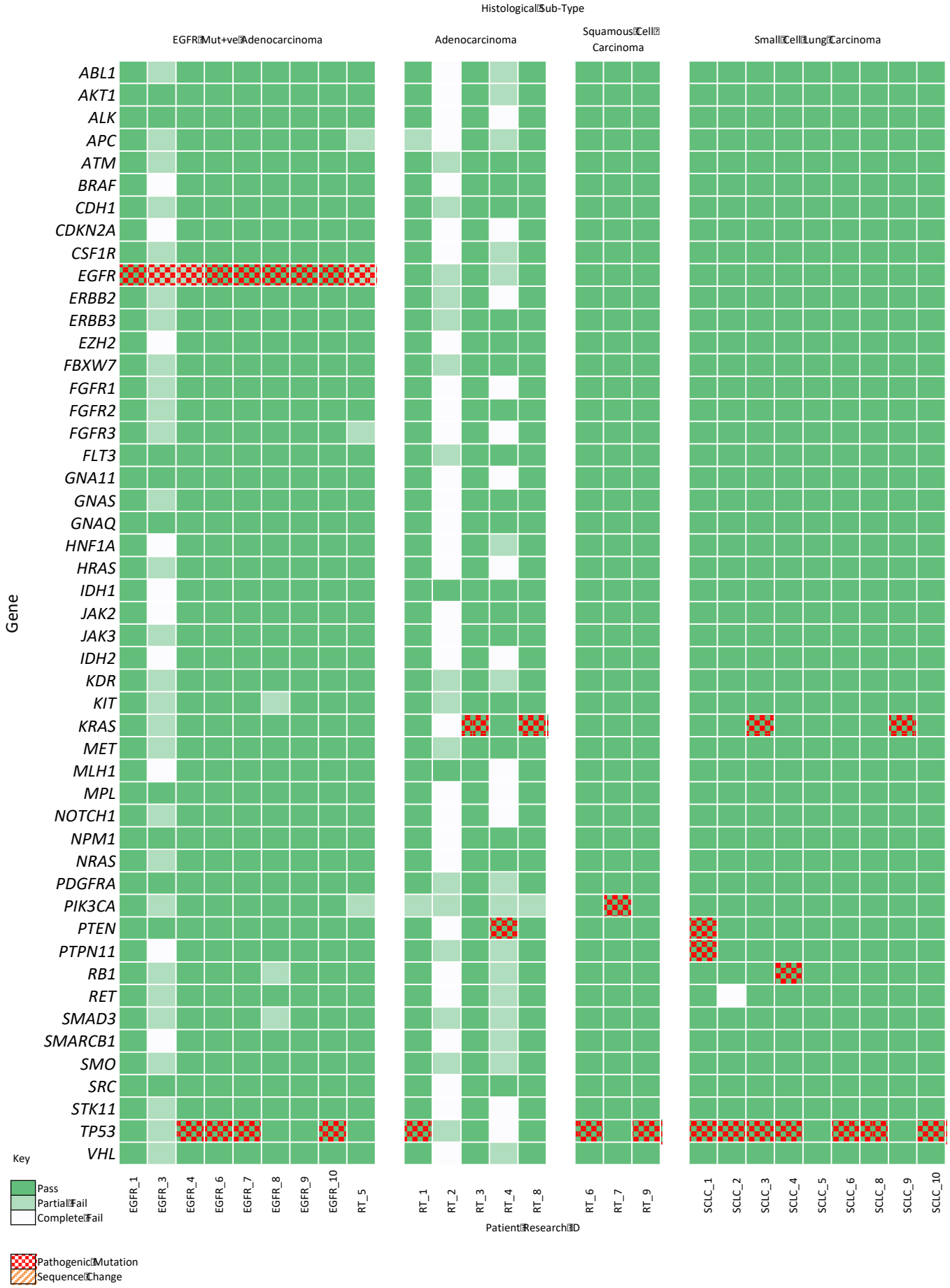

Figure 4-1 Heat map for sequence coverage and somatic genetic alteration using the 50 gene cancer hotspot panel. Genes received a 'pass' if the minimum required sequence depth was achieved in all gene regions covered by the panel. Partial fail if only some of the gene regions achieved the minimum required sequence depth. Complete fail if none of the gene regions achieved the minimum required sequence depth. Known pathogenic mutations are indicated by red stippling.

4.5 Patients with *EGFR*$^{mut+ve}$ Adenocarcinoma of the Lung

Eleven patients with *EGFR*$^{mut+ve}$ adenocarcinoma of lung were recruited to this study. For each of these patients, the *EGFR* mutation status of tumour DNA had been previously characterised, in routine clinical practice. This had been achieved using pyrosequencing for the common point mutations (*EGFR* Exon 18: c.2155G>A p.G719S, c.2155G>T p.G719C, c.2156G>C p.G719A, c.2159C>T p.S720F, Exon 20: c2294T>C p.V765A, c.2303G>T p.S768I, c.2305G>T p.V769L, c.2369C>T T790M, Exon 21: c.2573T>G p.L858R, c.2582T>A p.L861Q); and fragment length analysis for *EGFR* Exon 19 3 – 24 base pair deletions, and *EGFR* exon 4 – 9 base pair 20 insertions.

Table 4-2 summarises the histopathological characteristics of the tumour tissue sample for each of the 11 *EGFR*$^{mut+ve}$ adenocarcinoma patients, as well as the specific *EGFR* mutation sequence changes discovered by NGS. The most common source of tumour tissue for these patients was FFPE tumour tissue obtained from CT guided transcutaneous biopsy, confirming that for these patients, there is commonly a limited amount of tumour tissue available for genetic analysis.

The most commonly demonstrated *EGFR* mutations detected were *EGFR* c.2573T>G, p.L858R (44%) and *EGFR* exon 19 deletions (56%), consistent with reported mutational frequencies in the literature(133, 134). Where there was sufficient tumour tissue available for *EGFR* gene analysis, there was 100% concordance for *EGFR* mutation detection, between NGS, determined here, and prior pyrosequencing/fragment length analysis.

An additional benefit of NGS, over fragment length analysis, in the case of *EGFR* exon 19 deletions, is that NGS provides specific detail of the deleted sequence. As a result, it was possible to detect a novel, complex deletion/insertion *EGFR* exon 19 mutation; *EGFR* c.2239_2251del12>CCATTG, p.L747_T751>PL (Patient EGFR_1) which has not been previously reported in the literature. Knowledge of the specific deleted sequence is vital knowledge in the design of sequence-specific probes for subsequent ddPCR detection of the mutation.

There appears to be an approximate correlation between the percentage of tumour cells contained within the tissue sample (determined by histopathological examination), and the mutation alternate allele frequency (%) (The frequency of mutations compared to background 'wild-type' alleles) (Table 4-2). This is to be expected if all of the nucleated tumour cells contained within the tissue sample contain the mutation.

Table 4-2 Somatic Variant Discovery by NGS (50 cancer gene hotspot panel) in FFPE tumour tissue with previously characterised somatic mutations in the *EGFR* gene. The minimum required sequence depth is determined by the percentage of nucleated tumour cells seen on a haematoxylin and eosin (H&E) stained FFPE tumour tissue slide.

					EGFR Change Detected by NGS			
Patient Research ID	**Disease Stage**	**Tumour Sample**	**% Tumour Cells on H&E**	**Known *EGFR* mutation**	**DNA Change**	**Predicted Protein Change**	**Actual Sequencing Depth at Position**	**Alternate Allele Frequency (%)**
EGFR_1	IV	CT Guided Lung Biopsy	50	*EGFR* Exon 19, 12 bp del.	c.2239_2251del12>CCATTG	p.L747_T751>PL	40487	33.7
EGFR_2	IV	NA	-	*EGFR* p.G719A	Insufficient FFPE Tumour DNA for analysis			
EGFR_3	IV	CT Guided Lung Biopsy	75	*EGFR* p.L858R	c.2573T>G	p.L858R	1156	47.8
EGFR_4	IV	Surgical Resection of Primary Tumour	30	*EGFR* p.L858R	c.2573T>G	p.L858R	30125	17
EGFR_5	IV	NA	-	*EGFR* Exon 19, 15 bp del.	Insufficient FFPE Tumour DNA for analysis			
EGFR_6	IV	CT Guided Lung Biopsy	10	*EGFR* Exon 19, 15 bp del.	c.2235_2249del15	p.E746_A750	13071	23.8
EGFR_7	IV	EBUS LN Cytology	20	*EGFR* p.L858R	c.2573T>G	p.L858R	34448	20.4
EGFR_8	IV	CT Guided Lung Biopsy	30	*EGFR* Exon 19, 15 bp del.	c.2235_2249del15	p.E746_A750	5589	42
EGFR_9	IIIA	CT Guided Lung Biopsy	60	*EGFR* Exon 19, 12 bp del.	c.2238_2249del15	p.L747_T751>P	22565	18.5
EGFR_10	IV	CT Guided Lung Biopsy	60	*EGFR* Exon 19, 15 bp del.	c.2236_2250del15	p.E746_A750	68326	70.6
RT_5	IIIB	Surgical Resection of Primary Tumour	30	*EGFR* p.L858R	c.2573T>G	p.L858R	4877	37.1

4.6 Non *EGFR* Mutation Detection in Adenocarcinoma Samples

In the patients with *EGFR*$^{mut+ve}$ adenocarcinoma of the lung, 44% had a concurrent somatic mutation within the *TP53* gene (Table 4-3). This has not been previously reported in the literature. It is unknown whether mutant *TP53* in *EGFR*$^{mut+ve}$ adenocarcinoma patients has any clinical implications for these patients.

Table 4-3 Concurrent mutations detected in *EGFR*$^{mut+ve}$ adenocarcinoma samples. AAF = Alternate Allele

Patient ID	% Tumour Cells	AAF of EGFR mutation (%)(NGS)	Gene Mutation in addition to EGFR	Concurrent Gene Mutation			AAF of Concurrent Mutation (%)*
EGFR_1	50	33.7	No				
EGFR_3	75	41.1	No				
EGFR_4	30	17.0	Yes	*TP53*	c.655C>T	p.P219S	23
EGFR_6	10	23.8	Yes	*TP53*	c.764T>A	p.I255N	7.0
EGFR_7	20	20.4	Yes	*TP53*	c.259A>G	p.W220C	16.2
EGFR_8	30	42.0	No				
EGFR_9	60	18.5	No				
EGFR_10	60	70.6	Yes	*TP53*	c.844C>T	pR282W	36.3
RT_5	30	37.1	No				

***Frequency (%) determined by NGS**

There were 5 patients with adenocarcinoma of the lung where the *EGFR* mutational status was unknown prior to NGS. Samples were available for NGS for 4 of these patients. The histopathological characteristics of each sample, and the mutations detected are summarised in Table 4-4. In these tumour samples, mutations detected were *KRAS* (c.35G>C, p.G12A and c.35G>T, p.G12V), *PTEN* (c.494G>A, p.G165E) and *TP53* (c.1006G>T, p.E336TER). The presence of the *KRAS* mutations were able to be confirmed with pyrosequencing and COLD-PCR followed by pyrosequencing. There was insufficient DNA remaining after NGS, to confirm the presence of the other mutations in the other samples. No additional *EGFR* mutations were detected in these samples.

Table 4-4 Somatic Variant Discovery in previously untested Adenocarcinoma samples of Lung using NGS and the 50 gene, targeted cancer hotspot panel. CT = Computed Tomography, EBUS = Endobronchial Ultrasound Scan, LN = Lymph Node.

Patient Research ID	Stage	Tumour Sample	% Tumour Cells	Pre Test Pass/Fail	Somatic Mutation Detected?	Gene(s)	DNA Change	Predicted Protein Change	Actual Sequencing Depth Achieved at Change Position	Variant Allele Frequency (%)	Confirmed with Alternative Method
RT_1	IIIA	CT Guided Lung Biopsy	100	*Pass*	Yes	*TP53*	c.1006G>T	p.E336TER	2884	25.4	Insufficient DNA
RT_2	IIIA	EBUS guided LN biopsy/cytology	<1	*Fail*	Insufficient sequence coverage	-	-	-	-	-	-
RT_3	IIIA	EBUS guided LN biopsy/cytology	40	*Pass*	Yes	*KRAS*	c.35G>C	p.G12A	20371	59.7	Pyrosequencing
RT_4	IIIA	Thoracoscopic Biopsy	50	*Pass*	Yes	*PTEN*	c.494G>A	p.G165E	2800	3.2	Insufficient DNA
RT_8	IIIA	EBUS guided LN biopsy/cytology	40	*Pass*	Yes	*KRAS*	c.35G>T	p.G12V	7932	4.8	COLD PCR & Pyrosequencing

4.7 Mutation Detection in Squamous Cell Carcinoma

Sequence depth was adequate for all the genetic regions, for all 3 of the squamous cell carcinoma samples. It was possible to detect two mutations in the *TP53* gene (RT_6 and RT_9) and one in the *PIK3CA* gene (RT_7). There was enough remaining DNA in two samples (RT_7 and RT_9), after NGS, to confirm the presence of two of these mutations using a second technique. The presence of the *PIK3CA* mutation was confirmed by pyrosequencing, and the *TP53* c.713G>T, p.C238F was confirmed by Sanger Sequencing (Table 4-5). This was because the DNA for one sample (patient RT_6) was exhausted by NGS analysis. The histopathological characteristics of the squamous cell carcinoma samples, as well as the specific mutations are also detailed in Table 4-5.

Table 4-5 Somatic Variant Discovery in Squamous Cell Carcinoma samples of Lung using NGS and the 50 gene, targeted cancer hotspot panel. CT = Computed Tomography

Patient Research ID	Stage	Tumour Sample	% Tumour Cells	Pre Test Pass/Fail	Somatic Mutation Detected?	Gene(s)	DNA Change	Predicted Protein Change	Actual Sequencing Depth Achieved at Change Position	Variant Allele Frequency (%)	Confirmed with Alternative Method
RT_6	IIA	Bronchoscopic Biopsy	40	*Pass*	Yes	*TP53*	c.820G>C	p.V274L	6830	43.4	Insufficient DNA
RT_7	IB	CT Guided Lung Biopsy	50	*Pass*	Yes	*PIK3CA*	c.3140A>G	p.H1047R	7906	26.7	Pyrosequencing
RT_9	IIIB	Bronchoscopic Biopsy	40	*Pass*	Yes	*TP53*	c.713G>T	p.C238F	16013	41.8	Sanger Sequencing

4.8 Mutation discovery in patients with SCLC

Table 4-6 summarises the patient characteristics and NGS findings for SCLC patients. The sources of tumour tissue were bronchoscopic biopsy (50% of cases) aspiration cytology (30%), CT guided lung biopsy (one patient) and a needle core biopsy of an axillary tumour (one patient). This re-enforces the commonly encountered problem that patients with stage III or stage IV SCLC often only have very small amounts of tumour cellular material for genetic analysis. As previously stated, there was enough tumour DNA for NGS analysis in 9 of the 10 SCLC patients. One patient (SCLC_7) had a bronchoscopic biopsy, but there were no nucleated tumour cells remaining in the tissue block, as it is likely that previous histopathological examination had exhausted the area of the sample that did contain them.

A total of 10 mutations were found in all tested SCLC samples, 9 of which were confirmed with a second technique (Table 4-6). The most commonly found mutations in the small cell lung cancer samples were within *TP53* (78%), *KRAS* (22%), and *RB1* (11%). The presence of the *RB1* c.1700C>T, p.S567L mutation (patient SCLC_4) could not be confirmed, as there was insufficient DNA remaining after prior NGS and Sanger Sequencing. These are consistent with reported mutational gene frequencies for *TP53* and *KRAS* in SCLC, however, mutations within *RB1* are potentially under-represented in these patients, where they have been reported as high as 58%(135). This is likely due to the small sample size of SCLC patients included in the current research project.

For one patient (SCLC_5) no pathogenic mutations were identified. Given the good sequence 'depth' for all the genetic regions within the 50 gene cancer hotspot panel, therefore it can be concluded that the cancer in this patient was 'wild-type' for these genetic areas. It may be that, for this specific cancer sample, the pathogenic mutations exist in areas of these genes, or in other genes, not included by the gene panel. Another possibility is that different mutational mechanisms are responsible for the tumorigenesis in this patient's cancer, namely gene translocations or gene copy number variations, which cannot be detected by this gene panel and method of NGS analysis.

Table 4-6 Somatic Variant Discovery in SCLC samples using NGS and the 50 gene, targeted cancer hotspot panel. CT = Computed Tomography.

Patient Research ID	Stage	Tumour Sample	% Tumour Cells	Pre Test Pass/Fail	Somatic Mutation Detected?	Gene(s)	DNA Change	Predicted Protein Change	Actual Sequencing Depth Achieved at Change Position	Variant Allele Frequency (%)	Confirmed with Alternative Method
SCLC_1	IV	CT Guided Lung Biopsy	70	*Pass*	Yes	*TP53*	c.742C>T	p.R248W	34402	97.2	Sanger Sequencing
SCLC_2	IV	Bronchoscopic Biopsy	50	*Pass*	Yes	*TP53*	c.1001G>T	p.G334V	6083	31	Droplet Digital PCR
SCLC_3	IV	Pleural Fluid Cytology	75	*Pass*	Yes	*TP53*	c.844C>G	p.R282G	13681	40.2	Sanger Sequencing
						KRAS	c.34G>T	p.G12C	29246	18.1	Pyrosequencing
SCLC_4	IIIB	Lymph Node Aspirate Cytology	40	*Pass*	Yes	*TP53*	c.451C>A	p.P151T	32671	43.8	Sanger Sequencing
						RB1	c.1700C>T	p.S567L	16644	36.4	Insufficient DNA
SCLC_5	IV	Bronchoscopic Biopsy	70	*Pass*	No	-	-	-	-	-	-
SCLC_6	IV	Bronchoscopic Biopsy	60	*Pass*	Yes	*TP53*	c.592G>T	p.E198TER	23700	64.4	Sanger Sequencing
SCLC_7	IV	Bronchoscopic Biopsy	0	*Fail*	NGS not performed	-	-	-	-	-	-
SCLC_8	IV	Sub-Cutaneous Nodule Aspirate Cytology	95	*Pass*	Yes	*TP53*	c.641A>G	p.H214R	39757	98.9	Droplet Digital PCR
SCLC_9	IV	Needle Core Biopsy Right Axilla	100	*Pass*	Yes	*KRAS*	c.437C>T	p.A146V	41890	45.2	Pyrosequencing
SCLC_10	IIIB	Bronchoscopic Biopsy	60	*Pass*	Yes	*TP53*	c.892G>T	p.E298TER	7910	80.7	Sanger Sequencing

4.9 Custom 22-Gene Panel for Non-Small Cell Lung Cancer

An important aim of this thesis was to compare the commercially-available 50 cancer gene 'hotspot' panel, used above, with an in-house, custom, 22-gene, whole exon panel, specifically designed for NSCLC. Table 4-7 compares the target genetic regions contained within these two panels. The custom panel can target the same intragenic mutation hotspot regions for the genes contained within both panels, with the exception of *STK11*. Sub-optimal exon coverage of *STK11*, meant that two *STK11* mutations; *STK11* c.996G>A, p.W332* and *STK11* c.1062C>G, p.F354L (Exon 8) are detectable by the 50 gene hotspot panel, but not the custom 22 gene panel. Despite panel design attempts to improve *STK11* coverage, it was not possible to include exon 8 within the 22 gene custom panel. Some of the genes within the 22 gene custom panel, are not included by the 50 gene hotspot panel, namely *BRCA1*, *BRCA2*, *MAP2K1*, *MTOR* and *NF1*.

It was only possible to assess tumour tissue samples from 11 of the 29 patients using the custom 22-gene panel, specifically 8 of the *EGFR*$^{mut+ve}$ adenocarcinoma patients, and 3 of the squamous cell carcinoma patients, due to financial constraints of this research project. The custom 22-gene panel uses 3 primer pools per sample (see Methods and Materials – NGS Principles), compared with one primer pool used by the cancer hotspot panel. Consequently, the custom gene panel required 3 times as much input DNA, and 3 times more sequencing reagents compared to the 50 gene cancer hotspot panel, and is thus 3 times more expensive per sample analysed.

Table 4-7 Comparing target genetic regions between the 50 gene panel and the 22 gene panel. The final column denotes whether the regions contained within the 50 gene panel, are also included by the 22 gene panel. SNV = Single Nucleotide Variant. NA = Not applicable.

Gene	50 Gene Cancer Hotspot Panel	22 Gene Custom Panel	SNV Overlap
ABL1	Yes - Hotspots	No	
AKT1	Yes - Hotspots	Yes – Most Exons	Yes
ALK	Yes - Hotspots	No	
APC	Yes - Hotspots	No	
ATM	Yes - Hotspots	No	
BRAF	Yes - Hotspots	Yes – All Exons	Yes
BRCA1	No	Yes – All Exons	NA
BRCA2	No	Yes – Most Exons	NA
CDH1	Yes - Hotspots	No	
CDKN2A	Yes - Hotspots	Yes – Most Exons	
CSF1R	Yes - Hotspots	No	
CTNNB1	Yes - Hotspots	No	
DDR2	No	Yes – All Exons	NA
EGFR	Yes - Hotspots	Yes – All Exons	Yes
ERBB2	Yes - Hotspots	Yes – Most Exons	Yes
ERBB4	Yes - Hotspots	No	
EZH2	Yes - Hotspots	No	
FBXWT	Yes - Hotspots	No	
FGFR1	Yes - Hotspots	No	
FGFR2	Yes - Hotspots	Yes – All Exons	Yes
FGFR3	Yes - Hotspots	No	
FLT3	Yes - Hotspots	Yes – Most Exons	Yes
GNA11	Yes - Hotspots	No	
GNAS	Yes - Hotspots	No	
GNAQ	Yes - Hotspots	No	
HNF1A	Yes - Hotspots	No	
HRAS	Yes - Hotspots	No	
IDH1	Yes - Hotspots	No	
JAK2	Yes - Hotspots	No	
JAK3	Yes - Hotspots	No	
IDH2	Yes - Hotspots	No	
KDR	Yes - Hotspots	No	
KIT	Yes - Hotspots	Yes – All Exons	Yes
KRAS	Yes - Hotspots	Yes – All Exons	Yes
MAP2K1	No	Yes – Most Exons	NA
MET	Yes - Hotspots	Yes – Most Exons	Yes
MLH1	Yes - Hotspots	No	
MPL	Yes - Hotspots	No	

MTOR	No	Yes – All Exons	NA
NF1	No	Yes – Most Exons	NA
NOTCH1	Yes - Hotspots	No	
NPM1	Yes - Hotspots	No	
NRAS	Yes - Hotspots	Yes – All Exons	Yes
PDGFRA	Yes - Hotspots	Yes – All Exons	Yes
PIK3CA	Yes - Hotspots	Yes – All Exons	Yes
PTEN	Yes - Hotspots	Yes – All Exons	Yes
PTPN11	Yes - Hotspots	No	
RB1	Yes - Hotspots	No	
RET	Yes - Hotspots	No	
SMAD4	Yes - Hotspots	No	
SMARCB1	Yes - Hotspots	No	
SMO	Yes - Hotspots	No	
SRC	Yes - Hotspots	No	
STK11	Yes - Hotspots	Yes – Most Exons	No
TP53	Yes - Hotspots	Yes – All Exons	Yes
VHL	Yes - Hotspots	No	

Adequate sequence depth (>1000x) was achieved in 100% of the target gene regions of the 50 gene cancer hotspot panel for 23 of the 26 tumour tissue samples analysed with this gene panel. Conversely, adequate sequence depth could only be achieved for 39% of the target gene regions contained within the custom panel. Where sufficient sequence depth was achieved for accurate assessment of mutation status, there was some concordance between using the 22-gene custom and 50 gene hotspot panels, on the same patient tumour DNA samples (Figure 4-2). In Figure 4-2, all of the exons for most of the genes within the 22 custom panel failed to achieve 1000x sequence depth, due to individual amplicons repeatedly failing. These are summarized in Table 4-8. All of the other amplicons covering the exons for these genes did achieve 1000x sequence depth. This problem is commonly encountered with NGS gene panel design, and further work to optimize amplicon design to achieve the minimum required sequence depth, if that genomic region is of interest or relevance. This was not pursued within this thesis, as the individual amplicons that 'failed' were not considered to cover 'critical' mutation regions for the genes of interest for many of the genes. For *BRCA2* and *NF1*, there were 12 and 13 amplicons, respectively that consistently failed. Optimizing panel design to provide a greater sequence depth for these genetic regions would have utilized a further

'primer pool'. This increases the cost of the NGS pane/per sample, and increases the DNA requirement, which is a limiting factor from the samples utilized in this project.

The 50 gene cancer hotspot panel was able to detect 100% of the *EGFR* gene mutations in the 8 patient tumour DNA samples that were known to contain an *EGFR* mutation (as they had previously been characterised with pyrosequencing). In contrast, although an adequate sequence depth of the *EGFR* gene was consistently achieved with the 22-gene custom panel, it was only possible to the *EGFR* mutation in 1 one of the 8 *EGFR*$^{mut+ve}$ samples – specifically the *EGFR* p.L858R mutation in the tumour sample from patient EGFR_7. The custom panel failed to detect all of the *EGFR* exon 19 deletions. A retrospective, visual inspection of the NGS DNA sequence data revealed that this was due to a design fault of the gene panel. Therefore, in its current stage of development, the custom 22-gene panel is not suitable for clinical use in analysis of NSCLC patients' tumour DNA and subsequent treatment decision-making and/or determining eligibility for stratified clinical trials. Further work needs to be done so that it can detect *EGFR* exon 19 deletions, and achieve better sequence depth for many of the genetic regions included by the panel. Specifically, the custom 22 gene panel needs to be re-designed, so that EGFR exon 19 is covered by one amplicon (as with the 50 gene hotspot panel), rather than two adjacent ones (as with the current 22 gene custom panel). This further development was not possible in this project, owing to financial and time constraints.

The exact 3 pathogenic mutations discovered in the squamous cell tumour samples, were detected using both gene panels. Of note, the custom 22-gene panel was also able to detect an additional, concurrent somatic genetic alteration of unknown significance in *BRCA2* (*BRCA2* c.G8716C, p.E2906Q), in one patient with squamous cell carcinoma (patient RT_9). The *BRCA2* gene is not included in the 50 gene cancer hotspot panel, but may be of potential clinical relevance in lung cancer. Specifically, mutated *BRCA1/2* may act as a predictive biomarker of response to therapies, such as PARP inhibitors, for example, olaparib.

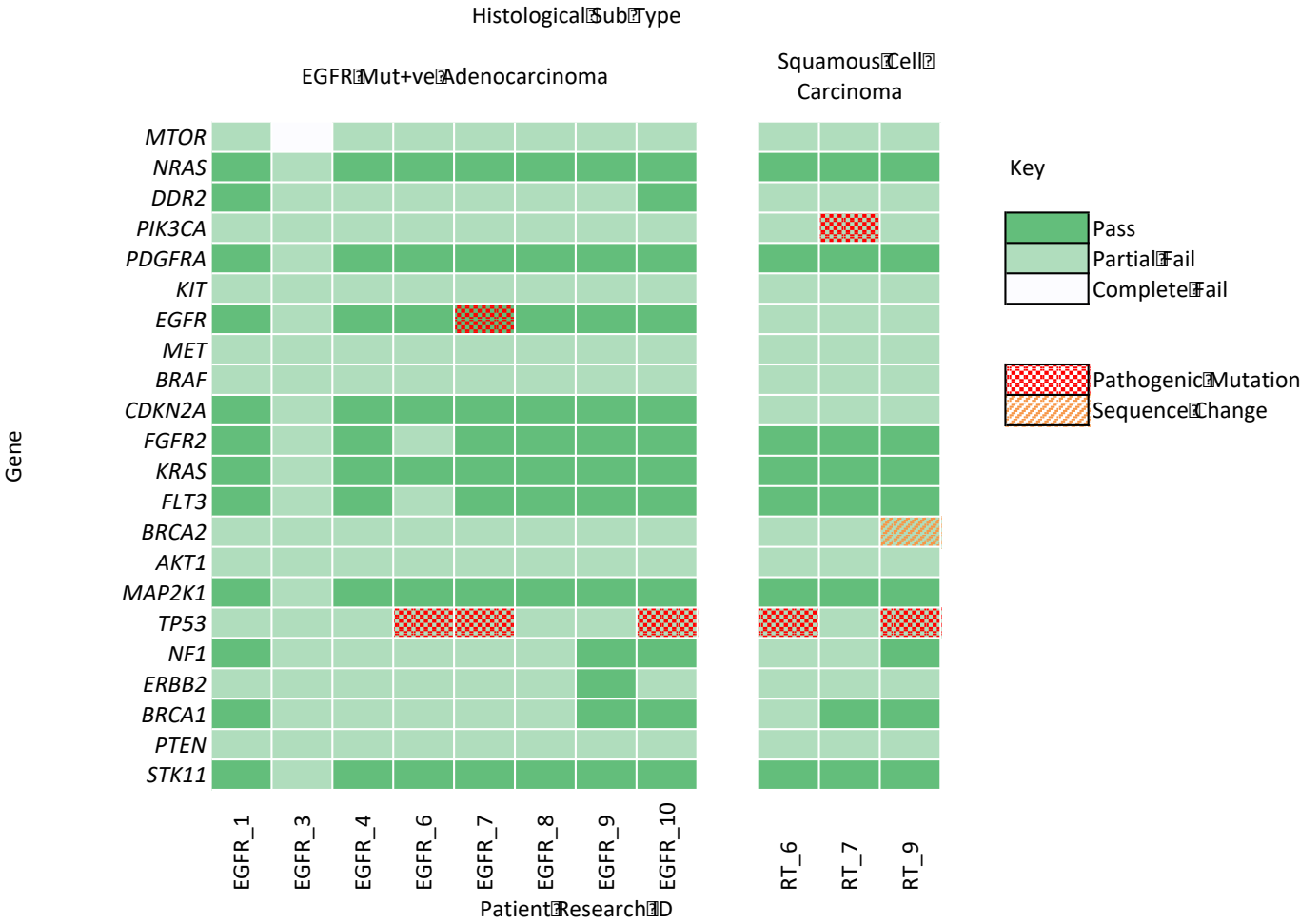

Figure 4-2 Heat map of sequencing coverage achieved, and somatic mutations identified, using the custom designed 22 gene panel for NSCLC. Genes received a 'pass' if the required minimum sequencing depth was achieved in all gene regions covered by the panel. Partial fail if only some of the gene regions achieved the minimum sequence depth. Complete fail if none of the gene regions achieved the required minimum sequence depth. Pathogenic mutations are indicated by red stippling, and genetic sequence changes of uncertain significance are indicated by orange stippling

Table 4-8. 22 gene custom panel amplicons regions that consistently failed to achieve the minimum 1000x sequence depth. The resulting 'missed genetic regions' are chromosomal location using genome hg19.

Gene	Number of 'failed' Amplicons	Missed Genetic Regions
MTOR	2	chr1:11182041-11182216, chr1:11199484-11199658
DDR2	2	chr1:162741866-162742017, chr1:162743266-162743439
PIK3CA	4	chr3:178916624-178916782, chr3:178937778-178937953 chr3:178937462-178937629, chr3:178947888-178948051
KIT	4	chr4:55561831-55561983, chr4:55592101-55592271 chr4:55569772-55569925, chr4:55594187-55594341
MET	2	chr7:116409589-116409760, chr7:116421907-116422072
BRAF	2	chr7:140439703-140439854, chr7:140549940-140550103
AKT1	1	chr14:105241881-105242055
TP53	2	chr17:7577130-7577279, chr17:7579486-7579631
ERBB2	2	chr17:37879658-37879823, chr17:37880211-37880374
BRCA1	1	chr17:41245638-41245813
PTEN	2	chr10:89653788-89653948, chr10:89725092-89725267
BRCA2	12	chr13:32899276-32899430, chr13:32953607-32953773 chr13:32903580-23903755, chr13:32953495-32953670 chr13:23912243-23912411, chr13:32945110-32945284 chr13:32912273-32912438, chr13:32937250-32937412 chr13:32913590-32913744, chr13:32920842-32921012 chr13:23913665-32913819, chr13:32918675-32918838
NF1	13	chr17:29508689-29508861, chr17:29679375-29679539 chr17:29527586-29527739, chr17:29676158-29676327 chr17:29528330-29528484, chr17:29670092-29670266 chr17:29548884-29549059, chr17:29665699-29665861 chr17:29559032-29559203, chr17:29661809-29661984 chr17:29585914-29586089, chr17:29657458-29657611 chr17:29592192-29592366

4.10 Clinical Implications

Increasingly, new treatments are being developed, which target specific aberrant tumour proteins that arise from tumour genetic alterations, such as the use of EGFR TKi in lung tumours that harbour *EGFR* mutations. Clinical trials of these new agents increasingly stratify patients on the basis of tumour molecular genetics profiles. Therefore it is crucially important to be confident in the ability of NGS, to achieve significant depth of sequence coverage, so that important genetic mutations can be accurately, reliably and reproducibly detected for clinical decision-making.

The results presented in this chapter demonstrate that it is possible to detect somatic mutations using NGS, in small lung cancer samples, using a commercially-available, targeted 50 gene cancer hotspot panel. In 88% of lung tumour samples, it was possible to perform NGS analysis, even on DNA derived from very small tumour tissue biopsy samples, or FFPE blocks made from cytological samples, where the tumour content may be very low, or the DNA is prone to significant degradation as a result of the fixation process. However, despite this, it has to be remembered that this gene panel only focuses on the pathogenic hot-spot regions within genes, and could potentially miss genetic sequence changes contained within exonic regions outside of the hotspots. Also, this panel does not include some genes that may become clinically relevant to lung cancer in the future, such as *BRCA1* and *BRCA2* mutations.

For these reasons, it is sometimes desirable to design a custom targeted panel, in attempt to include whole exonic regions of genes of interest, or to tailor a gene panel to a specific disease. The advantage of such a 'custom' gene panel approach is that it permits careful selection of genes and genetic regions to be included on the panel. We designed a custom panel to assay 22 genes which are implicated in non-small cell lung cancer, that each have a mutational frequency of >1% in NSCLC, and the mutational status of which may be potentially clinically relevant i.e. they predict response to targeted therapeutics currently in routine

clinical use, or in development in early phase clinical trials. In this results chapter, we have demonstrated the ability of a custom panel to detect somatic sequence changes in genes such as *BRCA2* that would otherwise have been missed using a generic, commercially panel. However, technical difficulties were encountered with the custom panel, whereby we did not achieve the required sequencing 'depth' in order to adequately detect, or confidently exclude, the presence of clinically-relevant somatic mutations. A further limitation of the custom gene panel is that it is more expensive. With the custom gene panel approach, it has been demonstrated that careful attention has to be given to panel design, as this was the main reason why *EGFR* exon 19 deletions could not be identified. Therefore it is recommended that a significant amount of time should be devoted to optimising the design of any custom targeted gene panel, to ensure it is robust and has practical clinical utility.

Compared with whole genome sequencing, targeted NGS sequencing method of molecular characterisation, allows sufficient 'depth' of sequencing to detect mutations within cancer DNA, that is often mixed with germline DNA derived from normal cells in the same sample. This can be at the expense of the inability to determine the presence or absence of other tumorigenic mutations observed in lung cancer, particularly larger-scale genetic changes, such as *ALK*, *ROS1* and *RET* translocations, and gene copy number changes, for example *MET* amplification. Therefore, there is still a requirement for the tumour tissue analysis using complementary methods such as FISH. Unfortunately, this may not always be possible, given the limited amount of tumour tissue available from biopsies from patients with advanced or metastatic lung cancer. Therefore, to increase the utility of a NGS platform for molecular characterisation of tumour tissue, future research is needed to be able to develop NGS assays that can simultaneously assess for point mutations, small insertions/deletions, gene copy number changes and gene translocations.

The results presented in this chapter also reflect that tumour tissue DNA is not always available or adequate. With routine genetic analysis of tumour tissue, there is a NGS failure rate, even when tumour DNA appears to be available, and

this can be due to DNA degradation resulting from the formalin fixation processing of tumour tissue samples. All of the samples in this project were FFPE tumour tissue or cytological samples. Future research into the molecular characterisation of tumour tissue may be improved by the use of fresh tumour tissue samples, but the research must include the feasibility of obtaining this tissue within the routine clinical setting, and whether it could subsequently be implemented as standard clinical care.

With conventional tumour sampling, it is not possible to fully characterise intra-tumour or intra-patient heterogeneity, especially where there is multi-focal metastatic disease, owing to sampling error/bias, unless a biopsy is potentially obtained from each tumour site. This is often either unacceptable to the patient or not clinically feasible. Likewise, monitoring clonal evolution of the tumour would require repeated, serial biopsies at different time points during a patient's cancer journey. New, or emerging sites of metastatic disease that have developed during anti-cancer therapy, may represent resistant sub-clones of tumour, and unless these can specifically be biopsied, a repeat biopsy of the original tumour may not capture this intra-patient tumour heterogeneity. Therefore, alternative sources of tumour DNA, such as circulating cell free tumour DNA (ctDNA) may provide DNA for analysis, to help to overcome the limitations of analysis of FFPE tumour tissue DNA. NGS of ctDNA in patients with lung cancer, alongside other molecular techniques, such as ddPCR, is the focus of the next results chapters.

5 Monitoring Disease Progression and Response to Therapy using Circulating Tumour DNA in Patients with Metastatic *EGFR*$^{mut+ve}$ Adenocarcinoma of the Lung

5.1 Chapter Overview

There is a pressing need to clarify the potential value of ctDNA, as a circulating, surrogate biomarker, in *EGFR*$^{mut+ve}$ adenocarcinoma of the lung, in different clinical settings. These include response to treatment, disease monitoring in periods of minimal residual disease, detecting disease progression and determining molecular mechanisms of therapeutic resistance. In this chapter we will explore the utility of ctDNA in the particular clinical circumstance of patients with *EGFR*$^{mut+ve}$ adenocarcinoma of the lung.

5.1.1 Chapter aims

For patients with locally advanced and metastatic *EGFR*$^{mut+ve}$ adenocarcinoma of the lung, this chapter aims to:

- Compare the sensitivity of NGS and ddPCR in assessing EGFR mutations in ctDNA
- Assess the capacity of ctDNA analysis to measure the kinetics of *EGFR* mutations in ctDNA, as the patient receives EGFR TKi therapy
- Investigate the use of ctDNA to monitor disease status/tumour burden and response to treatment, in patients with and without CNS metastases
- Assess whether NGS analysis of ctDNA can detect mutations not detected in the original diagnostic tissue biopsy
- Assess whether NGS analysis of ctDNA can detect mutations additional to mutant EGFR when the patient develops progressive cancer.
- Determine whether ddPCR can be used to detect the acquired *EGFR* c.2369C>T, p.T790M mutation, when a patient develops therapeutic resistance to EGFR TKi

5.2 Chapter Results

5.2.1 Patient and sample collection

The clinical course of the same eleven patients with locally advanced and/or metastatic, *EGFR*$^{mut+ve}$ adenocarcinoma of the lung (first identified in Chapter 4) were monitored. Blood samples for *EGFR* mutational analysis of ctDNA were taken at 3 month intervals, as each patient underwent therapy with an EGFR TKi.

Table 5-1 summarises the clinical characteristics for each patient, including stage of disease at the time of patient consent, sites of metastases, any prior anti-cancer therapies, and the type of EGFR TKi used. Relevant clinical parameters included the duration of EGFR TKi therapy, and whether the disease was progressing or responding at the end of the follow-up period for each patient.

At total of 39 blood samples were drawn for these patients for ctDNA analysis, over the research period, with a median of 4 blood samples per patient (range 2 to 6 blood samples). The median follow-up period, from the time of consent, was 375 days (range 173 to 697 days).

Patient Research ID	Sex	TNM	Stage	Sites of Metastases	Prior Treatment	EGFR TKI	Duration of TKI (Days)	Disease status at end of follow-up period
EGFR_1	F	T2aN2M1b	IV	Lung, Bone	1. Chemotherapy 2. EGFR TKI	Afatinib	373	Progressive disease
EGFR_2	F	T2aN1M1b	IV	Bone, Liver	1. Chemotherapy 2. EGFR TKI	Gefitinib	326	Progressive disease
EGFR_3	F	T2bN0M1b	IV	Liver	1. EGFR TKI	Gefitinib	634*	Partial Response
EGFR_4	M	T2aN0M1a	IV	Pleura	1. Surgical Resection of Primary Tumour 2. EGFR TKI	Gefitinib	697*	Partial Response
EGFR_5	F	T4N3M1b	IV	Liver, Lung	1. EGFR TKI	Gefitinib	497	Progressive disease (with brain metastases)
EGFR_6	F	T4N1M1b	IV	Bone, Meningeal	1. EGFR TKI	Afatinib	505*	Partial Response
EGFR_7	M	T2aN2M1b	IV	Bone, Liver	1. Palliative RT to Spine 2. EGFR TKI	Gefitinib	173	Progressive disease
EGFR_8	M	TxN2M1b	IV	Lung, Pleura	1. EGFR TKI	Afatinib	321	Progressive disease (with brain metastases)
EGFR_9	M	T4N0M0	IIIA	NA	1. RT to Primary Tumour 2. EGFR TKI	Afatinib	485*	Partial Response
EGFR_10	F	T1bN2M1b	IV	Bone	1. Palliative RT to Spine 2. EGFR TKI	Afatinib then gefitinib	375*	Partial Response
RT_5	F	T4N2M0	IIIB	NA	1. Incomplete surgical resection 2. Chemotherapy 3. RT to chest 4. EGFR TKI	Gefitinib then erlotinib	147*	Partial Response

Table 5-1 Summary of clinical characteristics for patients with *EGFR* mutation positive adenocarcinoma of the lung. Stage of disease based on International Association for the Study of Lung Cancer (IASLC) 7th Edition. * Patient still receiving EGFR TKI at time of final research blood sample. M = Male. F = Female.

5.2.2 Base-line mutational load

5.2.3 Comparison of NGS and ddPCR quantitative assessments

At the time of consent, the base-line levels of mutant *EGFR* in ctDNA, for the 11 patients described in Table 5-1 are summarised in Table 5-2. Using NGS with ITVC software, and ddPCR, it was possible to detect mutant *EGFR* in ctDNA in the blood samples from two of the 11 patients. It must be noted that these two samples contained the highest level of mutant *EGFR* within ctDNA (41.6 and 57.8% alternate allele frequency by ddPCR). Therefore it is expected that both techniques will detect mutant EGFR in ctDNA, since the alternate allele frequencies are well above the limits of detection of NGS and ddPCR.

In contrast, ddPCR, was able to detect mutant *EGFR* in an additional three patients, in whom the mutational abundance was significantly lower (0.42 – 0.57% alternate allele frequency). These alternate allele frequencies are below the lower limit of detection of NGS. This confirms the suggestion, from results chapter 2, that ddPCR is a more sensitive assay than NGS, when assessing for a limited number of specific mutations, at low mutational abundance.

The *EGFR* mutation was not detected in ctDNA in the baseline blood sample for 6 of the patients, using either methodology. Failure to detect mutant *EGFR* in ctDNA may be due to the use of anti-cancer therapies, prior to the first blood sample. At the time of consent and 'base-line' blood sample, two of these patients had received systemic chemotherapy within the preceding 3-week period. Two other patients were already taking EGFR TKi therapy, with responding disease based on radiology. One patient had prior surgical resection of the primary lung cancer with only microscopic residual disease within the pleura. One patient had prior chemotherapy and radiotherapy to the thorax. Even though patient numbers are small, it would not be unreasonable to suggest that a reason for failure to detect mutant *EGFR* within the baseline blood sample, is that successful therapy may have reduced the burden of disease to an extent that levels of mutant *EGFR* in ctDNA were below the level of detection of NGS and ddPCR,

described in Chapter 3. However, this does not mean that all anti-cancer treatments reduce the level of mutant *EGFR* in ctDNA to 'undetectable' levels. Patient EGFR_7 had received prior palliative radiotherapy to the spine, and it was still possible to detect mutant *EGFR* within the baseline blood sample.

Even though the proportion of mutant EGFR detection in baseline ctDNA samples may be low (18.2% positivity using NGS and 45.5% positivity using ddPCR), it may be that there was too little ctDNA to detect using the current technologies as a result of prior therapies. I.e. for the patients included in these results, the baseline blood samples did not represent patients that were truly treatment naive. Therefore a clinician must bear in mind that prior anti-cancer therapies received by a patient, will have a significant impact on the amount of ctDNA, and the chances of a positive result, if a diagnostic EGFR mutation test on ctDNA is required.

	FFPE				Base-line cfDNA sample		
	Pyrosequencing/Fragment Length Analysis	NGS				Mutation detection by NGS	Mutation detection by ddPCR
Patient Research ID	*EGFR* deletion or predicted protein change	DNA Change	Predicted Protein Change	Variant Allele Frequency (%)	Any prior anti-cancer treatment?	Ion Torrent Variant Caller Software	Alternate Allele Frequency (%) (95% Poisson Interval)
EGFR_1	EGFR Exon 19, 12 bp del.	EGFR Exon 19 c.2239_2251del12>CCATTG	EGFR p.L747_T751>PL	33.7	No	No	-
EGFR_2	EGFR p.G719A	Insufficient FFPE Tumour DNA for NGS analysis	-	-	No	No	-
EGFR_3	EGFR p.L858R	EGFR c.2573T>G	EGFR p.L858R	47.8	Yes	Yes	41.6 (38.1 - 45.1)
EGFR_4	EGFR p.L858R	EGFR c.2573T>G	EGFR p.L858R	17	No	No	-
EGFR_5	EGFR Exon 19, 15 bp del.	Insufficient FFPE Tumour DNA for NGS analysis	-	-	No	No	0.57 (0.0 - 1.91)
EGFR_6	EGFR Exon 19, 15 bp del.	EGFR Exon 19 c.2235_2249del15	EGFR p.E746_A750	23.8	No	No	-
EGFR_7	EGFR p.L858R	EGFR c.2573T>G	EGFR p.L858R	20.4	No	Yes	57.8 (55.2 - 60.3)
EGFR_8	EGFR Exon 19, 15 bp del.	EGFR Exon 19 c.2235_2249del15	EGFR p.E746_A750	42	Yes	No	0.42 (0.0 - 0.93)
EGFR_9	EGFR Exon 19, 12 bp del.	EGFR Exon 19 c.2238_2249del15	EGFR p.L747_T751>P	18.5	No	No	-
EGFR_10	EGFR Exon 19, 15 bp del.	EGFR Exon 19 c.2236_2250del15	EGFR p.E746_A750	70.6	No	No	0.53 (0.0 - 1.79)
RT_5	EGFR p.L858R	EGFR c.2573T>G	EGFR p.L858R	37.1	No	No	-

Table 5-2 Summary of *EGFR* mutations in FFPE tumour tissue samples and in cfDNA from base-line, pre-treatment blood samples.

5.2.4 Clinical factors in patients with mutant EGFR detected in baseline ctDNA

Two patients (EGFR_3 and EGFR_8) had not received any prior systemic anticancer therapies. Mutant *EGFR* was detected in ctDNA, by ddPCR and NGS, in the baseline blood sample of patient EGFR_3. This patient had liver metastases and there was much higher mutation abundance (41.6% alternate allele frequency). However, for the other patient (EGFR_8), with pulmonary and pleural metastases, but no extra-thoracic disease, there was a much lower abundance (0.42% alternate allele frequency, detected by ddPCR). NGS was not sufficiently sensitive to detect the mutation in ctDNA.

Two cases had received palliative radiotherapy to the spine, to treat vertebral body metastasis (EGFR_7 and EGFR_10). Mutant *EGFR* within ctDNA was detectable by ddPCR in the baseline blood sample (57.8% and 0.53% alternate allele frequency, respectively). Patient EGFR_10 had metastatic disease only in bones, whereas EGFR_7 also had liver metastases. It is impossible to conclude whether the differences in mutational abundance between these two patients are due to different sites of metastatic disease. For example, it is unknown whether more ctDNA is shed by liver metastases, compared to bone metastases.

One patient (EGFR_5) had their base-line blood sample taken, after the commencement of systemic therapy with gefitinib. The patient had been taking this therapy for over 420 days, and there was a clinical query at the time whether the patient was developing clinical progression of their cancer. The baseline alternate allele frequency for mutant *EGFR* in ctDNA for this patient was 0.57% alternate allele frequency, using ddPCR. Therefore it is unknown whether this represented a nadir of ctDNA during treatment, or whether this was increasing or decreasing as a result of disease response to therapy.

5.2.5 Clinical factors in patients with EGFR mutations in tumour biopsies which were not detected in baseline ctDNA

Mutant *EGFR* was not detected, using NGS or ddPCR, in the first blood sample in 6 patients. Two of these (patients EGFR_1 and EGFR_2) had received chemotherapy within 6 weeks, prior to base-line blood sampling. It may be fair

to argue that systemic chemotherapy can result in a decrease in ctDNA, though further research is needed to confirm this, and determine how long for. This is important, since a negative ctDNA test result should be interpreted with caution, in patients who have recently received chemotherapy.

Patient EGFR_4 had a baseline blood sample, for ctDNA analysis, after surgical resection of the primary lung cancer. At the time of surgery, biopsies were taken of the pleura, which confirmed microscopic adenocarcinoma. There was no detectable metastatic disease on pre- and post-operative CT scans. It is entirely possible that it was not possible to detect mutant *EGFR* within the baseline ctDNA sample due to a small volume of (microscopic) residual disease.

Patient EGFR_6 had leptomeningeal *EGFR*$^{mut+ve}$ adenocarcinoma of the lung, and had already commenced TKi (afatinib) therapy for 4 weeks by the time of their first research blood draw. Decreases in the amount of ctDNA in response to EGFR TKi are reported in the literature, and this may offer the explanation for why mutant *EGFR* was not detected in the first blood sample. However, little is known about whether leptomeningeal or CNS metastatic disease shed tumour DNA that can be detected and analysed in ctDNA in the peripheral circulation.

Patient RT_5, had lung surgery, followed by a course of post-operative chemotherapy, but despite this, developed malignant mediastinal lymphadenopathy. The first research blood draw was taken at this time point. The inability to detect mutant *EGFR* in ctDNA could be a reflection of the prior treatments received, or due to low volume of metastatic disease, or both.

Further studies will need to address the question of the duration of persistence, and rate of decline, of ctDNA after a course of chemotherapy. However, there is nothing in the literature about the effect of radiotherapy on ctDNA levels in *EGFR*$^{mut+ve}$ lung cancer patients. The results from these case studies suggest that it is possible to detect mutant *EGFR* within ctDNA, in a blood sample taken after a course of palliative radiotherapy to the spine. This will information will help the clinician to decide when it may be clinically appropriate to request a blood

sample for *EGFR* analysis on ctDNA. For example, to maximise the chances of detecting mutant *EGFR* in ctDNA, is when the patient has not received any recent anti-cancer therapies, apart from say, palliative radiotherapy to the spine.

5.2.6 Circulating Tumour DNA Kinetics During Treatment with EGFR TKi

5.2.6.1 Quantitative ctDNA and radiological responses to therapy

In the 5 patients with ddPCR-detectable *EGFR* mutations in ctDNA at base-line, we investigated the quantitative changes in alternate allele frequency of *EGFR* mutation, in ctDNA, over time.

Patient EGFR_3 had metastatic *EGFR*$^{mut+ve}$ adenocarcinoma of the lung, with detectable levels of mutant *EGFR* in the baseline ctDNA sample, as assessed by NGS (40.15% alternate allele frequency) and by ddPCR (41.6% alternate allele frequency). The patient was treated with gefitinib therapy and experienced a radiological response to therapy by the time of the CT scan at 70 days into treatment. This coincided with a reduction of mutant *EGFR* DNA in the blood, to undetectable levels (0% alternate allele frequency), using both NGS and ddPCR (Figure 5-1).

This patient (EGFR_3) had a durable radiological response to treatment with gefitinib, as demonstrated by the appearances of stable disease on a CT scan at day 254 compared with the CT scan at day 70. During this period, circulating mutant *EGFR* in ctDNA remained undetectable, suggesting a good correlation between ctDNA levels and radiological assessment and a potential role as a surrogate marker for response assessment.

Using NGS and the 50-gene hotspot panel, this patient's tumour also harboured base-line mutations in genes other than *EGFR* i.e. in exon 6 of *TP53* c.A583T, p.I195F, and in exon 12 of *NPM1* c.T877C, p.S293P. The remaining gene sequences assessed by the gene panel were wild-type. The mutational frequency of the *TP53* and the *NPM1* mutations occurred at lower levels (1.9% and 0.5% alternate allele frequencies, respectively), compared with the *EGFR* mutation.

These mutations may present in lower abundance in ctDNA as it is possible that these mutations are contained within sub-clones of the cancer, and represent a sub-population of cancer cells (intra-tumour heterogeneity). However, In order to confirm, or deny this hypothesis, biopsies from different parts of the primary tumour, and different metastatic sites would need to be assessed with NGS. This is the research strategy of some lab groups investigating tumour intra-patient tumour heterogeneity(136).

For this same patient (EGFR_3), NGS of serial ctDNA samples revealed that levels of mutant *TP53* and *NPM1* also decreased during treatment. Mutant *TP53* decreased to 0% alternate allele frequency by day 60 of treatment. Mutational levels of *NPM1* decreased from 0.5% to 0.24% alternate allele frequency by day 60 of treatment, and remained at that level by the time of the next blood test at 218 days. In the first instance, the changes in *NPM1* mutation level in ctDNA need to be confirmed with ddPCR, which was not performed due to lack of resources.

Each serial ctDNA sample was also assessed by ddPCR, for the emergence of the acquired resistance mutation *EGFR* c.2369C>T, p.T790M, as this accounts for therapeutic resistance in approximately 50% of $EGFR^{mut+ve}$ patients. In keeping with a continuing radiological response to therapy, *EGFR* c.2369C>T, p.T790M, was not detected in this patient, at any point during the assessment period.

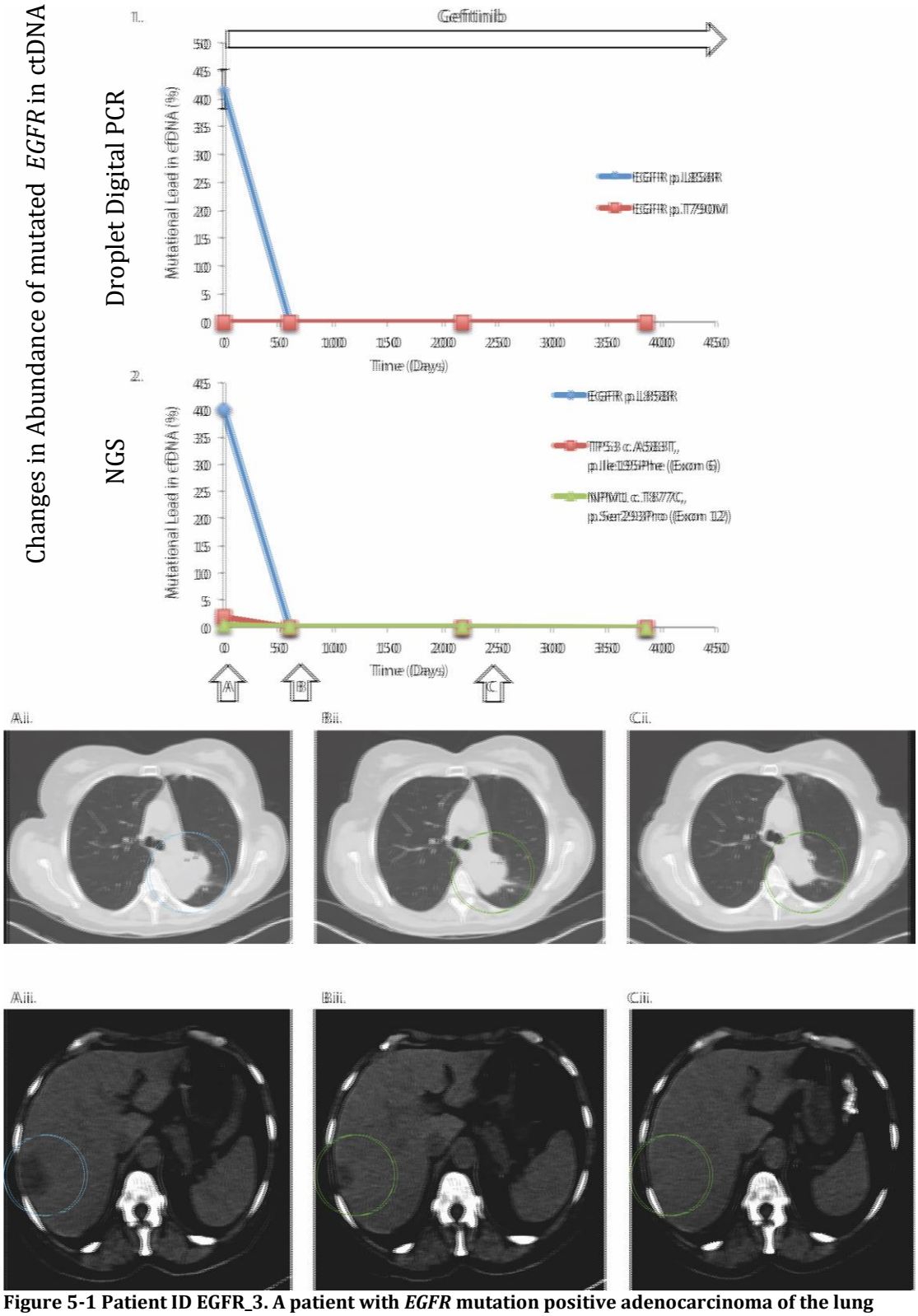

Figure 5-1 Patient ID EGFR_3. A patient with *EGFR* mutation positive adenocarcinoma of the lung responding to gefitinib. Changes in mutational *EGFR* load in ctDNA, using ddPCR (1.) and NGS (2.). The NGS data also show changes in mutant *TP53* and *NPM1* in ctDNA. 95% Poisson Error bars shown in (1.). CT scans are shown of primary tumour (A.i., Bi, Ci), and liver metastasis (A.ii., Bii and Cii) at base-line, day 105 and day 273). Blue circles identify baseline tumours. Green circles identify improving tumours.

Similar correlation between ctDNA kinetics and radiological response to treatment were also seen in patient EGFR_10 (Figure 5-2). This patient had metastatic *EGFR*$^{mut+ve}$ adenocarcinoma of lung with bone metastases. They received initial radiotherapy to a metastatic lesion in the spine to treat impending spinal cord compression, prior to systemic therapy with afatinib, subsequently converted to gefitinib due to hepato-toxicity. The base-line ctDNA sample was taken after radiotherapy and before the commencement of systemic therapy, when the abundance of mutated *EGFR* was 0.53% alternate allele frequency, assessed by ddPCR. The first CT scan to assess response to TKi treatment was performed at day 105 (relative to the start of radiotherapy). This revealed a partial radiological response to therapy, and coincided with a fall in levels of mutant *EGFR* in the blood, decreasing to undetectable levels (0% alternate allele frequency). A subsequent CT scan, performed at day 273, confirmed a sustained radiological response to therapy. The abundance of mutant *EGFR* in blood at the same time also remained at undetectable levels (0% alternate allele frequency by ddPCR). As with patient EGFR_3, the *EGFR* c.2369C>T, p.T790M mutation was assessed as a potential marker of therapeutic resistance, but this remained undetectable in this patient (0% alternate allele frequency, determined by ddPCR).

At the end of the follow-up period, 6 of the 11 *EGFR*$^{mut+ve}$ lung cancer patients (patients EGFR_3, EGFR_4, EGFR_6, EGFR_9, EGFR_10 & RT_5), were still receiving EGFR TKi, and still had undetectable levels of the original activating *EGFR* mutation in ctDNA, determined by ddPCR ("swimmer plot" in Figure 5-4), which corresponded with a period of responding or stable disease.

These findings are consistent with the potential utility of ctDNA as a surrogate marker of both response to EGFR TKi, and monitoring a period of sustained disease response, before the development of therapeutic resistance and disease progression.

Changes in Abundance of mutational EGFR

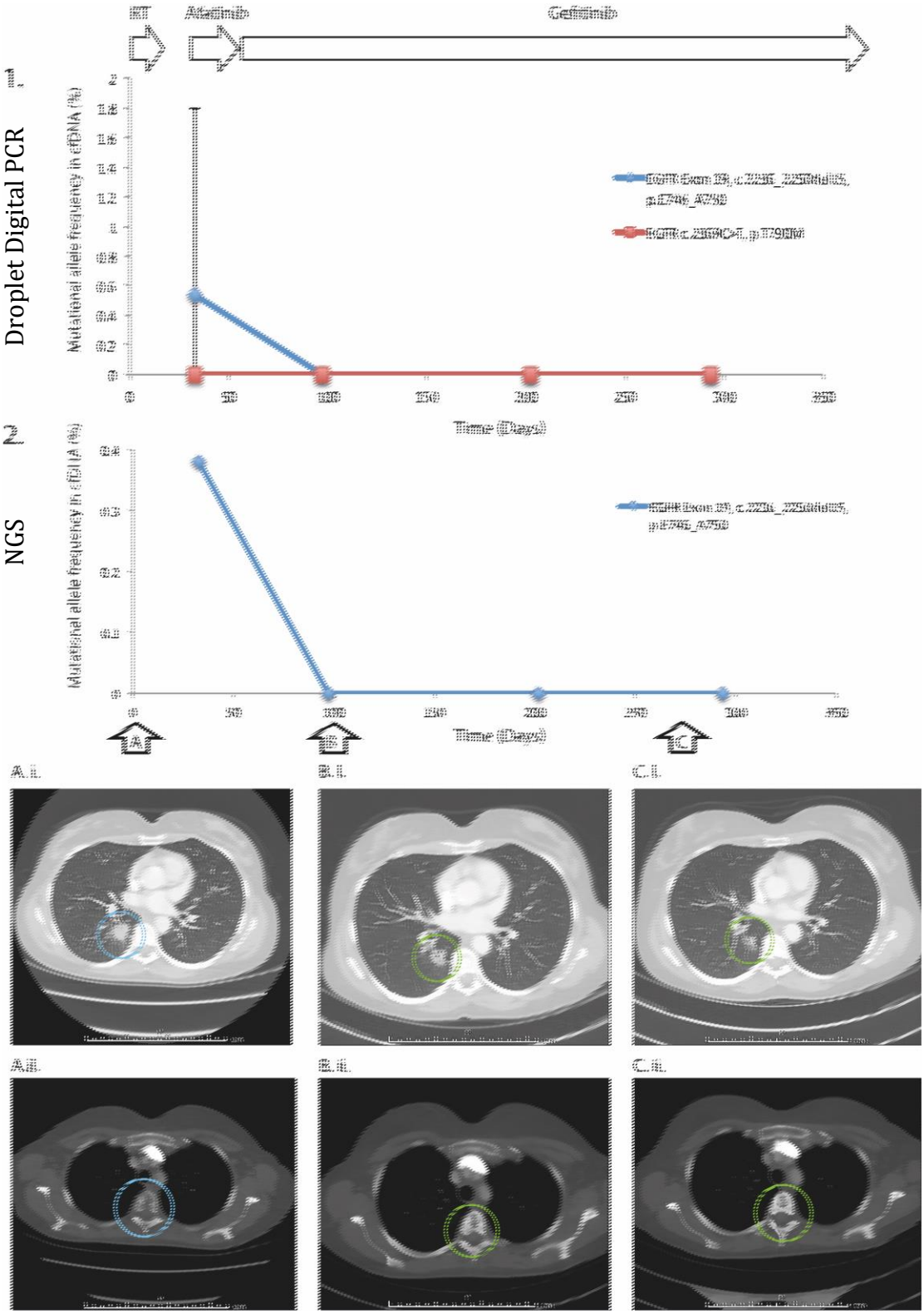

Figure 5-2 CtDNA and response to treatment. Patient EGFR_10 has T1bN2M1b (Bone) metastatic *EGFR*$^{mut+ve}$ adenocarcinoma of lung. CT scans show lung primary and spine lesion at diagnosis (A.i and A.ii.), after 105 days (B.i and B.ii.) and after 273 days (C.i and C.ii). Quantitative decreases in mutant *EGFR* load are shown by ddPCR (1.) and NGS (2.). Error bars represent 95% poisson error. Blue circles identify tumours at baseline. Green circles identify improving or stable tumours.

5.2.7 Quantitative changes in ctDNA and disease progression

By the end of the follow-up period, four of the patients (patients EGFR_1, 2 5 and 8) had undergone radiological or clinical progression of their lung cancer, while receiving EGFR TKi therapy. The mean duration of EGFR TKi treatment was 323 days (range 283 to 373 days), which is in keeping with the median progression free survival data published by the EGFR TKi clinical trials. Figure 5-3 is a "spider plot", illustrating over time, changes in abundance of the original activating *EGFR* mutations in ctDNA, for each of these 4 patients. In all but one of the four patients, it is possible to observe increasing levels of mutant *EGFR* in ctDNA earlier than the diagnosis of radiological or clinical disease progression. This raises the potential that it may be possible to detect lung cancer disease progression earlier, through increases in ctDNA, rather than waiting for scheduled radiological assessment of disease status. This has clear potential benefits for patients – this could give patients confidence to continue with therapy or give them the opportunity to discontinue treatment earlier, avoiding potential cumulative toxicities, and switch to alternative therapies that might be more effective.

To date, there are no established systemic biomarkers in routine clinical use in lung cancer. There is nothing in the literature about changes in therapy for lung cancer disease progression defined by ctDNA increases. This has huge potential for a clinical trial, of whether earlier intervention for patients developing disease progression, defined by increasing ctDNA levels, can improve clinical outcomes. This would be comparable disease monitoring using other systemic surrogate biomarkers, such as alpha feto-protein (AFP)/human chorionic gonadotrophin (hCG) for germ cell tumours, carcinoembryonic antigen (CEA) for colo-rectal cancer and Ca125 for ovarian cancer.

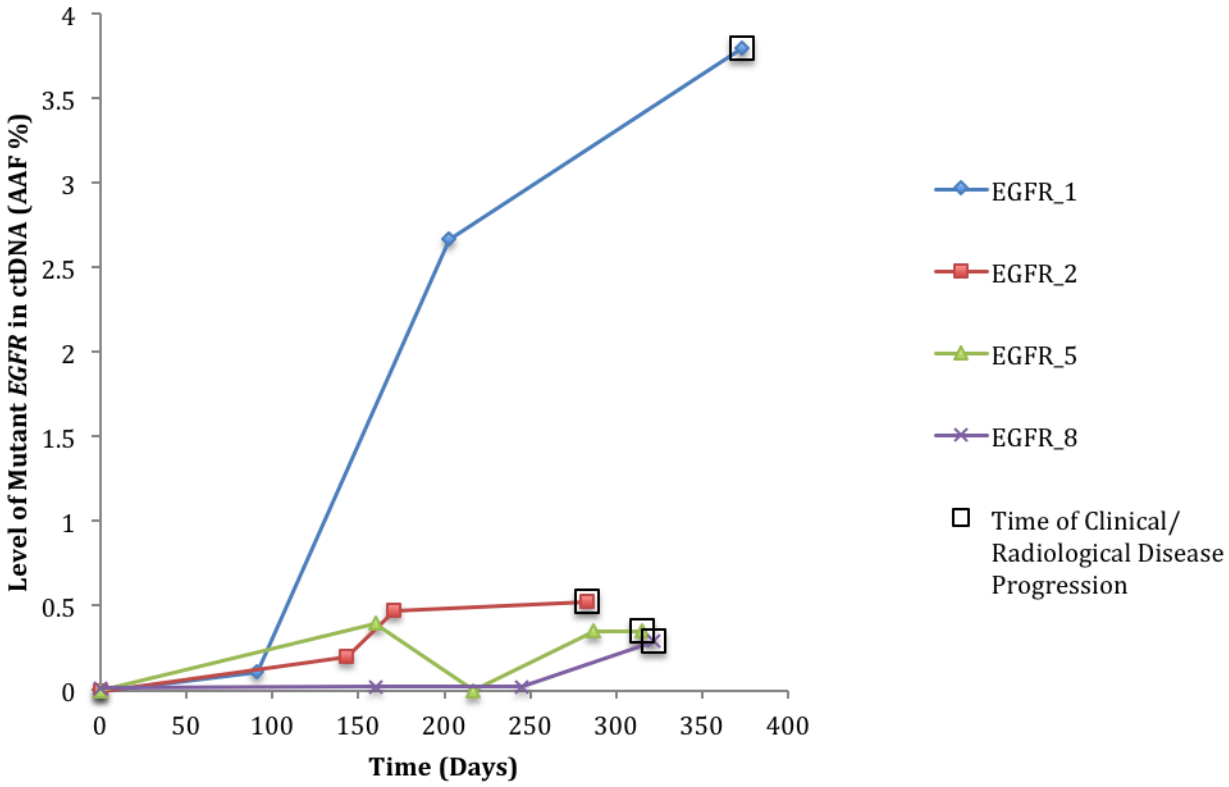

Figure 5-3 'Spider' plot showing changes with time (days), from first ctDNA sample, in mutational *EGFR* load in ctDNA from 4 *EGFR*$^{mut+ve}$ adenocarcinoma patients, assessed by ddPCR. The squares represent the time of the diagnosis of clinical or radiological disease progression. AAF = Alternate Allele Frequency.

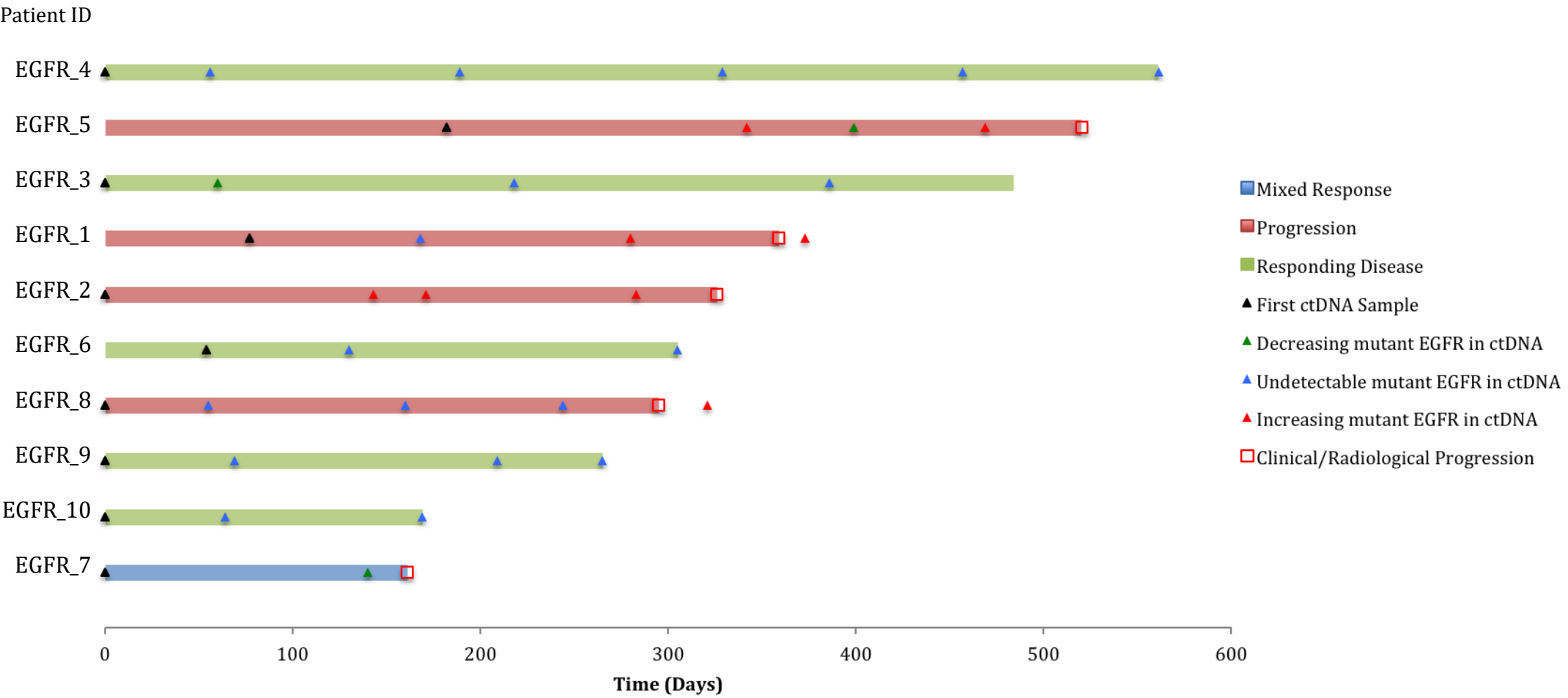

Figure 5-4 Swimmer plot showing ctDNA kinetics and radiological responses in 10 patients with *EGFR*$^{mut+ve}$ adenocarcinoma of the lung, receiving therapy with EGFR TKi. Radiological response to treatment, at the end of the follow-up period is indicated by different coloured bars. The triangles represent each blood test taken for ctDNA analysis, and the trend of abundance of mutational *EGFR* by ddPCR. Squares indicate the diagnosis of clinical or radiological disease progression while taking EGFR TKi therapy.

5.2.8 Qualitative analysis of ctDNA on disease progression

Patient EGFR_1 was treated with afatinib, and achieved a partial radiological response by the time of a CT scan on day 150 of therapy (arrow A in Figure 5-5). The base-line ctDNA blood sample, taken after the start of afatinib therapy, did not show detectable levels of the *EGFR* mutation in ctDNA, by ddPCR or NGS. Radiological disease progression was seen on a CT scan on day 350 (arrow B in Figure 5-5), which correlated with a quantitative increase in the abundance of the original *EGFR* mutation in ctDNA, as assessed using ddPCR reaching 4.72% alternate allele frequency, by day 373. At this time, the patient discontinued afatinib therapy, and received palliative radiotherapy to the left hemi-thorax. Interestingly, the level of the original *EGFR* mutation in ctDNA, determined by ddPCR, fell to 1.73% alternate allele frequency, after completion of radiotherapy, but had further increased to 14% alternate allele frequency, on repeat testing 84 days later, without any treatment in that period, suggesting a response to radiotherapy, followed by further disease progression.

Analysis of ctDNA using NGS and the 50 gene cancer hotspot panel, demonstrated an increasing abundance of a *CDKN2A* mutation on disease progression, as well as the *EGFR* mutation. The same *CDKN2A* mutation was present in the diagnostic biopsy, but not on the base-line ctDNA sample, taken during afatinib therapy (Figure 5-5). This mutation may have been present in ctDNA at the start of afatinib therapy, but levels declined to 0% prior to the time of the first blood sample for ctDNA analysis. NGS also demonstrated the appearance of a mutation that was not present in either the original tumour tissue biopsy or in base-line ctDNA sample, specifically a *TP53* c.637C>T, p.R213Stop, mutation (Figure 5-5.2 and 5-5.3). Therefore, in addition to tracking quantitative changes in levels of mutations in ctDNA, in response to different anticancer therapies, it is also possible to assess for the emergence of previously undetected mutations within ctDNA. These mutations may have been missed in the original tumour biopsy, due to tumour intra-heterogeneity, or they could develop during therapy, due to clonal evolution.

Detecting 'emerging' mutations in serial ctDNA samples, can offer insights into potential mechanisms of acquired resistance to therapy. Mutated TP53 is not currently a 'clinically actionable' target, since, unlike EGFR, there are no drugs which are effective against the aberrant protein derived from mutated forms of the gene. Therefore, *TP53* mutations are not currently assessed in routine lung cancer clinical practice. However, there are increasing numbers of molecularly-agents that are active against mutations which confer acquired resistance on tumours. For example, osimertinib is effective in the treatment of the acquired *EGFR* c.2369C>T, p.T790M mutation commonly seen in EGFR TKi resistance. Therefore, routine assessment, for the emergence of this *EGFR* mutation, by serial ctDNA analysis, could easily be incorporated into routine patient monitoring.

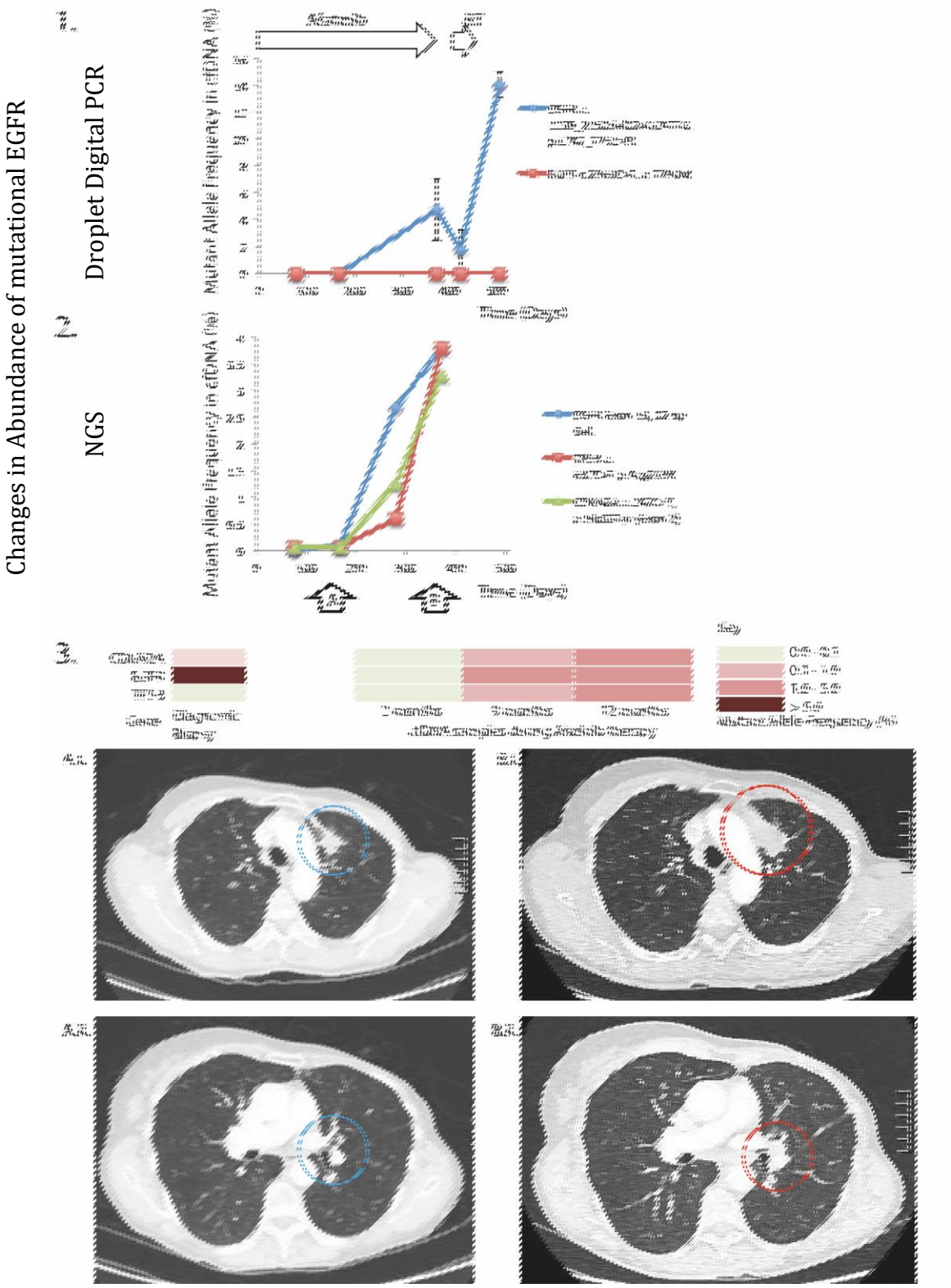

Figure 5-5 Patient EGFR_1. Molecular vs radiological assessments of disease progression during EGFR TKi therapy. Changes in levels of mutations in ctDNA are shown by ddPCR (1.) and NGS (2.), and a heatmap (3.).CT scan at day 150 of therapy (arrow A) revealing the lung primary (A.i) and the right hilar lymphadenopathy (A.ii). CT scan at day 350 (arrow B) revealing disease progression in both sites (B.i and B.ii). Blue circles highlight tumours after response to EGFR TKi therapy and red circles highlight growing tumours.

5.2.9 Screening for emergence of T790M using ddPCR

The presence of the *EGFR* c.2369C>T, p.T790M mutation, using NGS and ddPCR, was assessed in all 39 serial plasma samples from 10 patients. By the end of the follow-up period, six of the 10 patients had developed drug resistance/progressive disease on EGFR TKi therapy. In two of the patients who had progressive disease (33%), it was possible to detect the *EGFR* c.2369C>T, p.T790M acquired resistance mutation, using ddPCR, at alternate allele frequencies of 1%. It is expected that acquired *EGFR* c.2369C>T, p.T790M mutations emerge in approximately 50% of patients with $EGFR^{mut+ve}$ adenocarcinoma of the lung who develop resistance to EGFR TKi therapy.

For the other 4 patients with disease progression, it was not possible to demonstrate the *EGFR* c.2369C>T, p.T790M mutation. Obtaining a negative result for this mutation may be due to several reasons. Firstly, there may different mechanism of acquired resistance (expected in approximately 50% of $EGFR^{mut+ve}$ lung cancer patients). Secondly, a false negative result may also result from technical limitations, such as the limit of detection of current technologies, including droplet digital PCR (see Chapter 3). Thirdly, the progressive metastatic disease may not shed DNA into the peripheral venous circulation, and therefore cannot be assessed via ctDNA.

5.2.10 Disease progression within the CNS

Three $EGFR^{mut+ve}$ patients (EGFR_5,6 and 8) had CNS metastases. Patient EGFR_6 had meningeal disease at the time of diagnosis, while brain metastases were detected in patients EBFR_5 and 8, at the time of clinical/radiological progression of their lung cancer. It is impossible to exclude the presence of cerebral metastatic disease at diagnosis in these latter 2 patients, because baseline radiological imaging of the brain is not standard practice in the UK, unless there is a clinical indication, such as focal neurological signs.

At diagnosis, Patient (EGFR_5) had liver and lung metastases. A partial radiological response to primary palliative gefitinib therapy was seen in the primary lung tumour on CT scan at day 168 of therapy. At this time, it was also possible to detect mutant *EGFR* mutation (Exon 19, c.2236_2250del15, p.E746_A750) in ctDNA by ddPCR, at an alternate allele frequency of 0.57% (95% Poisson Error 0 – 1.91%), but not by NGS (Figure 5-6). The patient continued therapy with gefitinib for a further 259 days, at which time, the patient was admitted to hospital and treated for pneumonia. Neurological decline in this patient prompted a MRI scan of the head, shortly afterwards, (day 451) which revealed the presence of metastatic disease within the brain.

NGS and ddPCR analysis of ctDNA at the time of the MRI, revealed a low level of the original activating *EGFR* mutation. In addition, there may have been a concurrent low level *EGFR* c.2369C>T, p.T790M mutation detected in ctDNA by ddPCR (<0.5% alternate allele frequency). However, the Poisson error included 0%, which means this could represent a false-positive result. It is not known whether this patient had cerebro-metastatic disease at the time of diagnosis, and therefore, it is unknown whether this had progressed or responded during afatinib therapy. There is too little data from this patient to determine whether changes in levels of *EGFR* mutation or the potential appearance of *EGFR* c.2369C>T, p.T790M, in ctDNA represents tumour progression within the CNS.

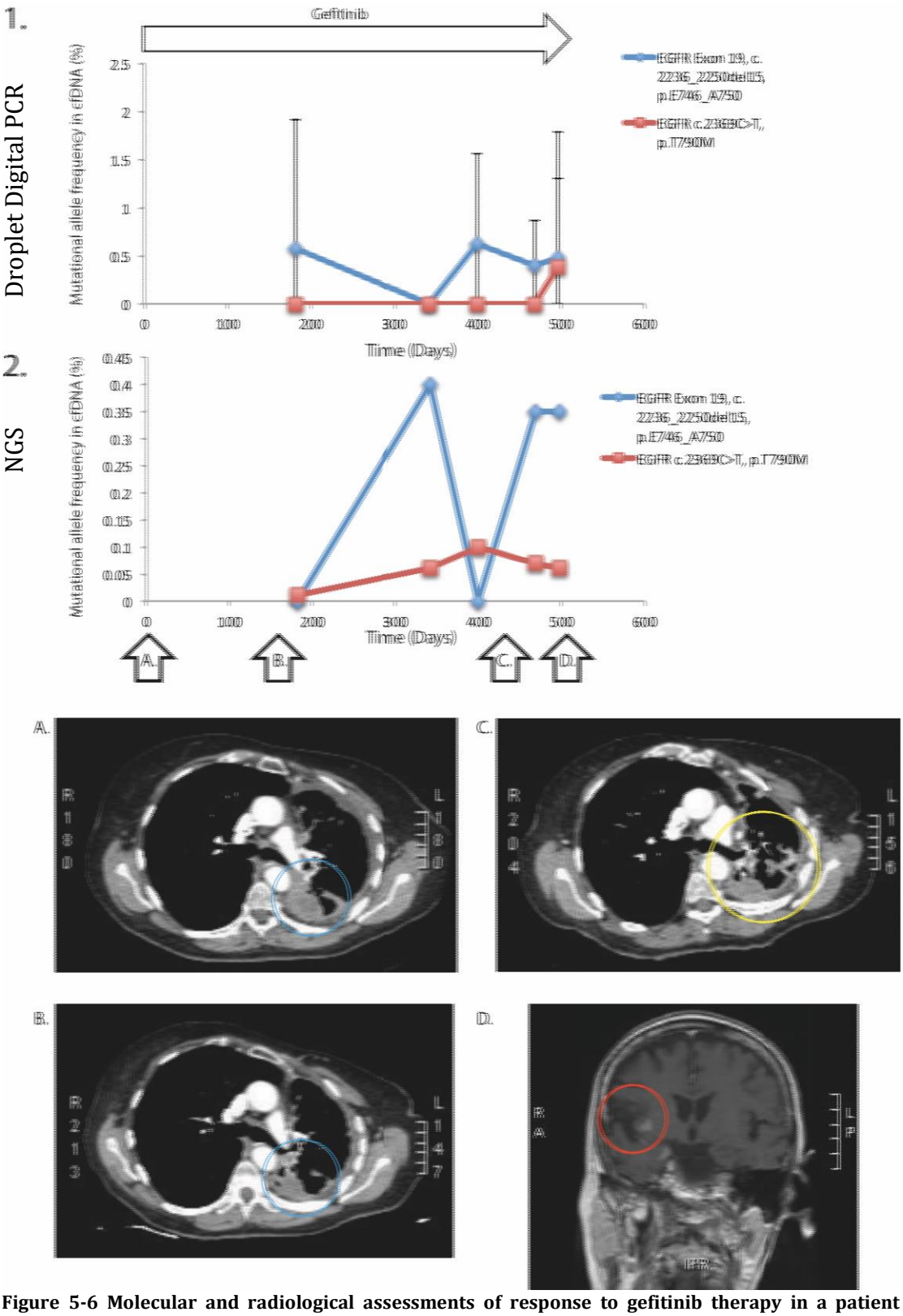

Figure 5-6 Molecular and radiological assessments of response to gefitinib therapy in a patient (EGFR_5) with metastatic CNS disease. Changes in mutant *EGFR* level in ctDNA, using ddPCR (1.) with 95% Poisson error bars, and NGS (2.).Baseline CT scan showing the primary lung tumour (A.). CT scan on day 168 (B.). CT scan at day 427 of therapy (C). Sagittal section from MRI head on day 451 (D.). Blue circles show lung tumour. Yellow circle show lung tumour and possible disease progression. Red circle shows brain metastasis.

A decrease in the level of the original activating *EGFR* mutation in ctDNA (Exon 19, c.2235_2249del15, p.E746_A750) was observed in one patient (EGFR_8), by day 56 of afatinib therapy (from 0.44% to 0% alternate allele frequency determined by ddPCR). Levels remained undetectable by ddPCR at the time of a CT scan (day 185), which also confirmed radiological response to treatment. This patient developed neurological deterioration prompting MRI imaging of the patient's brain on day 316. Radiological images showed a lesion in the left posterior occipital lobe, suspicious of a metastatic lesion, and hydrocephalus, consistent with malinant CNS disease. At this time, comparing ctDNA analysis using ddPCR and NGS, the abundance of the original *EGFR* mutation in ctDNA remained at 0% alternate allele frequency, utilising ddPCR, but there was an increase observed using NGS (from 0% to 0.3% alternate allele frequency) (Figure 5-7). This patient experienced clinical progression of their CNS disease, yet the increase in mutational level of the EGFR mutation in ctDNA was not consistent between the two platforms. Therefore there is not enough evidence with this case study, that changes in mutation level in ctDNA reflect CNS disease progression.

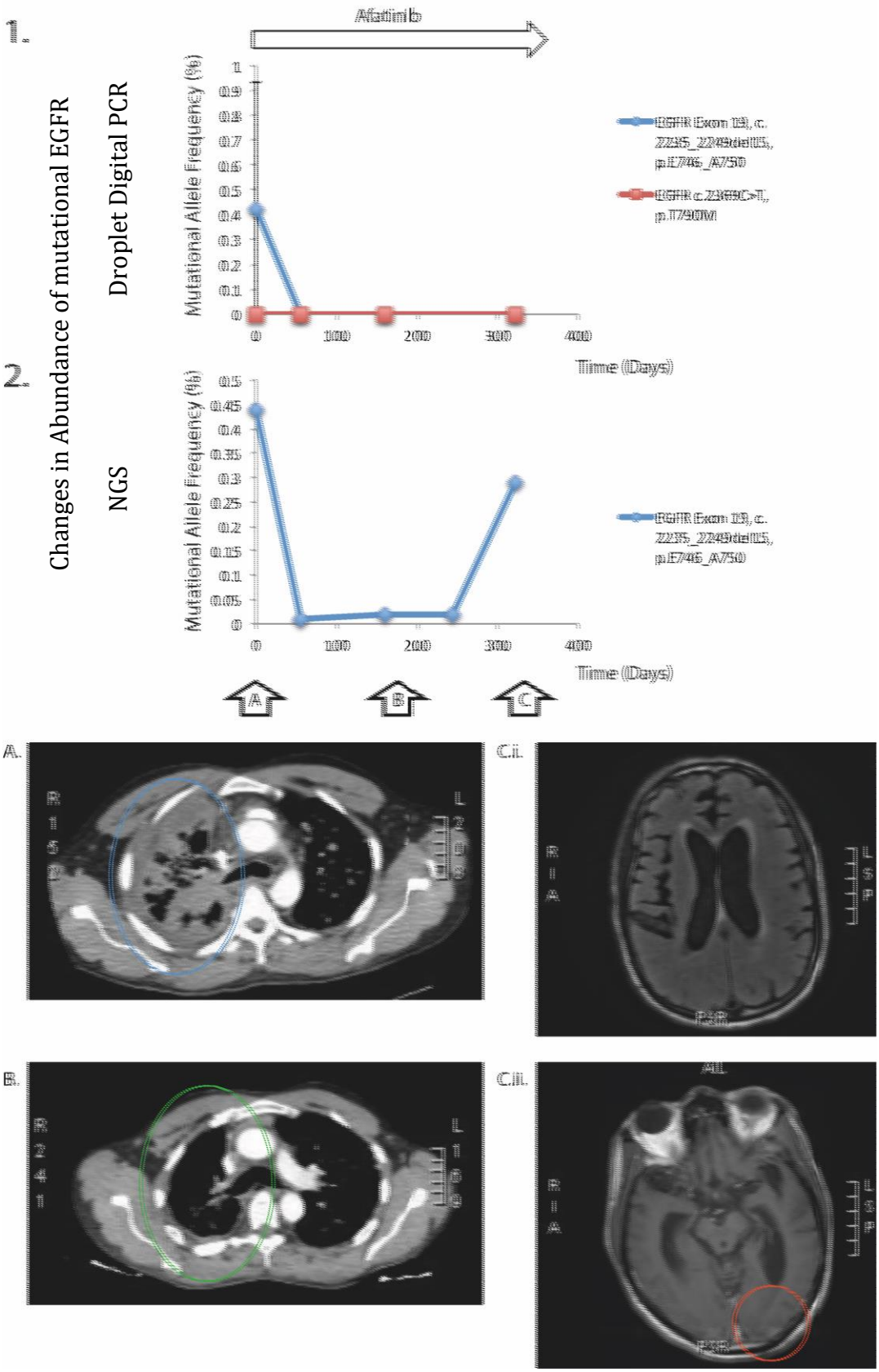

Figure 5-7 Patient EGFR_8. Molecular and radiological assessments of response to afatinib therapy in a second patient with metastatic CNS disease. Changes in mutant *EGFR* load in the blood are shown by ddPCR with 95% Poisson error bars (1.) and by NGS (2.).Baseline CT scan of lung tumour (blue circle) (A.). CT scan at day 185 (B.) showing radiological response to therapy (green circle). The two axial MRI images of the patient's brain on day 316 (arrow C), show hydrocephalus (C.i.) and a suspicious metastatic lesion in the left posterior occipital lobe (red circle) (C.ii.).

To help to further assess whether tumour-derived DNA may be able to cross the blood:brain barrier, permitting detection in blood samples as ctDNA, we investigated whether ctDNA is present in cerebro-spinal fluid (CSF) of a patient with CNS metastases. Patient EGFR_6 had leptomeningeal disease and bone metastases present at diagnosis. During the period of disease control with afatinib, the abundance of the original activating *EGFR* mutation contained within ctDNA remained at 0% alternate allele frequency, determined by ddPCR. Comparison of blood ctDNA and cell free DNA within CSF at the time of disease progression (day 420 of therapy) revealed the presence of the original activating *EGFR* mutation within CSF, but not within ctDNA. Both cell free DNA CSF, and ctDNA analysis were negative for the *EGFR* c.2369C>T, p.T790M resistance mutation, using ddPCR. This data provides a proof of principle that it is possible to identify *EGFR* mutations within cell free DNA contained within CSF, in an *EGFR*$^{mut+ve}$ lung cancer patient with CNS disease. It also raises the query of whether tumour DNA derived from CNS disease, fails to reach the peripheral venous circulation.

5.3 Clinical Implications

Data presented here demonstrate that it is possible to detect various *EGFR* mutations in ctDNA in patients with locally advanced and metastatic *EGFR*$^{mut+ve}$ adenocarcinoma of the lung. This has important implications for monitoring disease during therapy, and at disease progression, as well as in the diagnostic setting, where there is often little or no tumour tissue available for *EGFR* mutational analysis.

In this results chapter, patients with mutant *EGFR* detected in ctDNA experienced a clinical response to EGFR TKi. A blood test for ctDNA analysis is an attractive way to detect the presence of an *EGFR* mutation, especially due to the high failure rate of *EGFR* mutational analysis on a patient's original tumour biopsy, or lack of biopsy of the primary tissue altogether. Therefore, detecting *EGFR* mutations in ctDNA, could have the potential to increase the successful detection of *EGFR*$^{mut+ve}$ adenocarcinoma of the lung, and thereby supporting the use of EGFR TKi as a therapeutic option for these patients.

Understanding the clinical factors that influence the chances of detecting mutant *EGFR* within ctDNA is very important. The clinician needs to know the best time to take a blood sample from a patient, to optimise the chances of detecting mutant *EGFR* if it is present. The results in this chapter would suggest that it is possible to detect mutant *EGFR* within ctDNA when the patient is treatment naive, or when the patient has received prior palliative radiotherapy to the spine. If the patient has received recent radical radiotherapy, lung surgery or systemic chemotherapy, then the chances of detecting mutant *EGFR* in ctDNA could be less. Even though this is useful information for the clinician requesting *EGFR* mutation testing on ctDNA, these findings would need further validation in a larger study.

Data presented in this chapter also support the hypothesis that the kinetics of *EGFR* mutation load in serial ctDNA samples can be useful in monitoring therapy, both in assessing response and detecting progression, in conjunction with (and perhaps preferentially over) classical radiological assessment and may provide a means for

earlier and more sensitive detection of response to treatment and/or progressive disease/emergence of resistance to therapy.

These quantitative changes have been shown to correlate with radiological response to therapy with serial scans. An interesting future research question is to determine whether ctDNA kinetics could be a valid predictive biomarker for expected duration of response to EGFR TKi. This may influence clinical management and follow-up of these patients in that patients who are predicted to have a shorter duration of disease response may require closer monitoring for signs of disease progression.

Patients often undergo a period of effective disease control on EGFR TKi therapy. For the majority of the patients in this cohort, the duration of radiological stable disease was reflected in the consistent low level, or inability to detect, the mutant *EGFR* load in the blood. It would be interesting to investigate further whether consistently undetectable mutant *EGFR* levels, during a period of EGFR TKi therapy, may accurately reflect the period of disease control, and may help to potentially reduce the need for repeated CT scanning during this period.

We have also discovered that this is possible to detect increasing levels of mutated ctDNA up to 2 months before clinical disease progression. These finding have been confirmed by a larger study(137) (Zheng et al, 2016), who monitored mutant *EGFR* in ctDNA, in 117 patients with acquired resistance to EGFR TKi therapy. The authors reported that 47% of 117 *EGFR*$^{mut+ve}$ patients treated with EGFR TKi, developed the acquired *EGFR* c.2369C>T, p.T790M mutation, up to a median 2.2 months prior to clinical progression of their cancer(137). This has important implications for patient care. There may be the opportunity for an earlier change in therapeutic strategy, before the patient deteriorates clinically, as a result of disease progression. This interesting clinical question could be answered by a prospective clinical trial. Specifically whether an earlier change in therapeutic strategy, based on ctDNA progression, may offer superior patient outcomes, compared to those for whom therapy is changed upon radiological or clinical disease progression.

Previous research by Siravegna et al (2015) demonstrated that it is possible to observe the emergence of a *KRAS* mutation (a negative predictor of response) in ctDNA, in colorectal cancer patients treated with anti-EGFR (cetuximab) therapy. These mutations then become undetectable in ctDNA when cetuximab is stopped, and re-appear when treatment is recommenced(138). In line with this, the results in this chapter confirm that it is possible to observe the appearance of the EGFR TKi resistance mutation *EGFR* c.2369C>T, p.T790M, in patients with *EGFR*$^{mut+ve}$ adenocarcinoma of the lung. However, more research is required to determine whether the level of *EGFR* c.2369C>T, p.T790M mutation in ctDNA, decreases on cessation of EGFR TKi therapy. It is possible that ctDNA kinetics could be used in the assessment of continuous versus intermittent EFGR TKi therapy, and whether intermittent therapy may improve the duration of disease control.

NGS of ctDNA using the 50-gene cancer hotspot panel can reveal the emergence of new mutations, associated with disease progression, during or following EGFR TKi therapy. Detecting the acquired *EGFR* c.2369C>T p.T790M 'TKi resistance' mutation, in ctDNA has the advantage of avoiding the need for a repeat biopsy of the tumour tissue, with associated complications, and means that patients can be switched to osimertinib therapy. At present, a few diagnostic laboratories in the UK are implementing ctDNA analysis in routine clinical practice for some specific indications, namely screening for *EGFR* mutations in ctDNA at the time of diagnosis, for patients that either do not have a biopsy, or *EGFR* analysis of tumour tissue DNA has failed. These laboratories are also able to offer screening for the acquired *EGFR* c.2369C>T, p.T790M resistance mechanism, as this may more precisely define the therapeutic options for patients e.g. to direct the clinician towards recommending osimertinib, rather than gefitinib.

As more 'targeted therapies' become available for patients with locally advanced or metastatic lung cancer, detecting multiple potential genetic mutations in ctDNA, to direct these therapies, is of increasing importance. This can be achieved using NGS with a gene panel, provided the mutational level within ctDNA is above the limit of detection of NGS.

5.3.1 Progression of metastatic CNS disease

It has been possible to detect mutant *EGFR* in cell free DNA in CSF from a patient with progressing leptomeningeal disease, but not in ctDNA in peripheral venous blood. Subsequently, when a patient develops progressive disease within the CNS during therapy with an EGFR TKi, and it is not possible to detect the resistance conveying *EGFR* c.2369C>T, p.T790M mutation in ctDNA, it is worth considering testing cell free DNA within the CSF, provided a lumbar puncture is not contra-indicated. However, even though it is possible to detect mutant *EGFR* within cell free DNA within CSF, clinical response to osimertinib must be validated in a greater number of patients with CNS disease harbouring this mutation.

The inability to detect mutant *EGFR* in ctDNA in patients with CNS disease may be explained if the tumour DNA does not cross the blood-brain-barrier. However, for two patients in this results chapter (EGFR_5 and EGFR8) who had confirmed CNS metastatic disease at the time of clinical disease progression, there is not enough data to contribute to this hypothesis, and is the subject of ongoing research.

6 Circulating tumour DNA as a Prognostic Biomarker in Patients with Small Cell Lung Cancer

6.1 Chapter Overview

There is an important, unmet clinical need for valid surrogate biomarkers, which could offer prognostic information, and help to stratify SCLC patients between existing anticancer therapies, and in trials of new therapeutic approaches. In aiding careful patient selection and decision-making for individual patients, they may increase efficacy and simultaneously help to avoid potentially toxic treatment, especially as this type of lung cancer carries a very poor prognosis. We set out to determine the feasibility of ctDNA as a biomarker in patients with advanced small cell lung cancer (SCLC), as there is little in the literature on this topic.

6.2 Chapter Aims

1. To determine the feasibility of detecting and quantifying mutations contained within ctDNA, in patients with small cell lung cancer.
2. To monitor quantitative changes in ctDNA with time, as patients undergo anti-cancer therapies.
3. To compare base-line levels of ctDNA mutations with survival, and determine potential utility as a prognostic biomarker.
4. To determine feasibility of using base-line ctDNA levels as a basis for individual patient treatment stratification.

6.2.1 Chapter Objectives

1. Tumour tissue obtained from the diagnostic tumour biopsy of SCLC patients, will be analysed using NGS to detect somatic mutations.
2. Analysis of ctDNA, using ddPCR and NGS is performed on serial blood samples collected from patients with small cell lung cancer, as they undergo chemotherapy.

3. Levels of ctDNA are correlated with survival of the patients, and a cut-off ctDNA level can be determined using Receiver Operator Curves (ROC).
4. It is considered whether baseline levels of ctDNA can be used as a prognostic biomarker, and whether this information can be used to stratify patient treatments, in the case of small cell lung cancer, this can include whether patients should receive prophylactic cranial irradiation (PCI).

6.3 Chapter Results

6.3.1 Patient recruitment and clinical data

Ten patients with SCLC were recruited to this part of the study. Patients' clinical characteristics, including stage of disease, sites of metastases, therapeutic strategy and clinical outcome are summarised in Table 6-1. Eight of the 10 patients had stage IV disease and two had stage IIIB. One of the stage IV patients died before receiving any palliative chemotherapy (SCLC_3), but all of the other 9 patients underwent primary palliative chemotherapy. Carboplatin and Etoposide was the most common regimen used, with an average of 4 cycles (range 3 to 6 cycles). There is a 85-89% response rate to first-line treatment of SCLC(139), and all of the 9 patients who received chemotherapy, had an initial radiological response to treatment. In 8 of these 9 patients, initial radiological responses to chemotherapy were sustained at the time of repeat staging CT scan on completion of treatment.

At the end of chemotherapy, the most common subsequent treatment was consolidation radiotherapy to the thorax. Two patients with confirmed brain metastases at diagnosis also received therapeutic cranial irradiation, after completion of first line palliative chemotherapy. Prophylactic cranial irradiation (PCI) was a potential therapeutic option for the other 7 patients, but only 3 were treated in this way.

Although all of the treated patients achieved an initial radiological response to chemotherapy, the survival times were very variable (range 86 – 477 days). As is often the case in SCLC, the diagnosis of progressive or relapsed disease usually occurred shortly before death, therefore progression-free and overall survival are very similar, in these patients. Overall survival data were calculated from date of consent to this study (i.e. pre-treatment) to death.

Patient Research ID	TNM	Stage	Sites of Metastases	Treatment Intent	Chemotherapy	Initial Radiological Response	Subsequent Therapy	Cranial Irradiation	OS (Days)
SCLC_1	T2aN2M1b	IV	Brain	Palliative	Carboplatin/Etoposide	PR	Palliative RT to Brain	TCI	242
SCLC_2	T3N2M1b	IV	Bone, Liver	Palliative	Carboplatin/Etoposide	PR	Consolidation RT to chest, palliative RT to spine	PCI	165
SCLC_3	T2bN3M1b	IV	LN, Lung, Liver, Adrenal	Best Supportive Care	Nil	NA	NA	NA	2
SCLC_4	T4N2M0 (Oesophageal Encroachment)	IIIB	No Distant Metastases	Palliative	Cisplatin/Etoposide	PR	Consolidation RT to chest	PCI	477*
SCLC_5	T2aN3M1b	IV	Liver	Palliative	Carboplatin/Etoposide	PR	Consolidation RT to chest	PCI	201
SCLC_6	T2aN3M1b	IV	Adrenal	Palliative	Carboplatin/Etoposide	PR	RT for SVCO	NO	163
SCLC_7	T4N3M1b	IV	Bone, Liver, Possible Brain	Palliative	Carboplatin/Etoposide	PR	Nil	NO	171
SCLC_8	T4N3M1b	IV	Bone, Liver, Adrenal	Palliative	Carboplatin/Etoposide	PR	Nil	NO	86
SCLC_9	T3N1M1b	IV	Brain, Adrenal, Axillary LN	Palliative	Carboplatin/Etoposide	PR	Consolidation RT to chest	TCI	246*
SCLC_10	T4N2M0	IIIB	No Distant Metastases	Palliative	Carboplatin/Etoposide	PR	Consolidation RT to chest	NO	201*

Table 6-1 Summary of clinical characteristics for patients with Small Cell Lung Cancer. *Patients still alive at time of final data collection. LN = Lymph Nodes. NA = Not Applicable. OS = Overall Survival (Days). PCI = Prophylactic Cranial Irradiation. PD = Progressive disease. PFS = Progression Free Survival (Days). PR = Partial Response. RT= Radiotherapy. SVCO = Superior Vena Cava Obstruction. TCI = Therapeutic Cranial Irradiation.

6.3.2 Molecular Characterisation of Tumour DNA

Nine of the 10 patients had sufficient tumour tissue available for NGS analysis using the 50 gene cancer hotspot panel, used in previous chapters. The likely pathogenic mutations revealed are listed in Table 6-2. The most common somatic mutations observed were in the *TP53* gene, observed in 8 of 9 tumour samples. This is consistent with the previous literature(135). Pathogenic mutations in *KRAS* were observed in 2 patients, and one point mutation each of *RB1* and *SMAD4*.

Patient Research ID	Pathogenic Mutations	Alternate Allele Frequency in Tumour Tissue (%)	Potential Clinical Significance
SCLC_1	TP53 c.742C>T, p.R248W	97.2	1,2
SCLC_2	TP53 c.1001G>T, p.G334V	31.0	1,2
SCLC_3	TP53 c.844C>G, p.R282G	40.2	1,2
	KRAS c.34G>T, p.G12C	18.1	3
SCLC_4	TP53 c.451C>A, p.P151T	43.8	1
	RB1 c.1700C>T, p.S567L	36.4	2
SCLC_5	No Mutations Detected	-	4
SCLC_6	TP53 c.592G>T, p.E198TER	64.4	1,2
SCLC_7	*TP53 c.814G>T, p.V272L	40.6*	1,2
	*SMAD4 c.1081C>T, p.R361C	53.5*	
SCLC_8	TP53 c.641A>G, p.H214R	98.9	1,2
SCLC_9	KRAS c.437C>T, p.A146V	45.2	3
SCLC_10	TP53 c.892G>T, p.E298TER	80.7	1,2

Table 6-2 Summary of pathogenic mutations identified in tumour samples/ctDNA from patients with Small Cell Lung Cancer, identified by NGS. * Mutations detected in ctDNA, due to insufficient tumour tissue. 1. Sensitivity to PARP/WEE1 inhibitors. 2. Potential target for immunotherapy. 3. Clinical response to CDK4/6 inhibitor. 4. Not Determined.

For one patient (research ID SCLC_7) there was insufficient FFPE tumour tissue for DNA extraction and analysis. However, 2 pathogenic mutations were detected, using NGS, in a base-line ctDNA sample from this patient, one in *TP53* c.814G>T, p.V272L (alternate allele frequency of 40.6%), and another mutation in *SMAD4* c.1081C>T, p.R361C (alternate allele frequency 41.1%). This specific *SMAD4* mutation has not been previously documented in SCLC, but loss of function of the tumour suppressor gene, *SMAD4,* has previously been reported in non-small lung cancer, and this may be linked to increases in chemo-sensitivity to DNA topoisomerase inhibitors(140), such as etoposide. It is impossible to draw conclusions from a single patient, but it would be interesting to identify other patients whose tumours or ctDNA contain this mutation and determine whether this mutation may convey increased sensitivity to etoposide and have potential use as a predictive biomarker.

SCLC_7 had an initial tumour response to chemotherapy, based on initial plain chest radiograph imaging, rather than cross-sectional CT scans. However, a subsequent CT scan, on completion of treatment, demonstrated that, although the primary lung tumour had increased in size compared with the baseline CT scan, the liver metastasis had almost completely resolved. This is a prime example of intra-patient tumour heterogeneity. The best way to characterise whether there were different 'branch' mutations contained within the lung tumour, compared to the liver metastases, would be to perform NGS of tumour tissue samples from the lung and the liver. It is widely accepted that ctDNA could represent tumour DNA from all tumour sub-clones contained within the body. Had there been tumour tissue available for NGS analysis for SCLC_7, it may have been possible to determine whether any of the mutations, contained within ctDNA, were derived from different tumour sub-clones, and whether this may account for the differences in radiological response observed between the lung and liver tumours.

For one patient (patient ID SCLC_5), despite adequate NGS sequence analysis of the FFPE tumour tissue DNA, and baseline ctDNA, no pathogenic mutations were discovered, using the targeted 50 gene cancer hotspot panel. It is therefore

highly likely that the mechanism of tumorigenesis is the result of a 'driver' mutation that is not contained within this NGS gene panel.

6.3.3 Analysis of circulating tumour DNA

In addition to the FFPE tumour tissue samples, above, a total of 38 research blood samples were obtained, including a base-line blood sample from each of the 10 patients. Serial samples were also collected from 9 of these 10, during the course of treatment, for comparison with clinical data (radiological response and overall survival). There were between 1 and 7 serial blood tests per patient (median 4 samples per patient).

It was not always possible to obtain a research blood sample following the clinical or radiological diagnosis of disease progression, as patients were sometimes too unwell, or hospitalised, and blood sampling for research was either not appropriate or not feasible.

6.3.3.1 Comparison between NGS and ddPCR detection of mutations in ctDNA

NGS was useful to detect pathogenic mutations, but once known, it was more cost effective to track ctDNA kinetics using ddPCR rather than NGS. This was deemed acceptable as there was good correlation in alternate allele frequency between the two molecular platforms, as assessed by Bland-Altman plots (Figure 6-1), for both FFPE tumour tissue DNA and ctDNA.

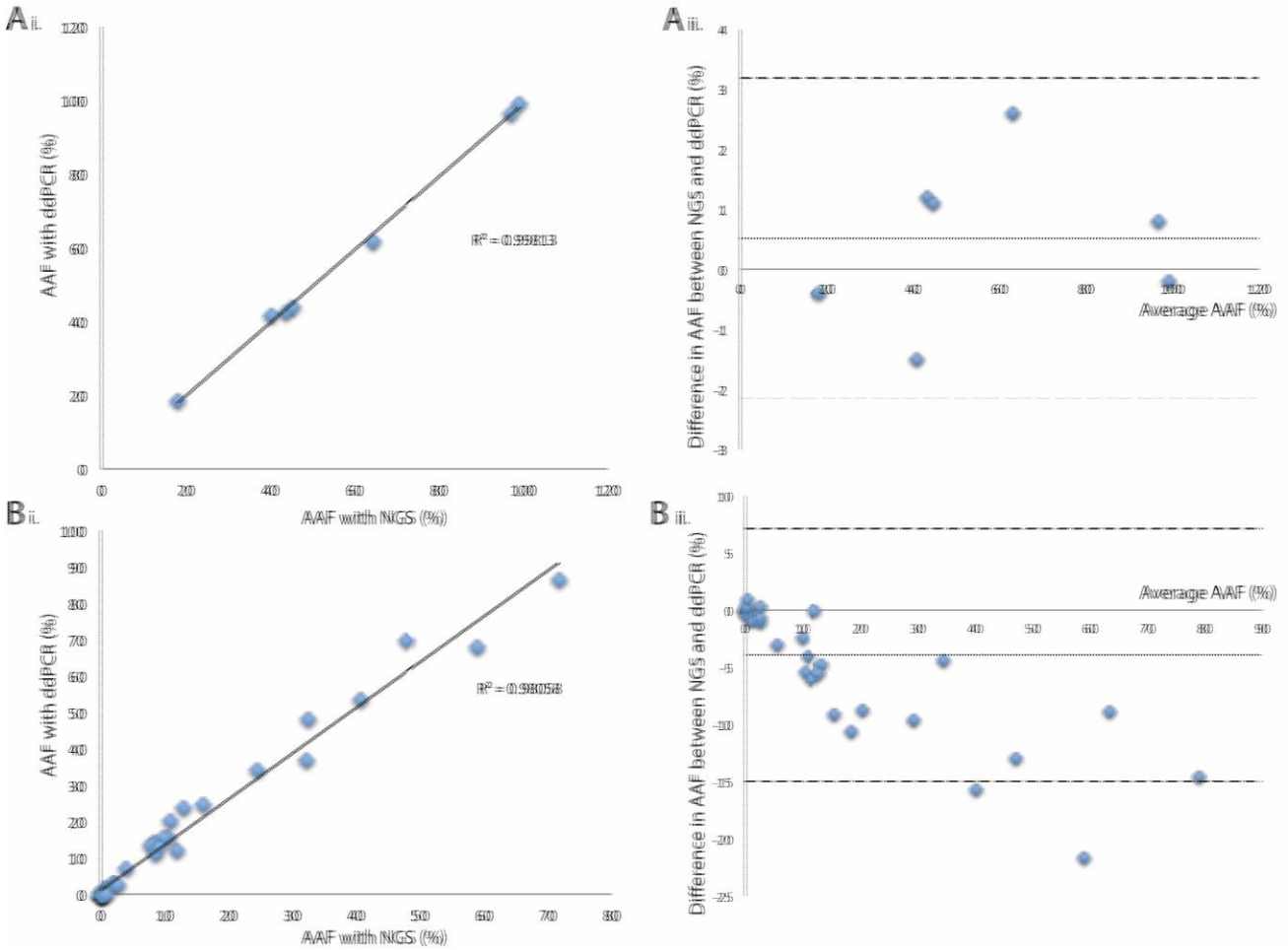

Figure 6-1 Correlation of alternate allele frequency (AAF) (%) of mutated DNA in FFPE tumour tissue (A) and ctDNA samples (B), as assessed by both ddPCR (i) and NGS (ii) approaches. Each data point represents one sample analysed.

6.3.4 Concordance between FFPE and baseline cfDNA

There was enough tumour tissue for NGS analysis for 9 of the 10 patients. There was not enough tumour tissue for NGS analysis for one patient (SCLC_7), since the FFPE tumour tissue had been exhausted by prior histopathological examination. Using NGS, mutations were detected with the 50 gene cancer hotspot panel in FFPE tumour tissue in 8 patients, and of these the same NGS detected the same mutations in baseline ctDNA in 6 patients. In one patient (SCLC_5), it was not possible to detect any mutations in the tumour tissue, or baseline ctDNA, using NGS. However, for the patient with no available tumour tissue (SCLC_7), it was possible to detect mutations in baseline ctDNA using NGS, as previously stated. Therefore, concordance for mutation detection between tumour tissue and baseline ctDNA, using NGS is 77.8%.

It was possible to detect the mutations in 7 of 8 patients where mutations could be assessed by ddPCR in the baseline ctDNA sample. This gives a concordance for mutation detection between tumour tissue and baseline ctDNA, using ddPCR, of 87.5%. The main difference between NGS and ddPCR, is that ddPCR was able to detect a mutation in baseline ctDNA in a patient that NGS was not able to.

6.3.5 Concordance between ddPCR and NGS for mutation detection in ctDNA

NGS analysis of FFPE tumour tissue, revealed 12 mutations that were assessable within the baseline ctDNA samples, using NGS and ddPCR. It was not possible to set up a ddPCR assay to assess for 2 of the mutations, owing to time and financial constraints of this research project. Of the 10 remaining mutations, ddPCR detected 9 of them, and NGS detected 8 of them.

For these 10 mutations, overall concordance in mutation detection in ctDNA, between ddPCR and NGS was 90% (Table 6-3). It was possible to detect 8 of the mutations, in baseline ctDNA, using ddPCR and NGS. For one patient SCLC_1 the *TP53* mutation was not detectable in base-line ctDNA either by NGS or by ddPCR. This patient had stage IV SCLC with brain as the only site of metastases. It is possible that the absence of the mutation in the ctDNA, by both NGS and ddPCR

may be due to an absence of ctDNA in the blood sample - the only site of metastatic disease in this patient was the brain and ctDNA may not be able to cross the blood: brain barrier into the peripheral circulation, as discussed in Chapter 5.

For one patient (SCLC_4) there was discordance in baseline ctDNA detection, between NGS and ddPCR. The *TP53* mutation was able to be detected using ddPCR, but not with NGS. The most likely explanation is that this *TP53* mutation occurred at an alternate allele frequency (0.69%) below the limit of detection of the ITVC software used for NGS (see Chapter 3). The reason for the low alternate allele frequency is that this patient had stage IIIB disease, where there may be lower levels of ctDNA compared to patients with stage IV disease.

			FFPE Tumour Tissue			Baseline ctDNA Sample					
			Next Generation Sequencing			Next Generation Sequencing			Droplet Digital PCR		
Patient ID	Stage	Gene	Mutation Detected	Coverage (x)	AAF (%)	ITVC	Coverage (x)	AAF (%)	Mutation Detected	AAF (%)	95% Poisson Error
SCLC_1	IV	TP53	Yes	34402	97.2	No	100414	0.13	No	0.0	
SCLC_2	IV	TP53	Yes	6083	31	Yes	62914	60.10	Yes	67.9	64.7 - 71.1
SCLC_3	IV	KRAS	Yes	29246	18.1	Yes	8473	32.30	Yes	36.7	35.6 - 37.9
	IV	TP53	Yes	13681	40.2	Yes	12437	71.10	Yes	86.3	85.8 - 86.8
SCLC_4	IIIB	TP53	Yes	32671	43.8	No	63381	0.54	Yes	0.69	0.0 - 1.73
		RB1	Yes	16644	36.4	No	22970	0.62	Not Tested	-	-
SCLC_5	IV	Nil Detected	Nil Detected	-	-	Nil Detected	-	-	NA	-	-
SCLC_6	IV	TP53	Yes	23700	64.4	Yes	25436	33.40	Yes	48.1	44.9 - 51.2
SCLC_7	IV	TP53	Not Available for Testing			Yes	23402	40.60	Yes	53.5	51.0 - 56.0
		SMAD4				Yes	30870	41.10	Not Tested	-	-
SCLC_8	IV	TP53	Yes	39757	98.9	Yes	18508	49.20	Yes	69.7	65.8 - 73.6
SCLC_9	IV	KRAS	Yes	41890	45.2	Yes	8735	8.20	Yes	8.73	11.1 - 12.8
SCLC_10	IIIB	TP53	Yes	7910	80.7	Yes	24113	3.20	Yes	0.783	0.0 – 2.63

Table 6-3 Comparing mutation detection between in FFPE tumour tissue DNA and baseline cfDNA by NGS and ddPCR. AAF = Alternate allele frequency. ITVC = Ion Torrent Variant Caller.

6.3.6 Baseline cfDNA mutational load as a Prognostic Biomarker

Receiver operating characteristic (ROC) analyses were performed to generate a cut-off value for baseline ctDNA level (alternate allele frequency (%)), to predict a prognosis of greater than, or less than 4 months, from the time of consent to enter the study (Table 6-4). The time-point of 4 months was chosen because this is the average time taken to complete a course of palliative chemotherapy, followed by a course of radiotherapy. Since it is preferable to correctly identify patients with a truly poor prognosis, the cut-off ctDNA level was chosen that gave the highest specificity while maintaining the optimum sensitivity. On the basis of these data, a ctDNA fraction of 44.3% can be suggested as a possible prognostic biomarker, with patients having a base-line ctDNA mutation load of less than 44.3% alternate allele frequency, belonging to a better prognostic. It would be important to validate this prospectively, in a larger group of patients.

	ROC Curve Analysis						
Stage of Disease	Area Under the Curve (AUC)	Indication of Test Suitability	95 % Confidence Interval (AUC)	Significance	Cut-off Baseline ctDNA AAF	Sensitivity (Optimum)	Specificity
IIIB and IV	0.929	Excellent	0.741 - 1.00	p = 0.079	44.30%	1.00	0.87
Only IV	0.900	Excellent	0.644 - 1.00	p = 0.121	44.30%	1.00	0.80

Table 6-4 Receiver Operating Characteristics (ROC) Curve analysis to determine a baseline ctDNA fractional abundance (%) to predict prognostic group for patients with Small Cell Lung Cancer.

Prognosis is poor in locally advanced and metastatic SCLC. If patients have more favourable prognoses, then more aggressive anticancer therapies can be considered, where clinically appropriate, as the benefits of treatments may outweigh the risks. However, it is desirable to be able to identify SCLC patients with a worse prognosis, to help to avoid the inappropriate use of toxic anti-cancer therapies, such as chemotherapy and cranial irradiation, where the risks may outweigh the benefits. The novel data presented here, suggest baseline ctDNA level may act as an independent prognostic biomarker in patients with stage IIIB and stage IV small cell lung cancer.

6.3.7 Treatment stratification according to baseline level of ctDNA

The base-line ctDNA level was compared to whether patients received PCI (Table 6-5). Using the proposed cut-off of 44.3% AAF in the base-line ctDNA, as a prognostic biomarker, could have potentially influenced the clinical management for two (patients SCLC_2 and 10) of the 7 patients that could have been considered for PCI. One patient (SCLC_2) received PCI, but using the 44.3% cut-off would have put them into a poorer prognostic group. In fact, this patient had an initial good clinical and radiological response to chemotherapy, and subsequently received consolidation thoracic radiotherapy (20 Grey in 5 fractions), palliative radiotherapy to a site of spinal metastatic disease (L2-5)(20 Grey in 5 fractions) and PCI (20 Grey in 5 fractions), but only survived 5.5 months from the time of consent. This was approximately 2.5 months following completion of treatment. Despite the initial radiological and clinical response to palliative chemotherapy, it is interesting to speculate whether the use of the 44.3% cut-off for this patient could have been spared the time burden and toxicities of the PCI, especially when pain control, such as the spinal metastatic disease, would have taken precedence. However, as mentioned, it would be vitally important to validate this biomarker in larger cohorts of prospective patients.

Conversely, the other patient (ID SCLC_10) would have been placed in the 'better prognostic' group, on the basis of base-line ctDNA level cut-off of 44.3%. They

also experienced a good radiological response to chemotherapy. This patient was living at 6.7 months from the time of consent (at the time of the last follow-up for this patient), and subsequent treatment modalities, such consolidation thoracic radiotherapy and PCI could be considered if clinically appropriate.

Patient Research ID	Did patient receive PCI?	Was ctDNA detected in baseline sample?	AAF (%)	Prognostic Group using baseline ctDNA AAF cut-off	Should PCI be considered based on baseline ctDNA cutoff	Potential Change to Clinical Management
SCLC_2	Yes	Yes	67.9	Poorer	PCI may not offer clinical benefit	Yes
SCLC_4	Yes	Yes	0.69	Better	PCI should be considered as clinically appropriate	No
SCLC_5	Yes	Unable to quantitate ctDNA			Unable to advise	No
SCLC_6	No	Yes	48.1	Poorer	PCI may not offer clinical benefit	No
SCLC_7	No	Yes	53.5	Poorer	PCI may not offer clinical benefit	No
SCLC_8	No	Yes	69.7	Poorer	PCI may not offer clinical benefit	No
SCLC_10	No	Yes	0.783	Better	PCI should be considered as clinically appropriate	Yes

Table 6-5 Comparing whether baseline ctDNA level, as a prognostic biomarker, would have potentially influenced clinical decision making regarding PCI.

6.3.8 Mutation detection in ctDNA as a surrogate for tumour tissue analysis

There was insufficient FFPE tumour tissue NGS anaylsis in one patient (research ID SCLC_7). NGS analysis was performed on baseline ctDNA, and the pathogenic mutations in *TP53* c.814G>T, p.V272L and in *SMAD4* c.1081C>T, p. R361C, were discovered with alternate allele frequencies of 40.6% and 41.1% respectively. Patients with treatment-naïve stage IIIB and stage IV SCLC had relatively high amounts of ctDNA in the blood (median alternate allele frequency of 32.85%, range 0.13% - 71.1%), therefore increasing the chances of mutation detection, compared to metastatic NSCLC (Chapter 3). Thus illustrating that ctDNA analysis is an exciting, alternative source of easily accessible tumour DNA for molecular analysis in patients with SCLC. It is therefore also an attractive surrogate for analysing tumour tissue samples.

6.3.8.1 Kinetics of ctDNA in SCLC patients treated with chemotherapy

Although decreases in ctDNA during chemotherapy have been observed in patients in various cancer types, there is little in the literature about the dynamics of ctDNA in patients with small cell lung cancer. We therefore

investigated the possibility of using NGS and ddPCR assessments of changes in levels of mutations in ctDNA, to monitor response/resistance to therapy in our SCLC patients.

A somatic *KRAS* c.437C>T, p.A146V mutation was discovered in the tumour tissue DNA from Patient ID SCLC_9. Changes in abundance of this mutation were monitored, using both NGS and droplet digital PCR methodologies, in serial ctDNA samples, taken at various time-points during palliative chemotherapy, and subsequent consolidation radiotherapy to the chest (Figure 6-2). Findings were similar to those observed in *EGFR*$^{mut+ve}$ NSCLC, in Chaper 5 – after commencing chemotherapy, ctDNA levels fell to undetectable levels, as assessed by both methods. This compared well with radiological assessment of response to treatment, as assessed by plain film radiography (chest X-ray), and confirmed by a CT scan on completion of treatment. Using ddPCR, *KRAS* mutation levels in ctDNA levels remained undetectable, until the end of the period of follow-up. However, there was a slight increase in mutant *KRAS* level using NGS, not observed by ddPCR, towards the end of therapy.

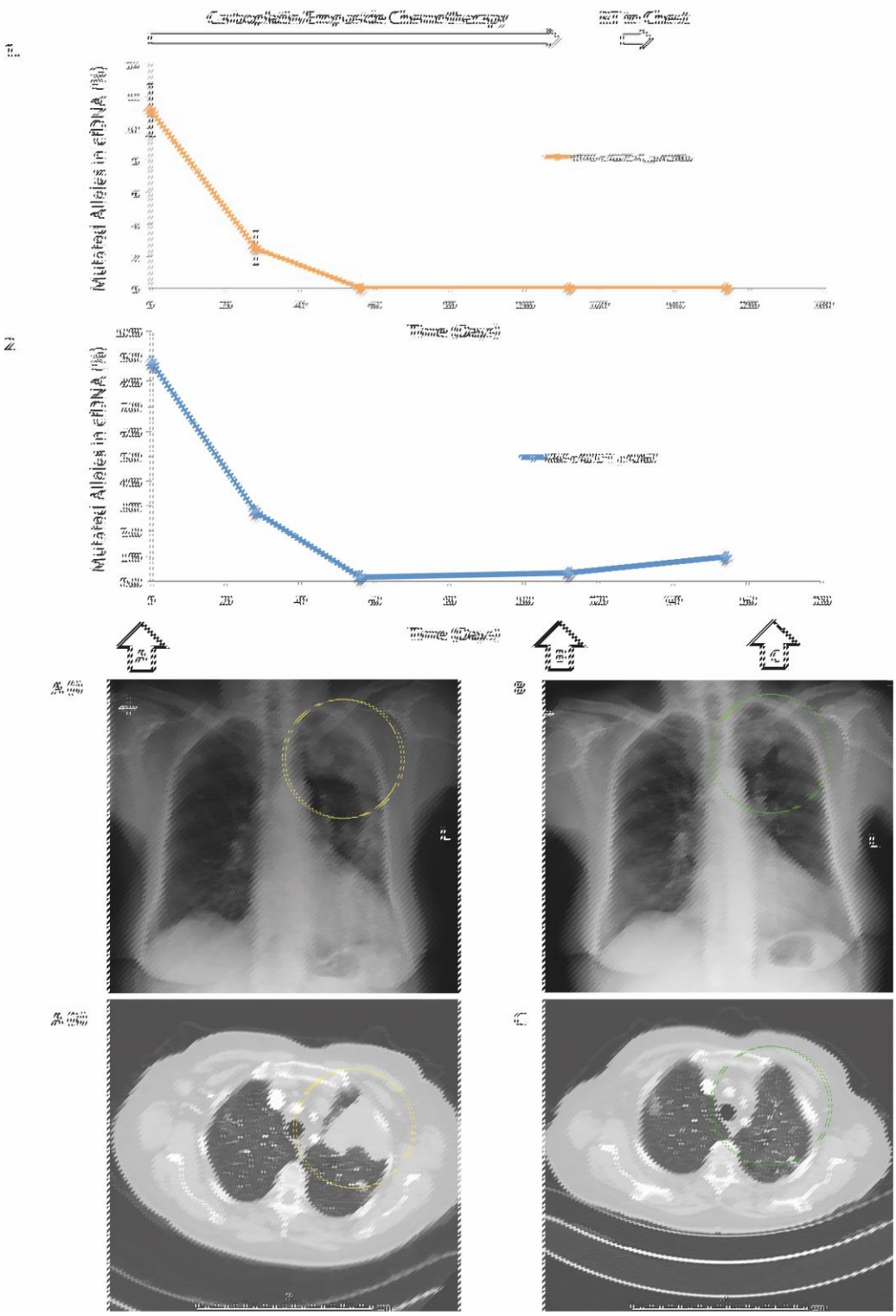

Figure 6-2 ctDNA kinetics and correlation with radiological response in a chemotherapy-sensitive, metastatic small cell lung cancer (Patient ID SCLC_9). Changes in mutant *KRAS* levels in serial ctDNA samples using ddPCR (95% Poisson error bars shown)(1.), and NGS (2.). A(i) Baseline chest radiograph (A(i)) and baseline CT scan (A(ii)), showing the primary lung tumour (yellow circles). Plain chest radiograph at 112 days (B). Post-treatment CT scan of chest (day 168) (C). Green circles highlight responding disease.

Quantitative changes in ctDNA abundance with time were also assessed in a patient (SCLC_7) who achieved an initial radiological response to palliative chemotherapy (Figure 6-3). Similar patterns of decline in a *TP53* mutation in ctDNA were observed with both ddPCR and NGS, during chemotherapy, with a plateau being reached at approximately 10-15% variant allele frequency, after cycle 1 of chemotherapy. NGS also showed a similar pattern for a second, *SMAD4* mutation, in this same patient. Towards the end of the course of 6 cycles of chemotherapy, there was a slight increase in levels of both mutations in ctDNA as assessed by NGS and, in the *TP53* mutation, by both methods. Although a CT scan after completion of chemotherapy demonstrated that the liver metastases had almost completely resolved, the rise in ctDNA mutation levels may be explained by an increase in size of the primary lung lesion, compared to the baseline CT scan. Unfortunately, a corresponding ctDNA sample was not taken at the same time as the CT scan. However, it is tempting to speculate that, by analogy with our results for NSCLC patients in Chapter 5, the slight rises in ctDNA mutations seen in the midst of the course of chemotherapy may have been an indication of early, sub-clinical disease progression in the primary tumour, outweighing a reduction in the burden of disease in the liver, which would not manifest itself, until later radiological investigations.

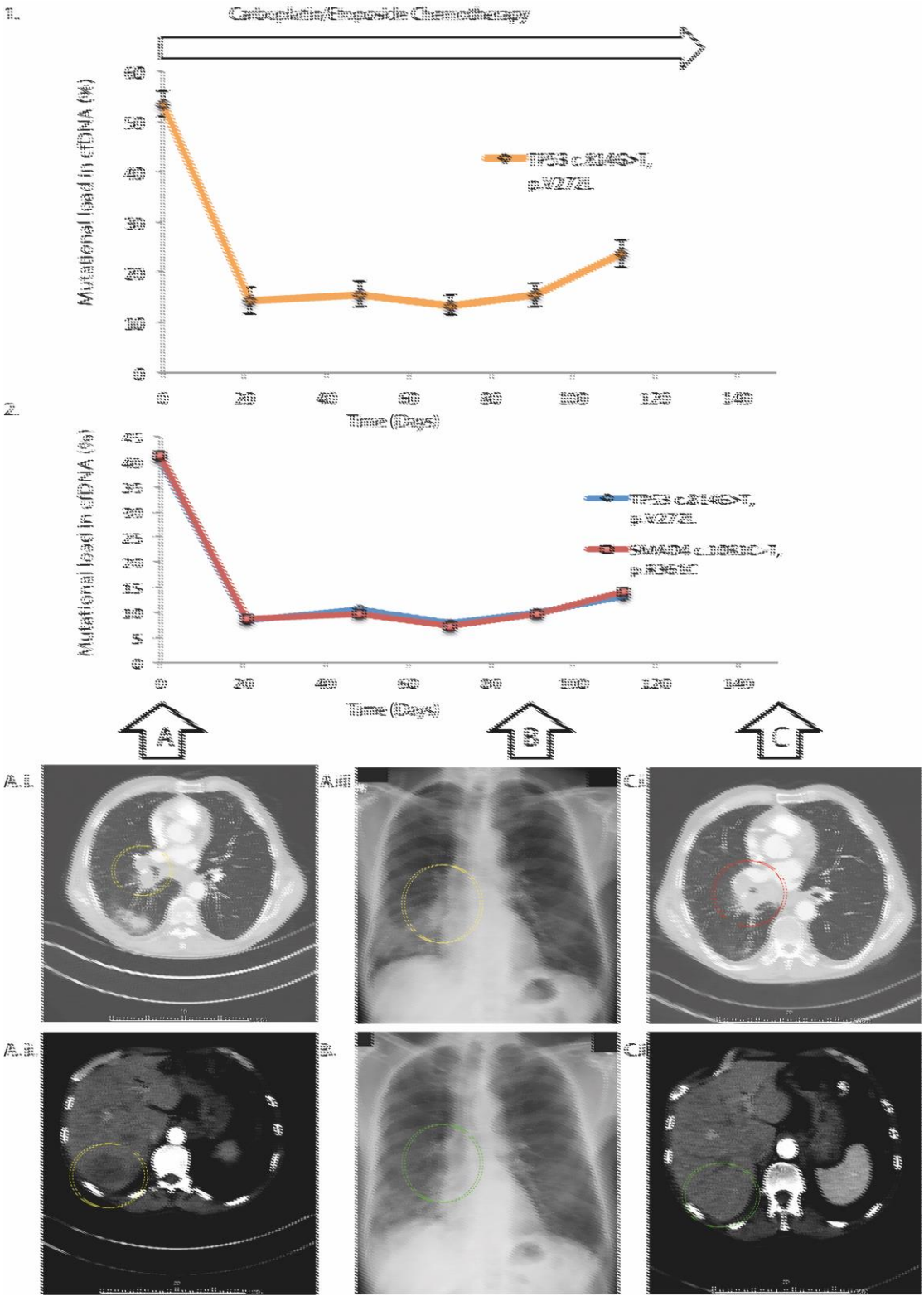

Figure 6-3 ctDNA kinetics and correlation with radiological response in a second patient with metastatic small cell lung cancer (Patient ID SCLC_7). Changes in mutational frequency in ctDNA determined by (1) ddPCR and (2) NGS. Baseline chest x-ray (A(iii)), and baseline CT scan illustrating the primary lung tumour (A(i)), and hepatic metastases (A(ii)) (tumours highlighted by yellow circles). The primary lung tumour on a chest x-ray at day 91 (B) (green circle highlights responding tumour). CT scan at 149 days with responding tumour in the lung (red circle)(C.i) and responding tumour in the liver (green circle) (C.ii.).

All of the patients who received chemotherapy obtained an initial radiological response to treatment. This compared with initial decreases in the mutational levels in ctDNA, compared to baseline levels (Figure 6-4), though the degree of the initial decrease was variable. For all patients, the greatest and most rapid decrease in ctDNA is seen during the first cycle of chemotherapy, between day 0 (baseline) and day 21. As for *EGFR*$^{mut+ve}$ adenocarcinoma patients treated with EGFR TKi (Chapter 5 above), ctDNA decline preceded radiological response to chemotherapy, suggesting that it may be possible to utilise ctDNA as a predictive biomarker of response to treatment, and to assess response/resistance both earlier and more frequently than by radiological imaging. This might inform individual patient treatment choices, such as the optimum number of cycles of chemotherapy before discontinuing or switching to subsequent anti-cancer treatments.

One interesting phenomenon observed, also illustrated on Figure 6-4, is that patients appear to cluster in 2 groups, based on the extent of ctDNA decrease. Patients with overall survival greater than 6 months all showed a smaller absolute decline in ctDNA mutation levels. No such clustering was seen when considering relative decreases in ctDNA mutation levels. Although this finding needs to be validated in a much larger series, this could be of great clinical significance in determining a SCLC patients' disease response to chemotherapy, and prognosis. This is especially interesting since radiological response to treatment is not of prognostic value in itself. Therefore, provided this finding could be validated, ctDNA mutation kinetics could be an important development as a prognostic biomarker.

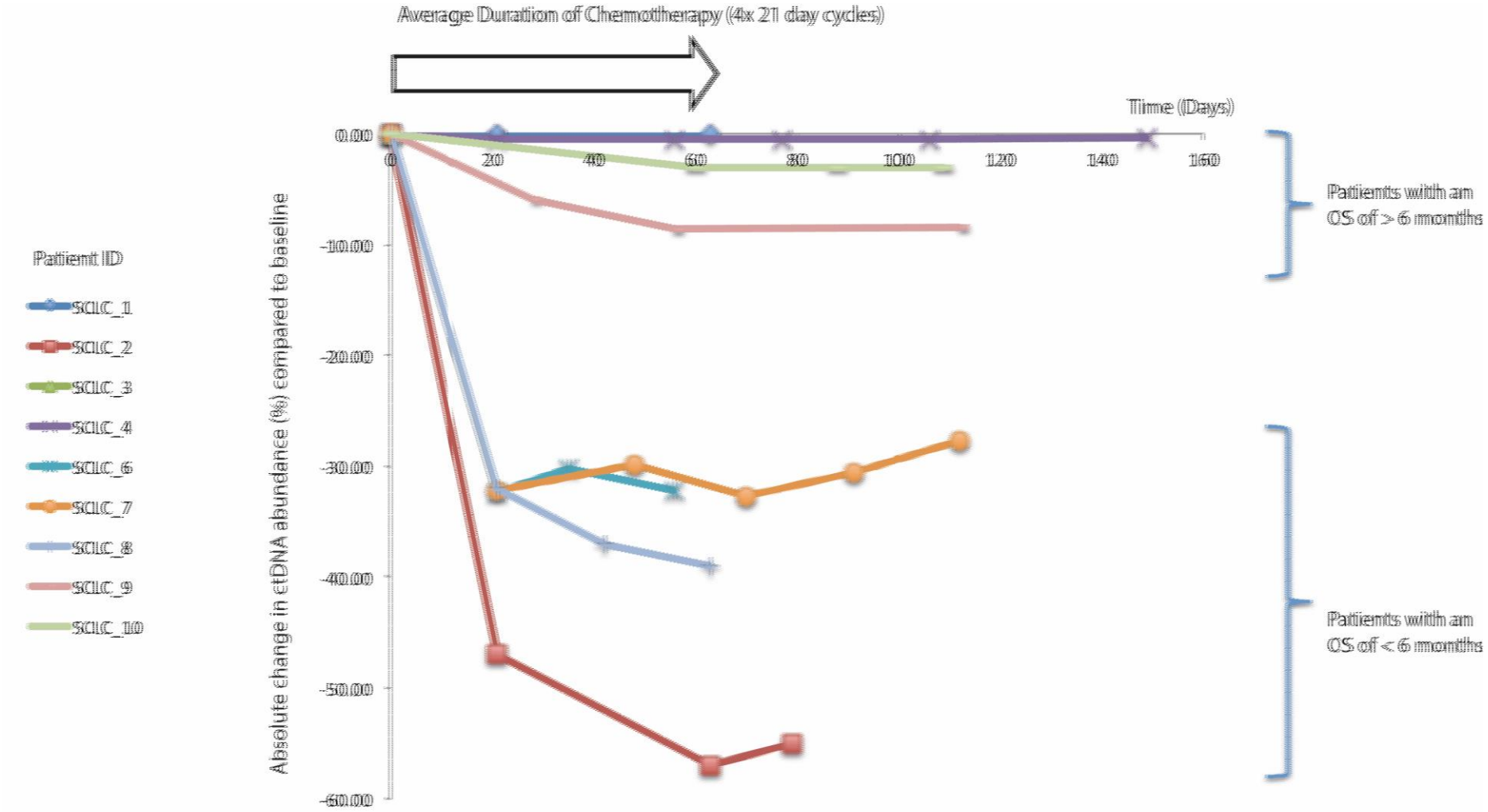

Figure 6-4 Quantitative changes in ctDNA fraction over time in patients with SCLC, treated with palliative chemotherapy. OS = Overall Survival

6.4 Clinical Implications

In this chapter, we have shown that it is possible to use NGS with a targeted gene panel to detect somatic mutations in tumour tissue DNA, from patients with locally advanced and metastatic SCLC, provided there is enough tumour tissue available for analysis. Furthermore, the same approach can reliably identify and quantitate the level of these mutations in ctDNA. Concordance between NGS and ddPCR for mutation detection in ctDNA analysis was high, at 90%. This supports the hypothesis that ctDNA can be used as an easily-accessible alternative to tumour DNA, for molecular analysis of SCLC and NSCLC patients, using NGS and ddPCR. In one patient (SCLC_7), where no tumour tissue was available for analysis, it was possible to detect pathogenic mutations in ctDNA from a baseline blood sample.

There is little in the way of SCLC prognostic scoring, and treatment stratification, in routine clinical practice in the UK. One previously reported prognostic indicator is the Manchester score(141), which groups patients into 'good', 'medium' and 'poor' prognostic groups, based on a two-year overall survival statistic. Since the vast majority of patients with advanced SCLC do not survive for 2 years, there is a need to develop improved biomarkers that may help facilitate earlier clinical decisions. In this chapter we chose 4 months as an appropriate prognostic 'milestone' for patients with locally advanced or metastatic small cell lung cancer, because this is the time taken to complete a course of 4 cycles of palliative chemotherapy and subsequent thoracic and/or cranial irradiation. These treatments involve multiple hospital visits and are associated with significant toxicities. It is therefore crucial to patients' quality of life to consider only anti-cancer therapies for which potential benefits can be expected to outweigh the potential risks. This means that although patients should be considered for any available palliative anti-cancer therapies, accurate prognostic biomarkers may help to avoid any toxic therapies, from which the patient may gain limited clinical benefit. Equally, patients with a prognosis of less than 4 months and who are not demonstrating an appropriate response to

therapy, on the basis of mutational ctDNA decline, may also have the option of discontinuing chemotherapy earlier than the planned four cycles of chemotherapy, if the toxicities are too great, and/or if the patient is not gaining any symptomatic improvement despite receiving chemotherapy.

It is proposed, on the basis of data presented here, that the baseline level of ctDNA could serve as a prognostic biomarker, when a poorer prognosis is defined as an overall survival of less than 4 months following baseline ctDNA testing. The baseline ctDNA level of 44.3% alternate allele frequency is suggested as a cut-off level for determining a patient's prognostic group (greater or less than 4 months overall survival). This would have to be validated as part of a larger, prospective clinical study, in which it could also be assessed whether it could be used to stratify patients to routine treatments such as whether patients should receive PCI, following a course of palliative chemotherapy, or clinical trials, such as shorter versus standard courses of chemotherapy. It would also be valuable to prospectively compare this prognostic biomarker to prognostic scores already in use for patients with SCLC, such as the Manchester Prognostic Index.

One potential limitation of the putative use of quantitative baseline ctDNA data as a prognostic indicator in SCLC is the ability to detect a pathogenic mutation in ctDNA. Previous studies show that approximately 10% of cases of small cell lung cancer will have mutational mechanisms other than point mutations in *TP53* or *RB1* genes(142). Using a targeted gene panel and NGS analysis of ctDNA may therefore not detect mutations in all cases of SCLC, and may miss other mutational mechanism such as translocations and gene copy number aberrations. This may have been the situation for one patient in this chapter, (patient ID SCLC_5). For the purposes of future clinical research, further molecular genetic analysis should be performed on the tumour tissue from these patients, to elucidate novel pathogenic mechanisms, and to determine whether these mechanisms can be detected in ctDNA.

It could be argued that the level of ctDNA may be a surrogate for volume of disease, which in itself could be a prognostic indicator. However, volume of disease is not routinely measured in clinical practice. There are problems with radiological inter-interpreter variability, and there remain practical problems such as how can diffuse disease be measured. There may also be clinically significant differences in patients with the same apparent total volume of disease, but differently distributed e.g. one has a large primary tumour and the other has several small metastases. Therefore ctDNA could offer a measureable surrogate for active disease volume, which may be more clinically relevant than absolute radiological disease volume. Future research should therefore address the correlation between ctDNA levels and absolute disease volume (with standard radiology such as CT), and metabolically active disease volume (such as with PET scanning).

An additional benefit of quantitative monitoring of mutations in ctDNA, is that it can be achieved via serial, minimally invasive blood tests, without requiring repeat radiology, and consequent exposure to ionising radiation, to assess disease progress. As demonstrated here, a fall in ctDNA to undetectable levels commonly correlates well with a radiological response. It may be that patients who do not achieve undetectable levels of ctDNA despite chemotherapy, may have minimal residual disease that harbours an underlying resistant sub-clone of the cancer.

The absolute decrease in mutational alternate allele frequency (%) in ctDNA, during chemotherapy, may be another prognostic indicator, though this also needs to be validated in a larger group of SCLC patients. An interesting clinical question is whether the number of chemotherapy cycles can be individualised for these patients based, on initial changes in ctDNA levels on therapy plus objective assessments of toxicity.

None of the patients in this results chapter received second line palliative therapies (as is often the case, in advanced SCLC clinical practice). One area for significantly more research is to determine the optimal timing of repeat blood

samples during the minimal residual disease phase, and whether it is possible to detect disease relapse earlier (as is the case with patients with *EGFR*$^{mut+ve}$ adenocarcinoma of the lung in Chapter 5, above). Clinical studies could be designed to determine whether patients with advanced SCLC survive longer, if they receive earlier intervention with second line palliative therapies, compared to waiting for patients to develop clinical or radiological disease progression. Also, much more work is needed to determine molecular mechanisms of therapeutic resistance in small cell lung cancer, to enable to development of therapies that could overcome these.

Therefore ctDNA analysis in small cell lung cancer, as with NSCLC, certainly provides an exciting, and easily accessible, alternative source of tumour DNA for analysis. Kinetics of ctDNA in response to systemic anticancer therapies may provide new prognostic and predictive biomarkers for lung cancer patients that help the clinician to optimise the benefits of anti-cancer treatments, while minimising the risks.

7 Discussion and Future Research

Lung cancer has been the focus of this thesis since it remains the most common cause of cancer death. Patients are commonly diagnosed at advanced stages of disease, where prognosis is poor. There is an urgent need for developing biomarkers for lung cancer, for use in routine clinical practice and clinical trials. Biomarkers would be particularly helpful in lung cancer for earlier diagnosis, prognosis, and patient stratification to different anti-cancer treatments, and for monitoring response and resistance to treatment.

There is often a very limited amount of tumour tissue available from patients with locally advanced or metastatic lung cancer. Thus, other sources of tumour DNA are attractive alternatives to permit genetic analyses, such as circulating, cell-free tumour DNA (ctDNA). The last decade, has seen the development of important new genetic technologies, such as next generation sequencing (NGS) and droplet digital PCR (ddPCR), which can be, applied to tumour DNA analysis. This thesis set out to apply these genetic technologies to 'real-life' tumour tissue samples and ctDNA from patients with *EGFR*$^{mut+ve}$ adenocarcinoma of the lung and advanced small cell lung cancer (SCLC).

7.1 Summary of Key Conclusions

In Chapter 3, we assessed the ability of a targeted 50-gene panel to detect known cancer mutations, in DNA from cell lines and patients, using the Ion Proton NGS platform and two different bioinformatics tools (Ion Torrent Variant Caller (ITVC), and VarScan with Annovar). Cancer mutations were successfully identified in patients' ctDNA, against a background of wild-type cell free DNA (expressed as an alternate allele frequency, AAF). VarScan with Annovar was able to identify the presence of mutations at a 10-fold lower AAF than ITVC, but at the expense of a 10-fold increase in the total number of genetic variants, many of which are likely to represent false-positive genetic sequence changes or 'sequence artifacts'.

We also confirmed that the lower limit of mutation detection, using ddPCR (in the order of 0.1% alternate allele frequency), was much lower than NGS (5%). However, a limitation of ddPCR is that it can only detect sequence-specific mutations, whereas NGS can provide sequence data for multiple genetic regions simultaneously. Furthermore, the lower limit of mutation detection in ddPCR was determined by the amount of DNA available for analysis, quite often a limiting factor of total cell free DNA.

In assessing factors which can influence the amount of DNA available for ddPCR and NGS, and in keeping with the results from Rothwell et al (2016)(110), we found no statistically significant difference between total cell free DNA yields from SCLC patients' plasma samples extracted at different time-points, up to 4 days after collection collected in CellSave Tubes (Janssen). In these SCLC patients, significantly more total cell free DNA was observed in patients with stage IV small cell lung cancer, compared with stage III SCLC, and stage IV NSCLC.

In Chapter 4, NGS and the 50 gene cancer hotspot panel were successful in identifying mutations in lung tumour tissue samples, despite there often being only a very small amount of tumour material present from the currently used tumour sampling methods used for patients with advanced lung cancer. Two different NGS panels (the 50 gene commercial panel, and an, in-house custom 22-gene panel) were compared mutations in FFPE tumor tissue from patients with Adenocarcinoma of lung, Squamous Cell Carcinoma of lung and Small Cell Lung Cancer (SCLC). The 50-gene panel demonstrated a superior ability to detect *EGFR* mutations, specifically *EGFR* exon 19 deletions, and fewer 'gene fails' were observed. However, the 50 gene hotspot panel is a much more focused panel, and did not identify mutations genes that were detected with the 22 gene panel, suggesting a value for custom gene panels. However, more validation work on a custom design panel is required to reduce false negative results seen here.

In Chapter 5, we were able to detect and monitor real-time changes in mutational level of mutant *EGFR* in ctDNA, in patients with *EGFR*$^{mut+ve}$ adenocarcinoma of the lung, throughout their clinical course. We detected a

novel *EGFR* mutation sequence change. It was possible to detect emergence of mutations in ctDNA, as the tumour developed treatment resistance. Quantitative increases in mutations in ctDNA were detected, several weeks before the diagnosis of clinical and/or radiological disease progression, suggesting a potential opportunity to detect therapeutic resistance at an earlier stage. This raises the question of whether an earlier change in therapeutic strategy, based on ctDNA changes, may improve lung cancer patient outcomes. This would have to be the topic for a clinical trial.

Similarly, in Chapter 6, we were able to monitor changes in mutation levels in ctDNA from patients with advanced small cell lung cancer, during systemic anti-cancer therapy. It is proposed that the base-line mutational 'load' of ctDNA may be valuable as a prognostic biomarker for these patients, and this may have important clinical implications, such as facilitating patient selection for potentially toxic treatments, such as prophylactic cranial irradiation (PCI). Further work will be required to validate baseline mutation level in ctDNA as a prognostic biomarker.

7.2 Molecular Characterisation of Lung Tumour DNA

7.2.1 Tumour Tissue

Genetic characterisation of tumour DNA is vitally important in defeating cancer. Many somatic genetic changes in lung cancer have practical clinical implications, such as guiding the use of EGFR tyrosine kinase inhibitor (EGFR TKi) drugs against lung tumours that harbour sensitising *EGFR* mutations. Therefore it is vitally important to ensure that modern technologies, such as NGS and ddPCR, can be translated into a clinically robust, and meaningful, genetic analysis platform. To be useful, it is crucial that NGS can detect clinically relevant or important genetic mutations.

We have confirmed that the commercially available, off-the-shelf, 50-gene cancer hotspot panel (LifeTechnologies), can be used reliably to detect clinically important genetic point mutations and small insertions or deletions (INDELs)

relevant to lung cancer. However, this gene panel only focuses on known pathogenic hot-spot regions within genes, and could potentially miss other, disease-causing genetic sequence changes, outside of these hotspot regions. Also, this panel does not include some genes that may become clinically relevant to lung cancer in the future, for example, with the potential use of PARP inhibitors for *BRCA1* and *BRCA2* mutant lung cancers(143). Hence, a need for custom gene panels for the various types of lung cancer.

When selecting mutations for inclusion in a custom lung cancer gene panel, clinical relevance must be paramount. The SMP-2-MATRIX clinical trial(144) has developed a 28-gene panel for NSCLC which includes mutations which are considered 'potentially clinically actionable'. In a similar fashion, a custom 22 gene panel for NSCLC cancer was designed specifically for use in this research project (Chapter 4). Technical difficulties were encountered with this panel, for example, the 22-gene panel was unable to detect a key lung cancer mutation, namely *EGFR* exon 19 deletions, and there was a high 'genetic sequencing failure rate' for other genes within the panel. Therefore it is recommended that a significant amount of additional research time has to be devoted to validating any custom targeted gene panel to ensure it is robust and able to detect all of the clinically relevant genetic mutations, before it can be applied to clinical samples.

The lung cancer tissue samples obtained from patients in this thesis reflect availability of tumour tissue in advanced lung cancer in 'real-life'. Often, the amount of tumour tissue is very small, resulting in small amounts of extractable DNA. This is provided the tumour sample is not exhausted by prior histopathological examination. Furthermore the DNA is often degraded by the sample formalin fixation process, further reducing the success of NGS analysis. Despite these limitations, the results in this thesis have shown that it is possible to achieve successful NGS analysis of FFPE tumour tissue, largely owing to the relatively small amount of input DNA required for the NGS technique used. To improve successful NGS analysis, DNA from fresh tissue samples could be used, but this is not always possible in patients with advanced lung cancer. Also, NGS of ctDNA can also overcome some of the limitations of tumour tissue DNA

analysis, especially since this thesis shows it is possible to demonstrate mutations in ctDNA, where no tumour tissue is available.

Another concern with molecular characterisation of FFPE tumour tissue, is that small biopsy samples acquired from locally advanced, or metastatic lung cancers, only represent a small proportion of any one tumour site, at any one moment in time. It is not, therefore, possible to fully characterise intra-patient tumour heterogeneity, especially when metastatic disease exists. Likewise, when a tumour progresses, a repeat biopsy of a lung cancer, or a site of metastatic disease, is not often clinically feasible. This thesis proposes that, despite low levels of total cell free DNA in the peripheral circulation; it is possible to detect mutations contained within ctDNA using techniques such as NGS and ddPCR. Results in this thesis have shown that analysis of ctDNA may reveal mutations contained by different tumour sub-clones, which will help to characterise intra-patient tumour heterogeneity. In conjunction with the TRACERx study, NGS of differential tumoural regions has better delineated phylogenetic tumour sub-clonal evolution. Longitudinal ctDNA profiling in patients with treated early stage NSCLC who relapse, can determine the phylogenetic evolution of relapsed and metastatic tumour sub-clones(145). While molecularly characterizing tumour sub-clones with relapse/metastatic potential, it can also facilitate earlier detection of relapse, and research into future therapeutic strategies based on individual sub-clones.

7.2.2 Circulating Cell Free Tumour DNA

Despite detection and molecular characterisation of ctDNA promising to facilitate personalised cancer therapy, current limitations to analytical sensitivity include low quantities of circulating cell free DNA (cfDNA), and technical and clinical factors.

Several NGS methods and bioinformatics applications have been assessed in the literature to address this problem of large numbers of sequencing artefacts generated by NGS, in order to increase the confidence in detecting true, low-level NGS sequence variants. However, there are no widely accepted consensus

technology and analysis protocols that are ready for use in widespread clinical practice, amongst the many different, currently-available NGS technologies, and analysis platforms.

Droplet digital PCR (ddPCR) is a molecular technique that currently offers the ability to detect low levels of mutant alleles, for a specific genetic variant. In Chapter 3, the amount of input DNA alters the confidence of detecting low level variants. For example, if 5ng input DNA is used, then it may only be possible to detect mutant alleles as low as 1% abundance, but if 20ng of input DNA is used, then it may be possible to detect mutant alleles as low as 0.5% or even 0.1% allelic frequency. This is important, given that the amount of total cell free DNA available from a plasma sample is usually very low, for example 1ng/µl. This has to be taken into account when writing clinical reports which will be used in treatment decision-making for individual patients, since it is important to establish a threshold, or limit of detection, as a factor of the amount of input DNA.

In keeping with previously published data(110), data in Chapter 3 show that if blood is collected in specialized blood collections tubes (CellSave tubes) there is no detrimental increase in the amount of total cell free DNA, if the sample takes up to 96 hours to be processed. This is important to maximize the chances of detecting cancer mutations within ctDNA. We can therefore advise that these specialised blood collection tubes can facilitate blood samples to be collected from patients in any clinic within the UK, and sent to a central lab for processing.

The absolute concentrations of total cell free DNA that can be extracted from plasma samples for these lung cancer patients in this thesis is low. At the time of the research conducted in this project, there were few platforms available for extracting circulating nucleic acids. It may be that more platforms are developed that are able to obtain better yields of circulating cell free DNA from a patient sample, but these need to be evaluated in future research.

Results in this thesis suggest that patients with stage IV small cell lung cancer generally have a much higher amount of cell free DNA, and a greater chance of detecting cancer mutations in ctDNA, compared to stage IV adenocacinoma. This could reflect a higher volume of 'more active' disease, PET FDG avidity can predict ctDNA detection in NSCLC(145). A clinical study correlating baseline ctDNA level with baseline PET-FDG avidity in SCLC could be designed.In this thesis, the number of patient samples with squamous cell carcinoma of the lung were too small to enable a meaningful comparison of availability of cell free DNA. Therefore, future work is required to determine whether there is more or less total cell free DNA in patients with advanced squamous cell cancer of the lung. This can further inform whether there is a greater or lesser chance of detecting cancer mutations in ctDNA, in lung cancer patients with a squamous histology. Subsequently, it has been reported that there is ctDNA detection is associated with histological subtype. This is where there is more chance of finding ctDNA in lung squamous cell cancers (97%), compared to lung adenocarcinomas (71%). 94% of stage I squamous cancers were positive for ctDNA, and only 13% of stage I lung adenocarcinomas were positive for ctDNA. This has been suggested that squamous cell cancers are more necrotic, and shed more tumour DNA into the circulation(145).

The greatest impact on longer term survival for lung cancer patients may come from developments in biomarkers for lung cancer screening and management of earlier stages of disease. The role of ctDNA as a biomarker in earlier stages of lung cancer must be further investigated, to see if it can facilitate earlier diagnosis and improve outcomes for lung cancer patients. Given the current technical limitations, it may be extremely difficult to detect and characterize ctDNA in early-stage lung cancers, assuming that there will be lower levels of ctDNA. Indeed, in NSCLC, ctDNA levels correlated to tumour size, such that a tumour volume of 10 cm^3 will yield a ctDNA variant allele frequency of 0.1%(145). To overcome this, increasingly more sensitive technologies are needed, which can identify mutations in ctDNA present at significantly lower level, in a background of wild-type/genomic DNA.

Despite the current limitations, analysis of ctDNA still offers the advantage of being a minimally invasive investigation (namely a blood tests), and serial samples that can be taken, and analysed in real-time, as a patient undergoes various anti-cancer therapies. The feasibility of serial blood sampling for ctDNA analysis has been clearly demonstrated with the several case studies presented in this thesis. As with these case studies, blood sampling ctDNA testing could easily be adopted in the routine clinic.

7.3 Role of ctDNA in stratified medicine

There are numerous clinical trials, and more anticipated, which utilize molecular characterization of tumour tissue to stratify lung cancer patients for treatment with molecularly targeted anti-cancer agents. This is in an effort to increase clinical efficacy of treatment and to reduce unnecessary treatment toxicity, especially if a particular anti-cancer treatment has no clinical efficacy.

We are fortunate, in UK lung cancer research, to have access to the Stratified Medicine Project 2 (SMP2) - the largest molecular characterization project of this sort, to date, funded by CRUK. SMP2 aims to molecularly profile tumour tissue from patients with non-small cell lung cancer, using NGS sequencing technology. On the back of this, there exists the UK MATRIX trial, an umbrella trial, examining the clinical efficacy of several targeted anticancer therapies, given to lung cancer patients harbouring specific somatic mutations in specific genes(144).

Nevertheless, SMP2 is prone to the same problems as those reported in this thesis. Conventional CT guided biopsy and/or bronchoscopic biopsies yield very small quantities of tumour tissue, which are often depleted or exhausted during histological diagnosis, severely limiting tumour DNA extraction and analysis from tissue samples. As a result, many samples fail the pre-NGS quality control (QC) test. Even if there is apparently sufficient tumour DNA, there is still a failure rate for comprehensive gene profiling using the targeted NGS platform. At the time of writing, the resulting fraction of patients that therefore have a

satisfactory NGS test result, that enables them the option of entry into clinical trials such as the MATRIX trial, is approximately 25%(146).

Sequencing of DNA obtained from Fresh Frozen tumour tissue samples is advised in the Genomics England, 100,000 genomes project, in order to yield better quality DNA, and improve the success of sequencing, and minimising the problems with DNA degradation which result from the FFPE fixation process. However, this is not routine practice in the NHS, and other limitations remain – particularly those arising from small biopsies, or small proportions of tumour cells within a biopsy sample(147).

A current limitation of ctDNA analysis is the risk of a false negative result when assessing for the presence of a somatic mutation, especially if the mutation is present at a level below the limit of detection of the technology used. This is even more problematic when trying to be confident that a particular gene is 'wild-type', which is required by some of the current proposed therapeutic strategies in the MATRIX trial.

7.4 Applications of ctDNA in Lung Cancer Clinical Practice

7.4.1 CtDNA in Patients with EGFR^{mut+ve} Adenocarcinoma of the Lung

We have demonstrated, in this thesis, that for patients with *EGFR*$^{mut+ve}$ adenocarcinoma of the lung, whether locally advanced or metastatic, it is possible to detect the *EGFR* mutation in ctDNA. It is known that mutant *EGFR* detected in ctDNA can act as a predictive biomarker for response to targeted, systemic EGFR TKi therapy, and this is especially important where it is not possible to obtain this information from a tumour tissue sample.

Data presented in this thesis show that increases and decreases in mutant *EGFR* levels in ctDNA can be observed during systemic anti-cancer therapy with EGFR TKi therapy. We observed decreases, down to undetectable levels, within 2 months of starting therapy. Levels then remain low, or undetectable, while the patient often experiences a sustained, on-going [radiological] response to

treatment. A quantitative increase in ctDNA was also observed, at the time of disease progression, due to acquired resistance to EGFR TKi therapy. In fact, it seems reasonable to conclude that levels of ctDNA may increase several weeks before the diagnosis of clinical or radiological progression of lung cancer (Chapter 5). A subsequent, recent publication has also reported a lead time of ctDNA detection in NSCLC of 70 days (range 10-346 days), before radiological confirmation of disease (CT or plain chest radiograph)(145). This may represent a potential window of opportunity to switch therapeutic strategies in these patients, earlier in their treatment course, and in a period before patients clinically deteriorate and when they are still able to tolerate subsequent therapies. This is an interesting area of future clinical research. For example, a randomised controlled trial, involving early (e.g. on ctDNA progression) versus delayed (e.g. on clinical disease progression) switch in therapeutic strategy. Clinical trial such as this, could further address the feasibility of ctDNA mutation analysis replacing radiological disease monitoring as the 'gold-standard' for monitoring response to therapy.

In relation to targeted therapies, it is possible to analyse ctDNA to detect known molecular mechanisms of acquired resistance to therapy. A prime example of this is the example of the acquired secondary *EGFR* c.2369C>T p.T790M resistance mutation, seen in patient case studies presented here. This is the expected resistance mechanism expected in approximately 50% of patients with *EGFR*$^{mut+ve}$ adenocarcinoma, treated with first line EGFR TKi. Osimertinib is a third generation oral EGFR TKi which can inhibit both the original activating *EGFR* mutation and the *EGFR* c.2369C>T p.T790M mutation. Clinical efficacy has been demonstrated for this drug, and the license of the drug allows it to be prescribed for patients who have an *EGFR* c.2369C>T p.T790M mutation demonstrated via ctDNA analysis. This has the advantage of negating the need for an invasive repeat biopsy at the time of disease progression. However, is understood that other resistance mechanisms include secondary mutations in other genes, and increase in gene copy number. Further work is needed to investigate whether ctDNA can be used to detect other molecular mechanisms of acquired therapeutic resistance. For example, research into whether increases in

gene copy number can be detected within ctDNA, compared to gene copy number changes in tumour samples from the same patients. Some research groups have started to explore whether it is possible to detect gene copy number changes in ctDNA, in a variety of clinical settings. These include: Taqman copy number assessment of androgen receptor (*AR)* and *CYP17A1* genes and response to abiraterone therapy in prostate cancer patients(148); OncoScan® platform for gene copy number profiling in cfDNA in patients with neuroblastoma(149); and massively multiplexed PCR for gene copy number variations in patients with breast cancer(150). Future research might similarly examine the possibility to detect gene copy number variation in ctDNA in patients with lung cancer.

7.4.2 Limitations of ctDNA as a biomarker for patients with *EGFR*$^{mut+ve}$ adenocarcinoma of lung

Despite these exciting new applications of ctDNA testing, with a potential direct impact on patient care, it is not possible to detect *EGFR* mutations in ctDNA, in all patients harbouring an *EGFR* mutation. There are clinical factors that may influence the ability to detect *EGFR* mutations in ctDNA. These include stage of disease, volume of disease, and whether the patient is treatment naïve. From data presented in Chapter 5, it is possible to detect mutant *EGFR* in ctDNA, when patients are treatment naïve, or have received only minimal treatments, such as palliative radiotherapy to focal lesions in the spine. It was not possible to detect mutant *EGFR* in ctDNA in lung cancer patients who have recently received surgical resection, radical radiotherapy, or systemic chemotherapy. Undetectable levels of mutations in ctDNA in these situations may have resulted from the effects of these treatments on the cancer. If serial levels subsequently remain undetectable, then this may represent a period of disease stability or 'minimal residual disease'. This phenomenon was also demonstrated in the case studies presented in Chapter 5.

From data presented in Results Chapter 4, it remains to be seen whether it is possible to capture tumour DNA from cerebral metastasis in the peripheral blood circulation. For patients with *EGFR*$^{mut+ve}$ adenocarcinoma of the lung who

develop brain metastases, *EGFR* mutational analysis can be considered on cell free DNA from cerebro-spinal fluid (CSF). Further research is needed to determine whether or not mutations in ctDNA are derived from CNS metastases, including the potential comparison with mutational analysis of cell free DNA in CSF.

7.4.3 NHS clinical testing service for ctDNA

Our data, presented here, and the data of others, raise the prospect that NHS molecular genetics diagnostic laboratories might implement *EGFR* analysis of ctDNA, as a diagnostic service for patients with lung cancer. A routine service of this sort will be facilitated by blood collection tubes containing preservative, such as CellSave, in which ctDNA remains stable in blood samples stored at room temperature for an acceptable period (4 days). This would allow blood samples taken from patients in clinics at distant locations to be transported, without the need for specialised transportation requirements, to a central lab for processing and testing, without diminishing the reliability of data obtained.

Indeed, at the time of writing this thesis, the NHS All Wales Molecular Genetics Laboratory, Cardiff, has recently been able to successfully establish a national diagnostic test for *EGFR* mutations on ctDNA from peripheral blood samples, effective since April 2016. It currently accepts samples for testing from Wales and South-West England. An *EGFR* mutation array which includes the *EGFR* p.L858R mutation, the top 20 *EGFR* exon 19 deletions (together covering 85 % of the most common *EGFR* sensitizing mutations), and the *EGFR* c.2369C>T p.T790M resistance mutation, is assessed using ddPCR platform.

There are currently two main testing routes for this service. The first is a ctDNA *EGFR* analysis at diagnosis, when no tumour tissue is available for genetic testing, or *EGFR* testing on DNA extracted from tumour tissue has failed. If a test is positive, then a clinical report can be given to the clinician, who may consider EGFR TKi therapy, such as gefitinib. If the test is negative, a clinical report can be issued, stating that this may be a false negative result, and/or that the presence

of other, rarer *EGFR* sensitizing mutations cannot be excluded. In this situation, the clinician may wish to consider a new or repeat biopsy for the patient to obtain more definitive assessment of mutation status.

The second route is for patients that have developed clinical or radiological disease progression while taking EGFR TKi therapy. The *EGFR* c.2369C>T p.T790M resistance mutation can be tested for, along with the original sensitizing *EGFR* mutation, as an internal control, thus helping to confirm the presence of ctDNA. If an *EGFR* c.2369C>T p.T790M mutation is detected on ctDNA, this can be reported to the clinician, who may wish to consider 3rd generation EGFR TKi therapy, such as osimertinib. If *EGFR* c.2369C>T p.T790M mutation is not detected on ctDNA, it is recognized that this may be due to a false negative result. Again, the clinician may wish to consider repeat blood sampling and repeat ctDNA *EGFR* testing, or a new or repeat tumor biopsy for mutation analysis, if clinically appropriate.

This service through the All Wales Molecular Genetics Laboratory is ISO accredited, and subject to a new NEQAS programme for ctDNA testing. It seems likely to be the forerunner of similar ctDNA-based regional or national mutation testing services for routine clinical use. Such services will be relatively cheap, and a relatively quick turnaround time, of within 5 working days from receipt of the blood sample, making this testing pathway applicable in a clinically meaningful timeframe. This is especially important for patients with lung cancer, where clinical management decisions have to be made on an urgent basis, due to the relatively advanced stage of their disease, their relatively poor performance status and high incidence of co-morbidity, at presentation.

One current limitation to a testing service of this sort is that it is not a comprehensive *EGFR* mutational panel. An NGS platform could be adopted to assess for mutant *EGFR* in ctDNA in the NHS diagnostic labs, but this technology is currently limited by a poorer lower limit level of detection. Therefore, it may be more clinically appropriate to assay for specific mutations, with a much more sensitive molecular technique. It is not currently possible to increase the number

of *EGFR* assays beyond the current 85% most common EGFR mutations, using ddPCR, since the small amount of available total cell free DNA, after extraction, is often a limiting factor. It is hoped that, as technologies advance, reduced amounts of input DNA will be required for expanded, reliable mutation detection in ctDNA.

When establishing a clinical diagnostic service, another consideration is that of which technology to use. There are increasing number of ctDNA testing platforms, one study compares two amplification refractory mutation detection systems (Cobas-ARMS and ADx-ARMS), droplet digital PCR and NGS. It reports that the Cobas-ARMS, droplet digital PCR and NGS are more sensitive, but the ADx-ARMS gives better quantification of alternate allele frequency when >1%(151). Globally, there is currently no consensus technology to use. Challenges in platform choice have to include the context of treatment evaluation and offer the best chance of drug resistance detection.

7.4.4 CtDNA in patients with Small Cell Lung Cancer

In this thesis, it has been demonstrated that ctDNA can be detected, and quantitatively and qualitatively analysed in patients with locally advanced and metastatic small cell lung cancer. Quantitative decreases mutations in ctDNA were observed when a patient with small cell lung cancer received standard palliative chemotherapy. However, there remains a need to predict the course of disease, in order to influence decision-making on additional therapeutic interventions. For example, one current therapeutic strategy for patients with small cell lung cancer is prophylactic cranial irradiation, after initial systemic anti-cancer therapy. However, PCI has potential significant toxicities. Selection for PCI is currently based on individual clinical assessment, and there are no established biomarkers that help the clinician to stratify patients to PCI, to optimize clinical benefit from this intervention, and minimize exposure of patients to unnecessary treatment toxicity.

Data presented here suggest that baseline ctDNA load may be an independent prognostic indicator, for shorter term prognosis of SCLC patients, which could help decide on whether PCI or other treatments is/are in the best interests of the patient. In this thesis, a suggested cut-off level for poorer (<4 months) prognosis is a ctDNA burden of 44.3% alternate allele frequency. Patients who are predicted to have a poorer prognosis, despite a radiological or clinical response to chemotherapy, could be spared the potential inconvenience and treatment toxicity of PCI.

It is vitally important that future research should validate baseline ctDNA load as a prognostic biomarker, in a larger group of patients, prospectively, using the same methodology as in Chapter 6. At the same time, baseline ctDNA load could also be correlated with radiological volume of disease, and other prognostic indicators, such as the Manchester Prognostic Score, even though this provides an estimate of prognosis at the 2 year time point.

Clinical trials have shown an improvement in outcome for SCLC patients, with newer treatment strategies, such as concurrent chemo-radiotherapy, in earlier stages of disease. It would be useful to explore the use of ctDNA as a biomarker in earlier stages of disease, especially for patient stratification to more aggressive treatment regimes, or subsequent toxic therapies such as PCI.

As with most cancers, earlier disease detection and treatment may help to improve patient survival outcomes. Eventually, ctDNA may help as a biomarker to help screen and detect small cell lung cancer at an earlier stage. In order to do so, however, newer, more sensitive genetic technologies will need to be developed to detect lower levels of ctDNA in earlier stages of disease.

7.4.5 Areas for future research and development

As technologies advance, it may be possible in the future to detect ever-decreasing amounts of ctDNA from a blood sample. Therefore it is of increasing important to try to place low-level mutation detection in ctDNA in clinical context. For example, what constitutes minimal residual disease (MRD), or

response to a particular anticancer therapy? The topic of MRD after initial anti-cancer therapy has attracted much interest, for example tracking MRD in acute myeloid leukaemia, in order to help evaluate treatment efficiency, patient risk stratification and long-term outcome prediction(152). Methods commonly employed in monitoring MRD in AML include multi-colour flow cytometry and PCR based methods, but this has to be constantly re-evaluated in light of technologies delivering improved levels of detection(153).

In addition to ctDNA, there exist other exciting circulating biomarkers, including circulating exosomes (which contain both tumour derived DNA and RNA), micro-RNAs and platelet-associated circulating tumour RNA. Research into these as circulating biomarkers are in their infancy in lung cancer research, but an example includes circulating exosomal micro-RNA as prognostic biomarkers in lung cancer(154), and the micro-RNA, miR-126 in exosomes and serum in lung cancer patients(155).

Circualting Tumour Cells (CTCs) may also provide an interesting source of liquid tumour sample. Global gene expression analysis of CTCs in lung cancer is possible, and may provide clues into the mechanisms for metastasis, and provide potential therapeutic targets. For example *NOTCH1* overexpression as a driver in lung cancer CTCs(156). It is possible to take CTCs from lung cancer patients and introduce them into mice models, which develop metastatic lung cancer tin which therapeutic response and resistance can be investigated. These mice models could also facilitate pre-clinical drug development(157). The clinical context of CTCs in lung cancer has been investigated by authors such as Chudasama, for example, the presence of CTCs after lung surgery is not necessarily linked with a worse prognosis(158).

As biomarkers, including ctDNA, in lung cancer emerge, and their clinical utility is better understood, offers the potential of an era of earlier and more personalised treatment regimens for individual lung cancer patients, with better outcomes.

8 Appendix - Patient Information Sheet

Research Study

Patient Information Sheet

Feasibility Study to Obtain and Analyse Circulating Cell Free Tumour DNA in Patients with Lung Cancer, in South Wales

INFORMATION SHEET FOR POTENTIAL PARTICIPANTS WHO HAVE LUNG CANCER

You are being invited to take part in a Cardiff University research project. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

1 What is the purpose of the study?

Increasingly, we are discovering more and more genes that are implicated in lung cancer. A future goal is to tailor each patient's treatments, based on which genes are altered within each individual patient's tumour. These may be slightly different, for different patients. However, until recently, we have only been able to look for changes in these genes by analysing DNA taken from tumour samples that have be acquired by a biopsy, or a surgical resection specimen. It may now possible to do this with a simple blood test. We wish to see if we can isolate cancer DNA from blood samples, in Cardiff, from patients in South Wales, and analyse it to detect the different changes, within the different genes that are causing the cancer to grow. It may also be possible to monitor changes in cancer DNA in a blood test, as a patient undergoes any type of anti-cancer treatment. In the future, this may provide us with invaluable information, for example, by allowing us to predict how useful a specific treatment may be, or whether any other treatments may be more effective in treating the cancer.

2 Why am I suitable for this research?

We are looking to take blood samples from patients with a diagnosis of lung cancer, that has been confirmed with a biopsy. This research project requires extra blood samples, which can

conveniently be taken in addition to any routine blood tests that you may need as part of your treatment plan. Where possible, we are looking for blood samples before you start any treatment for your lung cancer. We are also interested in taking further blood tests during any treatment you may have, and after the treatment has finished.

3 Do I have to take part?

No, it is entirely up to you to decide whether or not to take part, and giving extra blood samples for research, is purely voluntary. If you do decide to take part you should keep this information sheet and you will be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect your future medical care in any way.

4 What samples are required from me for the research?

Each time you are happy for us to take extra blood for this research, we would require up to two 10 mls vials of venous blood, which can be taken at the same time as a routine blood test. We would also request a small section of your tumour tissue that is already stored in the pathology lab. This is provided that there is enough remaining after all of the required routine tests, that may be required by your clinical team, have been performed adequately.

Therefore we do not require additional biopsies, or surgery, for the purposes of this study. Once obtained, the cancer DNA from the tumour can also be analysed. This helps us to identify the any cancer DNA that may be present in the blood.

5 How long am I involved in the study?

Your treating clinical team will provide you with the opportunity to allow us to take further blood samples at different time points, during your treatment, and after any treatment has completed, provided you remain happy to do so. You will only be asked at routine clinic appointments. The frequency of research blood tests may vary, between every few weeks to every few months, depending on the type of treatment you receive.

6 Will I get any results?

As the samples are being used solely for the purposes of research within Cardiff University, you will not be directly informed of any results. The genetic tests that we perform are using techniques that are not yet validated NHS diagnostic techniques. The results will not have any impact on your treatment.

7 What are the benefits for me taking part?

Giving a blood sample for research is relatively physically easy, and the research sample can be taken alongside a routine blood test, for your convenience.

All the necessary tests that will influence your treatment, will be performed by your NHS team. We must stress that this research project is using experimental lab techniques, and they are not validated, diagnostic NHS tests. We are also analysing several genes, for which we currently do not fully understand the implications of changes in them on your treatment, which is why they need to be researched. Therefore the results of this research study will not influence the best treatment of your cancer.

8 Are there any risks?

When allowing a blood sample to be taken for research, there are no additional physical risks than a routine blood test.

9 How many patients are we hoping to involved?

As a feasibility study, we only require at approximately 10-20 patients. Once we have recruited enough patients, we can then develop future research projects that look at how this information can be used in the future to further help patients with lung cancer.

10 What happens if I am happy to take part in this research?

If you are interested in participating in this research study, then we will ask you to sign a consent form. You can keep a copy of the consent form, and a copy of the consent form will be kept with your NHS case-notes. We will then arrange the best time to take the blood

samples, taking into account if you are due a routine blood test and co-ordinate with any planned treatments that you may receive.

The blood sample taken for research will be processed in a laboratory, so that we can extract any DNA that is present. This DNA will consist of 'normal' DNA, and may also consist of DNA that has come from the cancer. Likewise, tumour DNA will be extracted in the same laboratory, from a small sample of your tumour that we will request from the pathology laboratory.

Once extracted, only the DNA will remain for the purposes of research by the Cardiff University research team. The DNA samples will be stored in the Institute of Medical Genetics, at the University Hospital of Wales, Cardiff, for the duration of the research project. Here, the Cardiff University research team can use different lab techniques to identify whether specific lung cancer gene mutations can be detected.

11 What happens when the research project finishes?

The project aims to collect samples from patients over the next 2 – 3 years. When we extract DNA, it is usually all used up by the different lab techniques that we will use for this research. Any unused or residual research DNA samples, after all research test have been performed, will be disposed of in accordance with standard protocol within the Institute of Medical Genetics, UHW, Cardiff, within 12 months of completion of this research project.

12 Will the identity of my samples be kept confidential?

If you consent to take part in the research the investigators may look at your medical records to aid with analysing the results. Your records may also be looked at by regulatory authorities to check that the study is being carried out correctly. Your name, however, will not be disclosed outside the participating hospitals or health professionals.

Once within the Institute of Medical Genetics laboratory, the extracted DNA from your samples will be allocated a unique identification number. The DNA will be analysed within Cardiff University using the unique identification numbers, rather than any patient identifiable information.

13 What will happen to the results of the research?

The results of this research will be published in a medical or scientific journal. You will not be identified in any publications, nor will any patient identifiable information be used.

14 Who is organising and funding the research?

This research project is jointly funded by Stepping Stones Appeal – Velindre Fundraising, which specifically aims to support lung cancer research in South Wales, and Cancer Research UK (CRUK).

15 Who Do I Contact if I have any concerns?

You can discuss your concerns at any time, by contacting your treating clinical team, namely your consultant, or their registrar, or one of the specialist nurses who will usually have provided you with their contact details. You can also use the contacts listed below. We will consider all concerns seriously, and we can provide you with opportunities to discuss any concerns further, in clinic, so that we can adequately address them.

Cardiff and Vale University Health Board also have a formal complaints procedure, should you feel that you wish to make a formal complaint about your care, including if you are taking part in any research studies. A letter of complaint can be sent to the complaints department via fax (02920 336365), or e-mail to concerns@wales.nhs.uk , or they can be telephoned on (02920) 744095 for advice about how to proceed with a complaint. More information can be found at www.cardiffandvalehb.wales.nhs.uk/complaints

16 What do I do if I no longer want to participate in this research study?

Even if you give consent to take part in this research study, you can withdraw your consent and participation at any time, without giving any reason. This will not detract from your clinical care in any way. To no longer take part, then just inform your treating consultant, or a member of their team including any of the specialist nurses that you have the contact information for. Or you can contact the contacts listed below. After which, no further samples, or clinical information, will be taken or used for research purposes.

17 What happens if I can no longer give consent to provide research blood samples?

In any event that you are unable to provide consent, for any reason, no research samples will be taken. However, any samples that you have donated, while you were consenting, will have been anonymised, and as a result they can continue to be processed and analysed for research purposes. All research DNA samples will also continue to be stored for the duration of this research project, after which time, they will be appropriately destroyed.

18 Further information

If you have any further questions or queries about this research project, then this can be obtained from myself or;

Dr Jason Lester
Velindre Cancer Centre
Whitchurch, Cardiff. CF14 2TL
Tel :02920 615 888

Or

Dr Rachel Butler
Institute of Medical Genetics
UHW, Cardiff. CF14 4XW
Tel: 02920 742 641

Thank you for taking time to read this information sheet, and considering taking part in research.

Dr Daniel Nelmes
Clinical Research Fellow, Institute of Medical Genetics, Cardiff.

9 Appendix - Patient Consent Sheet

Patient Identification Number for this trial:

CONSENT FORM

Title of Project:

Feasibility Study to Obtain and Analyse Circulating Cell Free Tumour DNA in Patients with Lung Cancer, in South Wales.

Name of Main Researcher: **Dr Daniel Nelmes**

Please initial all boxes

1. I confirm that I have read and understand the information sheet dated 28.01.2015 (version C&V 2.5) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected. ☐

3. I understand that relevant sections of my medical notes will be consulted, and data collected during the study, by Cardiff University study investigators, where it is relevant to my taking part in this research. Clinical data used for this research purposes will be anonymised. I give permission for these individuals to have access to my records. ☐

4. I understand that DNA extracted from my samples will be stored in the institute of Medical Genetics, UHW, Cardiff, only for the duration of this research project. They will only be accessed by the Cardiff University Research team, for the purposes of this research only. ☐

5. Results obtained from analysing the samples are for research purposes only. I understand that no results will be fed back to myself, or my treating clinical team. ☐

6. I understand that results from this research may be published in medical or scientific journals, but they will not contain any patient identifiable information. ☐

7. I agree to take part in the above study. ☐

Name of Participant	Date	Signature
Name of Person taking consent.	Date	Signature

Consent form date of issue: 28.01.2015
Consent form version number: C&V 2.5

10 References

1. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/incidence.
2. Statistics C-LCI. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/incidence. Accessed 1 Jun 2017.
3. Dela Cruz CS, Tanoue LT, Matthay RA. Lung cancer: epidemiology, etiology, and prevention. Clin Chest Med. 2011;32(4):605-44.
4. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JH, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. J Thorac Oncol. 2015;10(9):1243-60.
5. McPhail S, Johnson S, Greenberg D, Peake M, Rous B. Stage at diagnosis and early mortality from cancer in England. Br J Cancer. 2015;112 Suppl 1:S108-15.
6. Kale MS, Mhango G, Gomez JE, Sigel K, Smith CB, Bonomi M, et al. Treatment Toxicity in Elderly Patients With Advanced Non-Small Cell Lung Cancer. Am J Clin Oncol. 2015.
7. Früh M, De Ruysscher D, Popat S, Crinò L, Peters S, Felip E, et al. Small-cell lung cancer (SCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2013;24 Suppl 6:vi99-105.
8. Vansteenkiste J, De Ruysscher D, Eberhardt WE, Lim E, Senan S, Felip E, et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2013;24 Suppl 6:vi89-98.
9. Wozniak AJ, Gadgeel SM. Adjuvant therapy for resected non-small cell lung cancer. Ther Adv Med Oncol. 2009;1(2):109-18.
10. Group NM-AC. 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. 2008;26(28):4617-25.
11. Schiller JH, Harrington D, Belani CP, Langer C, Sandler A, Krook J, et al. Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. N Engl J Med. 2002;346(2):92-8.
12. Scagliotti GV, Parikh P, von Pawel J, Biesma B, Vansteenkiste J, Manegold C, et al. Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. J Clin Oncol. 2008;26(21):3543-51.
13. Paz-Ares LG, de Marinis F, Dediu M, Thomas M, Pujol JL, Bidoli P, et al. PARAMOUNT: Final overall survival results of the phase III study of maintenance pemetrexed versus placebo immediately after induction treatment with pemetrexed

plus cisplatin for advanced nonsquamous non-small-cell lung cancer. J Clin Oncol. 2013;31(23):2895-902.
14. Fossella FV. Docetaxel in second-line treatment of non-small-cell lung cancer. Clin Lung Cancer. 2002;3 Suppl 2:S23-8.
15. Sorensen SF, Zhou W, Dolled-Filhart M, Georgsen JB, Wang Z, Emancipator K, et al. PD-L1 Expression and Survival among Patients with Advanced Non-Small Cell Lung Cancer Treated with Chemotherapy. Transl Oncol. 2016;9(1):64-9.
16. Herbst RS, Baas P, Kim DW, Felip E, Pérez-Gracia JL, Han JY, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet. 2016;387(10027):1540-50.
17. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. N Engl J Med. 2016;375(19):1823-33.
18. TA(428) N. Pembrolizumab for treating PD-L1-positive non-small cell lung cancer after chemotherapy. 2017.
19. TA(447) N. Pembrolizumab for untreated PD-L1 positive metastatic non-small cell lung cancer. 2017.
20. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer#heading-Zero.
21. http://www.cancerresearchuk.org/about-cancer/type/lung-cancer/treatment/statistics-and-outlook-for-lung-cancer.
22. Chute JP, Chen T, Feigal E, Simon R, Johnson BE. Twenty years of phase III trials for patients with extensive-stage small-cell lung cancer: perceptible progress. Journal of Clinical Oncology. 1999;17(6):1794-801.
23. Thatcher N, Faivre-Finn C, Lorigan P. Management of small-cell lung cancer. Annals of Oncology. 2005;16 Suppl 2:ii235-9.
24. Sundstrom S, Bremnes RM, Kaasa S, Aasebo U, Hatlevoll R, Dahle R, et al. Cisplatin and etoposide regimen is superior to cyclophosphamide, epirubicin, and vincristine regimen in small-cell lung cancer: results from a randomized phase III trial with 5 years' follow-up. Journal of Clinical Oncology. 2002;20(24):4665-72.
25. Sundstrom S, Bremnes RM, Kaasa S, Aasebo U, Aamdal S, Norwegian Lung Cancer Study G. Second-line chemotherapy in recurrent small cell lung cancer. Results from a crossover schedule after primary treatment with cisplatin and etoposide (EP-regimen) or cyclophosphamide, epirubicin, and vincristin (CEV-regimen). Lung Cancer. 2005;48(2):251-61.
26. Schiller JH, Adak S, Cella D, DeVore RF, 3rd, Johnson DH. Topotecan versus observation after cisplatin plus etoposide in extensive-stage small-cell lung cancer: E7593--a phase III trial of the Eastern Cooperative Oncology Group. Journal of Clinical Oncology. 2001;19(8):2114-22.
27. Slotman B, Faivre-Finn C, Kramer G, Rankin E, Snee M, Hatton M, et al. Prophylactic cranial irradiation in extensive small-cell lung cancer.[Reprint in Ned Tijdschr Geneeskd. 2008 Apr 26;152(17):1000-4; PMID: 18549175]. New England Journal of Medicine. 2007;357(7):664-72.

28. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.
29. Furrukh M. Tobacco Smoking and Lung Cancer: Perception-changing facts. Sultan Qaboos Univ Med J. 2013;13(3):345-58.
30. Cheng L, Alexander RE, Maclennan GT, Cummings OW, Montironi R, Lopez-Beltran A, et al. Molecular pathology of lung cancer: key to personalized medicine. Modern Pathology. 2012;25(3):347-69.
31. Marks JL, McLellan MD, Zakowski MF, Lash AE, Kasai Y, Broderick S, et al. Mutational analysis of EGFR and related signaling pathway genes in lung adenocarcinomas identifies a novel somatic kinase domain mutation in FGFR4. PLoS One. 2007;2(5):e426.
32. https://www.mycancergenome.org/content/disease/lung-cancer/. accessed 13.01.2018.
33. Thatcher N, Chang A, Parikh P, Rodrigues Pereira J, Ciuleanu T, von Pawel J, et al. Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). Lancet. 2005;366(9496):1527-37.
34. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med. 2004;350(21):2129-39.
35. Skov BG, Høgdall E, Clementsen P, Krasnik M, Larsen KR, Sørensen JB, et al. The prevalence of EGFR mutations in non-small cell lung cancer in an unselected Caucasian population. APMIS. 2015;123(2):108-15.
36. Sharma SV, Bell DW, Settleman J, Haber DA. Epidermal growth factor receptor mutations in lung cancer. Nat Rev Cancer. 2007;7(3):169-81.
37. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med. 2009;361(10):947-57.
38. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. 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. 2012;13(3):239-46.
39. NICE. http://guidance .nice.org.uk/TA258.
40. NICE. http://guidance .nice.org.uk/TA162.
41. NICE. http://guidance .nice.org.uk/TA192.
42. Shepherd FA, Rodrigues Pereira J, Ciuleanu T, Tan EH, Hirsh V, Thongprasert S, et al. Erlotinib in previously treated non-small-cell lung cancer. N Engl J Med. 2005;353(2):123-32.
43. Sequist LV, Yang JC, Yamamoto N, O'Byrne K, Hirsh V, Mok T, et al. Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. J Clin Oncol. 2013;31(27):3327-34.

44. Huang L, Fu L. Mechanisms of resistance to EGFR tyrosine kinase inhibitors. Acta Pharm Sin B. 2015;5(5):390-401.
45. Jänne PA, Yang JC, Kim DW, Planchard D, Ohe Y, Ramalingam SS, et al. AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. N Engl J Med. 2015;372(18):1689-99.
46. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007;448(7153):561-6.
47. Chia PL, Mitchell P, Dobrovic A, John T. Prevalence and natural history of ALK positive non-small-cell lung cancer and the clinical impact of targeted therapy with ALK inhibitors. Clin Epidemiol. 2014;6:423-32.
48. Chuang JC, Neal JW. Crizotinib as first line therapy for advanced ALK-positive non-small cell lung cancers. Transl Lung Cancer Res. 2015;4(5):639-41.
49. Shaw AT, Kim DW, Nakagawa K, Seto T, Crinó L, Ahn MJ, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. N Engl J Med. 2013;368(25):2385-94.
50. Casaluce F, Sgambato A, Sacco PC, Palazzolo G, Maione P, Rossi A, et al. Resistance to Crizotinib in Advanced Non-Small Cell Lung Cancer (NSCLC) with ALK Rearrangement: Mechanisms, Treatment Strategies and New Targeted Therapies. Curr Clin Pharmacol. 2016;11(2):77-87.
51. Slebos RJ, Kibbelaar RE, Dalesio O, Kooistra A, Stam J, Meijer CJ, et al. K-ras oncogene activation as a prognostic marker in adenocarcinoma of the lung. N Engl J Med. 1990;323(9):561-5.
52. Kosaka T, Yatabe Y, Endoh H, Kuwano H, Takahashi T, Mitsudomi T. Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. Cancer Res. 2004;64(24):8919-23.
53. Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba II, et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. J Natl Cancer Inst. 2005;97(5):339-46.
54. Mulloy R, Ferrand A, Kim Y, Sordella R, Bell DW, Haber DA, et al. Epidermal growth factor receptor mutants from human lung cancers exhibit enhanced catalytic activity and increased sensitivity to gefitinib. Cancer Res. 2007;67(5):2325-30.
55. Wang SE, Narasanna A, Perez-Torres M, Xiang B, Wu FY, Yang S, et al. HER2 kinase domain mutation results in constitutive phosphorylation and activation of HER2 and EGFR and resistance to EGFR tyrosine kinase inhibitors. Cancer Cell. 2006;10(1):25-38.
56. Cappuzzo F, Bemis L, Varella-Garcia M. HER2 mutation and response to trastuzumab therapy in non-small-cell lung cancer. N Engl J Med. 2006;354(24):2619-21.
57. Cappuzzo F, Varella-Garcia M, Shigematsu H, Domenichini I, Bartolini S, Ceresoli GL, et al. Increased HER2 gene copy number is associated with response to gefitinib therapy in epidermal growth factor receptor-positive non-small-cell lung cancer patients. J Clin Oncol. 2005;23(22):5007-18.

58. Bang YJ, Van Cutsem E, Feyereislova A, Chung HC, Shen L, Sawaki A, et al. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. Lancet. 2010;376(9742):687-97.
59. Swain SM, Baselga J, Kim SB, Ro J, Semiglazov V, Campone M, et al. Pertuzumab, trastuzumab, and docetaxel in HER2-positive metastatic breast cancer. N Engl J Med. 2015;372(8):724-34.
60. Ji H, Wang Z, Perera SA, Li D, Liang MC, Zaghlul S, et al. Mutations in BRAF and KRAS converge on activation of the mitogen-activated protein kinase pathway in lung cancer mouse models. Cancer Res. 2007;67(10):4933-9.
61. Bader AG, Kang S, Zhao L, Vogt PK. Oncogenic PI3K deregulates transcription and translation. Nat Rev Cancer. 2005;5(12):921-9.
62. Majewski IJ, Mittempergher L, Davidson NM, Bosma A, Willems SM, Horlings HM, et al. Identification of recurrent FGFR3 fusion genes in lung cancer through kinome-centred RNA sequencing. J Pathol. 2013;230(3):270-6.
63. Milowsky MI, Dittrich C, Durán I, Jagdev S, Millard FE, Sweeney CJ, et al. Phase 2 trial of dovitinib in patients with progressive FGFR3-mutated or FGFR3 wild-type advanced urothelial carcinoma. Eur J Cancer. 2014;50(18):3145-52.
64. Kong-Beltran M, Seshagiri S, Zha J, Zhu W, Bhawe K, Mendoza N, et al. Somatic mutations lead to an oncogenic deletion of met in lung cancer. Cancer Res. 2006;66(1):283-9.
65. Onozato R, Kosaka T, Kuwano H, Sekido Y, Yatabe Y, Mitsudomi T. Activation of MET by gene amplification or by splice mutations deleting the juxtamembrane domain in primary resected lung cancers. J Thorac Oncol. 2009;4(1):5-11.
66. Schrock AB, Frampton GM, Suh J, Chalmers ZR, Rosenzweig M, Erlich RL, et al. Characterization of 298 Patients with Lung Cancer Harboring MET Exon 14 Skipping Alterations. J Thorac Oncol. 2016;11(9):1493-502.
67. Takeuchi K, Soda M, Togashi Y, Suzuki R, Sakata S, Hatano S, et al. RET, ROS1 and ALK fusions in lung cancer. Nat Med. 2012;18(3):378-81.
68. Drilon A, Wang L, Hasanovic A, Suehara Y, Lipson D, Stephens P, et al. Response to Cabozantinib in patients with RET fusion-positive lung adenocarcinomas. Cancer Discov. 2013;3(6):630-5.
69. Jänne PA, van den Heuvel MM, Barlesi F, Cobo M, Mazieres J, Crinò L, et al. Selumetinib Plus Docetaxel Compared With Docetaxel Alone and Progression-Free Survival in Patients With KRAS-Mutant Advanced Non-Small Cell Lung Cancer: The SELECT-1 Randomized Clinical Trial. JAMA. 2017;317(18):1844-53.
70. Ou SH, Kwak EL, Siwak-Tapp C, Dy J, Bergethon K, Clark JW, et al. Activity of crizotinib (PF02341066), a dual mesenchymal-epithelial transition (MET) and anaplastic lymphoma kinase (ALK) inhibitor, in a non-small cell lung cancer patient with de novo MET amplification. J Thorac Oncol. 2011;6(5):942-6.
71. Hyman DM, Puzanov I, Subbiah V, Faris JE, Chau I, Blay J-Y, et al. Vemurafenib in Multiple Nonmelanoma Cancers with BRAF V600 Mutations. New England Journal of Medicine. 2015;373(8):726-36.

72. Planchard D, Kim TM, Mazieres J, Quoix E, Riely G, Barlesi F, et al. Dabrafenib in patients with BRAF(V600E)-positive advanced non-small-cell lung cancer: a single-arm, multicentre, open-label, phase 2 trial. Lancet Oncol. 2016;17(5):642-50.
73. Bergethon K, Shaw AT, Ou SH, Katayama R, Lovly CM, McDonald NT, et al. ROS1 rearrangements define a unique molecular class of lung cancers. J Clin Oncol. 2012;30(8):863-70.
74. Meldrum C, Doyle MA, Tothill RW. Next-generation sequencing for cancer diagnostics: a practical perspective. Clin Biochem Rev. 2011;32(4):177-95.
75. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013;45(10):1113-20.
76. Greaves M, Maley CC. Clonal evolution in cancer. Nature. 2012;481(7381):306-13.
77. Jamal-Hanjani M, Quezada SA, Larkin J, Swanton C. Translational implications of tumor heterogeneity. Clin Cancer Res. 2015;21(6):1258-66.
78. Swanton C. Intratumor heterogeneity: evolution through space and time. Cancer Res. 2012;72(19):4875-82.
79. Bedard PL, Hansen AR, Ratain MJ, Siu LL. Tumour heterogeneity in the clinic. Nature. 2013;501(7467):355-64.
80. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502(7471):333-9.
81. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the Evolution of Non-Small-Cell Lung Cancer. N Engl J Med. 2017;376(22):2109-21.
82. Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature. 2010;463(7278):184-90.
83. Wistuba II, Gazdar AF, Minna JD. Molecular genetics of small cell lung carcinoma. Seminars in Oncology. 2001;28(2 Suppl 4):3-13.
84. George J, Lim JS, Jang SJ, Cun Y, Ozretic L, Kong G, et al. Comprehensive genomic profiles of small cell lung cancer. Nature. 2015;524(7563):47-53.
85. Rudin CM, Durinck S, Stawiski EW, Poirier JT, Modrusan Z, Shames DS, et al. Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. Nature Genetics. 2012;44(10):1111-6.
86. Santarpia M, Daffinà MG, Karachaliou N, González-Cao M, Lazzari C, Altavilla G, et al. Targeted drugs in small-cell lung cancer. Transl Lung Cancer Res. 2016;5(1):51-70.
87. Alexandre J, Brown C, Coeffic D, Raban N, Pfisterer J, Mäenpää J, et al. CA-125 can be part of the tumour evaluation criteria in ovarian cancer trials: experience of the GCIG CALYPSO trial. Br J Cancer. 2012;106(4):633-7.
88. Ettinger DS, Wood DE, Akerley W, Bazhenova LA, Borghaei H, Camidge DR, et al. NCCN Guidelines Insights: Non-Small Cell Lung Cancer, Version 4.2016. J Natl Compr Canc Netw. 2016;14(3):255-64.

89. Milbury CA, Li J, Liu P, Makrigiorgos GM. COLD-PCR: improving the sensitivity of molecular diagnostics assays. Expert Rev Mol Diagn. 2011;11(2):159-69.
90. Milbury CA, Zhong Q, Lin J, Williams M, Olson J, Link DR, et al. Determining lower limits of detection of digital PCR assays for cancer-related gene mutations. Biomol Detect Quantif. 2014;1(1):8-22.
91. Sahnane N, Gueli R, Tibiletti MG, Bernasconi B, Stefanoli M, Franzi F, et al. Pyrosequencing for EGFR mutation detection: diagnostic accuracy and clinical implications. Diagn Mol Pathol. 2013;22(4):196-203.
92. Cruz-Rico G, Avilés-Salas A, Segura-González M, Espinosa-García AM, Ramírez-Tirado LA, Morales-Oyarvide V, et al. Diagnosis of EML4-ALK Translocation With FISH, Immunohistochemistry, and Real-time Polymerase Chain Reaction in Patients With Non-Small Cell Lung Cancer. Am J Clin Oncol. 2015.
93. Lorenz J, Blum M. Complications of percutaneous chest biopsy. Semin Intervent Radiol. 2006;23(2):188-93.
94. Vanderlaan PA, Yamaguchi N, Folch E, Boucher DH, Kent MS, Gangadharan SP, et al. Success and failure rates of tumor genotyping techniques in routine pathological samples with non-small-cell lung cancer. Lung Cancer. 2014;84(1):39-44.
95. Hiltermann TJ, Pore MM, van den Berg A, Timens W, Boezen HM, Liesker JJ, et al. Circulating tumor cells in small-cell lung cancer: a predictive and prognostic factor. Ann Oncol. 2012;23(11):2937-42.
96. Hanssen A, Loges S, Pantel K, Wikman H. Detection of Circulating Tumor Cells in Non-Small Cell Lung Cancer. Front Oncol. 2015;5:207.
97. Liao GJ, Gronowski AM, Zhao Z. Non-invasive prenatal testing using cell-free fetal DNA in maternal circulation. Clin Chim Acta. 2014;428:44-50.
98. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res. 2001;61(4):1659-65.
99. Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. Nat Rev Cancer. 2011;11(6):426-37.
100. Hohaus S, Giachelia M, Massini G, Mansueto G, Vannata B, Bozzoli V, et al. Cell-free circulating DNA in Hodgkin's and non-Hodgkin's lymphomas. Ann Oncol. 2009;20(8):1408-13.
101. Diaz LA, Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. J Clin Oncol. 2014;32(6):579-86.
102. Heitzer E, Ulz P, Geigl JB. Circulating tumor DNA as a liquid biopsy for cancer. Clin Chem. 2015;61(1):112-23.
103. Dawson S-J, Tsui DWY, Murtaza M, Biggs H, Rueda OM, Chin S-F, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. New England Journal of Medicine. 2013;368(13):1199-209.
104. Parkinson CA, Gale D, Piskorz AM, Biggs H, Hodgkin C, Addley H, et al. Exploratory Analysis of TP53 Mutations in Circulating Tumour DNA as Biomarkers of Treatment Response for Patients with Relapsed High-Grade Serous Ovarian Carcinoma: A Retrospective Study. PLoS Med. 2016;13(12):e1002198.

105. Murtaza M, Dawson SJ, Tsui DW, Gale D, Forshew T, Piskorz AM, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature. 2013;497(7447):108-12.
106. Diaz LA, Jr., Williams RT, Wu J, Kinde I, Hecht JR, Berlin J, et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. Nature. 2012;486(7404):537-40.
107. Chabon JJ, Simmons AD, Lovejoy AF, Esfahani MS, Newman AM, Haringsma HJ, et al. Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients. Nat Commun. 2016;7:11815.
108. Ma C, Wei S, Song Y. T790M and acquired resistance of EGFR TKI: a literature review of clinical reports. J Thorac Dis. 2011;3(1):10-8.
109. Chia PL, Do H, Morey A, Mitchell P, Dobrovic A, John T. Temporal changes of EGFR mutations and T790M levels in tumour and plasma DNA following AZD9291 treatment. Lung Cancer. 2016;98:29-32.
110. Rothwell DG, Smith N, Morris D, Leong HS, Li Y, Hollebecque A, et al. Genetic profiling of tumours using both circulating free DNA and circulating tumour cells isolated from the same preserved whole blood sample. Mol Oncol. 2016;10(4):566-74.
111. Group ECMCEaNCRINBICS. Cell-free DNA Consensus Meeting. Charels Wilson Building, University of Leicester2014.
112. Rothé F, Laes JF, Lambrechts D, Smeets D, Vincent D, Maetens M, et al. Plasma circulating tumor DNA as an alternative to metastatic biopsies for mutational analysis in breast cancer. Annals of Oncology. 2014;25(10):1959-65.
113. Sandmann S, de Graaf AO, Karimi M, van der Reijden BA, Hellström-Lindberg E, Jansen JH, et al. Evaluating Variant Calling Tools for Non-Matched Next-Generation Sequencing Data. Sci Rep. 2017;7:43169.
114. Lanman RB, Mortimer SA, Zill OA, Sebisanovic D, Lopez R, Blau S, et al. Analytical and Clinical Validation of a Digital Sequencing Panel for Quantitative, Highly Accurate Evaluation of Cell-Free Circulating Tumor DNA. PLoS One. 2015;10(10):e0140712.
115. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med. 2012;4(136):136ra68.
116. Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med. 2014;20(5):548-54.
117. Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F, et al. Integrated digital error suppression for improved detection of circulating tumor DNA. Nat Biotechnol. 2016;34(5):547-55.
118. Scientific T. https://tools.thermofisher.com/content/sfs/manuals/MAN0006735_AmpliSeq_DNA_RNA_LibPrep_UG.pdf.

119. Bollag G, Clapp DW, Shih S, Adler F, Zhang YY, Thompson P, et al. Loss of NF1 results in activation of the Ras signaling pathway and leads to aberrant growth in haematopoietic cells. Nat Genet. 1996;12(2):144-8.
120. Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature. 2005;434(7035):917-21.
121. Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature. 2005;434(7035):913-7.
122. Xu H, Yu S, Liu Q, Yuan X, Mani S, Pestell RG, et al. Recent advances of highly selective CDK4/6 inhibitors in breast cancer. J Hematol Oncol. 2017;10(1):97.
123. Young RJ, Waldeck K, Martin C, Foo JH, Cameron DP, Kirby L, et al. Loss of CDKN2A expression is a frequent event in primary invasive melanoma and correlates with sensitivity to the CDK4/6 inhibitor PD0332991 in melanoma cell lines. Pigment Cell Melanoma Res. 2014;27(4):590-600.
124. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22(3):568-76.
125. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research. 2010;38(16):e164.
126. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.
127. Sozzi G, Conte D, Leon M, Ciricione R, Roz L, Ratcliffe C, et al. Quantification of free circulating DNA as a diagnostic marker in lung cancer. J Clin Oncol. 2003;21(21):3902-8.
128. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Research. 2012;22(3):568-76.
129. Kiyoi H, Naoe T. FLT3 mutations in acute myeloid leukemia. Methods Mol Med. 2006;125:189-97.
130. Levis M. FLT3 mutations in acute myeloid leukemia: what is the best approach in 2013? Hematology Am Soc Hematol Educ Program. 2013;2013:220-6.
131. Munroe DJ, Harris TJ. Third-generation sequencing fireworks at Marco Island. Nat Biotechnol. 2010;28(5):426-8.
132. Schwarzenbach H, Stoehlmacher J, Pantel K, Goekkurt E. Detection and monitoring of cell-free DNA in blood of patients with colorectal cancer. Ann N Y Acad Sci. 2008;1137:190-6.
133. Irmer D, Funk JO, Blaukat A. EGFR kinase domain mutations - functional impact and relevance for lung cancer therapy. Oncogene. 2007;26(39):5693-701.
134. Suh JH, Johnson A, Albacker L, Wang K, Chmielecki J, Frampton G, et al. Comprehensive Genomic Profiling Facilitates Implementation of the National Comprehensive Cancer Network Guidelines for Lung Cancer Biomarker Testing and

Identifies Patients Who May Benefit From Enrollment in Mechanism-Driven Clinical Trials. Oncologist. 2016;21(6):684-91.
135. Dowlati A, Lipka MB, McColl K, Dabir S, Behtaj M, Kresak A, et al. Clinical correlation of extensive-stage small-cell lung cancer genomics. Ann Oncol. 2016;27(4):642-7.
136. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366(10):883-92.
137. Zheng D, Ye X, Zhang MZ, Sun Y, Wang JY, Ni J, et al. Plasma EGFR T790M ctDNA status is associated with clinical outcome in advanced NSCLC patients with acquired EGFR-TKI resistance. Sci Rep. 2016;6:20913.
138. Siravegna G, Mussolin B, Buscarino M, Corti G, Cassingena A, Crisafulli G, et al. Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. Nat Med. 2015;21(7):795-801.
139. Evans WK, Shepherd FA, Feld R, Osoba D, Dang P, Deboer G. VP-16 and cisplatin as first-line therapy for small-cell lung cancer. J Clin Oncol. 1985;3(11):1471-7.
140. Ziemke M, Patil T, Nolan K, Tippimanchai D, Malkoski SP. Reduced Smad4 expression and DNA topoisomerase inhibitor chemosensitivity in non-small cell lung cancer. Lung Cancer. 2017;109:28-35.
141. Kawahara M, Fukuoka M, Saijo N, Nishiwaki Y, Ikegami H, Tamura T, et al. Prognostic factors and prognostic staging system for small cell lung cancer. Jpn J Clin Oncol. 1997;27(3):158-65.
142. Karachaliou N, Sosa AE, Rosell R. Unraveling the genomic complexity of small cell lung cancer. Transl Lung Cancer Res. 2016;5(4):363-6.
143. Spigel DR. PARP inhibitors in lung cancer. J Thorac Oncol. 2012;7(16 Suppl 5):S392-3.
144. Middleton G, Crack LR, Popat S, Swanton C, Hollingsworth SJ, Buller R, et al. The National Lung Matrix Trial: translating the biology of stratification in advanced non-small-cell lung cancer. Ann Oncol. 2015;26(12):2464-9.
145. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature. 2017;545(7655):446-51.
146. Shaw E. http://www.ncri.org.uk/wp-content/uploads/2017/01/The-CRUK-Programme-Dr-Emily-Shaw.pdf. 2017.
147. Genomics England Sample Handling Guidance v2.2 (July 2016). https://wwwgenomicsenglandcouk/information-for-gmc-staff/sample-handling-guidance/.
148. Salvi S, Casadio V, Conteduca V, Burgio SL, Menna C, Bianchi E, et al. Circulating cell-free AR and CYP17A1 copy number variations may associate with outcome of metastatic castration-resistant prostate cancer patients treated with abiraterone. Br J Cancer. 2015;112(10):1717-24.

149. Chicard M, Boyault S, Colmet Daage L, Richer W, Gentien D, Pierron G, et al. Genomic Copy Number Profiling Using Circulating Free Tumor DNA Highlights Heterogeneity in Neuroblastoma. Clin Cancer Res. 2016.
150. Kirkizlar E, Zimmermann B, Constantin T, Swenerton R, Hoang B, Wayham N, et al. Detection of Clonal and Subclonal Copy-Number Variants in Cell-Free DNA from Patients with Breast Cancer Using a Massively Multiplexed PCR Methodology. Transl Oncol. 2015;8(5):407-16.
151. Xu T, Kang X, You X, Dai L, Tian D, Yan W, et al. Cross-Platform Comparison of Four Leading Technologies for Detecting. Theranostics. 2017;7(6):1437-46.
152. Matsuda K, Sugano M, Honda T. PCR for monitoring of minimal residual disease in hematologic malignancy. Clin Chim Acta. 2012;413(1-2):74-80.
153. Hauwel M, Matthes T. Minimal residual disease monitoring: the new standard for treatment evaluation of haematological malignancies? Swiss Med Wkly. 2014;144:w13907.
154. Liu Q, Yu Z, Yuan S, Xie W, Li C, Hu Z, et al. Circulating exosomal microRNAs as prognostic biomarkers for non-small-cell lung cancer. Oncotarget. 2017;8(8):13048-58.
155. Grimolizzi F, Monaco F, Leoni F, Bracci M, Staffolani S, Bersaglieri C, et al. Exosomal miR-126 as a circulating biomarker in non-small-cell lung cancer regulating cancer progression. Sci Rep. 2017;7(1):15277.
156. Mariscal J, Alonso-Nocelo M, Muinelo-Romay L, Barbazan J, Vieito M, Abalo A, et al. Molecular Profiling of Circulating Tumour Cells Identifies Notch1 as a Principal Regulator in Advanced Non-Small Cell Lung Cancer. Sci Rep. 2016;6:37820.
157. Lallo A, Schenk MW, Frese KK, Blackhall F, Dive C. Circulating tumor cells and CDX models as a tool for preclinical drug development. Transl Lung Cancer Res. 2017;6(4):397-408.
158. Chudasama D, Barr J, Beeson J, Beddow E, McGonigle N, Rice A, et al. Detection of Circulating Tumour Cells and Survival of Patients with Non-small Cell Lung Cancer. Anticancer Res. 2017;37(1):169-73.