The role of intratumour heterogeneity and chromosomal instability in cancer

Mariam Jamal-Hanjani

University College London

PhD Supervisor: Professor Charles Swanton

A thesis submitted for the degree of

Doctor of Philosophy

University College London

August 2015

Declaration

I, Mariam Jamal-Hanjani, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Increasing evidence supports the existence of intratumour heterogeneity in many solid and haematological tumour types, with potential clinical implications for both cancer diagnosis and treatment. Multi-region whole-exome sequencing of surgically resected non-small cell lung cancer (NSCLC) tumours demonstrated intratumour spatial and temporal heterogeneity in the mutational burden, copy number aberrations, and mutational signatures identified in these tumours. Furthermore, heterogeneity of mutations, including driver mutations, was also demonstrated in pre-invasive lung adenocarcinoma *in situ* lesions, suggesting that clonal evolution may be a feature of the early stages of cancer development. Whilst deciphering the clonal landscape of tumours may rely on multi-region and repeated tissue sampling, this remains challenging outside the context of clinical studies, and is not routine clinical practice. A non-invasive alternative may be the use of circulating biomarkers, such as circulating cell-free tumour DNA (cfDNA). Truncal and branch mutations were identified in cfDNA from patients with early stage NSCLC using different approaches. The detection of low frequency branch mutations, which are predicted to be subclonal in origin and may be potentially involved in the emergence of therapeutic resistance and tumour progression, were difficult to identify in cfDNA. Further studies are required to develop effective strategies for clonal and subclonal mutation detection in cfDNA, and to determine the utility of such biomarkers in representing the tumour genomic landscape, and in tracking tumour evolution in time.

Chromosomal instability (CIN), describes an increased rate of numerical and structural chromosome aberrations, and is a known driver of intercellular genetic tumour heterogeneity. CIN has been shown to be associated with drug resistance and poor clinical outcome in several cancer types. However, in oestrogen receptor (ER)-negative breast cancer it has previously been shown that extreme CIN is associated with improved clinical outcome, consistent with a negative impact of CIN on cellular fitness and growth. This paradoxical relationship was further validated in a large breast cancer cohort study, in which extreme CIN was associated with improved outcome in patients with ER-negative cancer (p trend = 0.03). A similar relationship was seen in ER-negative/human epidermal growth factor receptor (HER2)-negative cancers (p trend = 0.007). Identifying such patients may help distinguish good from poor prognostic groups, and therefore support treatment and risk stratification.

We are yet to uncover the true extent of intratumour heterogeneity and CIN in different cancer types, their relevance to clinical outcome, and how we may be able to overcome or exploit these features for the therapeutic gain and benefit of patients with cancer. Longitudinal studies employing serial tissue and circulating biomarker sampling have the potential to address these questions, and to truly define the breadth of genetic diversity in different tumour types and its relevance to patient outcome.

Acknowledgement

I would like to thank my supervisor, Charles Swanton, for his support and encouragement throughout my PhD. I would like to thank the Translational Cancer Therapeutics Laboratories at the UCL Cancer Institute and the Francis Crick Institute. In particular, I would like to thank Maria Cerone, Nnennaya Kanu, Elza de Bruin, and Seema Shafi for their support and encouragement. I would like to especially thank Su Kit Chew, Maria Cerone, Gareth Wilson, Nicolai Birkbak, Nnennaya Kanu, and Samra Turajlic for their help in proofreading this thesis. The work presented in this thesis would not have been possible without collaboration. I would like to thank Roger A'Hern, Pat Gorman, Eva Gronroos, Sharon Ngang, Pantelis Nicola, and Lamiah Rahman for their help in the TransTACT project. I would also like to thank our external collaborators, Agena Bioscience (Michael Mosko and Anders Nygren, San Diego, CA, USA), Illumina (Claire Fielding and Mark Ross, Cambridge, UK), and Natera (Robert Pelham, San Carlos, CA, USA), for their processing of cfDNA samples using their existing techniques. I am grateful to Cancer Research UK for funding my clinical research fellowship during the last three years.

I would like to thank my family for their continued love and support. To my husband, Adam, thank you for being by my side through it all, and for your patience and endless optimism, without which I would not have made it this far.

Table of Contents

Table of figures

List of tables

Abbreviations

ARMS - amplification refractory mutation system

BEAMing - beads, emulsion, amplification, and magnetics

CDX - CTC-derived explants

CEP2 - centromeric probe chromosome 2

CEP15 - centromeric probe chromosome 15

CIN - chromosomal instability

cfDNA - circulating free DNA

CTC - circulating tumour cell

ER - oestrogen receptor

FISH - fluorescence *in situ* hybridisation

HER2 - human epidermal growth factor receptor

LOH - loss of heterozygosity

MASA - mutant allele specific amplification

MCD – modal centromere deviation

ME - mutant enriched

NGS - next-generation sequencing

NICE - National Institute for Health and Care Excellence

NSCLC - non-small cell lung cancer

OS - overall survival

PCR - polymerase chain reaction

PEACE - Posthumous tissuE donAtion in CancEr

PFS - progression-free survival

PR - progesterone receptor

qPCR - quantitative PCR

RFLP - restriction fragment length polymorphism

SSCP - single-strand conformation polymorphism

SNP - single nucleotide polymorphism

TACT - Taxotere as Adjuvant Chemotherapy Trial

TAM-Seq - tagged-amplicon deep sequencing

TMA - tissue microarray

TRACERx - TRAcking Cancer Evolution through therapy (Rx)

VAF - variant allele frequency

WES - whole-exome sequencing

WGS - whole-genome sequencing

1 Introduction

The existence of distinct subpopulations of cancer cells within a tumour harbouring different behavioural phenotypes, including tumourigenicity, ability to metastasise and to develop resistance to treatment, has been recognised for many years (Heppner & Miller 1983). Recent advances in massively parallel sequencing technologies have enabled the analysis of the clonal architecture of both primary and metastatic tumours (Russnes et al. 2011). These genomic studies have demonstrated the complex and heterogeneous landscape in cancer (Lawrence et al. 2014; Lawrence et al. 2013), and its potential implications for treatment response and prognosis (Lawrence et al. 2013; Kandoth et al. 2014; Alexandrov, Nik-Zainal, Wedge, Aparicio, et al. 2013). Using whole-genome sequencing (WGS) or whole-exome sequencing (WES), these studies have demonstrated that tumours consist of somatic events, defined by mutations and copy number alterations, which can occur early in tumour evolution (events present in all cancer cells), or can occur later in tumour evolution (events present in some, but not all cancer cells) (Swanton 2014). In addition, these studies have shown spatial heterogeneity, branched evolution and mutational patterns that can vary over time (Nik-Zainal et al. 2012; de Bruin et al. 2014), as well as in response to cancer therapies (Walter et al. 2012; Boeckx et al. 2014; Diaz et al. 2012). Chromosomal instability (CIN), which describes an increased rate of change in chromosome number and structure, is also a potential mechanism generating intratumour heterogeneity (Gisselsson et al. 2000; McGranahan et al. 2012). CIN can create significant cell-to-cell heterogeneity, such that chromosomally unstable tumours may promote tumour evolution under selection pressures, such as chemotherapy, resulting in drug resistance and poor prognosis (Duesberg et al. 2000; Lee et al. 2011; McGranahan et al. 2012; McClelland et al. 2009; Swanton et al. 2009). Given the existence of such heterogeneity in tumours, the efficacy of therapies targeting somatic driver events, even in combination, may be limited in terms of achieving disease cure.

1.1 Intratumour heterogeneity in cancer

Intratumour heterogeneity, describes the existence of distinct cellular populations with specific phenotypic features within tumours and has long been recognised (Bloom & Richardson 1957; Heppner & Miller 1983). This has been described in several adult solid tumours, including lung (de Bruin et al. 2014; Zhang et al. 2014), breast (Nik-Zainal et al. 2012; Shah et al. 2012; Navin et al. 2012; Yates et al. 2015), ovarian

(Bashashati et al. 2013; R. F. Schwarz et al. 2015; Tone et al. 2014), pancreatic (Campbell et al. 2010; Yachida et al. 2010), kidney (Gerlinger et al. 2012; Gerlinger et al. 2014), colorectal (Thirlwell et al. 2010; Sottoriva et al. 2015), brain (Snuderl et al. 2011; Sottoriva et al. 2013; Suzuki et al. 2015; A. Kumar et al. 2014), thyroid (Le Pennec et al. 2015), oropharyngeal (Zhang et al. 2013), phaeochromocytoma (Flynn et al. 2015), and prostate cancers (Carreira et al. 2014; Haffner et al. 2013; Cooper et al. 2015), as well as hematological malignancies, such as chronic lymphoblastic leukemia (CLL) (Landau et al. 2013; Ojha et al. 2015), and acute lymphoblastic leukemia (Anderson et al. 2011). Other types of genetic heterogeneity have also been described in cancer biology. The most well known is inter-patient tumour heterogeneity, whereby no two patients with the same subtype of tumour behave the same clinically, with or without treatment. This may be related to host factors, such as tumour microenvironment, germ line mutations influencing treatment response, and the unique somatic mutations that can occur within the tumour of each individual patient (Vogelstein et al. 2013). Metastatic lesions at different secondary sites can arise from different cellular populations within a primary tumour, resulting in heterogeneity among metastases, known as intermetastatic heterogeneity (Vogelstein et al. 2013). In addition, since metastatic lesions can acquire new mutations and evolve independently with each cell division, heterogeneity within a metastasis can also exist, known as intrametastatic heterogeneity (Vogelstein et al. 2013) (**Figure 1** A-D).

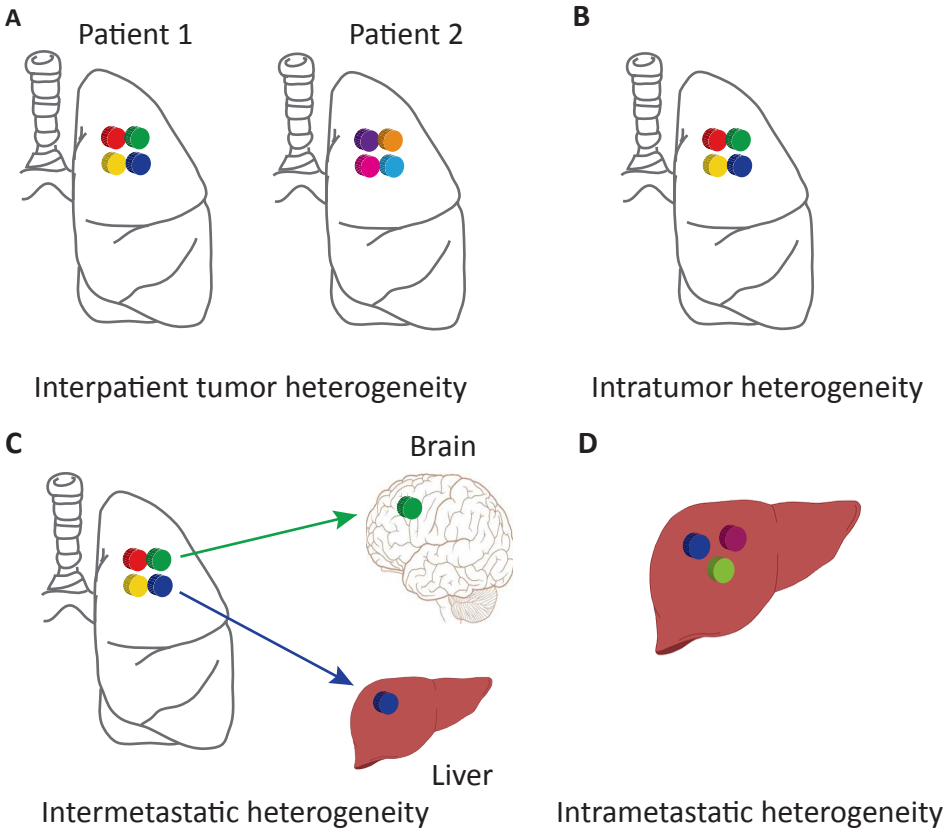


Figure 1 Types of intratumour heterogeneity

Reproduced from (Jamal-Hanjani, Quezada, et al. 2015). Inter-patient heterogeneity: the presence of unique subclones in the tumour of each patient (A), intratumour heterogeneity: the presence of multiple subclones within a primary tumour resulting in heterogeneity among tumour cells (B), intermetastatic heterogeneity: the presence of different subclones in different metastatic lesions of the same patient; some subclones may have been derived from the primary tumour and some may have emerged as a result of acquired alterations within each metastatic lesion (C), and intrametastatic heterogeneity: the presence of multiple subclones within a single metastatic lesion (D).

1.2 Intratumour heterogeneity and clonal evolution

Phenotypic heterogeneity observed in tumours results from both genetic and non-genetic causes. Spontaneous tumours can arise through Darwinian-like clonal evolution involving the acquisition of 'driver' events, such as genetic mutations or copy number aberrations, which are believed to affect cancer cell proliferation or survival, along with 'passenger' events, which are assumed to be phenotypically silent and without a selective fitness advantage (Abramyuk et al. 2010). Non-genetic causes of heterogeneity include epigenetic changes (Dick 2008), differentiation hierarchies as a result of cancer stem cells (Shipitsin et al. 2007), stochastic biochemical processes within individual cells and heterogeneous tumour microenvironments (Marusyk et al. 2012). Processes of genetic diversification promote tumour progression through clonal evolution so that tumours appear to be composed of evolved cell populations (McGranahan & Swanton 2015). The linear model of somatic tumour evolution is that of clonal succession, where a series of clonal expansions are triggered by the acquisition of driver events conferring fitness gain, outcompeting and outgrowing other clones (Nowell 1976). This model implies that tumours are homogenous for functionally significant mutations, and whilst some tumours are found to evolve through linear steps (Anderson et al. 2011), there is also increasing evidence for the existence of genetically distinct clonal subpopulations with substantial genetic divergence coexisting within different regions of the same primary tumour, between primary and secondary tumours, and within metastases (Navin et al. 2010).

1.3 Intratumour heterogeneity and cancer progression

Intratumour heterogeneity is not simply a tumour characteristic, it may also have the potential to forecast risk of tumour progression and therapeutic outcome (Swanton 2012). Using WES and copy number analysis to identify clonal and subclonal driver mutations, Landau and colleagues found that the presence of a subclonal driver was an independent risk factor for disease progression (Landau et al. 2013). Patients treated with cytotoxic chemotherapy were more likely to undergo clonal evolution, such that the extent of heterogeneity evolved during treatment. Using whole-genome genotyping single-nucleotide polymorphism (SNP) array analysis, Mengelbier and colleagues demonstrated intratumour diversity in a cohort of seven chemotherapy-treated childhood cancers, including nephroblastoma, neuroblastoma and rhabdoid tumours (Mengelbier et al. 2015). By analyzing metastatic samples, they found that this

diversity provided a substrate for further clonal evolution in the metastases. In a larger cohort of 44 patients with pre-operatively treated nephroblastoma, they found that intratumour diversity was more prevalent in high risk histological subtypes, and high stage disease. Furthermore, it was an independent predictor for poor cancer-specific survival (Mengelbier et al. 2015). Studies like these indicate the likely relevance of tumour heterogeneity in predicting adverse outcomes, and the evolution of tumour subclonal composition during treatment. It may not always be the case that the dominant clone dictates tumour growth and malignant potential. Using a mouse xenograft model, Marusyk and colleagues demonstrated that subclones, without a known fitness advantage, could drive tumour growth in a non-cell-autonomous manner by inducing tumour-promoting changes in the microenvironment (Marusyk et al. 2014). Furthermore, non-cell-autonomous subclones could be outcompeted by other subclones with greater proliferative potential resulting in tumour collapse. Observations where minority subclones influence progression of the tumour mass suggest challenges for predictive and prognostic biomarker discovery efforts, that have traditionally focused on identifying genomic alterations in the dominant clone. The mechanisms, by which subclonal alterations can impact tumour biology and phenotype, are yet to be determined and will require further functional genomic studies. Such studies will need to investigate the interactions, both synergistic and antagonistic, between subclones during tumour evolution. In addition, these studies should specifically assess how these relationships may promote or impede cancer progression, and contribute to therapeutic failure and drug resistance (Kleppe & Levine 2014). Heterogeneity in the tumour microenvironment may also influence the evolution and progression of tumours. Interactions between tumour and stromal cells, changes in the level of hypoxia or acidity, increased or decreased inflammatory cell infiltrate, and remodeling of the extracellular matrix may act as selection pressures and lead to increased phenotypic heterogeneity, potentially influencing treatment response, and therefore tumour evolution (Marusyk et al. 2012).

1.4 Intratumour heterogeneity and biomarker detection

Intratumour heterogeneity for both genetic and phenotypic biomarkers has been shown in several tumour types. This can complicate the detection of clinical biomarkers, such as those predictive of therapeutic response, due to sampling bias (Yap et al. 2012). Heterogeneous amplification of the *HER2* (human epidermal growth factor receptor 2)

gene in primary tumours has been shown in breast (Davila & Amazon 2010; Seol et al. 2012), mucinous ovarian (Anglesio et al. 2013), and oesophageal (Yoon et al. 2012) cancers. In primary gastric cancers, heterogeneous *HER2* amplification and HER2 expression has been shown within the same tumour, and between diagnostic biopsies and resected tumours (Yang et al. 2012). Glioblastomas (GBMs) have been shown to have mosaic patterns of *EGFR* (epidermal growth factor receptor) and *PDGFRA* (platelet-derived growth factor receptor, alpha polypeptide) amplification (Snuderl et al. 2011; Szerlip et al. 2012). In these tumours therapeutic response to single agent inhibitor therapy can be poor and may be associated with the heterogeneous amplification of such genes (Paulsson et al. 2011; Hegi et al. 2011; Stommel et al. 2007). In another GBM study, Sottoriva and colleagues demonstrated extensive intratumour heterogeneity at the copy number level, including heterogeneous copy number aberrations involving putative drivers of disease such as *PDGFRA, MDM4* (Mdm2-Like P53-Binding Protein) and AKT3 (v-akt murine thymoma viral oncogene homolog 3) amplification, and *PTEN* (phosphatase and tensin homolog) deletion (Sottoriva et al. 2013).

Significant biomarker discordance between primary and metastatic tumours in the same patient has been demonstrated in several tumour types, presumed to be the result of independent tumour growth associated with different phenotypes (Stoecklein & Klein 2010; Goswami et al. 2015). In breast cancer, discordance between single biopsies taken from primary and metastatic tumours has been shown for HER2 expression (Niikura et al. 2012), *HER2* amplification (Gancberg et al. 2002; Cottu et al. 2008), and oestrogen receptor (ER), progesterone receptor (PR) and HER2 status (Lindström et al. 2012). In colorectal cancer, mutational concordance in the genes *KRAS* (Kirsten rat sarcoma viral oncogene homolog)*, NRAS* (neuroblastoma RAS viral (v-ras) oncogene homolog)*, BRAF* (v-raf murine sarcoma viral oncogene homolog B)*, PIK3CA* (phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha), and *TP53* (tumor protein p53) has been demonstrated using single biopsies from primary and metastatic tumours (Vakiani et al. 2012). However, in patients with a history of more than one primary tumour who had received treatment, there was evidence for discordance in *TP53.* In a similar colorectal cancer study, a concordance rate of 78% was found for known recurrent mutations between primary and metastatic tumours (Vignot et al. 2015). In a study by Baldus and colleagues, by comparing two biopsies taken from the same primary tumour, a discordance rate of 8%, 5%, and 1% was

identified in activating mutations for *KRAS*, *BRAF*, and *PIK3CA*, respectively (Baldus et al. 2010). Mutational discordance was also identified between primary tumours and lymph node metastases in 31% (*KRAS*), 4% (*BRAF*), and 13% (*PIK3CA*) of cases. In non-small cell lung cancer (NSCLC), discordance for the *EGFR* mutation has also been demonstrated between primary and metastatic tumours (Gomez-Roca et al. 2009; Kalikaki et al. 2008). These studies demonstrate the sampling bias associated with single tumour biopsies, and the potential difficulties in identifying therapeutically relevant lesions (Yap et al. 2012). Furthermore, distinct subclonal populations appear to be unequally distributed over space and time, indicating that existing biomarkers are subject to change during disease progression (Greaves & Maley 2012). This may pose a challenge for therapeutic strategies if treatment allocation is based only on an archival primary tumour biopsy.

1.5 Intratumour heterogeneity, therapeutic resistance and alternative approaches

Heterogeneous tumours are composed of multiple subclones, some of which may be either intrinsically resistant to certain therapeutic interventions, or may acquire resistance under selection pressures, such as chemotherapy. Such selection processes allow these subclones to dominate a tumour mass and potentially drive disease progression (Misale et al. 2012; Engelman & Settleman 2008) (**Figure 2**). The selection of resistant subclonal populations, as a result of therapy leading to treatment resistance, has been shown in several solid tumour types including lung (Kosaka et al. 2006; Turke et al. 2010), colorectal (Diaz et al. 2012), gastrointestinal (Liegl et al. 2008) and brain (B. E. Johnson et al. 2014), as well as haematological cancers, such as chronic myeloid leukemia (Shah et al. 2002). In patients with *EGFR*-mutant NSCLC, the presence of the gatekeeper *T790M* resistance mutation pre-treatment has been associated with acquired resistance to the tyrosine kinase inhibitor (TKI) gefitinib as a result of the positive selection of cells harbouring the resistance mutation (Kosaka et al. 2006). The presence of subclones with *MET* (met proto-oncogene) amplification pre-treatment has also been associated with subsequent EGFR TKI resistance (Turke et al. 2010). In addition, a shorter progression-free survival has been demonstrated in patients with low frequency *T790M* mutant subclones before the onset of treatment (Su et al. 2012). In colorectal cancer, wild-type *KRAS* predicts sensitivity to anti-EGFR antibody therapies such as panitumumab. Diaz and colleagues showed that by

monitoring circulating tumour DNA in patients treated with panitumumab for initially *KRAS* wild-type tumours, the emergence of mutations in *KRAS* could be detected during the course of therapy resulting in acquired resistance (Diaz et al. 2012). They concluded that subclonal populations harbouring *KRAS* mutations existed pre-treatment, and that under the selection pressure of anti-EGFR therapy, resistant subclones can rapidly expand and repopulate the tumour. In chronic myeloid leukaemia and gastrointestinal tumours, resistance to imatinib due to mutations in the BCR-ABL (breakpoint cluster region-c-abl oncogene 1) fusion protein (Shah et al. 2002) and KIT (v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog) (Liegl et al. 2008), respectively, has also been demonstrated in the context of clonal evolution. It should be noted that not all cases of therapeutic resistance arise from genetic heterogeneity; non-genetic causes, such as stochastic epigenetic heterogeneity, may also play a role in the emergence of resistant clones under selection pressures (Jänne et al. 2009). These examples demonstrate that relapsed clones in metastatic tumours can often be traced back to pre-treatment low frequency subclones, suggesting that the extent of intratumour heterogeneity is a likely important determinant of therapeutic outcome. In addition, the presence of different subclones in the primary tumour lends itself to the development of multiple mechanisms of resistance in the same patient under selection pressures, such as targeted therapy in *EGFR*-mutant NSCLC (Sequist et al. 2011), anaplastic lymphoma receptor tyrosine kinase (ALK)-rearranged NSCLC (Katayama et al. 2012) or *BRAF*-mutant melanoma (Romano et al. 2013).

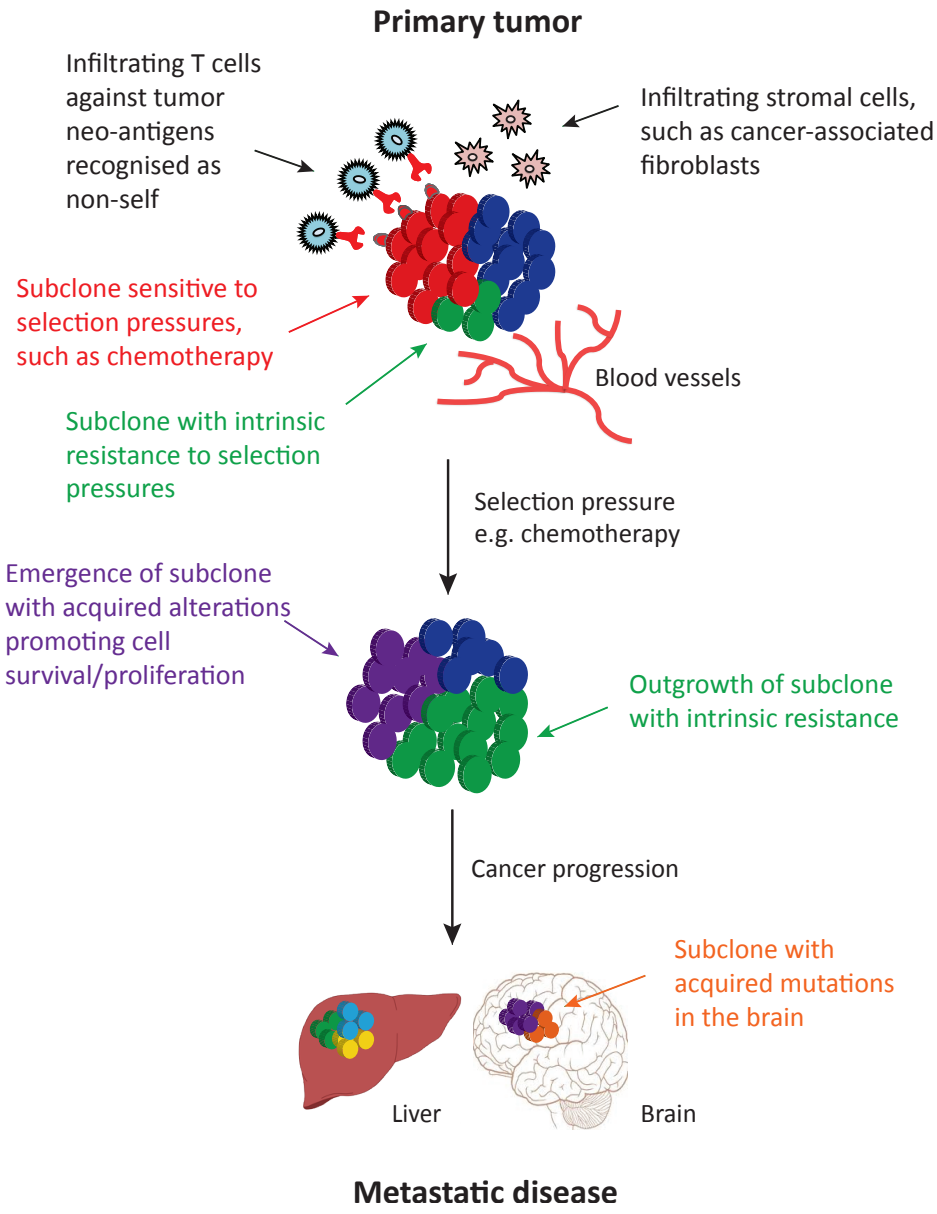


Figure 2 Intratumour heterogeneity and cancer progression

Adapted from (Jamal-Hanjani, Quezada, et al. 2015). Primary tumours consisting of different subclones may be subjected to various selection pressures, including chemotherapy and microenvironmental factors such as hypoxia, infiltrating stromal, and immune cells. Under the influence of such selection pressures, subclones with intrinsic resistance (green subclone) can outgrow a tumour mass, potentially leading to disease progression, and/or subclones can acquire somatic alterations (purple subclone), promoting cell survival, proliferation, and metastatic tumour formation. The outgrowth of some subclones (red subclone) may be constrained by selection pressures that they are sensitive to, for example, targeted therapy against a tumour subclone with a somatic alteration sensitive to therapy. Tumour infiltrating lymphocytes (TILs) may recognise neo-antigens presented on the surface of tumour cells as non-self, promoting enhanced T-cell activation and immune cell tumour infiltration.

Alternative therapeutic strategies may have the potential to overcome the challenges posed by clonal heterogeneity. Gillies and colleagues (Gillies et al. 2012) argue that subclonal populations that respond to initial therapy pass through an evolutionary bottleneck rendering them highly susceptible to a second therapy (Gerlinger & Swanton 2010), and that drug resistance in this instance, and the choice of a subsequent second therapy, could be anticipated. For example, combined therapy in *EGFR*-mutant NSCLC with an EGFR TKI and EGFR-specific antibody could prevent resistance associated with the expansion of a subclone harbouring a *T790M* mutation (Yelena et al. 2014). Approaches like this would require the development of biomarkers predicting likely resistance mechanisms in different patients, and such mechanisms could be targeted either in combination, or alternating, with standard treatment regimens (Komarova & Wodarz 2005). Treatment dosing schedules could also be adapted to prolong the suppression of resistant subpopulations, for example, drug holidays in androgen-dependent prostate cancer (Abrahamsson 2010) and melanoma (Thakur et al. 2013). Other adaptive approaches could involve combining standard treatment regimens with drugs targeting phenotypes known to contribute to tumour heterogeneity, such as altered tumour vasculature or altered glucose metabolism (Gillies et al. 2012).

Gatenby and colleagues have proposed the concept of adaptive therapy, whereby cancer treatment is continuously adjusted in order to maintain a fixed population of drug-sensitive cells, which can in turn suppress the growth of drug-resistant cells, allowing overall tumour burden to remain stable (Gatenby et al. 2009). Knowledge of the existence of low-frequency resistant subclones at diagnosis may indicate the use of combined therapeutic regimens targeting mechanisms of resistance. Evidence for phenotypic convergence within and across tumour types, suggests that genetic events driving resistance and disease progression may focus on either one or several signaling pathways that may be therapeutically targetable (Swanton 2014; Gerlinger et al. 2012; Gerlinger et al. 2014; Bolli et al. 2014; B. E. Johnson et al. 2014). In addition, parallel evolution of subclones during tumour progression suggests that there exist constraints to tumour growth, which may be potentially exploitable. By predicting the likely next step in the evolution of a tumour, preventative measures could be taken to potentially delay the onset of cancer progression (Swanton 2014). Such tactics will require a good evidence base from genomic studies in different tumour types involving longitudinal tissue sampling, and therapeutic trial designs will need to take into account

alternative approaches, such as adaptive therapy. The ultimate aim would be to develop predictive tools or evolutionary ‘rule books’ of tumour progression, so that tumour-specific treatment strategies can be initiated in advance of significant disease progression.

1.6 Evidence for intratumour heterogeneity in NSCLC

1.6.1 Spatial heterogeneity and branched evolution in NSCLC

Previous efforts to characterise the cancer genome in NSCLC have involved the analysis of copy number alterations (Weir et al. 2007; Tanaka et al. 2007), targeted sequencing of candidate cancer genes (Ding et al. 2008; Kan et al. 2010) and next-generation sequencing (NGS) of genomes and/or exomes (Liu et al. 2012; Lee et al. 2010; Govindan et al. 2012; Imielinski et al. 2012). By interrogating the mutational spectrum of tumours, these studies have demonstrated the complex and heterogeneous genomic landscape from point mutations to large structural variants, and the high mutational burden of smoking-related NSCLC. However, few studies in NSCLC have investigated the clonal and subclonal architecture of lung cancer tumours, and their evolution through disease progression. Furthermore, whilst it is known that different mutational processes, with specific mutation signatures, can contribute to the mutational burden in lung cancer, such as smoking (Pfeifer & Hainaut 2003; Lee et al. 2010; Pleasance et al. 2010) and the upregulation of APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) cytidine deaminases (Burns et al. 2013; Roberts et al. 2013; Alexandrov, Nik-Zainal, Wedge, Aparicio, et al. 2013), the temporal heterogeneity of such processes, and their potential contribution to the different stages of tumour evolution is unknown.

Recent work from our laboratory has demonstrated extensive intratumour heterogeneity in NSCLC, in terms of mutations and copy number alterations (de Bruin et al. 2014). By sequencing multiple regions sampled from resected primary NSCLC tumours, branched tumour evolution and intratumour heterogeneity in potential NSCLC driver mutations was demonstrated in a cohort of 7 NSCLC current or ex-smoker patients (5 adenoncarcinoma, 1 adeno-squamous cell carcinoma, and 1 squamous cell carcinoma). Mutational signatures were also investigated. Smoking-related mutations were found to decrease in late tumour evolution in all patients, accompanied by an increase in APOBEC-associated mutations in patients with lung adenocarcinoma. Intratumour heterogeneity in translocations, copy number alterations, and mutations associated with APOBEC cytidine deaminase activity was also demonstrated.

1.6.2 The lung TRACERx study

Each patient's cancer has a unique genomic landscape, comprised of populations of genetically distinct separated subclones, with the potential to undergo dynamic evolutionary processes throughout the disease course (Landau et al. 2013; Stratton et al. 2009). One of the major challenges in achieving the goal of precision medicine relies on obtaining an accurate view of this genomic landscape, in order to choose the appropriate therapeutic regimen (Yates & Campbell 2012). Intratumour heterogeneity poses a challenge in that a single tumour biopsy may not fully capture the current or future landscape of a tumour, and merely represents a 'snapshot' of the disease in space and time. Understanding how tumour clonal heterogeneity impacts upon clinical outcome, and how cancer subclones compete, adapt and evolve throughout the disease course and in relation to therapy, is an area of unmet clinical and scientific need. Lung TRACERx (TRAcking Cancer Evolution through therapy (Rx)) is a prospective multi-centre study in the United Kingdom aiming to recruit 842 patients with primary NSCLC, across stages I-IIIA over an accrual period of 4 years with a total 5-year follow-up per patient. Surgically resected primary NSCLC tumours and associated lymph nodes, surplus to diagnostic requirements, will be subjected to multi-region sampling and subsequent WES and/or WGS. In patients who develop recurrent disease, consent is obtained for a further biopsy, to assess how the tumour clonal structure has changed in response to therapy and after disease progression. The objectives of the TRACERx study are outlined below:

Primary objectives

- To determine the relationship between intratumour heterogeneity and clinical outcome, in terms of disease-free survival (DFS) and overall survival (OS)
- To determine the impact of adjuvant platinum-containing chemotherapy regimens on intratumour heterogeneity in relapsed disease

Secondary objectives

- To develop and validate an index of intratumour heterogeneity, which may be a predictive and/or prognostic biomarker in terms of DFS and OS
- To identify potential drivers of disease progression and drug resistance by tracking the tumour clonal dynamics throughout the disease course
- To define clonally dominant drivers of disease and to determine their role in targeted therapeutic response

- To determine the extent to which circulating biomarkers, such as circulating free DNA (cfDNA) and circulating tumour cells (CTCs), can represent the underlying mutational landscape of the primary and/or metastatic tumour
- To determine if cfDNA and CTCs can be used to track actionable mutations to guide therapeutic intervention, monitor residual disease and predict tumour recurrence
- To define the impact of intratumour heterogeneity on the tumour neo-antigen repertoire
- To determine the impact of distinct drivers of intratumour heterogeneity on tumour immune cell infiltration and function throughout the disease course
- To determine the correlation between measures of heterogeneity using clinical imaging and sequencing data.

I wrote and developed the protocol for this study under the guidance of my supervisor during the production of this thesis, with the input of the TRACERx consortium and the support of the UCL Cancer Trials Centre (Alan Hackshaw, Yentig Ngai, and Natasha Iles). This study started recruitment in April 2014, and is currently open in London, Manchester, Leicester, Birmingham, Aberdeen, and Cardiff.

1.7 Circulating biomarkers in cancer

Increasing disease burden and tumour size in patients with cancer is associated with increased cellular turnover and number of apoptotic and necrotic cells (Jahr et al. 2001; Stroun et al. 2001). Within a tumour mass, the process of phagocytic clearance of such cells is not as efficient as it is under normal physiological conditions, resulting in the accumulation of cellular debris, including tumour cells and tumour DNA, which can be released into the circulation. These circulating tumour cells and DNA may act as potential circulating biomarkers with predictive and prognostic value in cancer. Serial tumour sampling to track cancer progression, and to identify new targetable drivers of disease is practically challenging, and is currently not part of standard clinical practice. An alternative approach may be the use of circulating biomarkers, such as circulating cell-free tumour DNA (cfDNA) or circulating tumour cells (CTCs), which can be analysed in the peripheral blood of patients with cancer. Such sampling may offer a relatively simple and non-invasive method for the analysis of primary and metastatic tumours (Pantel et al. 2013), allowing the detection of therapeutic biomarkers to guide treatment stratification, the monitoring of residual disease, and detection of the emergence of molecular resistance (Diaz & Bardelli 2014).

1.7.1 Methods of detection in cfDNA

The detection of cfDNA derived from tumours has been challenging for several reasons, including the accurate quantification of mutant DNA fragments in a sample, distinguishing tumour cfDNA from normal circulating DNA, and identifying low levels of cfDNA in a sample. Standard sequencing technologies, such as Sanger sequencing (Jänne et al. 2006) or pyrosequencing (Ogino et al. 2005), rely on high levels of cfDNA, and are therefore able to accurately detect mutations in the tumour cfDNA of patients with significant disease burden. Advanced techniques, such as digital polymerase chain reaction (PCR) (Vogelstein & Kinzler 1999; Dawson et al. 2013; Yung et al. 2009), targeted NGS (Forshew et al. 2012) or WGS (Chan et al. 2013; Leary et al. 2012; Heitzer, Ulz, et al. 2013), pyrophosphorolysis-activated polymerisation (PAP) (Liu & Sommer 2000), beads, emulsion, amplification, and magnetics (BEAMing) (Dressman et al. 2003; Li et al. 2006; Diehl et al. 2008), mass spectrometry genotyping assay-mutant-enriched PCR (Brevet et al. 2011), and tagged-amplicon deep sequencing (TAm-Seq) (Forshew et al. 2012; Murtaza et al. 2014), have enabled the detection of mutations in cfDNA. In addition, the combination of PCR-based digital

approaches with NGS, has enabled the detection of rare mutations (Forshew et al. 2012; Kinde et al. 2011; Leary et al. 2010; Taly et al. 2013). Aside from single point mutations, amplifications, rearrangements, and aneuploidy have also been shown to be detected in cfDNA (Leary et al. 2012; Dawson et al. 2013; Murtaza et al. 2014; Schütz et al. 2015). Several studies have demonstrated the detection of genetic aberrations in the cfDNA of patients with cancer (**Table 1**).

Tumour type	Stage	Number of patients	Tumour-specific aberration	Technique	Reference
Colorectal cancer	Early to advanced	33	*APC*	BEAMing	(Diehl et al. 2005)
	Advanced	18	*APC, KRAS, PIK3CA, TP53*	BEAMing	(Diehl et al. 2008)
	Early to advanced	104	*APC, KRAS, TP53*	PCR-SSCP	(J.Y. Wang et al. 2004)
	Early to advanced	70	*KRAS*	ME-PCR	(Frattini et al. 2008)
Breast cancer	Early to advanced	72	*PIK3CA*	ARMS-Scorpion PCR	(Board et al. 2010)
	Early to advanced	34 (retrospective) and 51 (prospective)	PIK3CA	BEAMing	(Higgins et al. 2012)
	Advanced	30	*PIK3CA, TP53, structural variation*	TAm-Seq and digital PCR	(Dawson et al. 2013)
Ovarian cancer	Advanced	38	*TP53, PTEN, EGFR, BRAF, KRAS*	TAm-Seq Digital PCR	(Forshew et al. 2012)
	Early to advanced	63	*PIK3CA*	Fluorescent-PCR	(Kuhlmann et al. 2012)
Hepatocellular carcinoma	Early	4	SNV	WGS	(Chan et al. 2013)
Pancreatic cancer	Early to advanced	21	*KRAS*	MASA PCR	(Yamada et al. 1998)
	Early to advanced	44	*KRAS*	RFLP-PCR	(Castells et al. 1999)
Oral squamous-cell carcinoma	Early to advanced	64	Microsatellite loci	PCR	(Hamana et al. 2005)
	Early to advanced	20	Microsatellite loci	PCR	(Kakimoto et al. 2008)
NSCLC	Advanced	246	*KRAS*	ARMS-qPCR	(Nygaard et al. 2013)
Breast and osteosarcoma	Advanced	3	Genomic alterations	Nested-real time PCR	(McBride et al. 2010)
Colorectal and breast cancer	Advanced	10	Chromosomal alterations	WGS	(Leary et al. 2012)

Table 1 Detection of genetic aberrations in cfDNA

Adapted from (Crowley et al. 2013). Studies in which genetic aberrations have been detected in the cfDNA using different techniques. Abbreviations: ARMS, amplification refractory mutation system; BEAMing, beads, emulsion, amplification, magnetics; MASA, mutant allele specific amplification; ME-PCR, mutant enriched PCR; PCR-SSCP, single-strand conformation polymorphism PCR; qPCR, quantitative PCR; RFLP-PCR, restriction fragment length polymorphism PCR; SNV, single nucleotide variants; WGS, and whole-genome sequencing.

1.7.2 Insight into intratumour heterogeneity and clonal evolution using circulating biomarkers

Several studies have demonstrated that genetic aberrations present in tumours can be detected in cfDNA, and that these can be used to track tumour evolution, as well as the mutational burden, in patients over time (Murtaza et al. 2014; Diehl et al. 2008; Diehl et al. 2005; Dawson et al. 2013; Diaz et al. 2012; Carreira et al. 2014; Siravegna et al. 2015). Murtaza and colleagues performed serial sampling and exome-sequencing of plasma in six patients with advanced-stage breast, ovarian and lung cancer, during several lines of therapy (Murtaza et al. 2014). By quantifying the mutation allele frequencies in the plasma, they were able to correlate emerging therapeutic resistance with increased representation of mutant alleles. These included a *T790M* mutation following gefitinib therapy, a truncating mutation in *MED1* (mediator complex subunit 1) following tamoxifen and trastuzumab therapy, and a truncating mutation in *RB1* (retinoblastoma 1) following cisplatin chemotherapy. Bettegowda and colleagues detected cfDNA in a cohort of 640 patients covering a broad spectrum of both early- and late-stage cancers (Bettegowda et al. 2014; Diaz et al. 2012). Forty-seven percent of patients with stage I cancers of any type had detectable levels of cfDNA, supporting the potential use of cfDNA in the early detection of cancer. In a subset of patients with colorectal cancer, who subsequently relapsed after receiving anti-EGFR antibody therapy, they detected the emergence of several mutations in genes involved in the MAPK (mitogen-activated protein kinase) pathway, and therefore potential mechanisms of resistance. Diehl and colleagues used cfDNA to track changes in the plasma of colorectal cancer patients on therapy (Diehl et al. 2008). Fluctuations in cfDNA after surgery corresponded with the extent of surgical resection, and in patients with detectable post-operative tumour-derived cfDNA, the likelihood of relapse within 1 year was increased. Compared with the standard biomarker CEA (carcinoembryonic antigen), cfDNA was found to be more reliable as a predictor of tumor burden and outcome. Similarly, Dawson and colleagues showed that cfDNA was a more reliable indicator of tumour burden compared with the standard biomarker CA 15-3 (carcinoma

antigen 15-3) in patients with breast cancer (Dawson et al. 2013). Siravegna and colleagues used cfDNA to genotype tumours and track clonal evolution in a cohort of 100 colorectal cancer patients whilst on treatment with the anti-EGFR antibodies cetuximab and panitumumab (Siravegna et al. 2015). In patients with primary or acquired resistance to therapy, they showed an emergence and rise in *KRAS*-mutant clones. Upon treatment withdrawal, the number of *KRAS*-mutant clones declined, with an associated increase in drug sensitivity. This study eloquently demonstrated the dynamic nature of tumours under a treatment selection pressure, and how tumours adapt in response to intermittent drug dosing providing an explanation for the efficacy of targeted drug re-challenges in different tumour types (Hata, Katakami, Fujita, et al. 2013; Hata, Katakami, Kaji, et al. 2013; Seghers et al. 2012; Thakur et al. 2013), and a rationale for adaptive therapy strategies.

Concordance between the molecular landscape in the plasma and tumour has been previously shown. In a study by Rothé and colleagues, matched primary and metastatic tumours and cfDNA, were analysed using a targeted gene panel in a cohort of 17 breast cancer patients (Rothe et al. 2014). In 76% (13/17) cases, mutations found in cfDNA were concordant with those found in the tumour. In the study by Siravegna and colleagues discussed above, matched tumour and cfDNA samples were analysed for the mutational status of *KRAS*, *BRAF* and *NRAS* using droplet digital PCR (Siravegna et al. 2015). In 97/100 (97%) cases, the mutational status in the plasma was concordant with the tumour. Interestingly, in 8/100 (8%) cases there were mutations identified in the plasma that were not found in the tumour, suggesting that the plasma may be a more comprehensive representation of the molecular landscape of a heterogeneous tumour compared with a single tumour biopsy (Siravegna et al. 2015).

Similar to the studies in cfDNA, CTCs have also been used to study mutations and somatic copy number aberrations (SCNAs), and have been shown to correlate with clinical outcome (Antonarakis et al. 2014; Ni et al. 2013; Heitzer, Auer, et al. 2013; Hodgkinson et al. 2014; Lohr et al. 2014). In a large prospective study by Rack and colleagues, CTC enumeration in breast cancer patients was found to be an independent prognostic marker for disease-free survival (DFS) and overall survival (OS) (Rack et al. 2014). The worst prognosis was seen in patients with at least 5 CTCs per 30ml blood sample, and the presence of CTCs after chemotherapy was associated

with a poorer DFS and OS. One of the advantages of studying CTCs is that NGS can be used to study heterogeneity in driver events with single-cell resolution. This has been shown for EGFR expression and *KRAS/PIK3CA* mutation status in colorectal cancer patients (Gasch et al. 2013), and for ROS1 translocations in NSCLC patients (Pailler et al. 2015). Another advantage is the ability to study CTCs in the context of patient-derived mouse models. In a study by Hodgkinson and colleagues, CTCs were found to be tumourigenic in immune-compromised mice (Hodgkinson et al. 2014). CTCs derived from patients with small-cell lung cancer were implanted into mice resulting in CTC-derived explants (CDXs) (Hodgkinson et al. 2014). The genomic profiles of CTCs derived from patients and CDXs showed considerable similarity. Interestingly, CDXs reflected the donor patient's response to platinum and etoposide chemotherapy, suggesting that such mouse models could be used for therapy testing, and therefore potentially increase our understanding of potential resistance mechanisms.

In summary, the analysis of circulating biomarkers in blood samples obtained before and after therapy, and at points of disease progression, can provide a dynamic picture of the genomic landscape of each patient's tumour (Diaz & Bardelli 2014). Tracking this landscape at serial time points can improve our understanding of tumour evolution in response to therapy, and shed light on the mechanisms of acquired resistance at the molecular level.

1.8 Chromosomal instability in cancer

Chromosomal instability (CIN) is characterised by an increased rate of gain or loss of whole or fractions of chromosomes, and is a known driver of intercellular and intratumour heterogeneity (Gisselsson et al. 2000; McGranahan et al. 2012). It represents genomic instability at the karyotypic level, and the aberrant chromosomal characteristics that result from CIN include aneuploidy (an abnormal chromosomal number) and gross chromosomal rearrangements, which are hallmarks of solid tumours. CIN can be classified into structural and/or numeral CIN (**Figure 3**). Structural CIN, such as duplications, deletions and gene amplifications, may be precipitated by telomeric dysfunction (Stewénius et al. 2005; Perera et al. 2008; Bailey & Murnane 2006) or DNA replication stress (Stewénius et al. 2005; Gisselsson et al. 2001; Burrell et al. 2013; Dereli-Öz et al. 2011). Numerical CIN, a gain or loss of whole chromosomes, may be precipitated by defects in the mitotic checkpoint (Grabsch et al. 2003; Wang et al. 2002), the attachment of chromosomes to the mitotic spindle (Gregan et al. 2011), centromere amplification (Ganem et al. 2009), aberrant sister chromatid cohesion (Barber et al. 2008; Meyer et al. 2009; Zhang et al. 2008) and cytokinetic failure (Gisselsson et al. 2010). Increasingly it is recognised that structural CIN can precipitate numerical CIN and vice versa (Janssen et al. 2011; Crasta et al. 2012). **Figure 4** demonstrates the mitotic mechanisms of CIN.

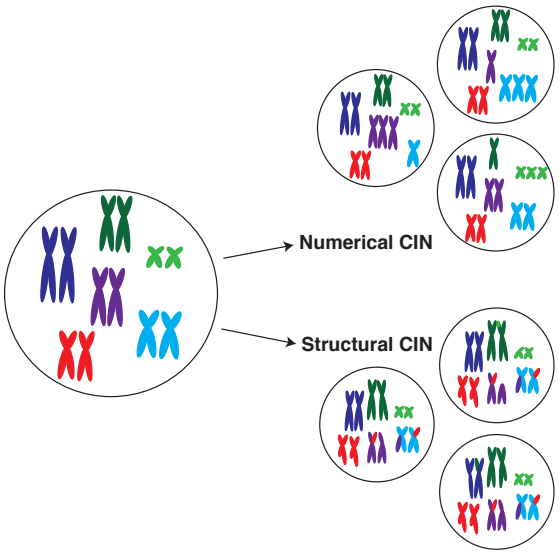


Figure 3 Numerical and structural CIN

Reproduced from (McGranahan et al. 2012). Schematic showing whole chromosome gains and losses (numerical CIN) and sub-chromosomal gains, losses, inversions and translocations (structural CIN).

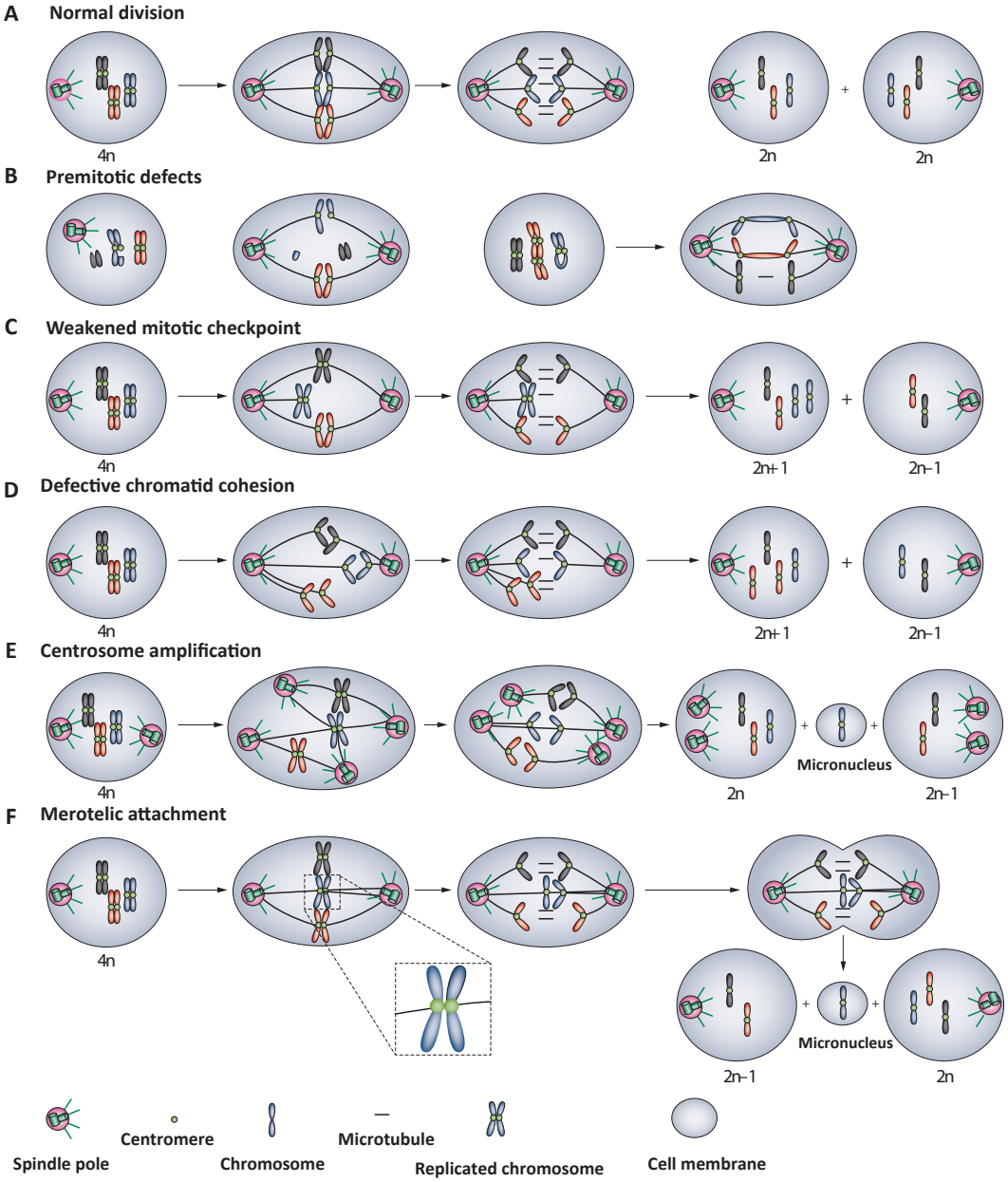


Figure 4 Mitotic mechanisms of CIN

Reproduced from (A'Hern et al. 2013). Normal cell division resulting in identical daughter cells (A), pre-mitotic defects resulting in acentric chromosomes and anaphase bridges (B), defects in mitotic checkpoint signaling (C), defects in sister chromatid cohesion (D), centrosome amplification resulting in a multipolar mitotic spindle (E), and abnormal attachment of chromosomes to the mitotic spindle resulting in merotelic attachments, which can involve cytokinesis-induced DNA damage to give rise to micronuclei (F).

1.8.1 Measuring CIN in tumour samples

Although aneuploidy is not always associated with the dynamic process of CIN, its measurement in tumour samples is the most reliable method in identifying likely

chromosomally unstable tumours. Determining the CIN status of a tumour has proven challenging given that karyotypic heterogeneity varies during the different stages of cancer progression (Roschke & Kirsch 2010), from one tumour region to another and, possibly, within different regions of the same tumour (Bergers et al. 1996; Gerlinger et al. 2012). Techniques employed to measure both the presence and the extent of CIN rely on determining cell-to-cell numeric chromosomal variability, such as fluorescence in situ hybridisation (FISH) using centromere-specific probes ((Roylance et al. 2011; Lingle et al. 2002; Farabegoli et al. 2001). Previous studies have demonstrated that using centromeric FISH for two chromosomes is sufficient enough to segregate diploid from aneuploid tumours (Fiegl et al. 2000; Takami et al. 2001), allowing the differentiation between unstable aneuploidy or CIN (aneuploid tumours with high clonal heterogeneity) and stable aneuploidy (aneuploid tumours with low clonal heterogeneity) (Lingle et al. 2002). Although these measurements are reasonable indicators of CIN, they are time consuming, labour intensive, and cannot measure the rate of change of ongoing chromosomal numerical and structural alterations (Geigl et al. 2008). A relationship between CIN and aberrations in the regulators of mitosis has also been shown before (Cahill et al. 1998; Wang et al. 2004; Sen et al. 1997; Diaz-Rodríguez et al. 2008; Yuan et al. 2006). Work within our laboratory has previously shown a significant association between direct centromeric CIN and the protein expression of the mitotic regulators AURKA (Aurora Kinase A) and SURVIVIN in breast cancer (Roylance et al. 2014). Such markers may also be of use as a proxy for CIN in breast cancer samples in the absence of more advanced molecular measurements. Alternative cell population-based methods involve the quantification of DNA content using DNA image and flow cytometry (Darzynkiewicz et al. 2010; Kronenwett et al. 2006; Habermann et al. 2009). Other techniques involve the assessment of combined populations of cells rather than individual cells, such as the measurement of CIN-associated gene expression signatures and copy number based scores (Habermann et al. 2009; Birkbak et al. 2011), which have been shown to correlate well with direct measures of CIN (Birkbak et al. 2011), and to be of prognostic value (Carter et al. 2006; Birkbak et al. 2011; Mettu et al. 2010; Chin et al. 2007; Smid et al. 2010).

1.8.2 CIN and response to taxane therapy

Microtubule-stabilising drugs, such as taxanes, initiate prolonged activation of the spindle assembly checkpoint, mitotic arrest and, in sensitive cells, promote cell death in mitosis or death preceded by multi-nucleation (Weaver & Cleveland 2005). Dominant

negative mutations or RNA interference of genes that regulate mitosis (such as BUBR1, BUB1, TTK/ MPS1, MAD2 and AURKB can augment the formation of aneuploid cells, compromise apoptotic death or promote mitotic exit in the presence of mitotic inhibitors such as paclitaxel (Cahill et al. 1998; Sudo et al. 2004; Anand et al. 2003; Nakayama et al. 2009; Inaba et al. 2005; Swanton et al. 2006; Swanton et al. 2007). These observations are consistent with taxane resistance observed *in vivo* in colorectal cancer, a tumour type in which CIN is highly prevalent, and in which taxanes have failed to demonstrate therapeutic benefit (Swanton et al. 2006). Work within our laboratory has previously shown in a cohort of 44 patients with ovarian cancer, that chromosomally unstable tumours (defined by their expression of the CIN70 signature, which reflects tumour aneuploidy status) (Carter et al. 2006) are relatively resistant to paclitaxel *in vivo* compared with tumours with lower expression of the CIN70 signature (Swanton et al. 2009). Furthermore, dysregulation of mitotic regulators, which results in aneuploidy and taxane resistance *in vitro* (Swanton et al. 2007) can predict failure to achieve a pathological complete response to paclitaxel combination chemotherapy in patients with breast cancer (Juul et al. 2010). These data indicate overlap between pathways involved in the maintenance of chromosomal stability and taxane sensitivity. Conversely, the molecular pathways that mediate tolerance of CIN might also result in relative taxane resistance (McClelland et al. 2009; Bouchet et al. 2007).

Several potential biomarkers have been reported to predict clinical response to taxane-based therapies, which are known to improve survival in women with both primary (Nowak et al. 2004) and metastatic (Ghersi et al. 2005) breast cancer. For example, the over expression or amplification of HER2 has been reported to predict sensitivity to taxane therapy (Konecny et al. 2004; Hayes et al. 2007), as has antigen Ki-67 expression in ER-positive breast cancers (Penault-Llorca et al. 2009). Paradiso and colleagues have shown that class III β-tubulin might be a predictive biomarker for taxane resistance (Paradiso et al. 2005). Other biomarkers, such as p53 and angiogenic-related markers, such as VEGF (vascular endothelial growth factor), have also been suggested, but the results have been inconsistent (Van Poznak et al. 2002; Khan & Wahl 1998; Lissoni et al. 2000; Grant et al. 2003). Predicting drug sensitivity can be challenging due to the polygenic nature of drug-resistance mechanisms, which although may be the result of multiple genetic and epigenetic aberrations, can still be perceived as a convergent phenotype (Burrell & Swanton 2014). For example, taxane resistance can be mediated *in vitro* by alterations in the expression of a number of

distinct genes, many of which have roles in transition through mitosis. A unifying mechanism of taxane sensitivity and resistance may help identify additional patient cohorts who will benefit from taxane-based therapies, and tumour CIN status might represent one such unifying predictor of taxane benefit.

1.8.3 CIN and clinical outcome in cancer

CIN has been shown to be associated with poor prognosis in several solid tumours, including lung (H. Nakamura et al. 2003; Carter et al. 2006; Choi et al. 2009; Mettu et al. 2010), breast (Kronenwett et al. 2004; Takami et al. 2001; Habermann et al. 2009; Smid et al. 2010), colorectal (Mettu et al. 2010; Walther et al. 2008), ovarian (Mettu et al. 2010), endometrial (Murayama-Hosokawa et al. 2010), endocrine pancreatic tumours (Jonkers et al. 2005), as well as diffuse large B-cell lymphoma (Bakhoum et al. 2011). This association may be a consequence of increased intratumour heterogeneity and cellular diversity allowing tumours to adapt to various microenvironmental selection pressures, such as chemotherapy (Gerlinger & Swanton 2010; Cahill et al. 1999), leading to drug resistance (Swanton et al. 2009; Duesberg et al. 2000; Li et al. 2005), and therefore disease progression (Takami et al. 2001; Jonkers et al. 2005). In keeping with this hypothesis, *in vitro* models have shown that CIN-positive cell lines develop multi-drug resistance at a higher rate compared with chromosomally stable cell lines (Duesberg et al. 2000), and *in vivo* models have suggested that CIN can promote early tumour relapse (Sotillo et al. 2010). Work within our laboratory has also shown that CIN-positive cell lines are intrinsically multi-drug resistant compared with chromosomally stable cell lines in colorectal cancer (Lee et al. 2011). In addition, chromosomally unstable tumours may also provide a substrate for the selection of copy number aberrations involving chromosomal segments containing genes associated with tumour proliferation, conferring a selective advantage, and therefore increased cellular fitness for tumour progression (Endesfelder et al. 2014). However, evidence is emerging that CIN may also have a negative impact on tumour fitness and growth. In yeast and mouse experimental studies, aneuploidy has been shown to result in decreased proliferation and cellular fitness (Torres et al. 2007; Williams et al. 2008). This may be related to the continuous gross numerical and structural chromosomal changes seen in CIN tumours resulting in the accumulation of deleterious genomic events affecting cancer cell survival (Loeb 2001). It therefore follows that a threshold of CIN may exist, such that up to a certain level of instability, tumour growth, adaptation and progression may be enhanced, but that beyond this level cancer cell survival may

be disadvantageous for tumour growth (Gerlinger & Swanton 2010; Komarova & Wodarz 2004). We have previously applied the CIN70 expression signature (Carter et al. 2006) to the gene expression profile of 265 patients with ER-negative breast cancer, and demonstrated that tumours with extreme CIN (upper quartile of CIN70 expression) were associated with improved prognosis relative to the other three quartiles (Birkbak et al. 2011). This paradoxical relationship between increasing CIN and improved prognosis was also seen in gastric, ovarian and non-small cell lung cancers (Birkbak et al. 2011). We further validated these findings in a small discovery cohort study of 246 patients with primary breast cancer using dual centromeric FISH as a surrogate measurement of CIN (Roylance et al. 2011). Whilst increasing CIN in ER-positive breast cancer was associated with worsening prognosis, in patients with ER-negative breast cancer, extreme CIN was paradoxically associated with improved prognosis.

1.9 Conclusion

This thesis explores the prevalence of intratumour heterogeneity in NSCLC using multi-region sequencing validation. It builds on previous data from our laboratory (de Bruin et al. 2014) using a much larger patient cohort representing both adenocarcinoma and squamous cell carcinoma histological subtypes, different tobacco smoke exposure histories, metastatic tumours and with clinical outcome follow-up data. Sequencing data were used to construct phylogenetic trees mapping the evolutionary history of each tumour. Spatial and temporal heterogeneity was seen in both mutations and copy number aberrations, such that branched evolution occurred in each tumour, with the presence of both early and late predicted drivers of disease. In a subset of NSCLC patients, cfDNA was analysed using three different approaches involving mass spectrometry, and a combination of multiplex PCR with NGS, to determine the extent to which cfDNA could represent the underlying genomic landscape of each tumour. A combination of multiplex PCR and NGS was the most promising method. Finally, having previously shown in our laboratory that extreme levels of CIN are associated with improved prognosis in ER-negative breast cancers (Roylance et al. 2011), this finding was further validated using direct centromeric FISH in a large prospective validation cohort of breast cancer patients (Jamal-Hanjani, A'Hern, et al. 2015).

2 Materials and Methods

2.1 Patient selection

2.1.1 NSCLC pilot cohort

Patients with NSCLC primary tumours eligible for curative surgical resection were identified by screening the surgical theatre lists for operations performed at the Heart Hospital, University College London Hospitals (UCLH) NHS Foundation Trust. These tumour samples were collected as part of the UCL/UCLH Biobank for Studying Health and Disease (UCLHRTB 10/H1306/42), and all study procedures were performed in accordance with Good Clinical Practice and national clinical research guidelines. This cohort consisted of eleven patients from whom follow-up data was also collected.

2.1.2 TRACERx cohort

The lung TRACERx study (Jamal-Hanjani et al. 2014) was approved in early 2014 by the Camden and Islington Research Ethics Committee (13/LO/1546, NIH study trial registration number NCT01888601; clintrials.gov), the Research and Development department at UCLH, and Cancer Research UK (C11496/A17786). This cohort consisted of 16 patients who were selected using the following eligibility criteria:

Inclusion criteria:

- Written Informed consent
- Patients ≥18-years of age, with early stage I-IIIA disease eligible for primary surgery
- Histopathologically confirmed NSCLC, or a strong suspicion of cancer on lung imaging necessitating surgery (e.g. diagnosis determined from frozen section in theatre)
- Primary surgery in keeping with NICE guidelines planned
- Agreement to be followed up in a specialist centre
- Performance status 0 or 1

Exclusion criteria:

- Any other current malignancy or malignancy diagnosed or relapsed within the past 5-years (other than non-melanomatous skin cancer and *in situ* cervical cancer)
- Psychological condition that would preclude informed consent

- Treatment with neo-adjuvant chemotherapy deemed necessary
- Adjuvant chemotherapy regimen other than platinum-based chemotherapy (if a patient is deemed suitable for adjuvant chemotherapy)
- Known human immunodeficiency virus, hepatitis B virus, hepatitis C virus or syphilis infection
- Sufficient tissue is unlikely to be obtained for the study based on pre-operative imaging.

2.1.3 TransTACT cohort and tissue microarray

The TACT trial (CRUK01/001) was a multi-centre open label phase III clinical study investigating the potential benefit of the addition of a taxane to adjuvant treatment in early breast cancer (Ellis et al. 2009). Patients in the study had either node-positive or high-risk node negative operable early breast cancer, and were randomly assigned to receive either FEC-T chemotherapy (fluorouracil 600 mg/m^2, epirubicin 60 mg/m^2, cyclophosphamide 600 mg/m^2 at 3-weekly intervals followed by docetaxel 100 mg/m^2 at 3-weekly intervals), or the control regimen, which consisted of either FEC for 8 cycles or epirubicin (100 mg/m^2 at 3-weekly intervals) for 4 cycles followed by CMF (cyclophosphamide 600 mg/m^2, methotrexare 40 mg/m^2, and fluorouracil 600 mg/m^2 at 4-weekly intervals) for 4 cycles. Accrual occurred between February 2001 and June 2003; 4162 patients were randomised into the TACT trial, 4124 patients were from centres within the UK and were approached for consent to collect tissue for research in a prospectively planned programme for translational biomarker evaluation within the TACT trial called the 'TransTACT' cohort. Archival paraffin-embedded breast cancer tissue blocks were used to create tissue microarrays (TMAs) containing cores, 0.6 mm in diameter and 4 mm in thickness, selected from representative tumour areas as determined by a consultant breast histopathologist from hematoxylin and eosin–stained sections.

2.2 Tumour sample processing

Tumour samples collected in theatre were marked with surgical clips to aid orientation of the anterior, superior and lateral aspects of the sample. Samples were transported on wet ice in a dry container without formalin fixative, to the UCLH pathology department.

2.2.1 Multi-region sampling of tumour tissue

All tumour samples were cut fresh in a class 2 cabinet. Each tumour sample was examined macroscopically by a pathologist to assess whether there was enough tissue surplus to diagnostic requirements, without compromising the bronchial or pleural resection margins. The sample was then dissected with new sterile scalpel blades on a clean surface to avoid DNA contamination. A cut (or more than one cut for tumours larger than 3-4cm) that fully bisected the tumour without compromising the integrity of the whole specimen was made. Multiple tumour regions, representing the macroscopic heterogeneity of the primary tumour, were excised from the newly exposed tumour surface. A fresh blade for every region sampled to avoid DNA carryover, and regions from the tumour, separated by 0.5-1cm, were collected as shown in **Figure 5**. Samples were then wrapped in foil and snap-frozen by immersion in liquid nitrogen for 30 seconds. Areas that were obviously necrotic, fibrosed or haemorrhagic were avoided to maximise viable tumour cellularity. Where possible, regions sampled reflected the macroscopic morphological heterogeneity of the tumour. The size of the tumours determined how many regions were collected, and on average the volume of each region was 10mm^3. Where possible, regions were also collected from lymph nodes with either obvious macroscopic or suspected tumour involvement.

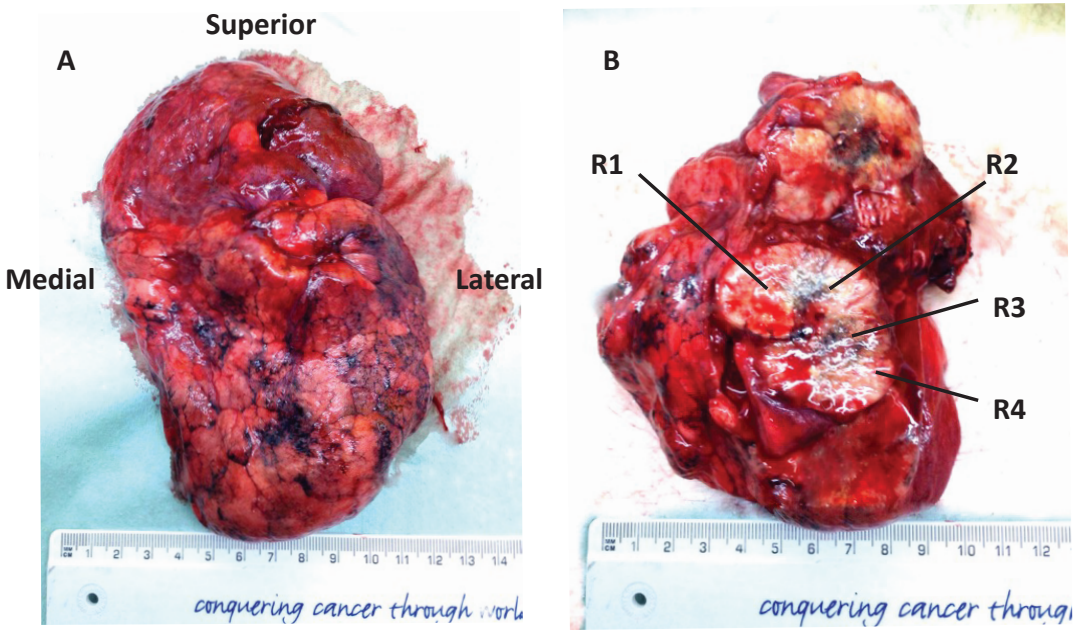


Figure 5 Primary lung tumour sample

Lung tumour after surgical resection in its entirety (A), and incised tumour with marked regions for multi-region sampling (B).

2.3 Plasma extraction from whole blood

Peripheral blood samples were collected from patients in standard phlebotomy tubes containing the anticoagulant ethylenediaminetetraacetic acid (EDTA). These samples were transported on wet ice between the surgical theatre and the laboratory, and processed within 1 hour of collection. The tubes were spun in a centrifuge at 1000g for 10 minutes at 4˚C, separating the plasma, buffy coat (leucocytes and platelets) and erythrocytes. The supernatant (plasma) was then removed using a disposable pipette, and further spun at 1000g for 10 minutes at 4^{o}C. The resulting supernatant was removed leaving behind any cellular debris, and stored at -80˚C in 1ml aliquots.

2.4 DNA extraction

2.4.1 Extraction from fresh frozen tissue

From each frozen tumour region stored at -80˚C, 30 mg was excised for DNA extraction. A 700µl mixture of β-mercaptoethanol (βME) and RLT Plus buffer (10µl βME per 1ml RLT Plus buffer) was added to each tissue sample prior to tissue homogenization using the TissueRuptor homogenizer (Qiagen). The resulting lysate was used to extract genomic DNA using the AllPrep DNA/RNA Mini kit (Qiagen), according to the manufacturer's instructions. DNA quality and concentration was assessed using a Qubit® 2.0.

2.4.2 Extraction from FFPE

Hematoxylin and eosin slides made from FFPE tumour tissue blocks were reviewed in order to demarcate regions of tumour compared to normal tissue, as shown in **Figure 6**. Using manual blade macrodissection from 10-40µm unstained slides tumour-rich areas of tissue were removed guided by the demarcated areas, added to 1.5ml of fresh xylene in centrifuge tubes and then vortexed to deparaffinise the tissue. This was centrifuged at full speed for 3 minutes, followed by the removal of the supernatant and the addition of 1.5ml of 100% ethanol. The mixture was centrifuged again and the process repeated with subsequent additions of 100%, 70% and 50% ethanol. The remaining tube with pellet was incubated at 40^{o}C using a heat block until no residual supernatant was present. Using the Qiagen DNeasy Blood and Tissue kit, 200µl of ATL

buffer and 20µl of proteinase were added to the tube with pellet, vortexed and centrifuged prior to incubation overnight at 56°C to aid tissue digestion. The following day, 200µl of AL buffer and 200µl of 100% ethanol were added to the pellet. The sample was vortexed and centrifuged briefly prior to transfer into a DNeasy spin column for centrifugation at 1000 rpm for 1 minute. This centrifugation was repeated with the addition of 650µl AW1 buffer to the column, and again with the addition of 650µl AW2 buffer, prior to DNA elution. DNA quality and concentration was assessed using a Qubit 2.0.

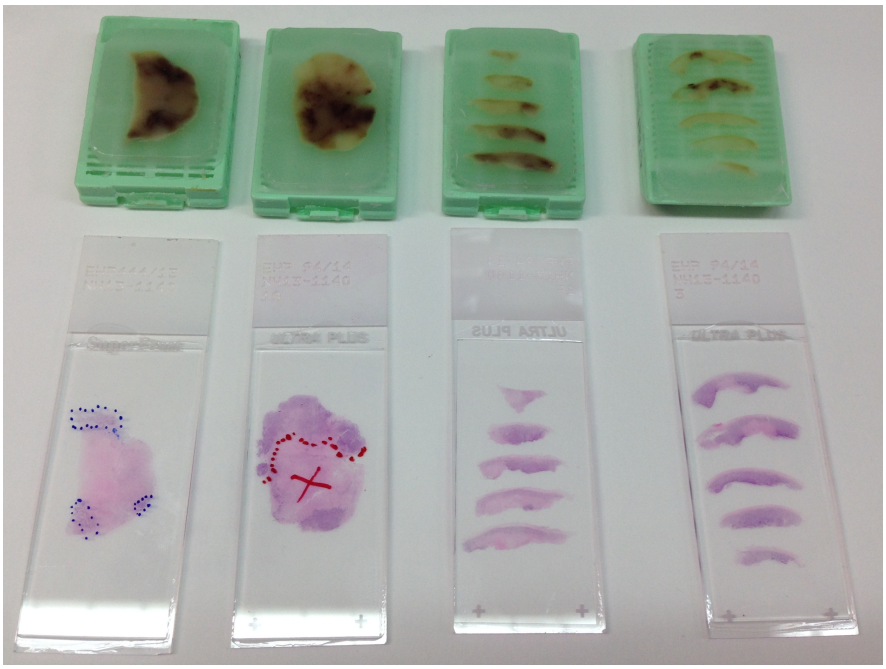


Figure 6 FFPE tumour blocks with corresponding H&E slides

H&E slides from FFPE tumour blocks with demarcated areas of tumour to guide manual blade macrodissection from unstained slides.

2.4.3 Extraction from whole blood

Whole blood stored at -80˚C was thawed at room temperature, and 400µl was added to a centrifuge tube containing 40µl proteinase K and 4µl of RNase A (100 mg/ml) stock solution. An average of 12µg of total DNA was extracted from 400µl of whole blood using the QIAamp DNA Blood Mini Kit (Qiagen), according to the manufacturer's instructions. DNA quality and concentration was assessed using a Qubit 2.0.

2.4.4 Extraction from plasma

Plasma samples were stored at -80˚C and thawed at room temperature prior to extraction. Each plasma sample (2ml) was added to a 50ml centrifuge tube containing

400µl of Proteinase K, and DNA was extracted using the QIAamp Circulating Nucleic Acid kit (Qiagen) and QIAvac 24 Plus vacuum manifold (Qiagen). To obtain higher concentrations, DNA was eluted in 20µl of elution buffer. DNA quality and concentration was assessed using a Qubit 2.0.

2.5 Multi-region whole-exome sequencing

Whole-exome sequencing (WES) was performed using DNA extracted from each tumour region and whole blood, where the latter was used for the germ line reference. For each tumour region and matched germ line, exome capture was performed on 1-2µg of genomic DNA using the Agilent SureSelect Human All Exon V4 kit or the Illumina® Nextera XT Index kit according to the manufacturer's protocol. Samples were paired-end multiplex sequenced on the Illumina® HiSeq platforms by Illumina, Inc. or the Illumina® Genome Analyzer II (GAII) and HiSeq 2500 platforms by the Advanced Sequencing Facility at the Francis Crick Institute. Genomic DNA was randomly fragmented by Covaris and run on a 2% agarose gel using a 100bp ladder at 110V for 30 minutes. A scalpol was used to excise fragments between 250 and 300 base pairs (bp) in length for purification using the Qiagen QIAquick® Gel Extraction kit. Adaptors were ligated to both ends of the fragments, adaptor-ligated templates were purified using Agencourt AMPure SPRI beads (Beckman Coulter), and fragments with an insert size of about 250bp were excised. Extracted DNA was amplified by ligation-mediated PCR, purified and hybridised to the SureSelect Biotinylated RNA Library (BAITS) for enrichment. Hybridised fragments were bound to streptavidin beads, whereas non-hybridized fragments were washed out after 24 hours. For all samples, each captured library was loaded on the Illumina platform, and paired-end sequencing was performed with 200bp read length at a mean sequencing depth of 347x. Section 8.3 shows the detailed coverage per region for each tumour.

2.5.1 Single nucleotide and indel variant calling from multi-region WES

Raw paired end reads in FastQ format generated by the Illumina pipeline were aligned to the full hg19 genomic assembly (including unknown contigs) obtained from GATK bundle 2.8 (McKenna et al. 2010), using Burrows-Wheeler Aligner (bwa-0.7.7 (Li & Durbin 2009) with a seed length of 72bp for data sequenced on the GAII and 100bp for data sequenced on the HiSeq. Up to 3 or 4 mismatches were allowed per read for the GAII or HiSeq respectively, all other settings were left as default. Picard tools v1.107

was used to clean, sort, and merge files from the same patient region and to remove duplicate reads (http://broadinstitute.github.io/picard). Quality control metrics were obtained using a combination of Picard tools (1.107), GATK (2.8.1) and FastQC (0.10.1) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). SAMtools mpileup (0.1.16) (Li et al. 2009) was used to locate non-reference positions in tumour and germ line samples. Bases with a phred score of <20 or reads with a mapping-quality <20 were skipped. BAQ computation was disabled and the coefficient for downgrading mapping quality was set to 50. Somatic variants between tumour and matched germ line were determined using VarScan2 somatic (v2.3.6) (Koboldt et al. 2012) utilising the output from SAMtools mpileup. Default parameters were used with the exception of minimum coverage for the germ line sample that was set to 10, minimum variant allele frequency (VAF) was changed to 0.01 and tumour purity was set to 0.5. VarScan2 processSomatic was used to extract the somatic variants.

The resulting single nucleotide variant (SNV) calls were filtered for false positives using the associated fpfilter.pl script from Varscan2, having first run the data through bam-readcount (0.5.1). Additionally, for those variants not subjected to Ion Torrent validation, further filtering was applied, whereby variants were only accepted if present in ≥ 5 reads, ≥ 5% VAF and a somatic p-value (as determined by VarScan) of ≤ 0.01 in at least one tumour region with germ line VAF ≤ 1%. If a variant was found to meet these criteria in a single region, then the VAF threshold was reduced to ≥ 1% in order to detect low frequency variants. All indel calls classed as ‘high confidence‘ by VarScan2 processSomatic underwent manual review prior to validation. All variants were annotated using ANNOVAR (Wang et al. 2010). Variants identified as non-silent mutations were manually reviewed using the Integrated Genomics Viewer (IGV) (Robinson et al. 2011), Thorvaldsdottir et al. 2013) and those showing an Illumina specific error profile (K. Nakamura et al. 2011) were removed from further analysis.

2.5.2 Manual review of variants

Exonic and/or splicing non-silent (non-synonymous, stop-gain or stop-loss) SNVs called using the above data platform were filtered prior to manual review using the following criteria:

- A VAF ≥ 5% in at least one tumour region

- A minimum variant count ≥ 5 reads in at least one tumour region
- A VAF ≤ 1% in the germ line

A variant was considered to be present in a tumour region provided the VAF was ≥ 1%. In addition, variants were further filtered using a somatic p-value of ≤ 0.01, which was derived based on a cohort of tumours with variants that had been manually reviewed and validated using high-depth sequencing. Non-silent (frameshift or non-frameshift) indels were all manually reviewed. All identified variants were manually reviewed using IGV (Robinson et al. 2011; Thorvaldsdottir et al. 2013) in order to determine which variants were likely to be real and therefore put forward for validation. Variants only present in poorly aligned reads and those showing an Illumina-specific error profile (K. Nakamura et al. 2011) were removed. Recurrent C>A mutations with a variant count of 4 and an average coverage of 60x, were found in the tumours prepared for WES using the Nextera XT Index kit. These mutations were subsequently deemed to be errors specific to the library preparation kit and were therefore excluded from further analyses.

2.6 Ampliseq™ Ion Torrent sequencing

2.6.1 Ion AmpliSeq custom validation

Ion AmpliSeq custom panels (Life Technologies) were designed by entering the genomic positions of all variants called in at least one region using the online designer (www.ampliseq.com). Multiplex PCRs were performed with tumour-specific primer pools on DNA from each region of the relevant tumour to ensure that primer products of an appropriate size were present prior to sequencing. PCR was carried out in a total volume of 15µl with 2µl (12ng) of DNA template supplemented with 13µl of a master mix consisting of 3µl of 5X Ion AmpliSeq HiFi Mix (Life Technologies), 7.5µl of 2X Ion AmpliSeq primer pool (Life Technologies), and 2.5µl of nuclease-free water. Reactions were initially incubated at 99˚C for 2 minutes followed 22 cycles of PCR at 99˚C for 15 seconds, 60˚C for 4 minutes. Presence of PCR products of the expected size were verified with 2µl of the PCR product (plus 8µl loading buffer) ran on a 2% agarose gel using a 100bp ladder at 110V for 30 minutes (**Figure 7**).

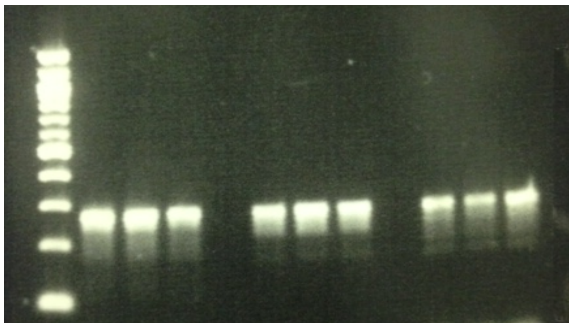


Figure 7 PCR amplification prior to validation

Agarose gel (2%) with a 100bp ladder (left hand side) and visible PCR products using primer pools for respective tumour and gem line DNA for 3 separate patients.

The Amplicon pools were used for the construction of barcoded sequencing libraries, and these were multiplex sequenced with 200bp read length on the Ion Torrent PGM™ sequencer (Life Technologies) at a mean sequencing depth of 1369x by the Advanced Sequencing Facility at the Francis Crick Institute. Section 8.3 shows the detailed coverage per region for each tumour. Sequence alignment to target regions from the hg19 genome was performed using the Ion Torrent Suite™ software. VAFs for each variant position having a phred score >20 were identified. All variants were manually reviewed again using IGV; a variant was considered absent if the VAF was < 1% for SNVs or 2% for indels, while having a read coverage ≥ 50x, or considered a germ line variant if the VAF was > 1% in the germ line. Variants that were absent in all tumour regions or identified as germ line variants were excluded from further analysis. Variants with read coverage <50x were considered inconclusive and regional distribution was extracted from exome or genome sequencing data. Some variants failed amplification, either because the AmpliSeq™ Designer could not generate suitable primers or because primer pairs failed in PCR, and these were also excluded from further analysis. Taking these failed variants into account, the overall validation rate was > 90% and these SNVs and indels were taken forward as validated mutations.

2.7 Intratumour heterogeneity index

Robust and accurate measures of intratumour heterogeneity are yet to be determined, but one potential measure developed in our laboratory is the pairwise intratumor heterogeneity index. This was calculated by determining the mean proportion of heterogeneous (branch) mutations relative to the total number of mutations for each possible pairwise comparison of all tumour regions. However, including ubiquitous

(truncal) mutations can introduce a bias against tumours that arise in the context of tobacco smoke exposure, since these tumours will generally have a greater number of somatic mutations generated by exposure to tobacco smoking (Pfeifer & Hainaut 2003), and therefore a greater number of truncal mutations. One alternative is to calculate the index in the absence of the truncal mutations, such that the relative lengths of the branches are determined indicating the degree of evolution that has taken place after clonal diversification (branching of the tree). This was calculated as the mean of the branch length for each pairing of tumour regions, divided by the sequencing depth after it had been adjusted for tumour purity. This measure was used to calculate an intratumour heterogeneity index for each tumour. A two-sided Wilcoxon rank-sum test was used to test for statistically significant correlations between the intratumour heterogeneity index and patient clinical variables between two groups of patients. Spearman's rank correlation coefficient was used to determine the strength and significance of the index with tumour stage and size.

2.8 Phylogenetic tree analysis

In the NSCLC pilot cohort, all non-silent mutations that were either validated or filtered were used for phylogenetic tree construction. In the TRACERx cohort, filtered synonymous and non-synonymous mutations were used for phylogenetic tree construction. Variants that were filtered met the following criteria:

- A VAF ≥ 5% in at least one tumour region
- A minimum variant count ≥ 5 reads in at least one tumour region
- A somatic p-value of ≤ 0.01
- A germ line VAF < 0.01
- Not detected in a pre-defined artefact blacklist region of the human genome

For each tumour, mutations that were present in all regions (ubiquitous) were considered to occur early in tumour evolution, and therefore present on the trunk of the phylogenetic tree. In contrast, mutations that were present in either one or some regions only (heterogeneous) were considered to occur later in tumour evolution, and therefore present on the branches of the phylogenetic tree (**Figure 8**). Trees were built using binary presence/absence matrices obtained from the regional distribution of variants within the tumour. The R Bioconductor package phangorn (1.99-7) (Schliep

2011) was used to generate unrooted trees using the parsimony ratchet method (Nixon 1999). Branch lengths were determined using the acctran function. Phylogenetic trees were redrawn in Adobe Illustrator with trunk and branch lengths proportional to the average number of mutations. Angles between branches were chosen only for convenience of display.

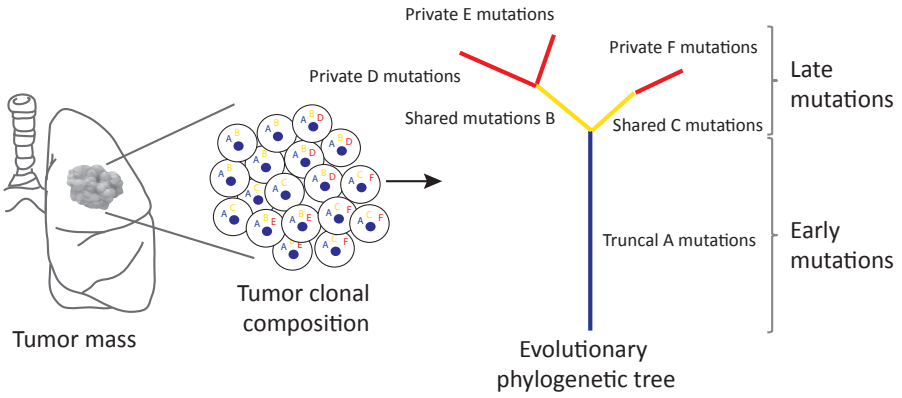


Figure 8 Phylogenetic tree construction

Reproduced from (Jamal-Hanjani, Quezada, et al. 2015). Primary tumours consist of different subclones with shared and private somatic alterations. Alterations shared by all tumour cells (A) occur early in tumourigenesis, represented by the blue trunk of the phylogenetic tree; alterations shared by tumour cells present in some regions of the tumour but not all (B and C) occur later in tumourigenesis, represented by the yellow branches of the tree; and private alterations (D–F) present in only one region of the tumour also occur later in tumourigenesis, represented by the red branches of the tree.

2.9 Identification and classification of predicted driver mutations

All identified non-silent variants were compared against a list of 582 potential driver genes. The driver gene list was consisted of all genes identified in the Catalogue of Somatic Mutations in Cancer (COSMIC) cancer gene census (downloaded June 2014) (Futreal et al. 2004), and those identified in a large scale pan-cancer analysis (using $q < 0.05$ as a cut-off for statistical significance) (Lawrence et al. 2013) and previous NSCLC sequencing studies (Lawrence et al. 2014; Cancer Genome Atlas Research Network, 2015). Any variants that were located within one of these genes underwent categorisation based on pre-set criteria. The COSMIC database and review of the existing literature was used to determine whether genes were considered to be oncogenes (OGs) or tumour suppressor genes (TSGs). Functional prediction scores associated with non-silent mutations were derived using the software packages SIFT

(P. Kumar et al. 2009), Polyphen-2 (Adzhubei et al. 2010) and MutationTaster2 (J. M. Schwarz et al. 2014)). If a variant in a TSG was deemed to be deleterious (either a stop-gain or predicted deleterious in two of the three computational approaches, the specific variant was classified as a category 1 (high confidence) driver mutation. If it failed to reach these criteria, the proximity to mutations annotated in COSMIC (data obtained February 2015) was determined. If ≥ 3 COSMIC mutations were located within 15bp, the variant was classified as a category 2 (putative) driver mutation. If not, it was classified as a category 3 (low confidence) driver mutation. If a variant was found in an OG, COSMIC was reviewed for the exact match to the specific variant. If an exact match was found ≥3 times, the variant was classified as a category 1 driver mutation. If an exact match was not found, the same criteria as described for the TSGs above was applied to classify the variant as a category 2 or 3. Finally, if the driver gene had not been classified as either an OG or TSG, all tests described above were applied and if it passed, the variant was classified as a category 2, otherwise a category 3 driver mutation. All the remaining variants were classified as a category 4 or 5 driver mutation (determined by the maximum variant count and maximum VAF being greater than 5 and 5%, respectively). These represented variants of unknown significance. In summary, the 5 driver categories were:

- Category 1 - high confidence driver mutations containing all inactivating mutations in TSGs or activating mutations in OGs
- Category 2 - putative driver mutations containing all mutations in driver genes located up to 15bp away from other mutations present in COSMIC or mutations meeting category 1 criteria but not annotated as a TSG or OG
- Category 3 - low confidence driver mutations containing all other non-silent mutations in genes that were present in the lists of cancer-related genes
- Category 4 and 5 – mutations of unknown significance

2.10 Mutational spectra plots

In the NSCLC pilot cohort, all non-synonymous mutations that were either validated or filtered, and all synonymous mutations that were filtered were used to create mutational spectra plots for each tumour. In the TRACERx cohort, filtered synonymous and non-synonymous mutations were used to create mutational spectra plots for each tumour. Variants with a VAF of ≥ 1% in all tumour regions were classified as truncal

mutations, and otherwise they were classified as branch mutations. Variants that were filtered met the following criteria:

- A VAF ≥ 5% in at least one tumour region
- A minimum variant count ≥ 5 reads in at least one tumour region
- A somatic p-value of ≤ 0.01
- A germ line VAF < 0.01
- Not detected in a pre-defined artefact blacklist region of the human genome

Chi-square tests were used to compare the mutation spectra of the six mutation types C>A, C>G, C>T, T>A, T>C, T>G (all substitutions are referred to by the pyrimidine of the mutated Watson-Crick base pair). A two-sided Fisher's exact test was used to compare the relative frequency of each mutation type between early and late variants.

2.10.1 Detecting a smoking signature pattern

In lung cancer it is well known that a substantial number of somatic mutations are generated by exposure to tobacco smoking (Pfeifer & Hainaut 2003). A significant association between a signature rich in C>A transversions and either a history of smoking or cancers associated with tobacco smoking (lung adenocarcinoma, squamous and small cell carcinomas, head and neck squamous, and liver cancers) has been previously demonstrated (Alexandrov, Nik-Zainal, Wedge, Aparicio, et al. 2013; Alexandrov, Nik-Zainal, Wedge, Campbell, et al. 2013). This pattern of mutation was investigated in all tumours.

2.10.2 Detecting an APOBEC signature pattern

To detect an APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) mutation pattern, previously established methods were used (Roberts et al. 2013; de Bruin et al. 2014). In brief, the enrichment E_{TCW} relating to the strength of mutagenesis at the TCW motif across the genome was calculated as follows:

$$E_{TCT} = \frac{mutations_{TCW} \times context_{CorG}}{mutations_{CorG} \times context_{TCW}}$$

where $mutations_{TCW}$ is the number of mutated cytosines (and guanines) falling in a TCW (or WGA) motif, $mutations_{CorG}$ is the total number of mutated cytosines or

guanines, $context_{TCW}$ is the total number of TCW (or WGA) motifs within a 41-nucleotides region centered on the mutated cytosines (and guanines) and $context_{CorG}$ is the total number of cytosines or guanines within the 41-nucleotides region centered on the mutated cytosines (or guanines). Only specific base substitutions were included (TCW to TTW or TGW, WGA to WAA or WCA, C to T or G, and G to A or C). Over-representation of APOBEC signature mutations in each sample was determined using a two-sided Fisher's exact test comparing the ratio of the number of cytosine-to-thymine or cytosine-to-guanine substitutions and guanine-to-adenine or guanine-to-cytosine substitutions that occurred in and out of the APOBEC target motif (TCW or WGA) to an analogous ratio for all cytosines and guanines that reside inside and outside of the TCW or WGA motif within 41-nucleotide region centered on the mutation cytosine (and guanine). P-values were corrected using Benjamin-Hochberg multiple testing correction, and a significance threshold of $q < 0.05$. For each sample, APOBEC mutation enrichment was determined for all mutations, early (truncal) mutations and late (branch) mutations separately. Comparisons between early and late APOBEC mutation enrichment was performed using a two-sided paired t-test.

2.11 Copy number analysis

All data analysis was performed using the R statistical package, version 3.0.2. Processed sample exome SNP and copy number data from paired tumour-normal was generated using VarScan2 (v2.3.6). Varscan2 copy number was run using default parameters with the exception of min-coverage (Robinson et al. 2011) and data-ratio. The data-ratio was calculated on a per sample basis as described in (Koboldt et al. 2012). Output from Varscan were processed using the Sequenza R package 2.1.1 (Favero et al. 2015) to provide segmented copy number data, and cellularity and ploidy estimates for all samples based on the WES data. The following settings were used: breaks.method = 'full', gamma = 40, kmin = 5, gamma.pcf = 200, kmin.pcf = 200. Manual verification was performed of the automatically selected models for ploidy and cellularity. Processed copy number data for each sample was divided by the sample mean ploidy, and log2 transformed. Gain and loss was defined as log2(2.5/2) and log2(1.5), respectively. Amplification was defined as log2(4/2). For calling copy number aberrations, segments smaller than 500kb or containing less than 5 SNPs were removed. A higher tumour purity was required for the copy number analysis pipelines compared with the variant calling, and therefore the following tumour regions were

excluded for the analysis: tumours LTX016 (R4, R5), LTX022 (R2, R3, R4, R6), LTX028 (R4, R8), LTX031 (R1, R2), LTX036 (R1, R2, R5, R7 and R8), and LTX038 (R7). In addition, tumour LTX030 was excluded from the copy number loss analysis since data from only one region was available, but included in the copy number gain and/or amplification analysis where two regions were analysed. Heat maps were created showing the distribution of potential driver copy number amplifications and deletions for each tumour region, based on recurrent amplified and deleted chromosomal segments identified in The Cancer Genome Atlas (TCGA) LUAD and LUSC cancer data. For each region, copy number amplification was determined as ≥ 2x ploidy, copy number gain was determined as ≥ 1 copy gain but less than the amplification threshold, and copy number loss was determined as ≤ 1 copy number relative to ploidy.

2.12 cfDNA analyses

2.12.1 Selection of mutations

Non-silent mutations identified by WES and subsequently validated by Ion AmpliSeq sequencing were considered for detection in cfDNA. The majority of the mutations investigated were SNVs, but in some cases indels were tested and in one case an EML4-ALK translocation was tested (collaboration with Illumina). All approaches involved multiplex PCR and PCR primers were designed using the specified genomic coordinates for the selected mutations. In order to address the question as to whether the heterogeneous genetic landscape of a tumour could be explored using cfDNA, both truncal and branch mutations were selected. cfDNA analyses involved collaborations with companies outside of our laboratory (as described below), whereby cfDNA extracted at diagnosis from 2ml of plasma for each patient was sent to these companies along with specified mutations for detection. The results of these experiments were analysed in our laboratory. Significant associations between VAFs for selected mutations in these analyses were tested for using the Mann-Whitney U test.

2.12.2 Multiplex PCR and MALDI-TOF mass spectrometry

In collaboration with Agena Bioscience (Michael Mosko and Anders Nygren, San Diego, CA, USA), the presence of mutations in cfDNA was analysed using a novel minor

variant detection technique called UltraSEEK, which is able to process multiple informative variants within a single reaction, and involves standard multiplex PCR followed by a mutation specific single base extension reaction and characterisation using Matrix Assisted Laser Desorption/Ionization-Time of Flight (MALDI-TOF) mass spectrometry (**Figure 9**).

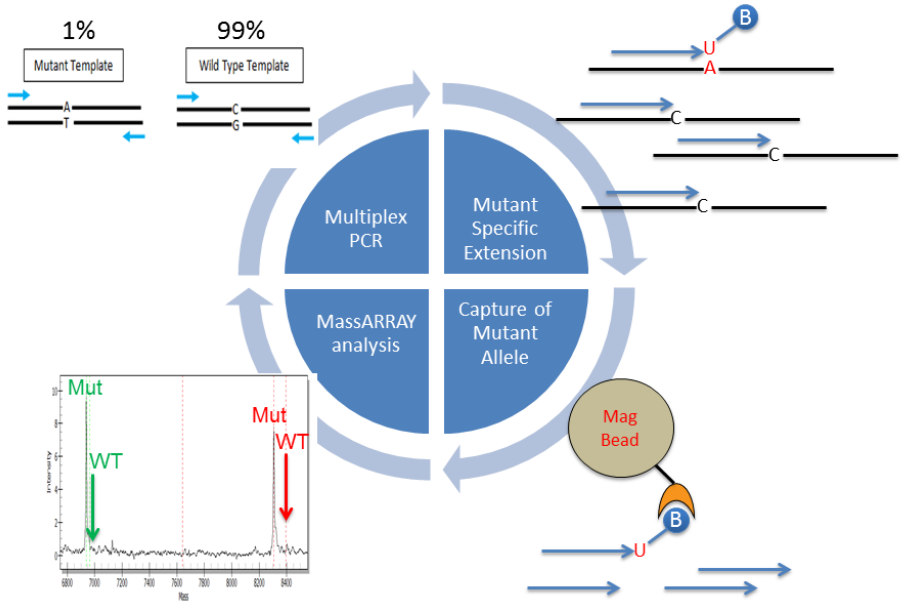


Figure 9 The UltraSEEK technique

The UltraSEEK technique consists of multiplex PCR followed by a mutation specific single base extension reaction. The extension reaction utilizes a single mutation specific chain terminator labelled with a moiety for solid phase capture. Captured, washed, and eluted products are interrogated for mass and mutational genotypes are identified and characterised using MALDI-TOF mass spectrometry.

2.12.2.1 Assay design

Each UltraSEEK assay consisted of three primers, two PCR primers and one single base extension primer. PCR primers were manually designed using the specified genomic coordinates of the selected mutations. **Table 2** summarises the chosen variants and primer designs. All amplicons were under 150bp in length to ensure cfDNA amplification success and mass tags were added to the 5' end of the primer to move unincorporated PCR primers out of the analytical mass window. Different extension probe lengths between 15 to 30 nucleotides long were used to ensure there were no conflicts between assays. Probe polarity was designed to ensure that the maximum number of assays shared the same variant nucleotide of extension. All oligonucleotides were obtained from Integrated DNA Technologies. The customised

UltraSEEK panel contained multiplex reactions consisting of assays for parallel detection of somatic mutations in selected genes. Control assays were also included to ensure functional biochemistry, a sufficient quality template, and the ability to distinguish between a wild type result and a failed assay. In addition, five capture controls were also present in each multiplex to verify that the bead capture, cleaning, and elution steps were successful.

Pt	Gene	Type	Genomic coordinates	Primer sequences		Probe sequence
				Forward	Reverse	
L003	BRAF	S	chr7:140624415	ACTCACCTCCTCCGGAATG	TCTGTTCAACGGGGACATGG	AGGCCGCGGCGCCGGCG
	CTNNB1	S	chr3:41266113	CCTCAGGATTGCCTTTACCA	TTTGATGGAGTTGGACATGG	CTGTGGTAGTGGCACCA
	EGFR	S	chr7:55259515	CCTCCTTCTGCATGGTATTC	AAACACCGCAGCATTGTCAAG	CACCCAGCAGTTTGGCC
	FOXP1	S	chr3:71179760	TTGCTGTAGATGGGTTCTGG	AAAGGGCAATGTTCCAGTGT	GCGGGGGTTGCCGAGTG
	IL7R	S	chr5:35871172	TTCCCGATAGACGACACTCA	TTTCCTTGGCTGCCCTTTAG	CGACACTCAGGTCAAAAG
	JAK1	S	chr1:65313277	GGGAAATATCCCTGTGGCTG	GATGGATTACAAGGATGACG	ATCACTTTTATCTTCTTCTCTT
	MLL	S	chr11:118375108	CCGCAGAGTCCACACAAGTA	TGAAGGCAGGATATTCCCAC	ATACTGAGCTCCTGAAAT
	MUC1	S	chr1:155159940	CTGGTCTGTGTTCTGGTTGC	CATGGAGTGCCTTCTACCG	CTGGCCATTGTCTATCT
L008	BRAF	S	chr7:140481402	TACCATGCCACTTTCCCTTG	GACGGGACTCGAGTGATGAT	ATTGGATCTGGATCATTTG
	MAGI2	S	chr7:77885584	AGATATAACAGATCGGCCGC	ACATTGTCATCATGGACGGG	GTGCCGTCTAGCTGACCAT
	MLL2	S	chr12:49420721	CTGCTGCTGACCATCCAGAA	ATGTCTCGCGGTACCTTGTC	GCTGCTCCATAAACTCTG
	PIK3CA	S	chr3:178936082	AAGAAAAAGAAACAGAGAATCTCCA	GCTAGAGACAATGAATTAAGGGAAA	CCTGCTCAGTGATTT
	RB1	S	chr13:49039501	TCAGAAGGTCTGCCAACACC	CCCCTCTCATTCTTTACTAC	AGAGAAAACACACACCTT
	TNK2	S	chr3:195615444	AGGCGGCAGAATGCAGCCA	TACTGTTGCAGCTGCACCTC	AGCTCCAGCAGCCAGCCTGTGC
	TP53	S	chr17:7578272	CATCCAAATACTCCACACGC	AGAGACGACAGGGCTGGTT	GTCTGGCCCCTCCTCAG
L011	BRAF	S	chr7:140453136	CCACAAAATGGATCCAGACA	TCTTCATGAAGACCTCACAG	CCACTCCATCGAGATTTC
	CHD2	S	chr15:93524686	GAAGTGTCAACAAGTGCAAC	CAATGCTGAGAACTCTCTCC	ACTTCTATCACAGTTTA
	CSMD3	S	chr8:113301756	ACTTCTCGCCAGACAGGACT	CATGCTGGTATAGGTTGTTCG	AGGAACGCCAGGGTGTC
	GPX6	S	chr6:28474130	TGCAAGCACTCACTTGAGAC	ATTGTGTTGGCCTTTCCCTG	TGTTCCTGGTTCTTGTTT
	HSP90AA1	S	chr14:102551757	CAGTTTGGTCTTCTTTCAGGTG	TGGCGTTGAGCACTAAATTG	CTTGCAGGTGAACCTATG
	LGALSL	S	chr2:64683447	TGGGATGATTTTCTTTTCAGC	CTTTGAGTTCGATTGCCACA	TCAGCTTGACCTGTGGG
	OR10Z1	S	chr1:158577038	TTATGGCTGTGCTTCCTTCG	GGGTCACTACAGTATAGGTC	CAAAGCCAGCTAGTCTCTTGAGAG
	TP53	S	chr17:7577046	TGCTTGCTTACCTCGCTTAG	ACAGAGGAAGAGAATCTCCG	AAGGGGAGCCTCACCAC
L013	EGFR	S	chr7:55241708	AAAATTCCCGTCGCTATCAA	GAGAAAAGGTGGGCCTGAG	GCCAACAAGGAAATCCTC
	EGFR	S	chr7:55242511	TCTCTTGAGGATCTTGAAGG	CCCCACCAGACCATGAGA	ATTCAAAAAGATCAAAGTGCTGG
	FBXO46	S	chr19:46216357	TTGGTGGGGTCAAGACAGC	AGCTCCATGAAGGTCAAGGG	GATAGCTCCAAGGCCAAG
	JAK2	S	chr9:5022084	AGATACACCTGAAGAACTGG	GAAATGGAGGGAACATCCAC	TGGAAATGCCAATTCTATG
	KMT2C	S	chr7:151947008	TGAACATACTGCTTACCAGCAA	ACTCACAGCGGTTCAAGTCC	GACTTTCTGAGGGATGACTC
	MSH2	S	chr2:47693816	TCTTCCTTACAGGTTACACG	TGGAATACTTTTTCTTTTTCTTCTTGA	GGACCCTGGCAAACA
	TP53	S	chr17:7579509	CTCTGGCATTCTGGGAGCTT	ATGATTTGATGCTGTCCCCG	TCACTGAAGACCCAGGT
L017	BRCA2	S	chr13:32914959	TTGGTTCCTAATACCAACTG	TGTTGAAGGTGGTTCTTCAG	CTATTAAAGTTTCTCCATATCTCT
	EIF2AK1	S	chr7:6098705	TCTTCGCGCTTGCG	ATCGGAGTGTGGC	CGGCCGGCGATG

				GACCC	AGTGCT	CAGGGG
	KRAS	S	chr12:25398284	CTGTATCGTCAAGGCACTCT	AGGCCTGCTGAAAATGACTG	GTGGTAGTTGGAGCTG
	NF1	S	chr17:29653134	ACACCAAGTATCATGAGCGG	CAGCAGGTAGTTTCTGTTGTTCA	GCTCAGCCAGTTGCCCAGGA
	NF1	S	chr17:29528088	TGTTGGGGTTTTTATAGAACCTG	GCAAGAAACAAGGCAGTCAA	CCAAGTAAGCCATTCTCA
	PAX8	S	chr2:113984793	ATGGCAGAGGAGGCATAGC	CTGTATTTTTCCAGGGCGAG	GCCCGGATACCCACC
	TP53	S	chr17:7577079	TGGTGAGGCTCCCCTTTCTT	AGCTTTGAGGTGCGTGTTTG	CGGCGCACAGAGGAA
	TP53	S	chr17:7577610	CCATGCAGGAACTGTTACAC	CTGCTTGCCACAGGTCTCC	GGCCTGTGTTATCTCCT
	TRIM67	S	chr1:231299607	GTTCTCCATTTCATGCTCGG	GGCTGCAAGAGCCCGGGAG	GGGGGCAGCACGGCC
	TRIP11	S	chr14:92471631	TGTTCCTTGATCGTGTTCAG	ACGACTCAGGGAAGAGCAGA	GTGTTACTGAACTAGCATCT
L019	AKAP9	S	chr7:91674380	GATCTATCTAGAATCCCAAGGACA	AGGTATGCACTCCAGAAAGC	TTTCTTGAACAGCTGCTGTTG
	ARID5A	S	chr2:97215990	GGTGCCCTTGCAGGTGA	CCGCCGTACCTCTCGTAGT	TTCCAGCGGCAGATCCC
	EML4-ALK	T	Chr2:29447613-29447682/Chr2:42497966-42498097	GTGTTAGCTCCTATTATCCTGTCC	TTTTTGAGATAGGGTCTCACTGT	CCAGGGTAGAAGGGAGA
	KNTC1	S	chr12:123026669	GTAAATATTGCTGCAAATC	TGTATTACAAACCTTCAGC	AAAAATATAATGTATTACTTACCTTGT
	LEPRE1	S	chr1:43232537	TTCCATGCTCAGGACCACC	CCAAGCCGAGGTCGAGTC	GGGGCATGGTGACGCCT
	MYCN	S	chr2:16082899	GCGGGTATTGCCGCCCCAG	TACTGAGGGCCTTGTGGTC	TAGCGGCCGCCTGGG
	POU2AF1	S	chr11:111228360	GGTGTAAGGTGTCCATGGG	TGTGTCTGCAGTGACAGAGG	CCTGTGTGTCGGCTGGCTCT
	XRRA1	S	chr11:74644897	GGGGCCTTAAGAGACAACTC	TTGGAGAACTTCAGGCCAC	GAAGCACCACTGTGTGAG

Table 2 Primer designs (UltraSEEK)

Genomic coordinates, forward and reverse primer sequences, and probe sequences for each mutation. Abbreviations: Pt, patient; S, SNV; T, translocation.

2.12.2.2 PCR amplification

In order to identify 1% of mutant copies in 10ng of cfDNA, the UltraSEEK technique required 30 mutant copies per reaction. The amount of functional template DNA was quantified in several of the cfDNA samples using the TaqMan® Copy Number Reference Assay (Life Technologies) targeting the telomerase reverse transcriptase (TERT) gene, according to the manufacturer's instructions. In L003, there were approximately a total of 25 copies/µl in 10µl, in L008 there were approximately a total 10 copies/µl in 10µl, and in L011 there were approximately a total 150 copies/µl in 19µl. The PCR was carried out in a total volume of 20µl with 10µl of DNA template supplemented with 10µL of a master mix consisting of 1x PCR Buffer supplemented with 1mM $MgCl_2$, 125µM dNTPs, 0.125U Uracil-DNA glycosylase (New England Biolabs), 4U Taq polymerase, and 100nM of each PCR primer. Reactions were initially incubated at 30˚C for 10 minutes followed by 94˚C for 2 minutes. 45 cycles of PCR were performed at 94˚C for 30 seconds, 56˚C for 30 seconds, and 72˚C for 1 minute. The PCR was completed with a final incubation of 5 minutes at 72˚C. 5µl of amplified products were conditioned with the addition of 2µl of 0.5U shrimp alkaline phosphatase (SAP) in 0.24X SAP buffer in a total volume of 7µl for 40 minutes at 37˚C followed by SAP enzyme denaturation for 10 minutes at 85˚C. All reagents used were obtained from Agena Bioscience.

2.12.2.3 Single base extension

Single base extension was performed by adding 2µl of a mastermix consisting of 0.2X extension buffer, 5.56µM extend mix, extension probes at various concentrations, and 0.14U iPLEX® Pro enzyme. Single base extension reactions were performed in a total volume of 9µL. Reaction parameters included an initial incubation at 94˚C for 30 seconds, followed by 40 cycles at 94˚C for 5 seconds with five nested cycles of 52˚C for 5 seconds then 80˚C for 5 seconds. The single base extension was completed with incubation at 72˚C for 3 minutes.

2.12.2.4 Control assays

The UltraSEEK panel contained two types of control assays as quality assurance metrics of the acquired data, so that only mutated alleles were detected. Without the presence of the wild-type allele, these controls differentiate a negative result from a failed reaction in the absence of a somatic mutation (**Figure 10**). To control for the

initial steps of the reaction including DNA addition, PCR and single base extension, a DNA-dependent control targeting Albumin (*ALB)*, a housekeeping gene highly unlikely to be mutated in tumours, was included. A second set of controls was added at the post-PCR step to give an indication of the success of the capture process. The capture controls exist at constant concentration across all reactions and were used for data normalisation and mutation calling. The intensity of the capture ions were used to create a capture control fit, and the intensities of each assay were normalised to this fit using a linear least squares regression model (**Figure 11**).

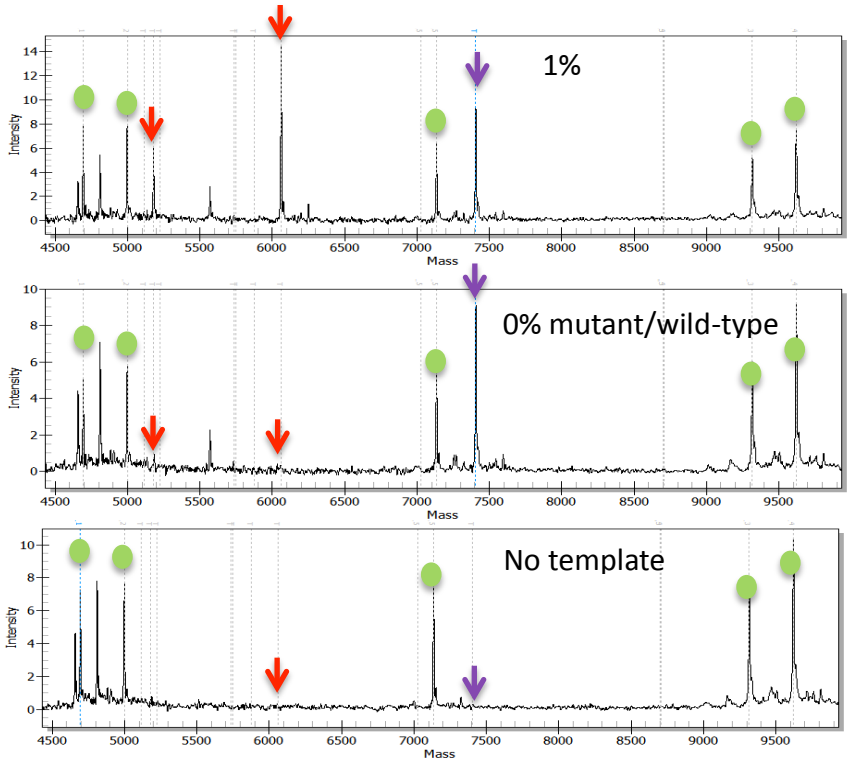


Figure 10 UltraSEEK controls

MALDI-TOF results of the UltraSEEK reaction with controls: capture controls (green), DNA process control (purple), mutation-specific assay peaks (red).

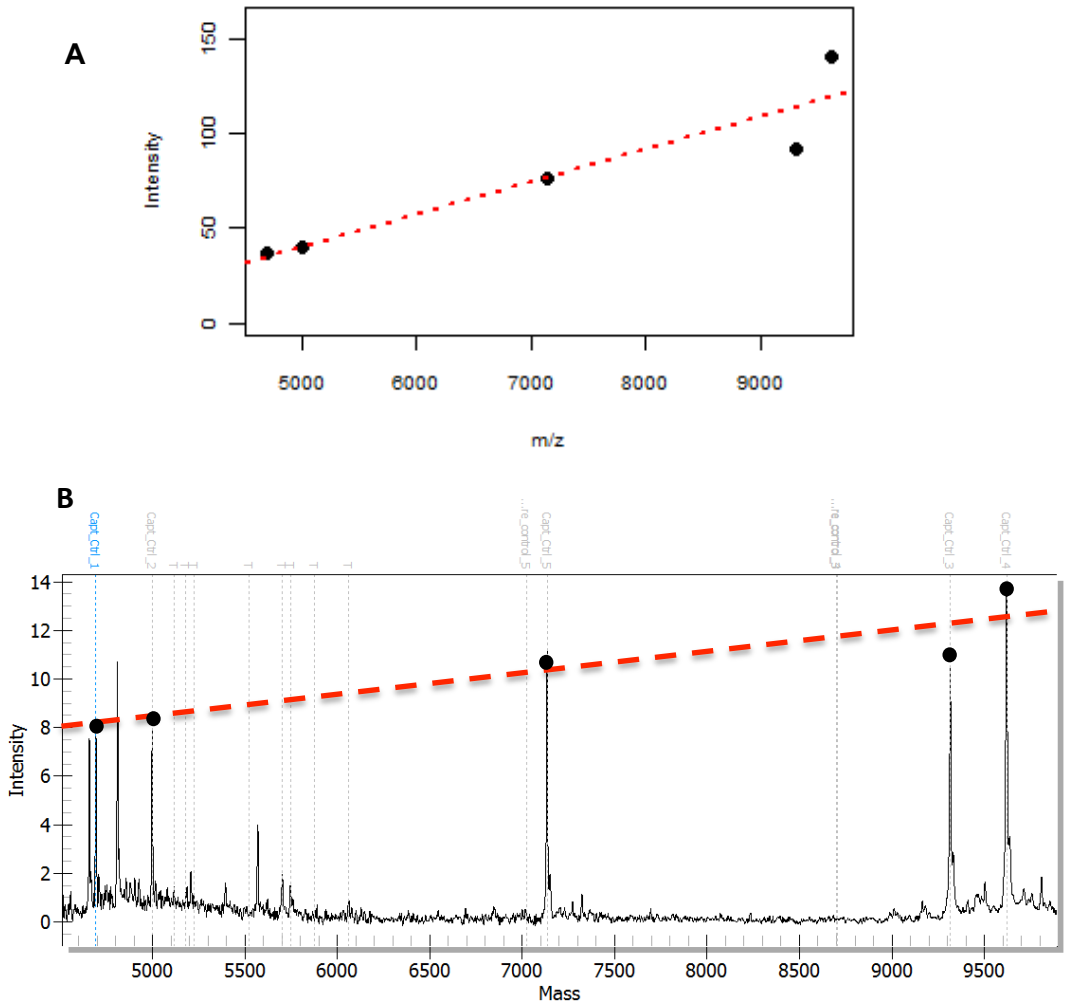


Figure 11 UltraSEEK capture control fit

Per-well capture control ion intensities (A), and linearity of the capture control response using a linear least squares regression model (B).

2.12.2.5 Capture and data acquisition

Prior to capture, the streptavidin coated magnetic beads were conditioned in binding and wash buffer. Two rounds of conditioning were performed on the beads and then they were re-suspended in the binding and wash buffer at a concentration of 1µg/µl. A total volume of 41µl of conditioned beads was added to 9µl reaction and capture was performed at room temperature for 30 minutes under constant rotation. Beads with captured products were pelleted using a magnet and the binding and wash solution was removed. The beads were washed once with 100µl of HPLC grade water, re-suspended with 13µl of elution solution, and incubated at 95˚C for 5 min. Eluted products were conditioned with 5µl (3mg) of anion exchange resin slurry. Finally, the analyte was dispensed onto a Spectrochip® II solid support using an RS1000

Nanodispenser. Data was acquired via MALDI-TOF mass spectrometry using the MassARRAY® 4 instrument.

2.12.2.6 Automated data analysis

Data analysis was performed using the Typer Software (Agena Bioscience). The software acquired raw peak intensity data for all assay products and a linear least squares function was used to fit intensities of the capture control assays and determine the per-well data quality, and the normalisation factor. The intensity of each assay was normalised to the linear fit of the internal controls. The average adjusted normalised intensity for each assay was calculated as the difference between the normalised intensity data point per assay and the average normalised intensity for all data points per assay. This did not change the relative distribution of the data for each assay. Mutations were detected using robust z-score (median absolute deviation (MAD)-based z-score). A robust z-score was calculated for each assay using the median and the MAD values previously established with known non-mutant samples. This data served as the historical baseline for the mutation detection and was used for each data analysis. Samples that exceeded the user-defined assay z-score cutoff (default of 10) and met the peak quality criteria (adjustable minimum peak intensity and the call probability of 0.8 and better) were labelled as containing the mutation by the analysis software and reported accordingly.

2.12.3 Multiplex PCR and targeted MiSeq™ sequencing

2.12.3.1 Assay design and PCR amplification

In collaboration with Illumina (Claire Fielding and Mark Ross, Cambridge, UK), the presence of mutations in cfDNA was analysed using a multiplex PCR and NGS (MiSeq sequencing) approach. PCR primers were designed both manually and separately using the software package Primer3 version 4.0.0 (Koressaar & Remm 2007; Untergasser et al. 2012). **Table 3** summarises the chosen variants and primer designs. Primers were tested and optimised individually beforehand using Promega control DNA (30ng input) and tumour DNA (30ng input). In general, the primers designed using Primer3 performed better with PCR amplification (**Figure 12**). Primers for the *AKAP9* and *KNTC1* mutations failed PCR amplification, and were therefore not included in the multiplex PCR reaction. Primers for the EML4-ALK translocation were designed manually and tested on DNA from a patient in the NSCLC pilot cohort who was known

to have the translocation previously identified by the UCLH pathology department using a diagnostic FISH assay. DNA from lymph node region LN2 was used, and PCR amplification was performed using different DNA concentrations (neat: 30ng, 1:100 and 1:10 dilutions) with Primer3 and manually designed primers, as well as an EML4-ALK primer set previously used by Illumina (**Figure 12**). The PCR products were analysed and quantified using the DNA 1000 kit (Agilent). The 1:10 dilution using the Primer3 designed primers (circled red in **Figure 12**) was taken forward for the multiplex approach.

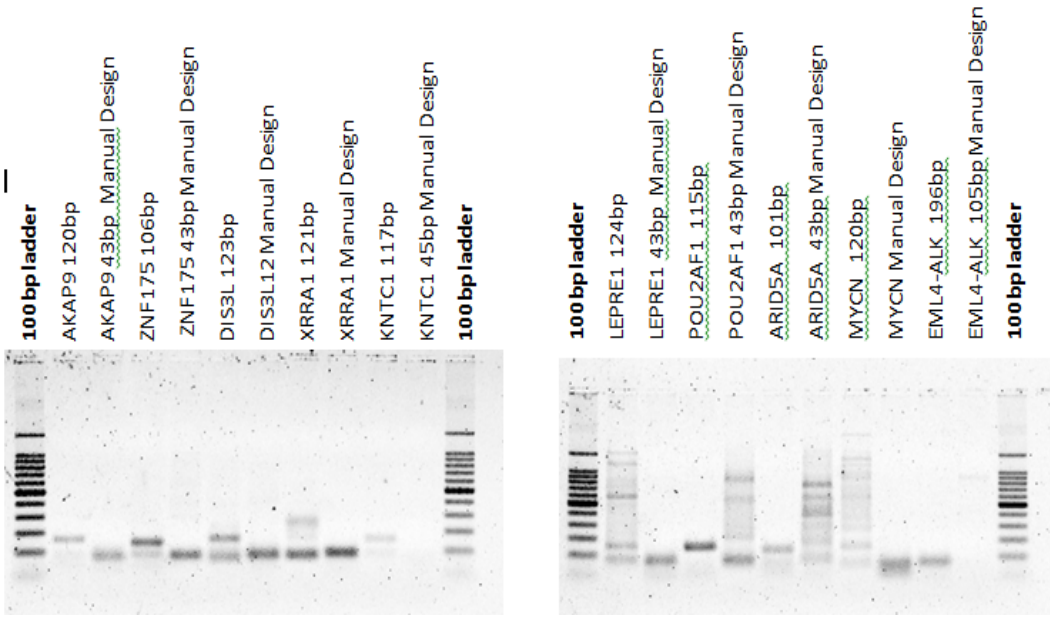


Figure 12 Primer optimisation using Promega control DNA

Electrophoresis gels showing PCR products using primers designed both manually and with Primer3 were tested on control DNA. Each well is labeled with the mutation gene name and amplicon size, and those labeled with 'ManualDesign' were for primers designed manually. PCR performed by Clare fielding, Illumina.

Figure 13 Testing the EML4-ALK translocation primers

Gel electrophoresis using different EML4-ALK primers with DNA from patient L019. Primers designed for the mutation in *ZNF1751* was used on DNA from patient L019 as a control sample. PCR performed by Clare fielding, Illumina.

The total amount of available cfDNA in the case analysed was 4.5ng. Provided PCR products were present with individual primers, all primers were combined into one pool to allow a multiplex approach using cfDNA and tested using a titration of Promega DNA from 25ng to 1ng inputs within a larger 65ul reaction to account for the volume required of cfDNA for 4.5ng input. Using the cfDNA sample, two rounds of PCR were conducted to amplify the mutations using the designed primers. PCR1 was carried out in a total volume of 65µl with 25µl of DNA template (4.5ng) supplemented with a mastermix consisting of 32.5µl Phusion (2x) (New England Biolabs), 6.5µl primer pool (100µM) and 1µl nuclease-free water (Life Technologies). Reactions were initially incubated at 98˚C for 1 minute followed by 30 cycles of 98˚C for 40 seconds, 64˚C for 20 seconds and 72˚C for 15 seconds. The PCR was completed with a final incubation of 5 minutes at 72˚C. PCR2 was performed using a 1:10 dilution of the PCR1 product (2µl PCR1 product plus 18µl nuclease-free water) as the DNA template. PCR2 was carried out in a total volume of 25µl with 2.5µl of DNA template supplemented with a mastermix consisting of 12.5µl Phusion (2x) (New England Biolabs), 1µl primer PE1.0 (10µM), 1µl primer PE2.0 index (10µM), and 8µl nuclease-free water (Life Technologies). Reactions were initially incubated at 98˚C for 1 minute followed by 10 cycles of 98˚C for 40 seconds, 64˚C for 20 seconds and 72˚C for 15 seconds. The PCR was completed with a final incubation of 5 minutes at 72˚C.

2.12.3.2 Targeted sequencing using the MiSeq platform

The amplified cfDNA from PCR2 was analysed and quantified using the Agilent DNA100 kit (Agilent). An initial concentration of 82.3nmol/l was diluted to 2nM, then denatured, and diluted again so that a final cfDNA concentration of 6pM was used for targeted multiplex sequencing on the MiSeq (Illumina) platform with a paired end 151bp read length. Results were analysed using the FASTQ files generated.

Gene	Type	Design	Genomic coordinates	Amplicon size	Primer sequences	
					Forward	Reverse
ZNF175	Indel	Primer3	Chr19:52084702-52084807	106	TGGACCCTGCCCAGAGATGC	AGGATCCAGCCAAGATTCCC
		Manual	Chr19:52084742-52084784	43	GGAGCTCTATAGCCATCTCT	TCAGTTGTGCTCACCCACTG
AKAP9	SNV	Primer3	Chr7:91674338-91674457	120	GCTAATAATAGACTTTTGAAGATCC	AGATGACTGGCTTTTACTAGATCT
		Manual	Chr7:91674359-91674401	43	ATCCTCTTAGAAGTTGTAAA	TTTCTTCAACAGCTGCTGTT
DIS3L2	Indel	Primer3	Chr2:233194568-233194690	123	CAAGATCCACCGCGCCTTCC	GAGCTGAAGTCCACGGGCAG
		Manual	Chr2:233194592-233194634	44	GCAGGCCCTGCTGCGCCGGC	CATCCTTGTTTGGGGCGGGG
POU2AF1	SNV	Primer3	Chr11:111228304-111228417	115	TGGAAGGTTCTGTGTCTGCA	ACTCGGTGTAAGGTGTCCAT
		Manual	Chr11:111228339-111228381	43	CCCTGTGTGCCGGCTGGCTC	AGGGTGGCCGGGGTGGGCTG
KNTC1	SNV	Primer3	Chr12:123026609-123029724	117	GCCGCCTCTGGAAGAACGTG	CTAGTAAATATTGCTGCAAATCA
		Manual	Chr12:123026646-123026690	45	GGGCAGCCCAGGCAGCACCA	ATAATGTATTACTTACCTTG
XRRA1	SNV	Primer3	Chr11:74644852-74644972	121	GATTAATTAAACTTTGTAGGGGCC	GAACTTCAGGCCACTCACATTAA
		Manual	Chr11:74644876-74644923	48	CCAGCGCTGAAGCACCACTGTGT	GGTGCACAGATCTGATGGC
LEPRE1	SNV	Primer3	Chr1:432325043232628	124	GTTGAAGCTGCTGACCACAC	GTAGGCTGCGGTCCCCTCGG
		Manual	Chr1:432325143232558	43	AGGATGGGGCATGGTGACGCC	TCCCCTCGGCGAAGAGCAGA
MYCN	SNV	Primer3	Chr2:160828516082975	120	GGTATTGCCGCCCCAGC	TGAATCGCTCAGGGTGTCC
		Manual	Chr2:160828716082920	43	CCCCGGGGGTCGCCCCTCCG	CTGGTCTGGCGGCCGCCTGG
ARID5A	SNV	Primer3	Chr2:972159397216035	101	GCCGCCTCTGGAAGAACGTG	CGCCCCGCCGTACCTCTCGT
		Manual	Chr2:972159697216011	43	GGGCAGCCCAGGCAGCACCA	GCGGCGCGTGCACGTGGCCG
EML4-ALK	Translocation	Primer3	F:Chr2:29447613-29447634 (ALK) R:Chr2:42497966-42497989 (EML4)	196	CCTATTATCCTGTCCCTTTGAG	GGATAAACAGAAAGGACGGAAGAC
		Manual	F:Chr2:29447656-29447675 (ALK) R:Chr2:42498016-42498035 (EML4)	105	CAGTGTGTGCTGCCATCTCC	CTTAATTCTAAAACCTATGT

Table 3 Primer designs (Illumina)

Chosen variants in L019 and primer sequences designed manually and using Primer3.

2.12.4 Multiplex PCR and targeted HiSeq sequencing

In collaboration with Natera (Robert Pelham, San Carlos, CA, USA), the presence of mutations in cfDNA was analysed using a combined multiplex PCR and NGS (HiSeq sequencing) approach.

2.12.4.1 Assay design, sequencing library preparation, and PCR amplification

Previously extracted cfDNA was quantified using the Universal SYBR Green Quantitative PCR protocol (Sigma-Aldrich), according to the manufacturer's instructions. DNA sequencing libraries for germ line and tumour DNA, cfDNA from each patient, and cfDNA from negative control samples from healthy individuals (non-pregnant females) were prepared by blunt end repair, ligation with standard sequencing adaptors, and library amplification using 50X Advantage® 2 Polymerase Mix (Clontech Laboratories, Inc.) in a volume of 40µl. Adaptor-ligated templates were purified using Agencourt AMPure SPRI beads (Beckman Coulter) eluted in 50µl volume. PCR primers were designed using the Primer3 software package (Untergasser et al. 2012). Interacting primers were estimated *in silico* and placed in separate amplification pools. **Table 4** summarises the chosen variants and multiplex primer designs. Multiplex PCR-NGS was performed with 2 assay pools. PCR was performed using AmpliTaq Gold® (Life Technologies), 25nM of each primer, and 6µl purified library in a 20µl reaction volume. The multiplex PCR product was diluted 1:10 and 1µl of the diluted PCR product was used as a template for sample bar-coding (BC-PCR), using Q5® High-Fidelity 1X PCR master mix (New England BioLabs), up to 96 sample barcodes, and 1µM forward and reverse primers in a 10µl total volume. For each of the two primer pools, the BC-PCR products were pooled (8µl of each library product and 2µl of each DNA product) and purified. The pool was quantified with Qubit® dsDNA HS assay kit (Life Technologies).

2.12.4.2 Targeted sequencing using the HiSeq 2500 platform and statistical analysis

Sequencing libraries were sequenced using a HiSeq 2500 (Illumina, San Diego, CA) with a paired end 50bp read length. For the cfDNA samples, sequencing resulted in a median depth of read of 62,399x per target and a median depth of read of 62,335x per call. Reads were aligned to the human genome (hg19) using BWA-mem v0.7.10 aligner (Li & Durbin 2009). Bases observed across reads were piled and counted with

SAMtools mpileup (Li et al. 2009). Target assays with < 100 reads in the tumour or < 10,000 reads in the plasma samples were tagged as failed assays and were not analysed further. A background error model was constructed using negative-control cfDNA samples from healthy individuals. These negative control samples were sequenced in the same sequencing run as positive test samples to account for run-specific artefacts. Noisy positions with a normal median VAF > 0.5% were removed. Outlier samples were iteratively removed from the model to account for noise and contamination. For each base substitution at every genomic locus, the depth of read weighted mean and standard deviation of the error were calculated. In matched tumour and cfDNA, variant calls with at least 5 variant reads, a z-score of 10 and a confidence of 99.9% (using a Bayesian likelihood fit that involved parameter estimation from the control samples) against the background error model were called as a candidate mutation.

Pt	Gene	Type	Genomic coordinates	Primer sequences	
				Forward	Reverse
L012	**BRIP1**	SNV	chr17:59924572	ACACGACGCTCTTCCGATCTAATTGTGTACTTCTGTTCCAAAGC	AGACGTGTGCTCTTCCGATCTCTCCTTAAACTTCATTTTGGTTCTGTGT
	CARS	SNV	chr11:3062181	ACACGACGCTCTTCCGATCTGTTGGCCCACAGCAATACC	AGACGTGTGCTCTTCCGATCTGTGTTCATACCTCAAGATGGGAAAAA
	CDKN2A	SNV	chr9:21971177	ACACGACGCTCTTCCGATCTGGCAGCGCCCGAGTG	AGACGTGTGCTCTTCCGATCTCGCCGTGGAGCAGCA
	CIC	SNV	chr19:42797381	ACACGACGCTCTTCCGATCTCACGGCCCGGAGCAG	AGACGTGTGCTCTTCCGATCTGCTCCTCAGCAGGTGGG
	FAT1	SNV	chr4:187519147	ACACGACGCTCTTCCGATCTACAACGGAGGCTTTGTTTGC	AGACGTGTGCTCTTCCGATCTAAATACACATACCTCTGACCAGTATATAA
	KDM6A	SNV	chrX:44921898	ACACGACGCTCTTCCGATCTACTATATTCTCTTTTTGTTCTTCTTCTAGCA	AGACGTGTGCTCTTCCGATCTGGATTTAAATTATTTCTATTTGCGCGGAG
	LEMD2	SNV	chr6:33756755	ACACGACGCTCTTCCGATCTGCCTCCTCCCGCAGC	AGACGTGTGCTCTTCCGATCTAAGCTGCGCCGCCTG
	MLLT4	SNV	chr6:168347475	ACACGACGCTCTTCCGATCTCCCAAGGCAGCTGAGGAC	AGACGTGTGCTCTTCCGATCTATAATAATTCAACTCAAAATGGGTCTCC
	NBEAL2	SNV	chr3:47044229	ACACGACGCTCTTCCGATCTGGCAACGCAGCACTCCA	AGACGTGTGCTCTTCCGATCTCGAGCTGGCGCCACAG
	NFE2L2	SNV	chr2:178098801	ACACGACGCTCTTCCGATCTGGCTGGCTGAATTGGGAGA	AGACGTGTGCTCTTCCGATCTTTTTCGCTCAGTTACAACTAGATGAA
	RASA1	SNV	chr5:86642517	ACACGACGCTCTTCCGATCTAAGATGGTTCCATGGGAAGATTTC	AGACGTGTGCTCTTCCGATCTAGCAAGTAAATATGTAAGTACCTGTCAT
	TLX1	SNV	chr10:102891574	ACACGACGCTCTTCCGATCTAGGAGCCGGTCAGAGGA	AGACGTGTGCTCTTCCGATCTGAGGTCCCGGCGGCC
	TP53	SNV	chr17:7578406	ACACGACGCTCTTCCGATCTAGCAGCGCTCATGGTGG	AGACGTGTGCTCTTCCGATCTACATGACGGAGGTTGTGAGG
	TP53	SNV	chr17:7578190	ACACGACGCTCTTCCGATCTACATAGTGTGGTGGTGCCC	AGACGTGTGCTCTTCCGATCTGCAAACCAGACCTCAGGCG
	CELSR3	SNV	chr3:48698933	ACACGACGCTCTTCCGATCTCGGATAAGGCCGCTCTGC	AGACGTGTGCTCTTCCGATCTCGCTGGAGCTGTTCAGCAT
L013	**EGFR**	SNV	chr7:55241708	ACACGACGCTCTTCCGATCTTGCCGAACGCACCGG	AGACGTGTGCTCTTCCGATCTGGAAACTGAATTCAAAAAGATCAAAGTG
	EGFR	SNV	chr7:55242511	ACACGACGCTCTTCCGATCTCACAGCAAAGCAGAAACTCACA	AGACGTGTGCTCTTCCGATCTATCTCCGAAAGCCAACAAGGAA
	FBXO46	SNV	chr19:46216357	ACACGACGCTCTTCCGATCTGGGTCAAGACAGCGCCTC	AGACGTGTGCTCTTCCGATCTGCACTGGGGCAGCGATAG
	HERC4	SNV	chr10:69793756	ACACGACGCTCTTCCGATCTCTATCTTTGGATGGGGACGCA	AGACGTGTGCTCTTCCGATCTTTTCATCATTAAGACCTAGCTGACCAA
	JAK2	SNV	chr9:5022084	ACACGACGCTCTTCCGATCTCACCTGAAGAACTGGATCTATTTGC	AGACGTGTGCTCTTCCGATCTTGGTGATATTTCTGGAAATGCCA
	KMT2C	SNV	chr7:151947008	ACACGACGCTCTTCCGATCTAGTCCACACTGAAGAGCAACA	AGACGTGTGCTCTTCCGATCTGTGTCAAGACTTTCTGAGGGATGA

	MSH2	SNV	chr2:47693816	ACACGACGCTCTTCCGATC TGGCTTGGACCCTGGCAA	AGACGTGTGCTCTTCCGA TCTCCAAACTGTGCACTG GAATCC
	MTOR	SNV	chr1:11292495	ACACGACGCTCTTCCGATC TGGATTCCTCTTTGTTGGCC AAAA	AGACGTGTGCTCTTCCGA TCTGCTCAATCAGGAAGC AGTAATACTC
	PLCG2	SNV	chr16:81942036	ACACGACGCTCTTCCGATC TGGCCCAGGATATACCCCC T	AGACGTGTGCTCTTCCGA TCTTGTGGAACCATTTCT CCCCAA
	TP53	SNV	chr17:7579509	ACACGACGCTCTTCCGATC TCTGGCATTCTGGGAGCTT CA	AGACGTGTGCTCTTCCGA TCTATGGTTCACTGAAGA CCCAGG
	AGAP2	SNV	chr12:58131938	ACACGACGCTCTTCCGATC TGCTCGAGTCGGTGCCTC	AGACGTGTGCTCTTCCGA TCTCAGTCTCGGAGCCTC TGG
L015	**ALK**	SNV	chr2:29940530	ACACGACGCTCTTCCGATC TTTCCATGTAAAATAATCAG GAGAAGGA	AGACGTGTGCTCTTCCGA TCTAGGTCATAGCTCCTT GGAATCAC
	ATHL1	SNV	chr11:290854	ACACGACGCTCTTCCGATC TGCAGCTGCAGGGCCT	AGACGTGTGCTCTTCCGA TCTCCAGGCTGAGGCTC AGG
	CD8B	SNV	chr2:87088972	ACACGACGCTCTTCCGATC TTGCGCGGCCAAGAGG	AGACGTGTGCTCTTCCGA TCTCGGGGCCAGGTGTC C
	FOXK1	SNV	chr7:4799190	ACACGACGCTCTTCCGATC TGGGGCTCCCATGATGCG	AGACGTGTGCTCTTCCGA TCTCCCAGGTCCAGCAC GG
	GABRG1	SNV	chr4:46060315	ACACGACGCTCTTCCGATC TGAAAGAACAACTGTCAGA ATGCATG	AGACGTGTGCTCTTCCGA TCTGACCTGAGCAGAAG AATGGGATAT
	KDM6A	SNV	chrX:44922755	ACACGACGCTCTTCCGATC TGCTGTGGCTGCTGGC	AGACGTGTGCTCTTCCGA TCTCTCATCACTGCCTAC AAACTCAG
	MLL2	SNV	chr12:49443815	ACACGACGCTCTTCCGATC TCCGGTGCACGTGGCT	AGACGTGTGCTCTTCCGA TCTCAGGGCTCACCATGT GAAGAA
	RHOXF1	SNV	chrX:119249598	ACACGACGCTCTTCCGATC TGCGGTGTGAACCACGAGA	AGACGTGTGCTCTTCCGA TCTGATCATGCCGCCATC GC
	ROS1	SNV	chr6:117687379	ACACGACGCTCTTCCGATC TGATCCAGAACAGCCGACC A	AGACGTGTGCTCTTCCGA TCTATTTCCCAGAATGCA CTGATGTAC
	SLC39A4	SNV	chr8:145638322	ACACGACGCTCTTCCGATC TCAGCCCCGCGTGCAG	AGACGTGTGCTCTTCCGA TCTGTGGACTCGCCCGC A
	TP53	SNV	chr17:7578254	ACACGACGCTCTTCCGATC TCCTCAGCATCTTATCCGAG TGG	AGACGTGTGCTCTTCCGA TCTCATCCAAATACTCCA CACGCAAA
	ZFHX4	SNV	chr8:77776735	ACACGACGCTCTTCCGATC TCCATCTAGGCCAGAAGCA GG	AGACGTGTGCTCTTCCGA TCTGCTAGAAGACTTAGA TAATTCTTTGGAAGT
	ZMYM4	SNV	chr1:35827319	ACACGACGCTCTTCCGATC TTGGTGCCATTGCTTTGTCA TAC	AGACGTGTGCTCTTCCGA TCTTCAAATATTAGAATTA AAGAAGAACCTTTGGA
	BRCA2	SNV	chr13:32914959	ACACGACGCTCTTCCGATC TACCAACTGTTGTTTGTCTT GTTG	AGACGTGTGCTCTTCCGA TCTCAGAAAATAATCACT CTATTAAAGTTTCTCCA
L017	**EIF2AK1**	SNV	chr7:6098705	ACACGACGCTCTTCCGATC TGCCGGCGATGCAGGG	AGACGTGTGCTCTTCCGA TCTGCGCTTGCGGACCC C
	KRAS	SNV	chr12:25398284	ACACGACGCTCTTCCGATC TGAATATAAACTTGTGGTAG TTGGAGCT	AGACGTGTGCTCTTCCGA TCTGCTGTATCGTCAAGG CACTCT
	NF1	SNV	chr17:29653134	ACACGACGCTCTTCCGATC TGTGCTCAGCCAGTTTCCC A	AGACGTGTGCTCTTCCGA TCTGCCTCAAAGGTAGCA AAAGGC
	NF1	SNV	chr17:29528088	ACACGACGCTCTTCCGATC TCCACATCTGCAGGCTGAC T	AGACGTGTGCTCTTCCGA TCTCCTGCTTTTTAATCC AAGTAAGCCA

	PAX8	SNV	chr2:113984793	ACACGACGCTCTTCCGATC TGCTGCCCGGATACCCAC	AGACGTGTGCTCTTCCGA TCTGCCCTGTCCGCTGG TG
	TP53	SNV	chr17:7577079	ACACGACGCTCTTCCGATC TACCGGCGCACAGAGGA	AGACGTGTGCTCTTCCGA TCTCTCCCCTTTCTTGCG GAGAT
	TP53	SNV	chr17:7577610	ACACGACGCTCTTCCGATC TTGGTGGTACAGTCAGAGC CA	AGACGTGTGCTCTTCCGA TCTATCTTGGGCCTGTGT TATCTCC
	TRIM67	SNV	chr1:231299607	ACACGACGCTCTTCCGATC TGACTGGGGGCAGCACG	AGACGTGTGCTCTTCCGA TCTGGGACACGTGGGGA ACTTG
	TRIP11	SNV	chr14:92471631	ACACGACGCTCTTCCGATC TCCTTGATAGTGTTACTGAA CTAGCATC	AGACGTGTGCTCTTCCGA TCTTGTTCCTTGATCGTG TTCAGTTG

Table 4 Primer designs (Natera)

Genomic coordinates, forward and reverse primer sequences for each mutation.

2.13 Fluorescence *in situ* hybridisation

2.13.1 Sample preparation and centromeric probe hybridisation

The SPOT-Light® Tissue Pretreatment kit (Invitrogen) was used for enzyme and heat pretreatment of FFPE slides prior to hybridisation. Each TMA slide was placed in a dry oven at 55˚C for 1 hour, after which it was deparaffinised twice using fresh xylene in a coplin jar for 5 minutes each time. Each slide was then rehydrated twice in 100% ethanol for 3 minutes. 250µl of 100% ethanol was added to each slide under a 22x50mm coverslip and placed on a heat block at 98˚C until the ethanol had evaporated. Each slide was then washed twice using distilled water in a coplin jar on a shaker for 3 minutes. 40ml of SPOT-Light Tissue Pretreatment solution (reagent 1) pre-heated to 98-100˚C was added to each slide and kept at 98-100^{o}C for 15 minutes in a water bath. The slides were then washed twice with distilled water in a coplin jar on a shaker for 3 minutes. 60µl of the digestion enzyme (reagent 2) was added to each slide at room temperature (RT), and a 22x50mm coverslip placed on top. This was incubated at RT for 5 minutes prior to washing twice with distilled water in a coplin jar for 3 minutes.

In order to capture underlying numerical CIN, rather than structural CIN resulting from intra-chromosomal rearrangements, chromosomes 2 and 15 were chosen based on a previous discovery cohort in our laboratory (Roylance et al. 2011), whereby infrequent copy number alterations were seen for these chromosomes in a series of breast tumours analysed by 1 Mb array comparative genomic hybridisation (aCGH). Dual colour FISH was carried out using centromeric probes CEP2 (D2Z1, Abbott Molecular) and CEP15 (D15Z1, Abbott Molecular) labelled with spectrum red and green, respectively. A solution of 1.5µl for each centromeric probe was mixed and diluted to give a final volume of 10µl per slide and placed on wet ice. 10µl of this mixture was placed on a 22x22mm coverslip, the slides were placed on the coverslip such that the mixture covered the TMA directly, and then the coverslip was sealed to the slide with rubber solution. This was placed on a heat block at 80˚C for 10 minutes. The slides were then placed inside a wet chamber and incubated overnight at 37˚C. The following day they were removed from the incubator, the rubber seal broken and the coverslip taken off. Each slide was placed inside a coplin jar containing a solution of 0.5X saline-sodium citrate (SCC) at 37˚C for 5 minutes, and then washed three times with 100ml of phosphate-buffered saline (PBSA) solution in a coplin jar on the shaker at RT. After

washing, 2-3 drops of ProLong® Gold antifade mounting media (Life technologies, 1:10,000 dilution) was added to a 22x50mm coverslip, each slide was placed on top of this to cover the TMA, and then stored at 4˚C.

2.13.2 Scoring of centromere signals

Slides were scanned on the Applied Imaging Ariol System (Applied Imaging), and images were captured using a 40x objective with seven 0.5mm z-stacks. Depending on the TACT study site, each TMA slide contained between 100-200 tissue cores. Each core was magnified so that forty nuclei per core with clear and discrete hybridisation signals for both chromosomes were chosen and scored manually (**Figure 14**). The number of centromeres in each nucleus for each chromosome was counted and recorded on scoring sheets.

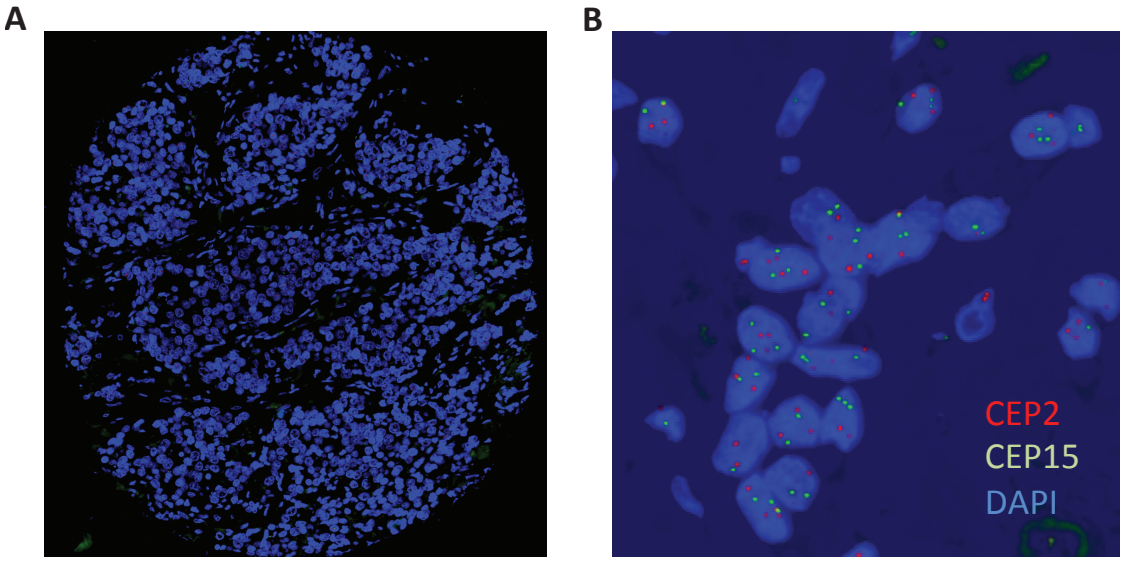


Figure 14 TMA slide from the TACT cohort

A single tumour core from a TACT TMA slide (A), and magnified nuclei from within a core with discrete CEP2 and CEP15 centromere signals (B).

2.14 Defining the MCD group

2.14.1 Calculating the CIN score

Using previously established methods (Roylance et al. 2011), a CIN score was derived by counting the numbers of centromeres for chromosomes 2 and 15 in 40 nuclei per core. The mean (counts for chromosomes 2 and 15 combined) percentage of cells deviating from the modal centromere number was used to define 4 CIN score groups

(Modal Centromere Deviation groups 1 to 4: MCD1, 0%–15%; MCD2, 15%–30%; MCD3 30%–45%; and MCD4, >45%). The MCD4 cohort was classified as having an MCD score greater than 45%, in keeping with the definition for "unstable aneuploidy" previously used in an analysis of 20 breast tumours (Lingle et al. 2002). To avoid false classification of CIN due to sectioning artefacts, and to control for bimodality in diploid tumours, all centromere counts equal to 1 were removed in deriving the CIN score.

2.15 Calculating the Shannon Diversity Index (SDI)

To confirm the validity of the MCD score approach in identifying tumours with the most extreme CIN, the MCD score was compared with the Shannon Diversity Index (SDI) (Maley et al. 2006), a measure of clonal heterogeneity, where centromere counts equal to 1 were included. The SDI was estimated for chromosomes 2 and 15 using the following formula:

$$H = -\sum_{i} p_i \ln (p_i)$$

where p_i is the frequency of the centromere number *i*.

2.16 CIN and clinical outcome statistical analysis

Blinded outcome data were analysed at the Institute of Cancer Research Clinical Trials and Statistics Unit, which was responsible for the TACT trial management and data analysis. The primary clinical endpoint was invasive disease-free survival as previously reported (Ellis et al. 2009). Patients who were alive and disease-free at the last follow-up were censored. Correlation between patient characteristics and biomarkers was examined using Spearman's rank correlation coefficient. For survival-related endpoints, Kaplan-Meier product limit curves were plotted and prognostic and predictive effects examined by use of Cox proportional hazards regression models. Time-to-event analyses were stratified by the control regimen and included all patients with available biomarker data on an intention-to-treat basis. With 382 events a hazard ratio (HR) of 1.4 would be detectable with approximately 85% power, and a two-sided 5% significance level would be achieved if patients were allocated into two approximately equal sized groups on the basis of biomarker values. A p-value of 0.05 was used to define statistical significance, but confidence intervals were also considered relevant. In order to define the relationship between MCD group and prognosis, the more sensitive method of calculating the value for p trend was used, which assesses whether a statistically significant tendency for trend in the HR exists as MCD group increases from 1 to 4 in a multivariate analysis (Peto et al. 1977).

3 Results 1: Intratumour heterogeneity in patients with NSCLC

3.1 Introduction

The NSCLC pilot (UCLHRTB 10/H1306/42) and TRACERx (13/LO/1546, NCT01888601) cohort consisted of patients with primary NSCLC eligible for curative surgical resection (stages IA to IIIA). Multiple tumour regions were collected from each primary tumour for whole-exome sequencing (WES) using DNA extracted from each tumour region. WES was also performed on whole blood for the germ line reference. Where possible, lymph node samples were also sequenced. All the variants identified by WES were manually reviewed, and subsequently validated in a subset of patients using Ion AmpliSeq sequencing at a mean sequencing depth of 1369x, and an overall validation rate of > 90% for SNVs and indels. In the NSCLC pilot cohort, variants identified in tumours were validated by Ion AmpliSeq sequencing, except for tumours L022 and L023. In the TRACERx cohort, tumours were not validated, but instead variants were filtered using thresholds derived from the NSCLC pilot cohort proven to confidently predict the presence or absence of a mutation in a given tumour region. As previously described (Section 2.5.2), variants that were filtered met the following criteria:

- A variant allele frequency (VAF) ≥ 5% in at least one tumour region
- A minimum variant count ≥ 5 reads in at least one tumour region
- A germ line VAF < 0.01
- Not detected in a pre-defined artefact blacklist region of the human genome

A variant was considered to be present in a tumour region provided the VAF was ≥ 1%. In addition, variants were further filtered using a somatic p-value of ≤ 0.01, which was derived based on the variants that had been manually reviewed and validated using high-depth Ion AmpliSeq sequencing. Ubiquitous (truncal) variants were those present in all tumour regions, and heterogeneous (branch) mutations were those present in one or several, but not all, tumour regions.

Work from our laboratory has previously shown extensive intratumour heterogeneity, in terms of mutations and copy number alterations, in a small cohort of 7 NSCLC patients

(de Bruin et al. 2014). Multi-region WES demonstrated branched tumour evolution with predicted drivers of disease occurring both early and late in tumour evolution. In this chapter, the genomic landscape of NSCLC is investigated in a larger cohort of 27 patients, including adenocarcinoma and squamous cell carcinoma histological subtypes with greater power to infer subclonal mutational events. In each tumour, the spatial heterogeneity of mutations, mutational signatures, and somatic copy number aberrations are assessed. In addition, there are three cases in which paired primary and metastatic tumours, have been sequenced. The bioinformatics pipeline was developed by Gareth Wilson and Richard Mitter (The Francis Crick Institute).

3.1.1 Baseline characteristics and histopathological variables

Table 5 and **Table 6** show the baseline characteristics of the patients in the NSCLC pilot and TRACERx cohort. All patients had either lung adenocarcinoma (LUAD) or squamous cell carcinoma (LUSC), and were eligible for curative surgical resection at diagnosis (stage IA-IIIA). The number of smoking pack-years was calculated by multiplying the number of cigarette packs smoked per day by the total number of years smoked. Median values are given since the clinical variables were not normally distributed. In the combined cohort of 27 patients, the median age was 67-years (range 47-79 years), 13/27 (48%) were female, 14/28 (52%) were male, 12/27 (44%) had LUAD, 15/27 (58%) had LUSC, 14/27 (52%) were current smokers, 10/27 (37%) were ex-smokers, and 3/27 (11%) were never-smokers. A detailed smoking history was taken from each patient, with regards to the number of smoking pack-years, and how long before surgery they had stopped smoking in the case of ex-smokers. The distribution of tumour stage was 6/27 (22%) for stage IA, 5/27 (19%) for stage IIA, 4/27 (15%) for stage IIIA, 9/27 (33%) for stage IB, and 3/27 (11%) for stage IIB. There was evidence of lymph node involvement in 5/27 (19%) of the patients, two of whom subsequently developed recurrent disease. The median tumour size was 30mm (range 10-67mm). In the pilot NSCLC cohort, in addition to the primary tumours, metastatic tumours were sequenced in patients L011, L017 and L023. These cases are discussed separately in Sections 3.5.1, 3.5.2, and 3.5.3.

3.1.2 Recurrence and clinical outcome

Patients were followed-up after surgery for evidence of disease recurrence. Follow-up data were collected up to June 2015. Four out of twenty-seven patients (15%)

developed recurrent disease during the follow-up period. Four patients (15%) died during the follow-up period, three of whom died as a result of disease progression. Patient LTX030 was diagnosed with brain metastases 4 months after surgery. Given the short time to recurrence, it is likely that this patient had occult metastatic disease at diagnosis, which was not detected since imaging of the brain was not clinically indicated at the time. **Table 7** and **Table 8** show the clinical outcome data, in terms of recurrence and survival, for the NSCLC pilot and TRACERx cohort.

Pt	Age	Gender	Smoking status (pack-years)	Smoking stopped (years)	Histo	Stage	Tumour location/ max diameter (mm)	Lymph node involve-ment
L011	49	F	Smoker (45)		LUAD	IB	LUL/55	No
L012	69	F	Smoker (40)		LUSC	IB	RLL/40	No
L013	68	F	Smoker (50)		LUSC	IB	RLLL/50	No
L015	68	M	Smoker (100)		LUSC	IA	LUL/30	No
L016	65	M	Smoker (40)		LUSC	IIIA	RLL/27	No
L017	61	F	Smoker (48)		LUAD	IIB	RUL/14 and RUL/20	No
L019	47	M	Never		LUAD	IIIA	LUL/25	Yes
L022	54	M	Smoker (35)		LUAD	IIIA	RUL/32	Yes
L023	50	F	Smoker (60)		LUAD	IIIA	RUL/10 and RML/25	No
L029	66	F	Ex-smoker (40)	2	LUAD	IA	LLL/15	No
L030	79	F	Ex-smoker (40)	10	LUAD	IA	LLL/24	No

Table 5 Baseline characteristics of the NSCLC pilot cohort

Baseline patient and tumour characteristics of the NSCLC pilot cohort. Abbreviations: Pt, patient; Histo, histology; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; RUL, right upper lobe; RLL, right lower lobe; LUL, left upper lobe; and LLL, left lower lobe.

Pt	Age	Gender	Smoking status (pack-years)	Smoking stopped (years)	Histo	Stage	Tumour location/ max diameter (mm)	Lymph node involve-ment
LTX001	68	F	Ex-Smoker (35)	20	LUAD	IB	RUL/26	No
LTX012	65	M	Ex-Smoker (35)	16	LUSC	IIA	RUL/35	Yes
LTX015	64	M	Smoker (51)		LUSC	IB	RLL/32	No
LTX016	67	M	Ex-Smoker (20)	15	LUSC	IIB	RLL/66	No
LTX019	65	M	Smoker (50)		LUSC	IIA	RML/50	Yes
LTX022	72	M	Smoker (50)		LUSC	IIA	LUL/30	No
LTX028	66	M	Ex-Smoker (40)	3	LUSC	IIA	RLL/55	No
LTX029	78	F	Ex-smoker (40)	28	LUAD	IB	RUL/22	No
LTX030	76	M	Smoker (50)	1	LUSC	IIA	RUL/45	Yes
LTX031	50	F	Smoker (30)		LUSC	IA	RUL/15	No
LTX033	73	M	Ex-Smoker (50)	10	LUSC	IB	RLL/35	No
LTX034	73	F	Never		LUAD	IA	RLL/30	No
LTX036	68	M	Smoker (30)		LUAD	IB	RUL/25	No
LTX038	76	M	Ex-Smoker (30)	26	LUSC	IIB	RUL/67	No
LTX051	67	F	Never		LUAD	IA	RUL/26	No
LTX058	73	F	Ex-Smoker (45)	13	LUSC	IB	RUL/44	No

Table 6 Baseline characteristics of the TRACERx cohort

Baseline patient and tumour characteristics of the TRACERx cohort. Abbreviations: Pt, patient, Histo, histology; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; RUL, right upper lobe; RLL, right lower lobe; LUL, left upper lobe; and LLL, left lower lobe.

Patient	Recurrence (months)	Current status	Cancer-related death?	Survival (months)
L011	Yes - brain (14)	Dead	Yes	19
L012	No	Dead	No	2
L013	No	Alive		33
L015	No	Alive		36
L016	No	Alive		35
L017	Yes - lung (4)	Alive		35
L019	Yes - liver (8)	Dead	Yes	15
L022	No	Alive		34
L023	No	Alive		33
L029	No	Alive		29
L030	No	Alive		29

Table 7 Clinical outcome in the NSCLC pilot cohort

For each patient, if applicable, the site of recurrence and time to recurrence is shown. The current status and the overall survival from the date of surgery are shown. In patients who died, whether the cause of death was cancer-related is noted.

Patient	Recurrence (months)	Current status	Cancer-related death?	Survival (months)
LTX001	No	Alive		15
LTX012	No	Alive		12
LTX015	No	Alive		11
LTX016	No	Alive		11
LTX019	No	Alive		12
LTX022	No	Alive		11
LTX028	No	Alive		10
LTX029	No	Alive		10
LTX030	Yes - brain (4)	Dead	Yes	4
LTX031	No	Alive		11
LTX033	No	Alive		11
LTX034	No	Alive		11
LTX036	No	Alive		11
LTX039	No	Alive		10
LTX051	No	Alive		9
LTX058	No	Alive		9

Table 8 Clinical outcome in the TRACERx cohort

For each patient, if applicable, the site of recurrence and time to recurrence is shown. The current status and the overall survival from the date of surgery are shown. In patients who died, whether the cause of death was cancer-related is noted.

3.2 Regional distribution of mutations and phylogenetic trees

Heat maps representing the regional distribution of non-silent SNVs and indels in tumours were created for the NSCLC pilot cohort as shown in **Figure 16** and **Figure 17**, and for the TRACERx cohort as shown in **Figure 18** and **Figure 19**. Where available, validated non-synonymous mutations were used to create the heat maps, and otherwise filtered non-synonymous mutations were used based on thresholds proven to confidently predict the presence or absence of a mutation, as previously described. In the NSCLC pilot cohort, phylogenetic trees were created using validated non-synonymous SNV and indel mutations. In the TRACERx cohort, they were created using filtered non-synonymous and synonymous SNV and indel mutations. Consistent with previous findings from our laboratory (de Bruin et al. 2014), intratumour heterogeneity was evident in all tumours, regardless of smoking status, tumour stage or histological subtype. Certain mutations were present in all tumour regions (shared, truncal), certain mutations were present in some, but not all, tumour regions (shared, branch), and certain mutations were present in one tumour region only (private, branch). This degree of heterogeneity in the spatial distribution of mutations is shown in the heat map for each tumour (**Figure 16** to **Figure 19**). Using this regional distribution, the temporal heterogeneity of each tumour was represented by phylogenetic trees demonstrating branched tumour evolution, with mutations present on the blue trunks of the trees occurring prior to tumour initiation or early in tumour evolution, and mutations present on the yellow and red branches of the trees occurring later in tumour evolution (**Figure 16** to **Figure 19**). These heat maps and trees depict the extent of spatial and temporal heterogeneity in these tumours, and demonstrate the variation in heterogeneity from one tumour to another. Tumours from patients who had never smoked often had shorter trunks compared to tumours from patients who were either current or ex-smokers ($p = 0.24$, **Figure 15**A). This was most likely related to the high mutational burden associated with smoking-related NSCLC (Hammerman et al. 2012; Govindan et al. 2012; Lawrence et al. 2013), supported by the fact that the proportion of C>A transversions (Pfeifer & Hainaut 2003), which have been shown to be associated with a history of smoking, was greater in the trunks of patients who were current smokers compared to patients who were either ex-smokers or had never smoked before ($p = 0.064$, **Figure 15**B).

Robust and accurate measures of intratumour heterogeneity are yet to be determined, but one potential measure developed in our laboratory is the pairwise intratumor

heterogeneity index. This was calculated by determining the mean proportion of heterogeneous (branch) mutations relative to the total number of mutations for each possible pairwise comparison of all tumour regions. However, including ubiquitous (truncal) mutations can introduce a bias against tumours that arise in the context of smoking, since these tumours will generally have a greater number of somatic mutations associated with exposure to tobacco smoke (Pfeifer & Hainaut 2003), and therefore a greater number of truncal mutations. One alternative is to calculate the index in the absence of the truncal mutations, such that the relative lengths of the branches are determined indicating the degree of evolution that has taken place after clonal diversification (branching of the tree). This was calculated as the mean of the branch length for each pairing of tumour regions, divided by the sequencing depth after it had been adjusted for tumour purity. This measure was used to calculate an intratumour heterogeneity index for each tumour.

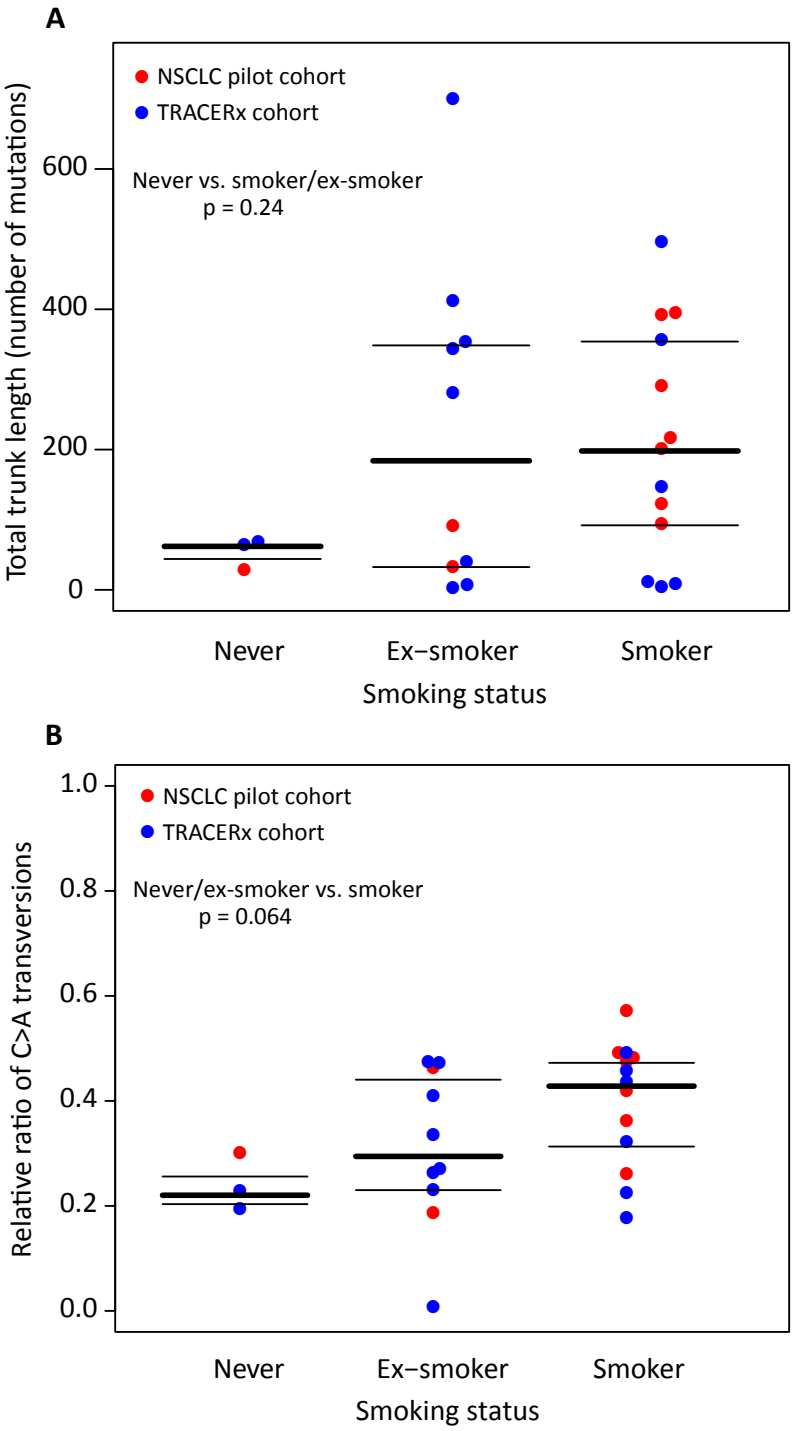


Figure 15 Phylogenetic tree trunk length and smoking status

The trunk length of phylogenetic trees in the NSCLC pilot and TRACERx cohort (A), and the relative ratio of C>A transversions in the trunk compared to branch mutations, associated with tobacco smoke exposure (B).

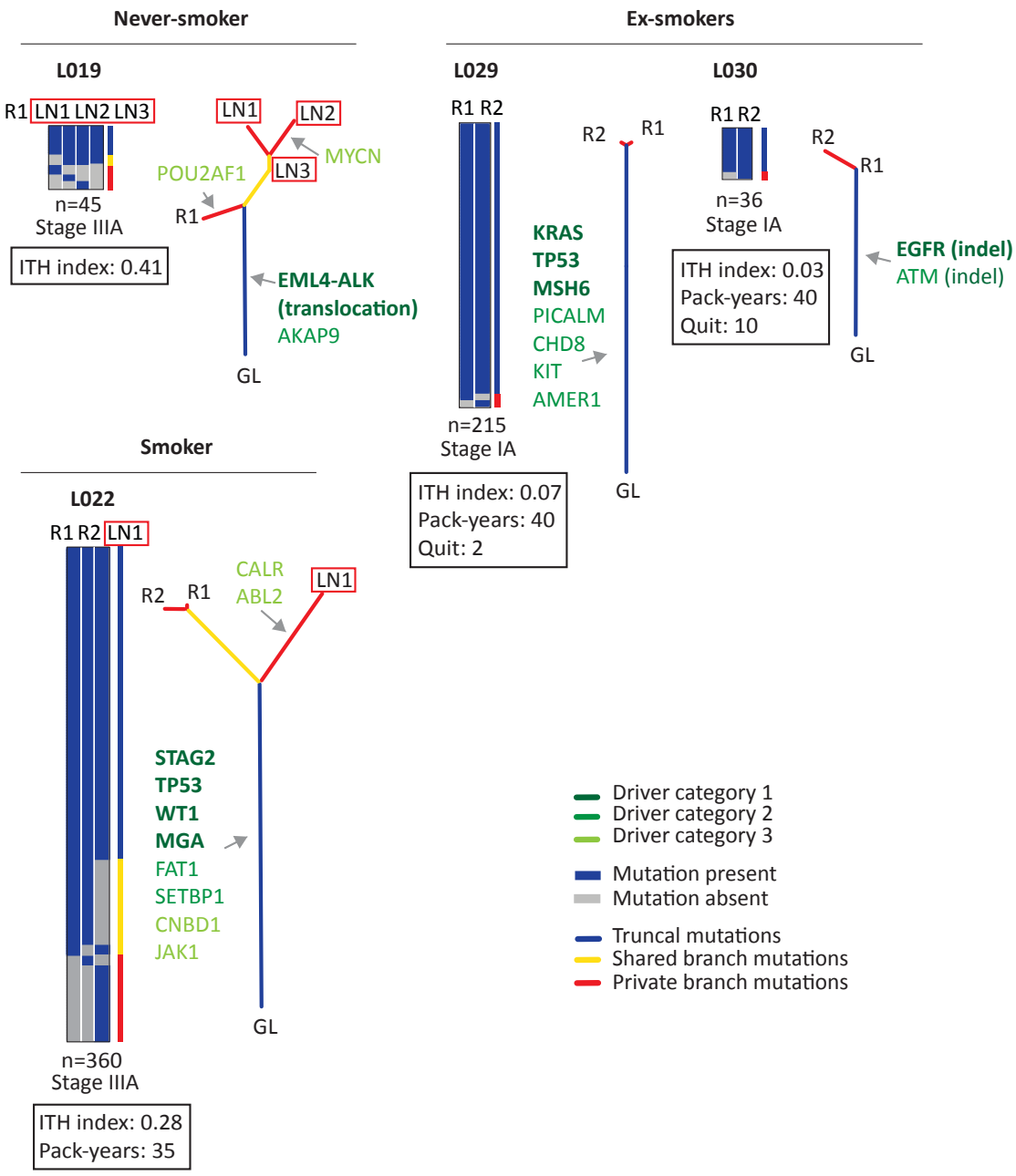


Figure 16 Heat maps and phylogenetic trees in LUADs from the NSCLC pilot cohort

Heat maps representing the regional distribution of mutations, where the presence of a mutation is indicated by blue and its absence by grey. Alongside each heat map is a vertical bar, where blue represents mutations that are present in all regions (truncal), yellow represents mutations that are present in some, but not all, regions (shared, branch), and red represents mutations that are present in one region only (private, branch). Underneath each heat map is the total number of non-synonymous mutations, the stage of the tumour, the intratumour heterogeneity index, the number of smoking pack-years, and how many years ago smoking was stopped (for ex-smokers). Phylogenetic trees annotated by genes representing predicted driver category 1 to 3 mutations, with arrows pointing towards the part of the tree on which they have been acquired. Ubiquitous mutations (present in all tumour regions) are shown on the blue trunks of trees, shared mutations (present in some, but not all, tumour regions) are shown on the yellow branches of trees, and private mutations (present in only one tumour region) are shown on the red branches of trees.

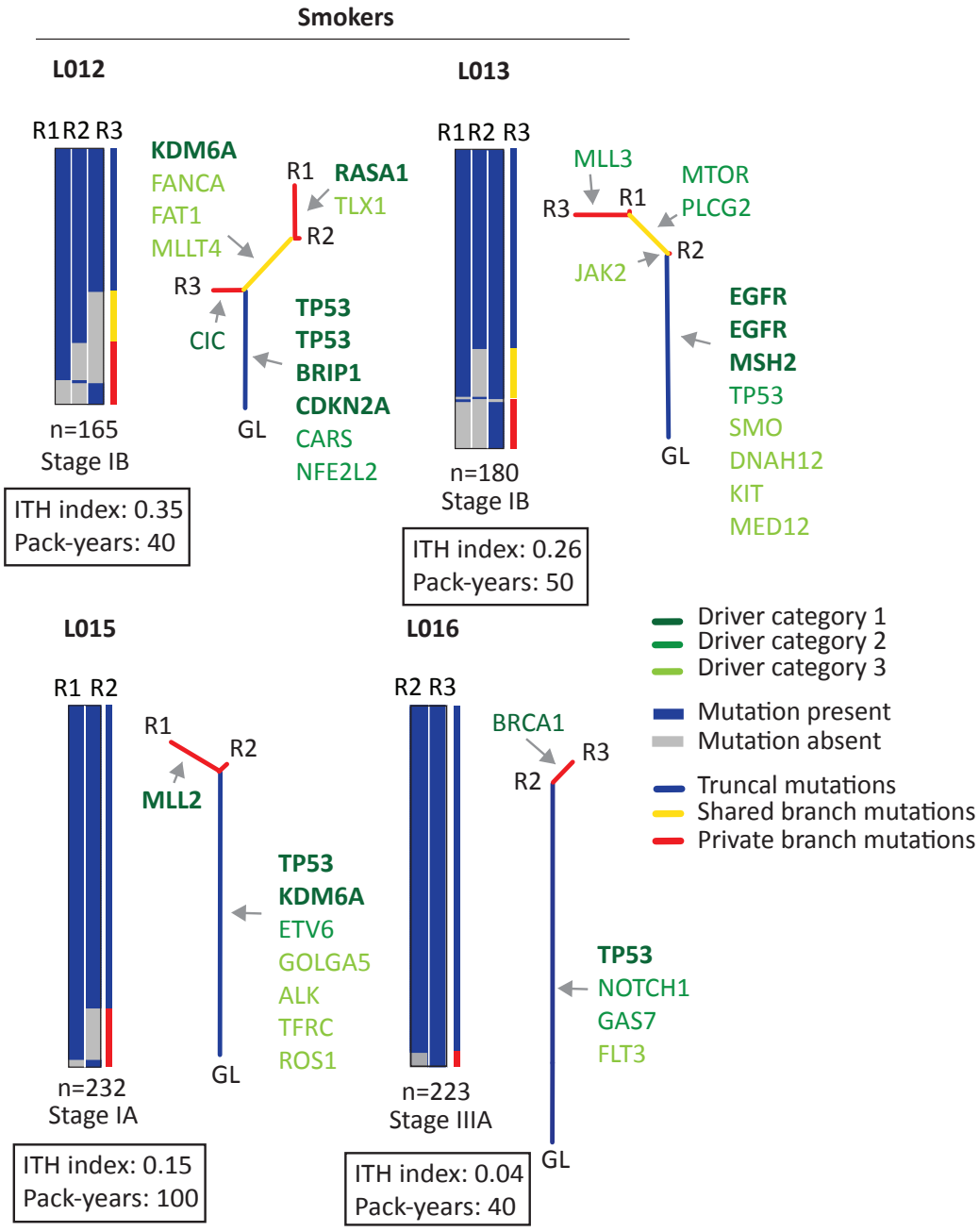


Figure 17 Heat maps and phylogenetic trees in LUSCs from the NSCLC pilot cohort

Heat maps representing the regional distribution of mutations, where the presence of a mutation is indicated by blue and its absence by grey. Alongside each heat map is a vertical bar, where blue represents mutations that are present in all regions (truncal), yellow represents mutations that are present in some, but not all, regions (shared, branch), and red represents mutations that are present in one region only (private, branch). Underneath each heat map is the total number of non-synonymous mutations, the stage of the tumour, the intratumour heterogeneity index, the number of smoking pack-years, and how many years ago smoking was stopped (for ex-smokers). Phylogenetic trees annotated by genes representing predicted driver category 1 to 3 mutations, with arrows pointing towards the part of the tree on which they have been acquired. Ubiquitous mutations (present in all tumour regions) are shown on the blue trunks of trees, shared mutations (present in some, but not all, tumour regions) are shown on the yellow branches of trees, and private mutations (present in only one tumour region) are shown on the red branches of trees.

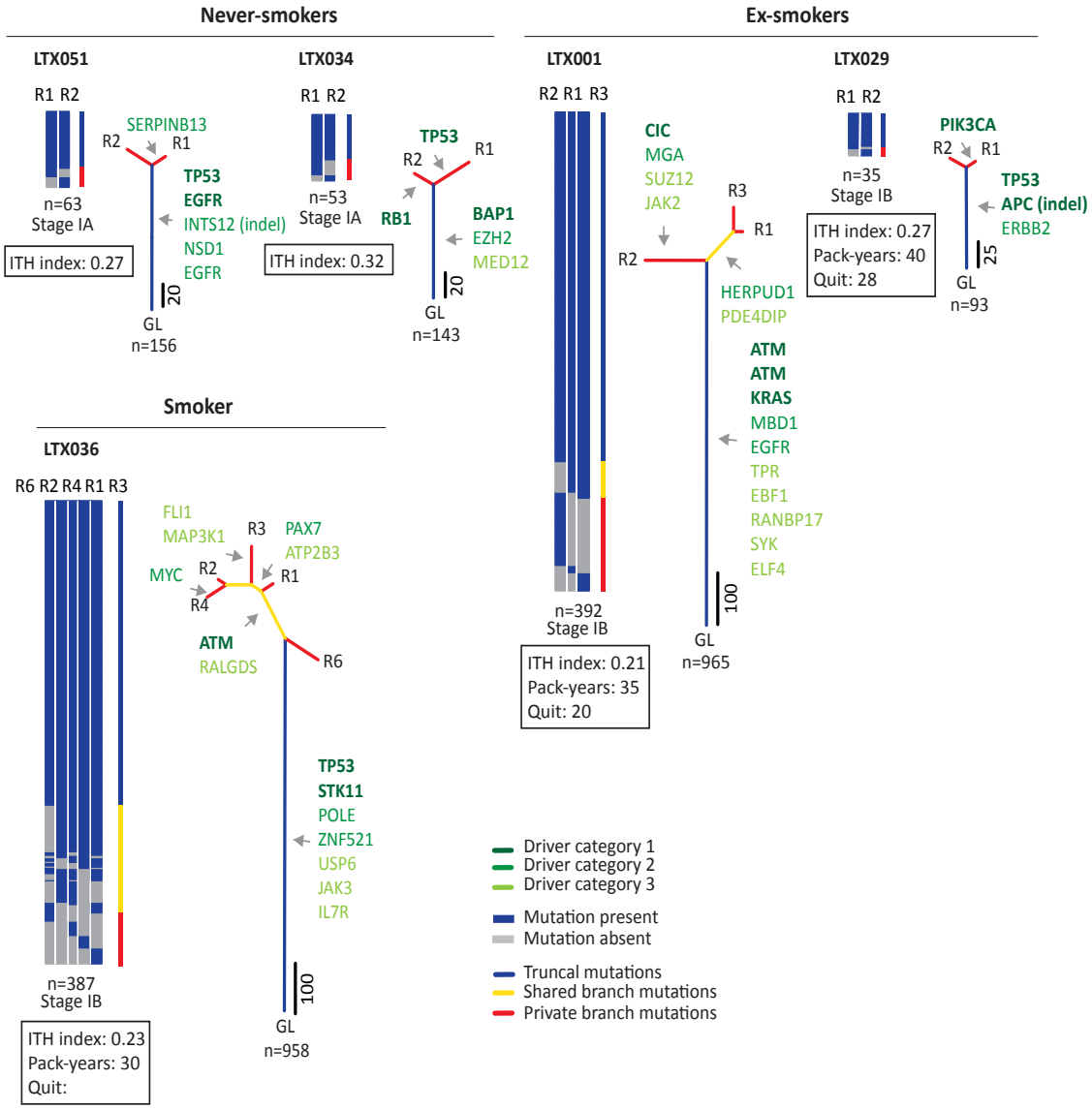
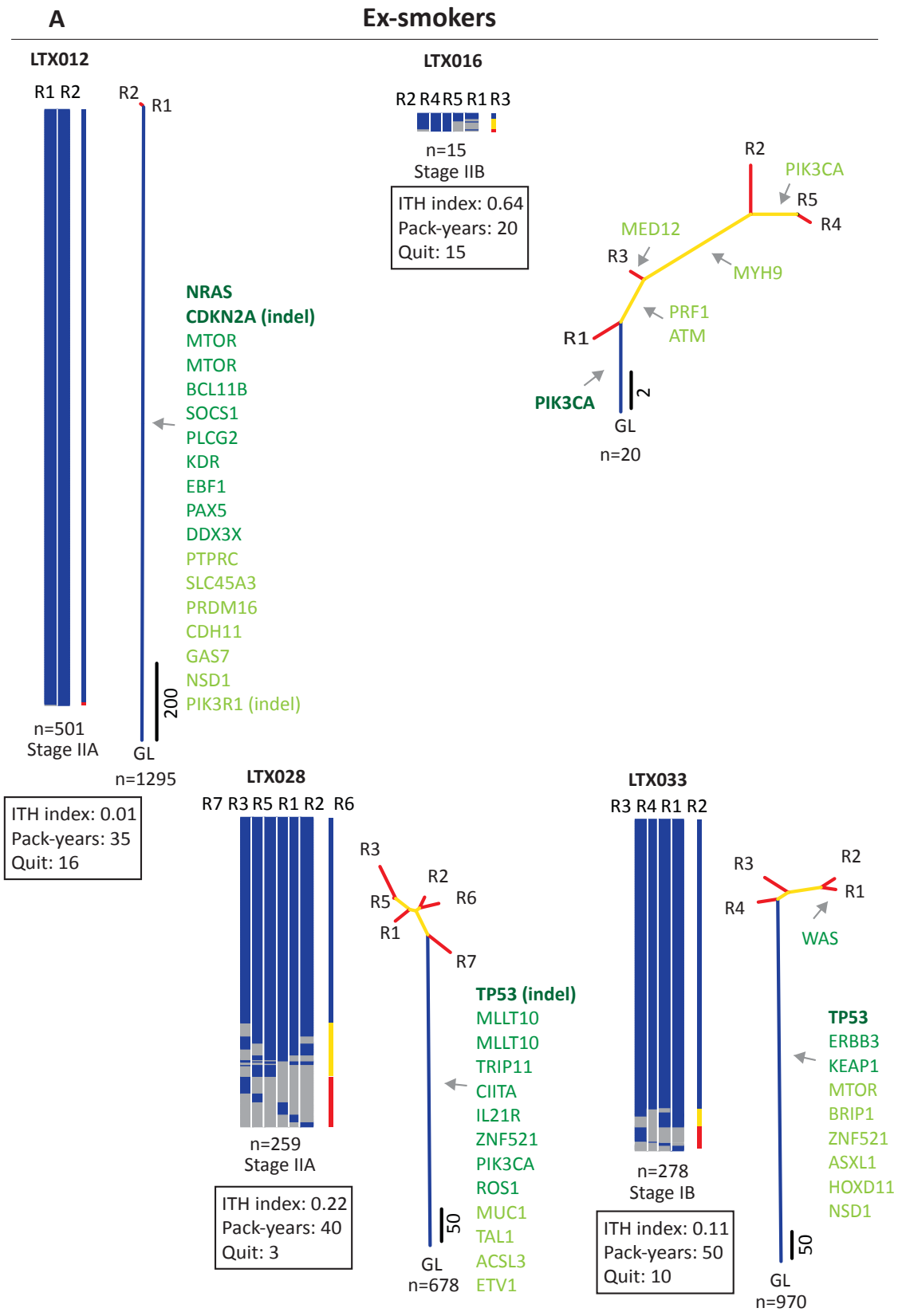
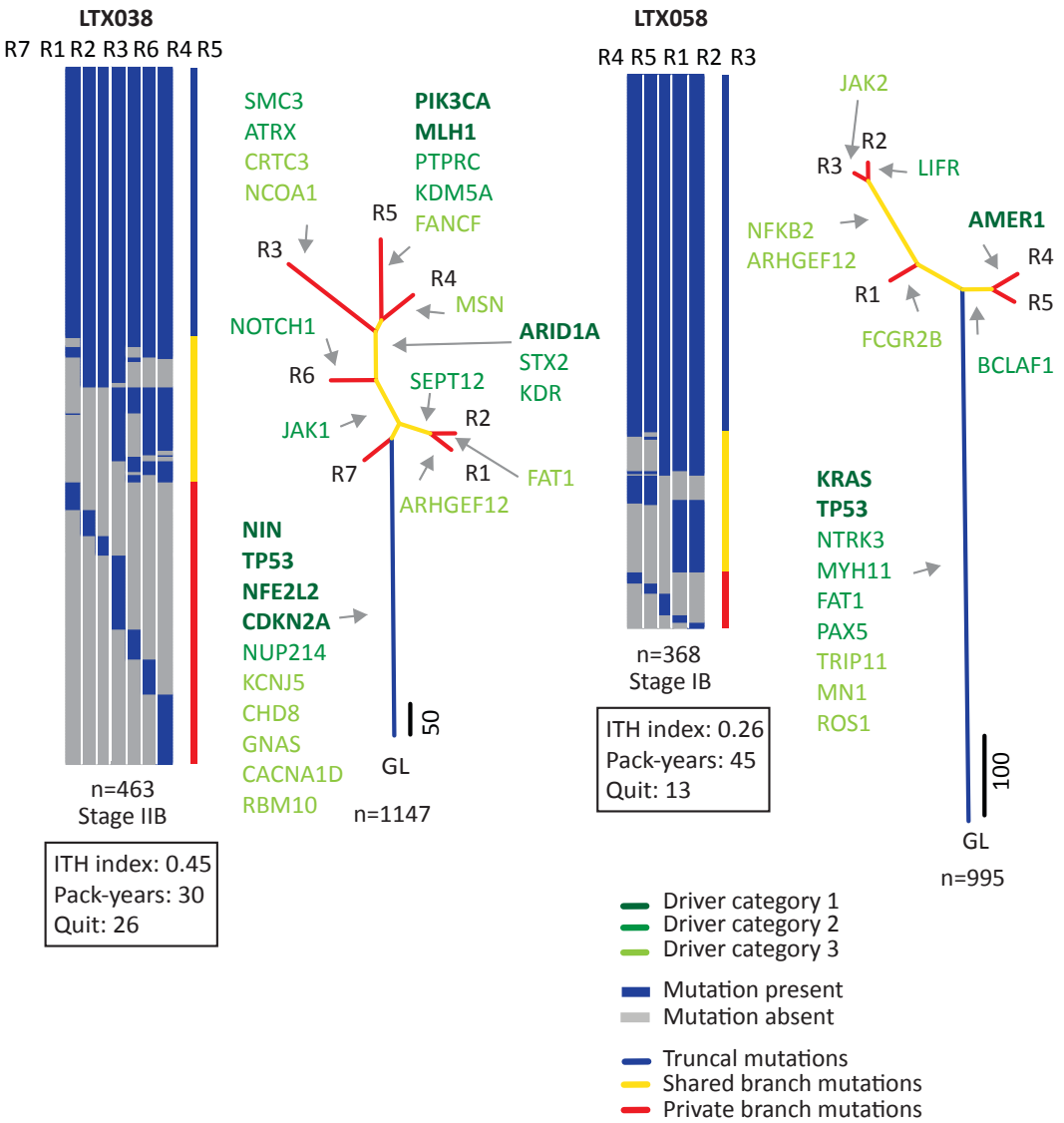


Figure 18 Heat maps and phylogenetic trees in LUADs from the TRACERx cohort

Heat maps representing the regional distribution of mutations, where the presence of a mutation is indicated by blue and its absence by grey. Alongside each heat map is a vertical bar, where blue represents mutations that are present in all regions (truncal), yellow represents mutations that are present in some, but not all, regions (shared, branch), and red represents mutations that are present in one region only (private, branch). Underneath each heat map is the total number of non-synonymous mutations, the stage of the tumour, the intratumour heterogeneity index, the number of smoking pack-years, and how many years ago smoking was stopped (for ex-smokers). Phylogenetic trees annotated by genes representing predicted driver category 1 to 3 mutations, with arrows pointing towards the part of the tree on which they have been acquired. Ubiquitous mutations (present in all tumour regions) are shown on the blue trunks of trees, shared mutations (present in some, but not all, tumour regions) are shown on the yellow branches of trees, and private mutations (present in only one tumour region) are shown on the red branches of trees. Underneath each tree is the total number of filtered non-synonymous and synonymous mutations.


A
Ex-smokers
LTX012
R1 R2
R2
R1
NRAS
CDKN2A (indel)
MTOR
MTOR
BCL11B
SOCS1
PLCG2
KDR
EBF1
PAX5
DDX3X
PTPRC
SLC45A3
PRDM16
CDH11
GAS7
NSD1
PIK3R1 (indel)
200
n=501
Stage IIA
GL
n=1295
ITH index: 0.01
Pack-years: 35
Quit: 16
LTX016
R2 R4 R5 R1 R3
n=15
Stage IIB
ITH index: 0.64
Pack-years: 20
Quit: 15
R2
PIK3CA
R5
R4
MED12
R3
MYH9
PRF1
ATM
R1
PIK3CA
2
GL
n=20
LTX028
R7 R3 R5 R1 R2 R6
n=259
Stage IIA
ITH index: 0.22
Pack-years: 40
Quit: 3
R3
R2
R5
R6
R1
R7
TP53 (indel)
MLLT10
MLLT10
TRIP11
CIITA
IL21R
ZNF521
PIK3CA
ROS1
MUC1
TAL1
ACSL3
ETV1
50
GL
n=678
LTX033
R3 R4 R1 R2
n=278
Stage IB
ITH index: 0.11
Pack-years: 50
Quit: 10
R3
R2
R4
R1
WAS
TP53
ERBB3
KEAP1
MTOR
BRIP1
ZNF521
ASXL1
HOXD11
NSD1
50
GL
n=970


LTX038
R7 R1 R2 R3 R6 R4 R5
n=463
Stage IIB
ITH index: 0.45
Pack-years: 30
Quit: 26
SMC3
ATRX
CRTC3
NCOA1
PIK3CA
MLH1
PTPRC
KDM5A
FANCF
R5
R3
R4
MSN
NOTCH1
ARID1A
STX2
KDR
R6
SEPT12
JAK1
R2
R7
R1
FAT1
ARHGEF12
NIN
TP53
NFE2L2
CDKN2A
NUP214
KCNJ5
CHD8
GNAS
CACNA1D
RBM10
50
GL
n=1147
LTX058
R4 R5 R1 R2 R3
n=368
Stage IB
ITH index: 0.26
Pack-years: 45
Quit: 13
JAK2
R2
R3
LIFR
NFKB2
ARHGEF12
AMER1
R4
R1
R5
FCGR2B
BCLAF1
KRAS
TP53
NTRK3
MYH11
FAT1
PAX5
TRIP11
MN1
ROS1
100
GL
n=995
Driver category 1
Driver category 2
Driver category 3
Mutation present
Mutation absent
Truncal mutations
Shared branch mutations
Private branch mutations

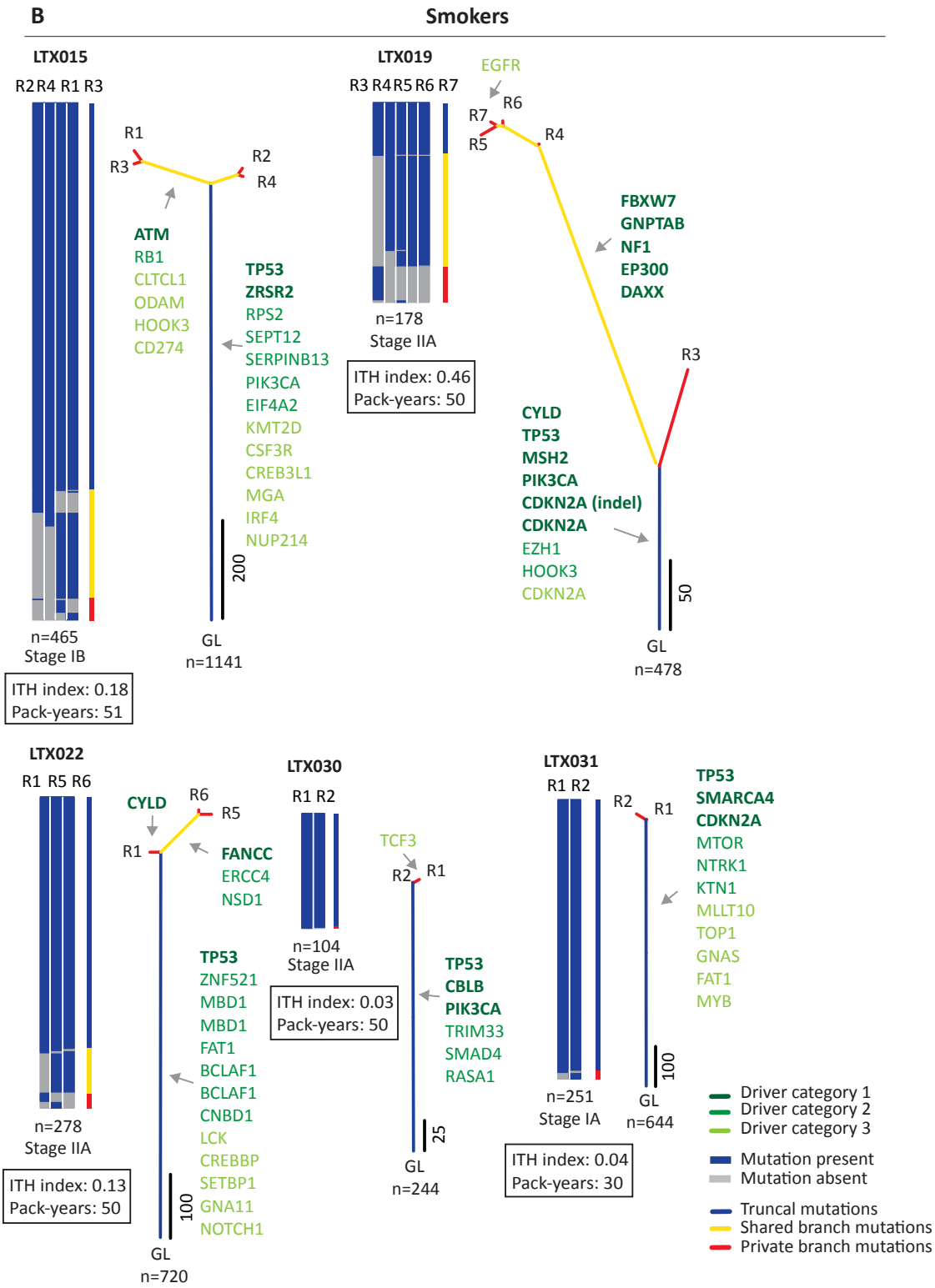

B
Smokers
LTX015
R2 R4 R1 R3
R1
R3
R2
R4
ATM
RB1
CLTCL1
ODAM
HOOK3
CD274
TP53
ZRSR2
RPS2
SEPT12
SERPINB13
PIK3CA
EIF4A2
KMT2D
CSF3R
CREB3L1
MGA
IRF4
NUP214
200
n=465
Stage IB
GL
n=1141
ITH index: 0.18
Pack-years: 51
LTX019
R3 R4 R5 R6 R7
EGFR
R6
R7
R5
R4
FBXW7
GNPTAB
NF1
EP300
DAXX
R3
CYLD
TP53
MSH2
PIK3CA
CDKN2A (indel)
CDKN2A
EZH1
HOOK3
CDKN2A
50
n=178
Stage IIA
GL
n=478
ITH index: 0.46
Pack-years: 50
LTX022
R1 R5 R6
CYLD
R6
R5
R1
FANCC
ERCC4
NSD1
TP53
ZNF521
MBD1
MBD1
FAT1
BCLAF1
BCLAF1
CNBD1
LCK
CREBBP
SETBP1
GNA11
NOTCH1
100
n=278
Stage IIA
GL
n=720
ITH index: 0.13
Pack-years: 50
LTX030
R1 R2
n=104
Stage IIA
ITH index: 0.03
Pack-years: 50
TCF3
R2
R1
TP53
CBLB
PIK3CA
TRIM33
SMAD4
RASA1
25
GL
n=244
LTX031
R1 R2
R2
R1
TP53
SMARCA4
CDKN2A
MTOR
NTRK1
KTN1
MLLT10
TOP1
GNAS
FAT1
MYB
100
n=251
Stage IA
GL
n=644
ITH index: 0.04
Pack-years: 30
Driver category 1
Driver category 2
Driver category 3
Mutation present
Mutation absent
Truncal mutations
Shared branch mutations
Private branch mutations

Figure 19 Heat maps and phylogenetic trees in LUSCs from the TRACERx cohort

Heat maps representing the regional distribution of mutations, where the presence of a mutation is indicated by blue and its absence by grey, in ex-smokers (A) and current smokers (B). Alongside each heat map is a vertical bar, where blue represents mutations that are present in all regions (truncal), yellow represents mutations that are present in some, but not all, regions (shared, branch), and red represents mutations that are present in one region only (private, branch). Underneath each heat map is the total number of non-synonymous mutations, the stage of the tumour, the intratumour heterogeneity index, the number of smoking pack-years, and how many years ago smoking was stopped (for ex-smokers). Phylogenetic trees annotated by genes representing predicted driver category 1 to 3 mutations, with arrows pointing towards the part of the tree on which they have been acquired. Ubiquitous mutations (present in all tumour regions) are shown on the blue trunks of trees, shared mutations (present in some, but not all, tumour regions) are shown on the yellow branches of trees, and private mutations (present in only one tumour region) are shown on the red branches of trees. Underneath each tree is the total number of filtered non-synonymous and synonymous mutations.

3.3 Predicted driver mutations

All identified non-silent mutations were compared against a list of 582 potential driver genes. The driver gene list consisted of all genes identified in the Catalogue of Somatic Mutations in Cancer (COSMIC) cancer gene census (downloaded June 2014) (Futreal et al. 2004), and those identified in a large scale pan-cancer analysis (using $q < 0.05$ as a cut-off for statistical significance) (Lawrence et al. 2013), and previous NSCLC sequencing studies (Lawrence et al. 2014; Cancer Genome Atlas Research Network, 2015). Any variants that were located within one of these genes underwent categorisation based on pre-set criteria, as previously described (Section 2.9). In summary, driver categories 1 to 3 were defined as:

- Category 1 - high confidence driver mutations containing all inactivating mutations in tumour suppressor genes (TSGs) or activating mutations in oncogenes (OGs)
- Category 2 - putative driver mutations containing all mutations in driver genes located up to 15bp away from other mutations present in COSMIC or mutations meeting category 1 criteria but not annotated as a TSG or OG
- Category 3 - low confidence driver mutations containing all other non-silent mutations in genes that were present in the lists of cancer-related genes

The phylogenetic trees in **Figure 16** to **Figure 19** are annotated with genes representing category 1 to 3 predicted driver mutations, demonstrating both spatial and temporal heterogeneity of driver mutations, with driver mutations occurring both early (present on the trunks of the trees) and late (present on the branches of the trees) in tumour evolution. Category 1 to 3 predicted driver mutations for each tumour are listed in Section 8.4.

The exome capture panel used in TRACERx has been adapted to detect common translocation events as part of the bioinformatics pipeline, but when the NSCLC pilot tumours were sequenced, detection of such events relied on histopathological assessment alone. The EML4-ALK translocation, a known driver event in NSCLC, was identified in patient L019 from the NSCLC pilot cohort by the UCL Advanced Diagnostics department using a dual colour FISH assay with a fusion probe for

ALK/EML4 t(2;2); inv(2) (Kreatech™ FISH probe, Leica Biosystems) (**Figure 21**). In collaboration with Illumina, the EML4-ALK translocation was identified in the tumour region R1 and all lymph nodes (LN1-LN3) based on PCR (**Figure 22**) and subsequent MiSeq sequencing, which showed that the number of mutant reads for the translocation was significantly higher in the tumour and lymph node regions compared to the germ line and control DNA (see Section 5.3 and **Table 21**). These data indicated that the translocation was likely to be present in all regions, and was therefore considered to be an early truncal driver event.

Across the cohort, there were 366 category 1 to 3 predicted driver mutations; 253/366 (69%) of these were truncal, and 113/366 (31%) of these were branch mutations (**Figure 20**). Amongst the branch driver mutations, 24/27 (89%) patients had a category 1 to 3 mutation, and 15/27 (56%) patients had at least one category 1 mutation. Overall, there were 93 category 1 driver mutations, which contained all inactivating mutations in TSGs or activating mutations in OGs, 69/93 (74%) of these were truncal (predicted to be clonal in origin) mutations, and 24/93 (26%) of these were branch (predicted to be subclonal in origin) mutations. Amongst all heterogeneous/branch category 1 to 3 driver mutations, at least 2 or more driver mutations occurred in the following recurrent genes: *ARHGEF12* (rho guanine nucleotide exchange factor 12, shown to have oncogenic properties in haematological malignancies), *ARID1A* (AT rich interactive domain 1A, a known TSG), *ATM* (ATM serine/threonine kinase, a known TSG), *BCLAF1* (BCL2-associated transcription factor 1, activating mutations found in cancers, in particular melanoma), *CIC* (capicua transcriptional repressor, a known TSG), F*ANCA* (Fanconi anemia, complementation group A, shown to have TSG properties in haematological malignancies), *HIST1H3B* (histone cluster 1, H3b, shown to have oncogenic properties in gliomas), *JAK2* (janus kinase 2, shown to have OG properties in haematological malignancies), *PIK3CA* (activating mutations found in several cancers, including lung), and *RB1* (a known TSG).

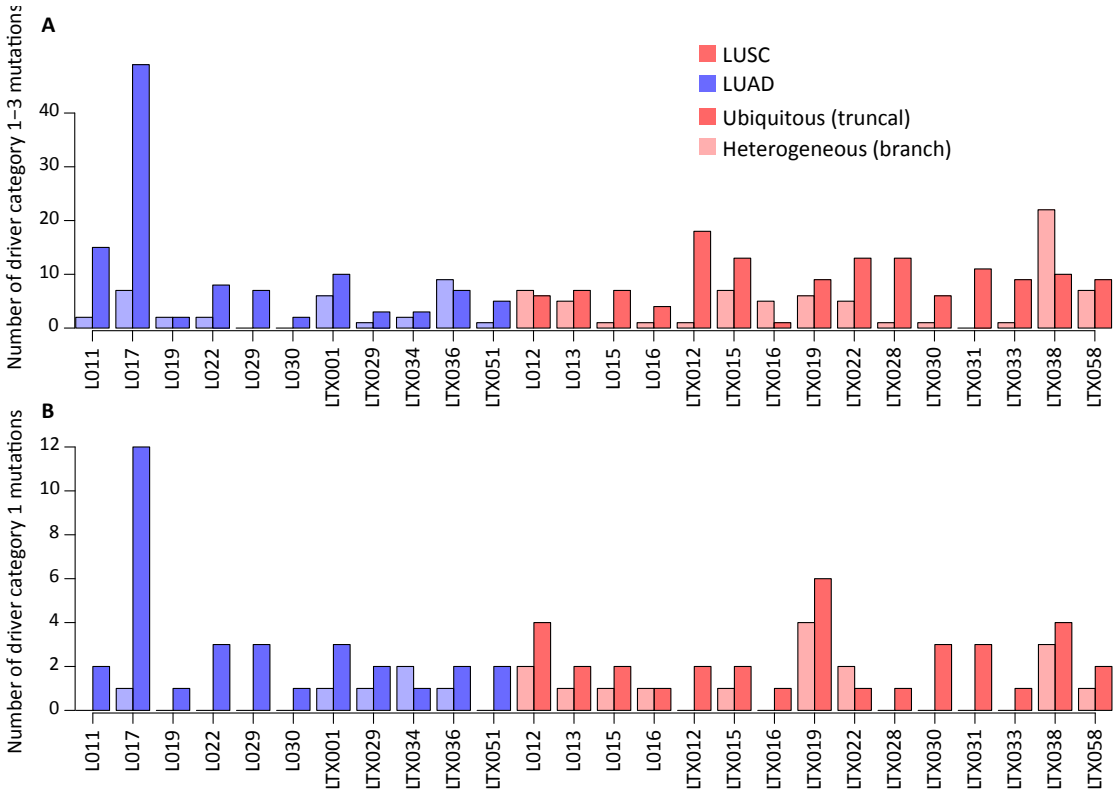


Figure 20 Truncal and branch predicted driver mutations
The number of category 1 to 3 predicted driver mutations (A), and the number of category 1 predicted driver mutations in the NSCLC pilot and TRACERx cohort (B).

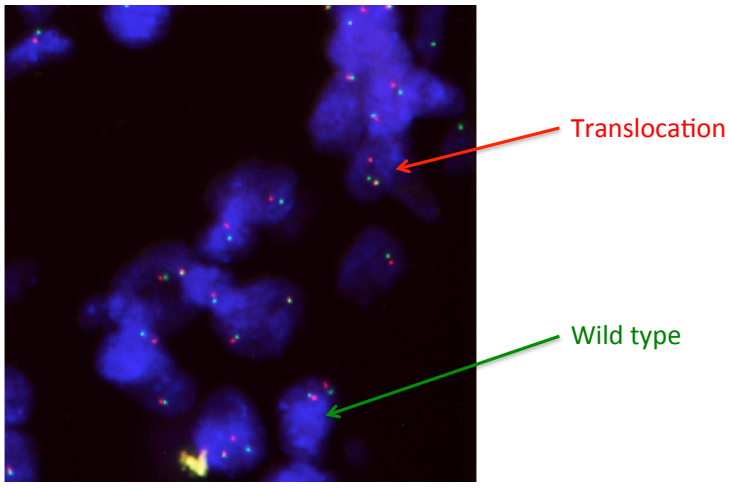


Figure 21 EML4-ALK FISH in L019

Dual colour FISH assay with a fusion probe for ALK/EML4 t(2;2); inv(2) (Kreatech™ FISH probe, Leica Biosystems) using a representative tumour sample from L019. Cells containing two sets of overlapping green and red signals represent wild type cells (green arrow), and cells containing one overlapping green and red signal and one separated green and red signal represent cells harbouring the translocation (red arrow). Image produced by UCL Advanced Diagnostics.

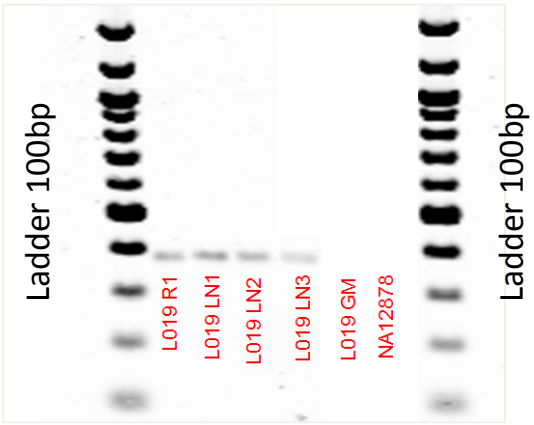


Figure 22 PCR amplification for EML4-ALK in L019

PCR amplification of DNA extracted from tumour region R1 and lymph nodes LN1 to LN3 in patient L019 showing the presence of PCR products in tumour regions R1 and lymph node regions LN1-LN3, but not in the germ line (GL) or control DNA (NA12878). PCR performed by Clare fielding, Illumina.

3.3.1 Category 1 and 2 driver mutations in genes relevant to NSCLC

Category 1 and 2 mutations were predicted to be either high confidence or putative cancer driver mutations. Several of these driver mutations had either existing or potential value in guiding NSCLC patient management. Several truncal category 1 mutations were seen in *KRAS*. In patient L023 there was a *KRAS* G12C mutation (42% frequency in *KRAS*-mutated NSCLC, (Forbes et al. 2011)), in patient L029 there was a *KRAS* G12A mutation (7% frequency in *KRAS*-mutated NSCLC, (Forbes et al. 2011)), in patients L017 and LTX001 there was a *KRAS* G12V mutation (20% frequency in *KRAS*-mutated NSCLC, (Forbes et al. 2011)), and in patient LTX058 there was a *KRAS* G12S mutation (5% frequency in *KRAS*-mutated NSCLC, (Forbes et al. 2011)). *KRAS* mutations occur in approximately 15-25% of LUAD tumours, and are uncommon in LUSC tumours (Brose et al. 2002; Riely et al. 2008). Over 90% of all *KRAS* mutations are associated with changes in the glycine residue coded by codon 12 of wild type *KRAS*. In NSCLC, G12C accounts for most of these mutations, followed by G12V and G12D (Forbes et al. 2011; Garassino et al. 2011). This is likely to be related to the association between NSCLC and tobacco smoke exposure, since G>C or G>T transversions of the guanine nucleotide residues located in the wild type *KRAS* codon 12 are known to be associated with smoking (Ahrendt et al. 2001), and are uncommon in NSCLCs amongst non-smokers (Riely et al. 2008). Although the prognostic

significance of *KRAS* mutations in NSCLC is yet to be determined (Mascaux et al. 2005; Schiller et al. 2001), they have been associated with decreased response to adjuvant chemotherapy in early-stage NSCLC (Tsao et al. 2007; Winton et al. 2005), and EGFR TKIs, such as erlotinib and gefitinib (Pao et al. 2005; Mitsudomi & Yatabe 2007). *KRAS*-mutated NSCLCs may benefit from the MEK (mitogen-activated protein kinase kinase) inhibitors, such as selumetinib, in combination with chemotherapy (Jänne et al. 2013).

Truncal category 1 driver mutations in *EGFR* were identified. In patient L013 there were two mutations identified in the kinase domain of *EGFR*. There was an exon 18 G719A mutation (0.6% frequency, (Forbes et al. 2011)), which is known to confer increased sensitivity to EGFR TKIs (Han et al. 2005; Lynch et al. 2004; Rosell et al. 2005; Taron et al. 2005). In addition to this, there was an exon 19 D761Y mutation, which was the first non-*T790M* secondary mutation identified to be associated with resistance to EGFR TKIs. This mutation was originally described in a patient with a known TKI sensitising *EGFR* mutation (L858R) with metastatic resistant disease (Balak et al. 2006). Although the frequency of non-*T790M* mutations in *EGFR*-mutated TKI-resistant tumours is below 5% (Kosaka et al. 2006; Balak et al. 2006; Bean et al. 2008), detecting this mutation in patients may have clinical implications, for example such patients may be monitored closer for the development of resistant disease whilst on EGFR TKI therapy. Three regions were deep-sequenced in this tumour; the VAFs for the G719A mutation were 21%, 25% and 55%, and the VAFs for the D761Y mutation were 20%, 17% and 48% in the respective tumour regions. With similar VAFs for each mutation in each tumour region, it is likely that these mutations were present in all the tumour cells for each region. The fact that one mutation conferred sensitivity and another conferred resistance, suggested that this tumour would have had intrinsic resistance to EGFR TKI therapy and that such drug resistance may have manifested early during treatment. This example illustrates the likely clonal nature of intrinsic drug resistance mechanisms (Swanton et al. 2014), which has previously been shown in other tumours, such a melanoma (Turajlic et al. 2014). In patient L030 there was an exon 19 deletion (48% frequency in *EGFR*-mutated NSCLC, (Mitsudomi & Yatabe 2010)), and in patient LTX051 there was an exon 21 L858R mutation (43% frequency, (Mitsudomi & Yatabe 2010)). Both of these mutations are known to confer increased sensitivity to EGFR TKIs. Truncal driver category 2 mutations in *EGFR* were also

identified in patients LTX051 (A86T), and LTX001 (G630E). These mutations were predicted to be putative driver mutations.

EGFR mutations occur within exons 18-21, and are usually heterozygous, with evidence of amplification in the mutant allele (Soh et al. 2009). Approximately 90% of these mutations are exon 19 deletions or exon 21 L858R point mutations (Ladanyi & Pao 2008). *EGFR* mutations are mostly found in female never-smokers with LUAD histology (Lynch et al. 2004; Paez et al. 2004; Pao et al. 2004). All the above patients were female, patient L030 was an ex-smoker (40 pack-years, quit 10-years ago) with LUAD histology, patient L013 was a smoker (50 pack-years) with LUSC histology, and patient LTX051 was a never-smoker with LUAD histology.

Several category 1 driver mutations were identified in the *PIK3CA* gene. The E545K mutation (26.7% frequency in *PIK3CA*-mutated NSCLC, (Forbes et al. 2011)) was found in patients LTX030 (truncal) and LTX038 (branch), and the *PIK3CA* E542K mutation (8.9% frequency in *PIK3CA*-mutated NSCLC, (Forbes et al. 2011)) was found in patients LTX019 (truncal) and LTX029 (branch). The E542K and E545 mutations occur within the highly conserved helical domain of the *PIK3CA* gene, and are frequently observed as PI3-kinase mutations (Kang et al. 2005). These mutations have been shown to increase the catalytic activity of PIK3CA proteins resulting in oncological transformation in *in vitro* studies (Kang et al. 2005). Patient LTX016 had a truncal *PIK3CA* V344G mutation, which has previously been described in glioblastoma multiforme tumours (Gallia et al. 2006). Truncal driver category 2 mutations in *PIK3CA* were also identified in patients L023 (A533V), LTX015 (C147R) and LTX028 (T727K). These mutations were predicted to be putative driver mutations.

Somatic *PIK3CA* mutations have been found in 1-3% of all NSCLC patients (Samuels et al. 2004; Kawano et al. 2006), and are more common in the LUSC compared to LUAD histological subtype (Kawano et al. 2006). The clinical characteristics of patients with *PIK3CA*-mutated NSCLC tumours are yet to be determined. However, NSCLC cell lines with activating *PIK3CA* mutations have been shown to be sensitive to dual PIK3CA/mTOR inhibitors (Zou et al. 2009), and the introduction of an activating *PIK3CA* mutation in *EGFR*-mutant lung cancer cell lines is associated with EGFR TKI resistance (Engelman et al. 2006). In addition, *PIK3CA* mutations have been found in *EGFR*-mutant lung cancers with acquired resistance to EGFR TKI therapy (Sequist et

al. 2011). In a cohort study of 79 patients with LUSC, Paik and colleagues performed WES on paired primary and metastatic tumours and identified *PIK3CA*-mutant tumours as an aggressive disease subset associated with brain metastases and poor outcome (Paik et al. 2015).

In patient LTX012 there was a truncal category 1 *NRAS* Q61K driver mutation (10-25% frequency, (Ohashi et al. 2013), (Forbes et al. 2011)). This mutation has been shown to drive aberrant pathway signalling, tumour cell survival, and tumour growth in melanoma *in vivo* studies (Li et al. 2012). Mutations in *NRAS* have been identified in approximately 1% of all NSCLCs (Brose et al. 2002; Ohashi et al. 2013), and are more commonly found in patients with LUADs and a history of smoking (Ohashi et al. 2013). In *NRAS*-mutant lung cancer cell lines, these mutations have been associated with sensitivity to MEK inhibitors, such as selumetinib and trametinib (Ohashi et al. 2013). Patient LTX012 had LUSC histology and was an ex-smoker.

3.3.2 The significance of predicted driver mutations in branches

Tumours with shared or private branch mutations in genes with diagnostic, therapeutic or prognostic relevance were identified using the TARGET (tumour alterations relevant for genomics-driven therapy) database (https://www.broadinstitute.org/cancer/cga/target, version 3, updated February 2015) as shown in **Table 9**. Many of these genes either predict sensitivity or resistance to a class of targeted drug therapies, and their detection at diagnosis, or at relapse, may have significant implications in terms of treatment stratification, and therefore patient outcome. Given that these subclonal mutations were heterogeneous as a result of branched tumour evolution, that is they were present in some parts of the tumour and not others, identification of these mutations in clinically relevant genes may have been restricted if only one tumour region had been sampled and sequenced. This is analogous to the potential sampling bias with a single biopsy, and further adds to the challenges posed by intratumour heterogeneity and the difficulties in adequately profiling a tumour with therapeutic intervention in mind.

Patient	Gene	Driver category	Clinical relevance
LTX015	ATM	1	Biallelic inactivation may predict sensitivity to PARP inhibitors
LTX016	ATM	3	
LTX036	ATM	1	
LTX036	BCL2	3	Translocation is prognostic/diagnostic in some haematological cancers and activating alterations may be sensitive to BCL2 inhibitors
L016	BRCA1	1	Biallelic inactivation predicts sensitivity to PARP inhibitors
LTX019	EGFR	3	Mutations may predict sensitivity to TKIs, and other mutations predict resistance to TKIs
LTX019	FBXW7	1	Mutations may predict sensitivity to MTOR inhibitors, and resistance to anti-tubulin chemotherapy
L023	GNAS	3	Mutations may be diagnostic of tumours arising in McCune-Albright syndrome, and may be targetable with JAK inhibitors
L013	JAK2	3	Mutations may predict sensitivity to JAK2 inhibitors
LTX001	JAK2	3	
LTX058	JAK2	3	
LTX038	KDR	2	Activating mutations may predict sensitivity to specific KDR inhibitors
L023	MAP2K1	3	Mutations may predict resistance to RAF and MEK inhibitors
LTX016	MED12	3	May predict resistance to targeted therapies
LTX038	MLH1	1	Prognostic/Diagnostic in colon cancer
LTX015	MPL	3	Mutations may predict sensitivity to JAK2 inhibitors
L013	MSH2	1	Prognostic/Diagnostic in colon cancer
L013	MTOR	2	Activating mutations may predict sensitivity to MTOR inhibitors, and secondary mutations may predict resistance to MTOR inhibitors
LTX036	MYC	2	May be prognostic/diagnostic in some cancer types
LTX019	NF1	1	Biallelic inactivation may predict sensitivity to PI3K pathway inhibitors (PI3K/AKT/MTOR). Biallelic inactivation may predict resistance to RAF and MEK inhibitors.
LTX038	NOTCH1	2	Activating mutations may predict sensitivity to Notch inhibitors

LTX016	PIK3CA	3	Mutations may predict sensitivity to PI3K pathway inhibitors (PI3K/(AKT/PKB)/MTOR). Mutations may predict resistance to anti-RTK therapy, including cetuximab, anti-EGFR TKIs, and trastuzumab and lapatinib
LTX029	PIK3CA	1	
LTX038	PIK3CA	1	
LTX015	RB1	2	May be prognostic or diagnostic in some tumour types. Loss may predict resistance to CDK inhibitors
LTX034	RB1	1	
L023	TET2	3	Prognostic in myelodysplastic syndrome
LTX034	TP53	1	Biallelic inactivation or mutation may be prognostic in some tumour types. May predict sensitivity to some cell cycle inhibitors and p53 specific gene therapies or immunotherapies.

Table 9 Clinically relevant genes from the TARGET database in branches

Genes with driver category 1 to 3 branch mutations in the NSCLC pilot and TRACERx cohort were screened against the TARGET database to identify clinically relevant genes. Abbreviations: PARP, poly ADP ribose polymerase; TKI, tyrosine kinase inhibitor; RTK, receptor tyrosine kinase; MEK, mitogen-activated protein kinase; JAK, jannus kinase; AKT/PKB, protein kinase-B; PI3K, phosphoinositide 3-kinase; and MTOR, mammalian target of rapamycin.

3.4 Early versus late mutation signatures in the TRACERx cohort

With greater sequencing depth and an adequate number of filtered non-synonymous and synonymous mutations, the mutational spectra and associated signature patterns for tobacco smoke exposure and APOBEC enrichment were explored in the TRACERx cohort only. A two-sided Fisher's exact test was used to compare the relative frequency of each mutation type between early and late mutations. In the LUAD histological subtype, 2/5 of the tumours (LTX001 and LTX036) had a statistically significant shift in the spectra of the six different mutation types (C>A, C>G, C>T, T>A, T>C, T>G), when comparing early with late mutations ($p < 0.05$, **Figure 23**A). In 3/5 of the tumours (LTX001, LTX034 and LTX036), there was a statistically significant difference in the proportion of C>A transversions in early compared with late mutations ($p < 0.05$). In tumours LTX001 and LTX036, there was a decrease in the proportion of C>A transversions in the late mutations, whereas in tumour LTX034 an increase was seen. However, patient LTX001 and LTX036 were ex- and current smokers, whereas patient LTX034 was a never-smoker in whom a smoking-associated mutation signature would not have been expected. There was no history of passive tobacco smoke or

other carcinogen exposure in this patient. There was a statistically significant increase in the proportion of C>T transitions and C>G transversions at TpC sites in late compared with early mutations, indicative of APOBEC cytidine deaminase activity (Roberts et al. 2013; Burns et al. 2013; Alexandrov, Nik-Zainal, Wedge, Aparicio, et al. 2013; de Bruin et al. 2014) in tumours LTX001 ($p < 0.01$) and LTX036 ($p < 0.01$) (**Figure 24**).

In the LUSC histological subtype, 7/11 of the tumours (LTX012, LTX015, LTX019, LTX028, LTX033, LTX038 and LTX058) had a statistically significant shift in the spectra of the six mutation types (C>A, C>G, C>T, T>A, T>C, T>G), when comparing early with late mutations ($p < 0.05$, **Figure 23**B). In 6/11 of the tumours (LTX012, LTX015, LTX019, LTX028, LTX038 and LTX058), all of which came from either ex- or current smokers, there was a statistically significant decrease in the proportion of C>A transversions in the late compared with early mutations ($p < 0.05$). There was a statistically significant increase in the APOBEC-associated mutation signature in tumours LTX019 ($p < 0.001$), LTX028 ($p < 0.05$), and LTX038 ($p < 0.05$) (**Figure 25**). The 95% confidence intervals for the APOBEC enrichment odds ratio was particularly wide in tumours LTX012 and LTX016 due to the small number of late mutations in LTX012, and the few mutations detected overall in LTX016.

Overall, there was a statistically significant shift in the mutational spectra of the six different mutation types in 9/16 (56%) tumours, a decrease in the smoking-associated signature in the late mutations in 8/16 (50%) tumours, and an increase in the APOBEC-associated mutation signature in the late mutations in 5/16 (31%) tumours. Since C>A transversions are known to be associated with the mutagenic effects of tobacco smoke (Pfeifer & Hainaut 2003), a decrease in their proportion in the late mutations implies a decrease in the overall mutational burden attributable to smoking in late tumour evolution. This suggests that the mutagenic effects of tobacco smoke may play a more prominent role in the early tumourigenesis of some NSCLCs. This association was not seen in all the tumours from patients who were either current or ex-smokers. In the case of tumours from patients who were ex-smokers, the time since smoking cessation took place did not appear to influence the presence or absence of a smoking-associated signature. A significant increase in APOBEC enrichment in the late mutations suggests a potential role for APOBEC cytidine deaminase activity in the late tumourigenesis of some NSCLCs (de Bruin et al. 2014; McGranahan & Swanton 2015).

Determining the temporal significance of such signatures in tumours, allows one to interpret the relevance of specific mutational processes and pathways that may be involved in the initiation and progression of tumours. In addition, it may identify potential therapeutic strategies aimed at, for example, limiting genetic diversity.

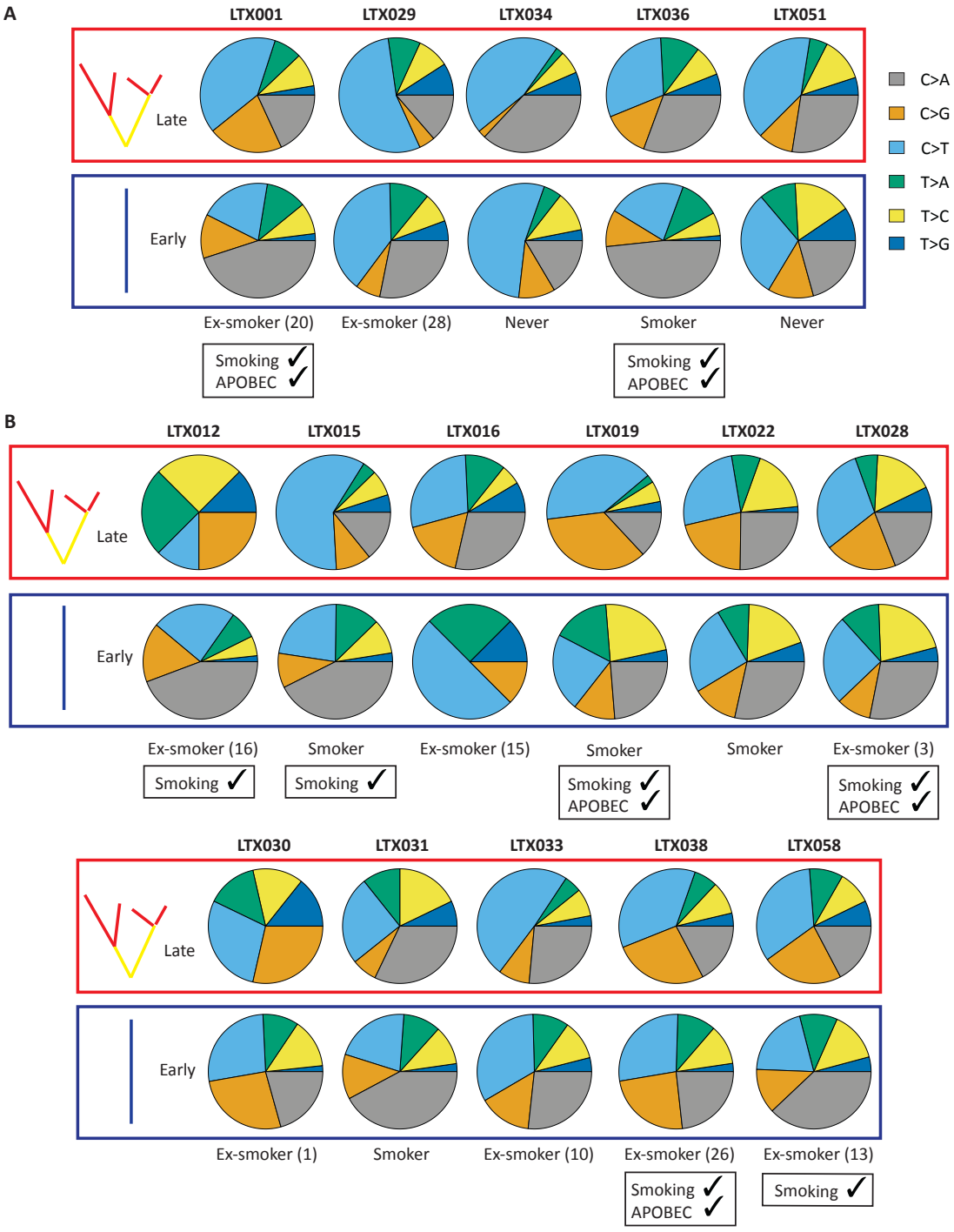


Figure 23 Mutational spectra in the TRACERx cohort

Proportion of early mutations (present on the trunk) and late mutations (present on the branches) accounted for by each of the six mutation types (C>A, C>G, C>T, T>A, T>C and T>G) in the LUADs (A) and LUSCs (B). Tumours for which there was a statistically significant difference between late compared with early mutations for signatures associated with smoking (decrease in C>A transversions in late mutations) and APOBEC deaminase activity (increase in C>T and C>G transversions at TpC sites in late mutations) are highlighted. In the case of ex-smokers, the number of years since smoking cessation is shown in brackets.

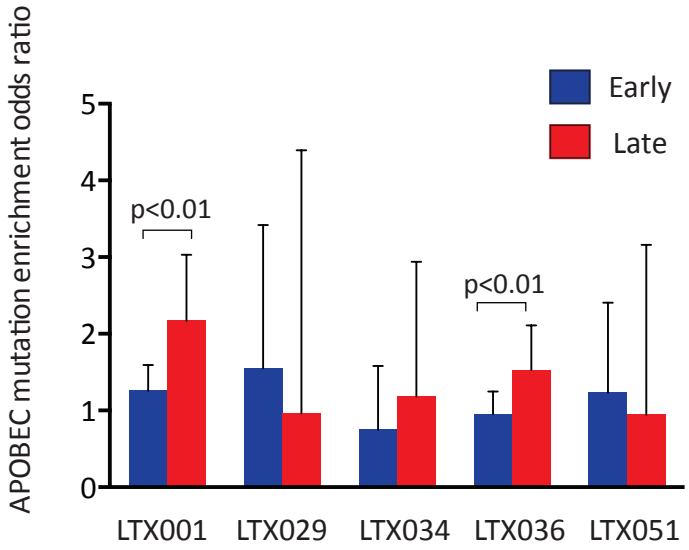


Figure 24 APOBEC mutation enrichment in LUADs

APOBEC mutation enrichment odds ratio for early (blue) versus late (red) mutations in LUADs. Statistically significant APOBEC enrichment in late compared with early mutations was seen in tumours LTX001 and LTX036 ($p < 0.01$). The 95% confidence intervals are indicated.

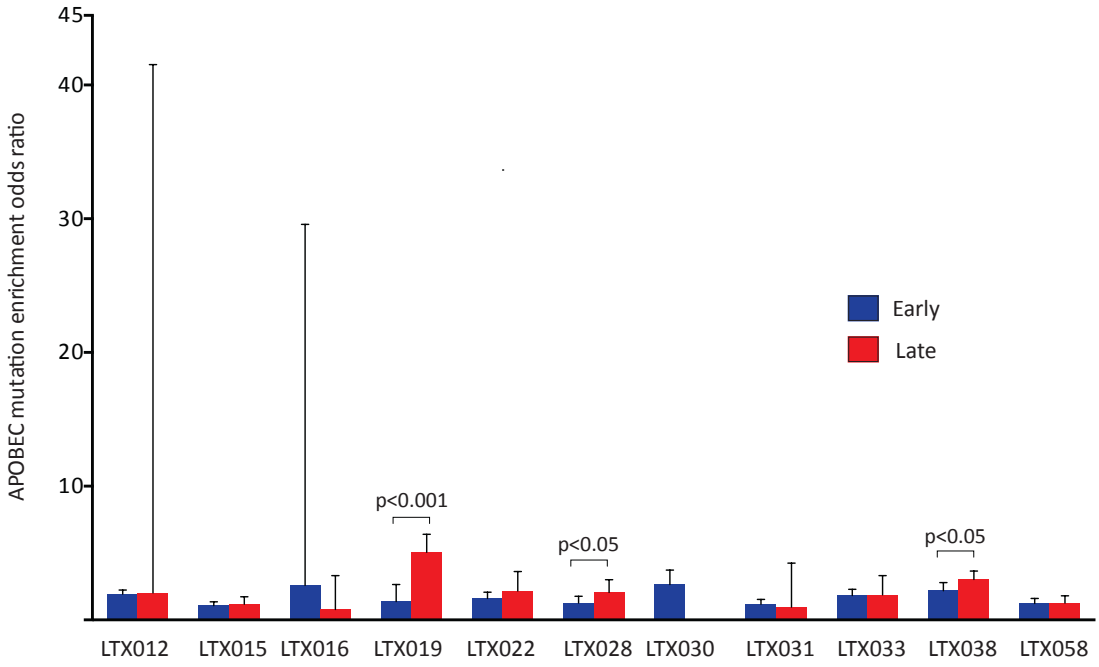


Figure 25 APOBEC mutation enrichment in LUSCs

APOBEC mutation enrichment odds ratio for early (blue) versus late (red) mutations in LUSCs. Note: there were too few mutations in LTX030 for an analysis of late events. Statistically significant APOBEC enrichment in late compared with early mutations was seen in tumours LTX019 ($p < 0.001$), LTX028 ($p < 0.05$) and LTX038 ($p < 0.05$). The 95% confidence intervals are indicated.

3.5 Patients with primary and metastatic tumours

3.5.1 Patient L011

Patient L011 was a 49-year-old female smoker (45 pack-years) diagnosed with LUAD who received adjuvant chemotherapy, but stopped treatment after one cycle due to renal toxicity. Fourteen-months after surgery, the patient presented with headaches and unsteadiness, and was diagnosed with a large occipital brain metastasis associated with increased intracranial pressure. The patient underwent a debulking craniotomy, whereby part of the brain metastasis was surgically resected. From this resection, four FFPE tumour blocks were collected, and DNA extracted from these spatially separated regions (M1 to M4) was used for WES. The patient died 5-months after surgery.

The heat map and phylogenetic tree demonstrate that mutations common to all four metastatic tumour regions M1 to M4 were shared with only region R3 from the primary tumour, suggesting that the metastasis originated from this region of the tumour (**Figure 26**). Intratumour heterogeneity was evident in both the primary and metastatic tumours. The intratumour heterogeneity index for the primary tumour was 0.05. **Figure 26**A shows a region in M1 outlined with a red box, which highlights a set of mutations in genes present on chromosomes 1 and 9. The copy number and mirrored B allele frequency (mBAF) profile for this region, relative to the other regions in the metastasis, suggest that there was a copy number neutral loss of heterozygosity event in chromosome 1 and chromosome 9. **Figure 27** shows the copy number and mBAF for region M1 relative to region M3. The incorporation of copy number aberrations in the analysis of SNV and indel mutations in such cases can identify absent mutations driven by copy number loss, and lead to a more accurate interpretation of branch mutations, and therefore the intratumour heterogeneity index. A statistically significant shift was seen in the mutation spectra of the six mutation types (C>A, C>G, C>T, T>A, T>C, T>G), comparing the early with late mutations for tumour regions R3 and M1 to M4 only ($p < 0.001$, **Figure 26**B). A significant decrease in the smoking-associated signature ($p<0.001$), and an increase in APOBEC enrichment (odds ratio 2.27, range 1.09 to 4.45, $p = 0.019$) was seen in the late compared with early mutations for tumour regions R3 and M1 to M4. Since tumour region R3 was associated with the brain metastasis, this suggests that APOBEC cytidine deaminase activity may have had a role to play in tumour progression, and therefore the metastatic process, perhaps by

initiating genetic diversity upon which selection could act. Although APOBEC enrichment was not seen in tumour regions R1 and R2, there were few late mutations in these regions, and therefore APOBEC enrichment cannot be confidently excluded (**Figure 28**).

This patient had a truncal *BRAF* V600E mutation (55% frequency in *BRAF*-mutated NSCLC (Forbes et al. 2011)). *BRAF* mutations occur in approximately 1-4% of all NSCLCs (Brose et al. 2002; Cardarella et al. 2013). *BRAF* mutations in NSCLC cell lines have been associated with decreased sensitivity to EGFR TKIs (Gandhi et al. 2009; Pratilas et al. 2008), and sensitivity to MEK inhibitors (Pratilas et al. 2008) and BRAF inhibitors (Sen et al. 2012). Increasing evidence also supports the role of BRAF inhibitors in patients with NSCLC, such as vemurafenib (Gautschi et al. 2012) and dabrafenib (Rudin et al. 2013). Temporal heterogeneity was seen in predicted driver mutations, in particular branch mutations were identified in *BCLAF1* (category 2, regions R3 and M1 to M4), and in *CACNA1D* (calcium channel, voltage-dependent, L type, alpha 1D subunit) (category 3, regions M1 and M2). There were three independent truncal category 2 driver mutations in the tumour suppressor gene *FAT1* (FAT atypical cadherin 1) (L1107R, I1601M, and a frameshift variant), suggesting that these mutations may have had a role to play in the early development of the primary tumour. Inactivating mutations in *FAT1* have been shown to be associated with aberrant Wnt (wingless-type MMTV integration site family) pathway signalling and tumourigenesis (L. G. T. Morris et al. 2013). Although these specific *FAT1* mutations have not been described before, they are located within 15bp away from other mutations that are present in COSMIC and have been identified in other cancers.

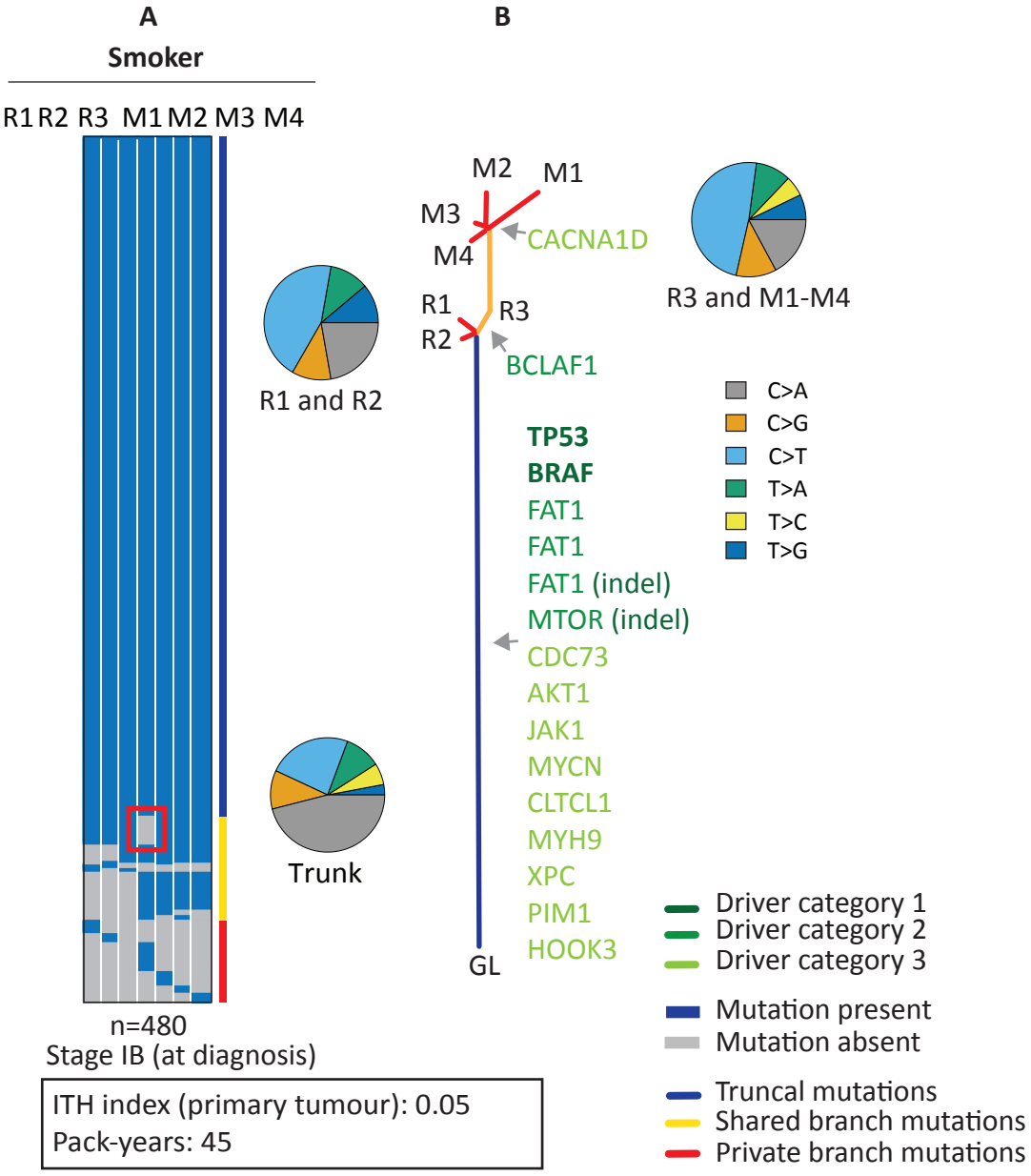


Figure 26 Heat map, phylogenetic tree and mutational spectra for L011

Heat map representing the regional distribution of mutations in the primary (R1-R3) and brain metastatic tumours (M1-M4) (A), and phylogenetic tree annotated with genes in which driver mutations have been predicted and trunk/branch-specific mutational spectra (B). The red box in region M1 represents a potential copy number neutral loss of heterozygosity event.

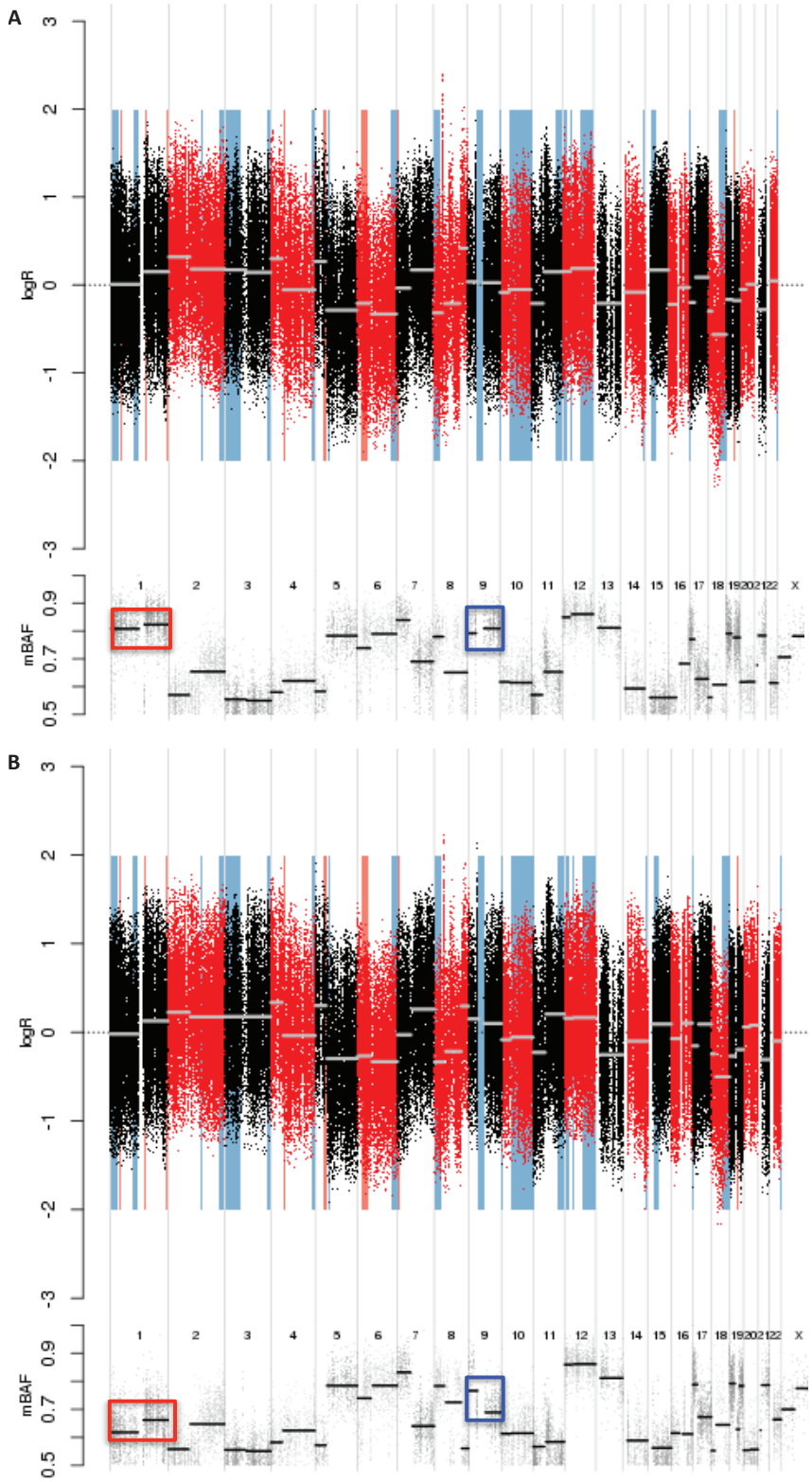


Figure 27 Copy number and mBAF profile for L011 tumour region M1 and M3

Chromosomal segments with copy number profile above and mirrored b allele frequency (mBAF) profile below for regions M1 (A) and M3 (B). Relative to region M3, the red box indicates increased mBAF for chromosome 1 in both the short p arm and long q arm, and the blue box indicates increased mBAF for chromosome 9 in the long q arm. The copy number does not change at these points, suggesting a copy number neutral loss of heterozygosity.

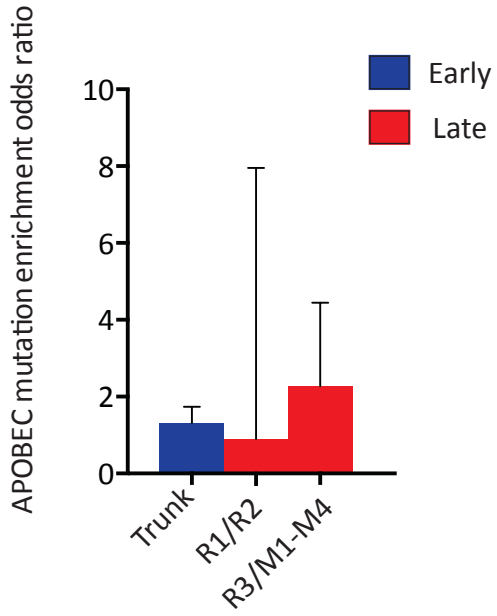


Figure 28 APOBEC enrichment in L011

APOBEC enrichment odds ratio for early (blue) and late (red) mutations in L011.

3.5.2 Patient L017

Patient L017 was a 61-year-old female smoker (48 pack-years) diagnosed with LUAD who initially presented with two tumours in the right upper lobe (RUL). The patient underwent surgical resection of both tumours and completed four cycles of adjuvant chemotherapy. Regions R1 and R2 were collected from the first tumour, and region R3 was collected from the second tumour. Four-months after surgery, the patient was diagnosed with disease recurrence in a single lymph node and was treated with radical radiotherapy to the mediastinum. Nineteen-months after surgery, the patient was diagnosed with disease recurrence in a peripheral lesion in the right lower lobe (RLL), treated again with surgical resection. From this resection, one region (M1) was collected, and DNA was extracted from fresh frozen tissue for WES. This patient was alive at the time of her last follow-up assessment (35-months after surgery) with no evidence of further disease recurrence.

Despite a distance of only 8mm between the two tumours, the heat map and phylogenetic tree supported the diagnosis of two separate synchronous primary tumours in this patient (**Figure 29**). Only tumour regions R1 and R2, collected from one

tumour, had mutations in common. Tumour region R3, which was collected from the second tumour, had only private mutations. Region M1 from the metastatic tumour had mutations in common with R3 only, suggesting that the metastasis originated from R3. Intratumour heterogeneity was evident in the primary tumour from which R1 and R2 were collected, with an intratumour heterogeneity index of 0.15 for the primary tumour, which consisted of regions R1 and R2 (**Figure 29**A). A statistically significant shift was seen in the spectra of the six mutation types (C>A, C>G, C>T, T>A, T>C, T>G), comparing the early with late mutations for both tumour regions R1 and R2 ($p < 0.01$), and R3 and M1 ($p < 0.01$) (**Figure 29**B). A significant decrease in the smoking-associated signature was seen in the late compared with early mutations in tumour regions R1 and R2 ($P < 0.001$) and R3 and M1 ($p < 0.001$), suggesting a role for tobacco smoke in early tumourigenesis. APOBEC enrichment in the late compared with early mutations was seen in tumour regions R1 and R2 (odds ratio 3.15, range 1.96 to 4.96, $p < 0.001$), but not in regions R3 and M1 (**Figure 30**).

Temporal heterogeneity in the predicted driver mutations was seen, in particular branch mutations in the tumour suppressor gene *BCOR* (BCL6 corepressor) (category 1, region R1), and the oncogenes *TPR* (translocated promoter region, nuclear basket protein) (category 3, R1)*, TRIP11* (thyroid hormone receptor interactor 11) (category 3, R1), *PAX8* (paired box 8) (category 3, R1), *HIST1H3B* (histone cluster 1, H3b) (category 3, R2), *HOXD13* (homeobox D13) (category 3, M1) and *RANBP17* (RAN binding protein 17) (category 3, M1). Relevant to NSCLC, a category 1 truncal *KRAS* G12V mutation (20% frequency in *KRAS*-mutated NSCLC, (Forbes et al. 2011)), known to confer decreased sensitivity to EGFR TKIs (Pao et al. 2005; Mitsudomi & Yatabe 2007), was also detected in this tumour. Interestingly, despite the existence of two separate lesions, there were four independent driver category 1 mutations in the tumour suppressor gene *ATRX* (α-thalassaemia/mental retardation syndrome X-linked); two on each tumour trunk (T1621A and S25X for R1/R2 and R2164S and D2144Y for R3/M1). Inactivating mutations in the *ATRX* gene are known to be associated with increased activity of the alternative lengthening of telomeres (ALT) pathway (Lovejoy et al. 2012; Heaphy et al. 2011; Bower et al. 2012; Schwartzentruber et al. 2012). Such increased activity can enable cancer cells to escape replicative senescence or apoptosis, promoting tumour cell proliferation (Bryan et al. 1997). In addition, there were two independent driver category 1 mutations in the *TP53* gene; one on each tumour trunk (E155X for R1/R2 and an exon 4 splice site mutation for

R3/M1). This suggests that the patient's germ line may have predisposed the patient to such mutations, and therefore have selected for the *ATRX* gene, or its combination with the *TP53* gene, as drivers of disease and tumour growth. Furthermore, the selection of such driver events may have also been influenced by the patient's microenvironment.

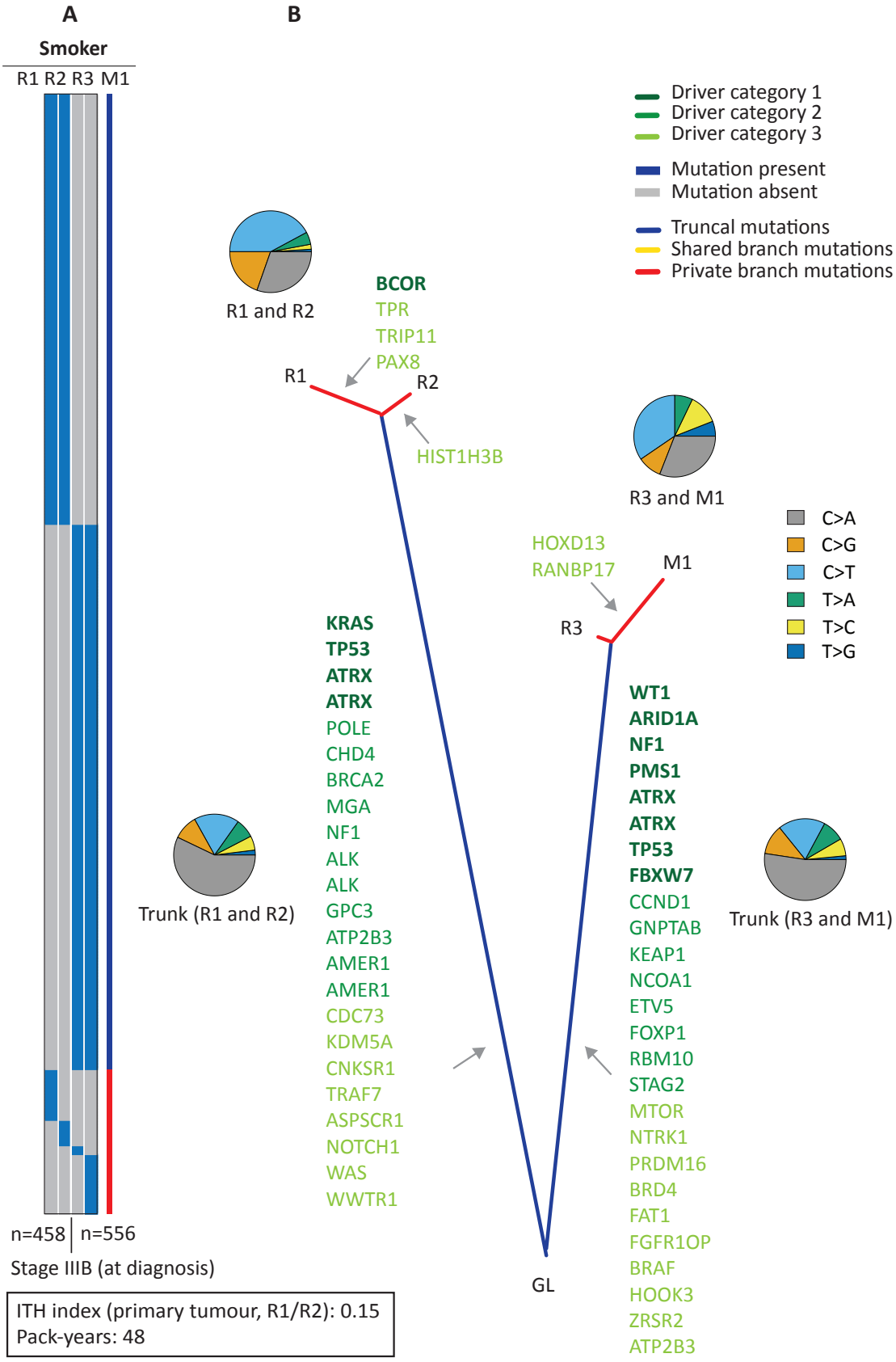


Figure 29 Heat map, phylogenetic tree and mutational spectra for L017

Figure 29 Heat map, phylogenetic tree and mutational spectra for L017

Heat map representing the regional distribution of mutations in the primary (R1-R3) and metastatic tumours (M1) (A), and phylogenetic tree annotated with annotated with genes in which driver mutations have been predicted and trunk/branch-specific mutational spectra (B). Both tumour trunks have independent category 1 driver mutations in *ATRX* and *TP53*.

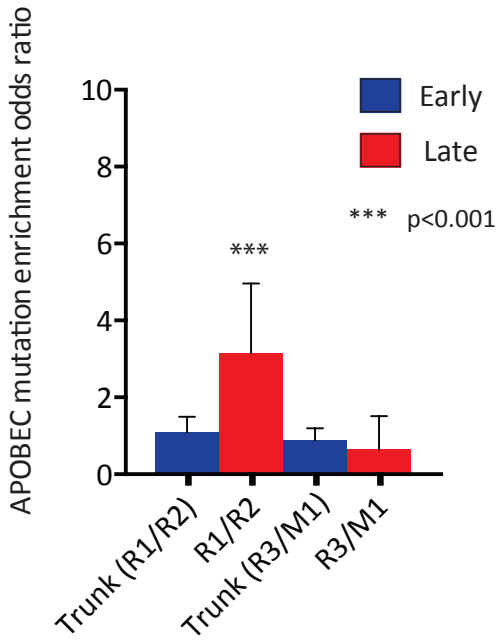


Figure 30 APOBEC enrichment in L017

APOBEC enrichment odds ratio for early (blue) and late (red) mutations in L017. Statistically significant APOBEC enrichment in late compared with early mutations was seen in tumour regions R1 and R2 ($p < 0.001$).

3.5.3 Patient L023

Patient L023 was a 50-year-old female smoker (60 pack-years) diagnosed with LUAD who initially presented with two separate tumours; one in the RUL and one in the right middle lobe (RML), from which tumour regions R2 and R1 were collected, respectively. The patient was alive at the time of her last follow-up with no evidence of further disease recurrence.

The heat map and phylogenetic tree identified common mutations between the two tumours, as well as mutations private to each region, suggesting that one tumour in one lobe had metastasised to the other in another lobe (**Figure 31**). A statistically significant shift was seen in the spectra of the six mutation types (C>A, C>G, C>T, T>A,

T>C, T>G), comparing the early mutations with late mutations ($p < 0.001$) (**Figure 31**B). There was no significant difference in smoking-associated signature or APOBEC enrichment signatures in the late compared with early mutations (**Figure 32**). Temporal heterogeneity in predicted driver mutations was seen; in particular branch mutations in the tumour suppressor genes *FANCA (*category 1, region R1) and *CASP8* (caspase 8, apoptosis-related cysteine peptidase) (category 1, R2) were identified. Private driver category 3 mutations were seen in R1 in the following genes: *GMPS* (guanine monphosphate synthase)*, TET2* (tet methylcytosine dioxygenase 2)*, PIM1* (pim-1 proto-oncogene, serine/threonine kinase)*, ARID1A* and *HNRNPA2B1* (heterogeneous nuclear ribonucleoprotein A2/B1), and in R2 in the following genes: *GNAS* (GNAS complex locus)*, CCNB1IP1* (cyclin B1 interacting protein 1, E3 ubiquitin protein ligase)*, MAP2K1* (mitogen-activated protein kinase kinase 1), and *SETBP1* (SET binding protein 1). Relevant to NSCLC, a truncal *KRAS* G12C mutation (42% frequency in *KRAS*-mutated NSCLC, (Forbes et al. 2011)), known to confer decreased sensitivity to EGFR TKIs (Pao et al. 2005; Mitsudomi & Yatabe 2007), was identified.

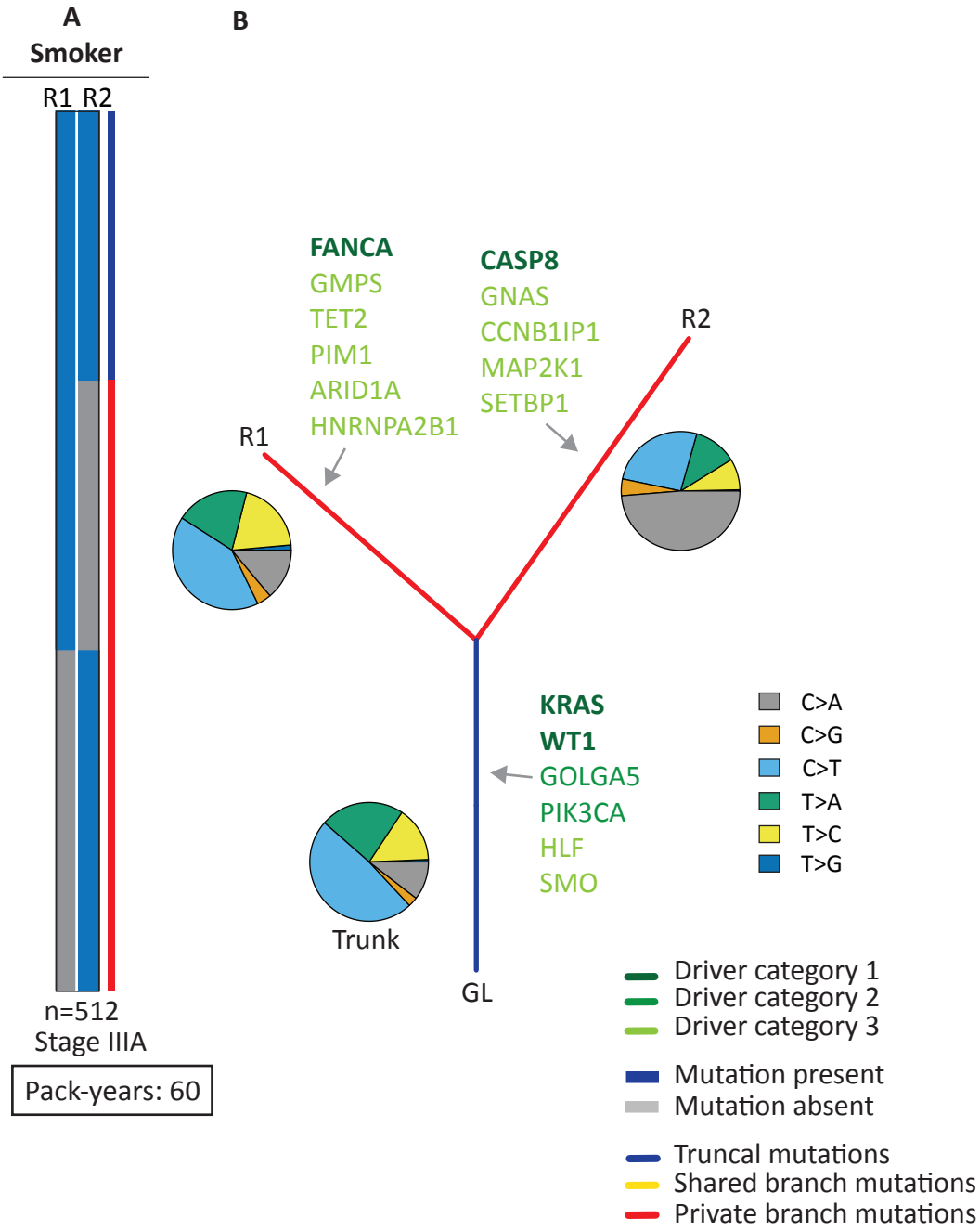


Figure 31 Heat map, phylogenetic tree and mutational spectra for L023

Heat map representing the regional distribution of mutations in the two primary tumours (R1 and R2) (A), and phylogenetic tree annotated with genes in which driver mutations have been predicted and trunk/branch-specific mutational spectra (B).

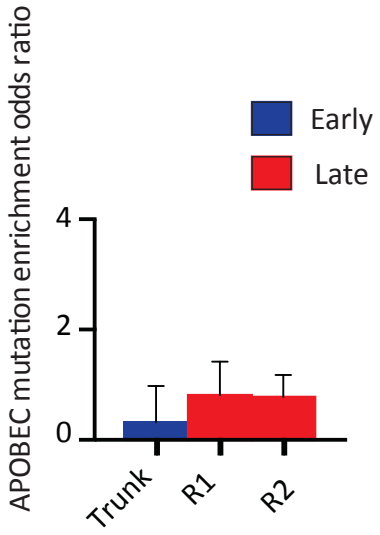


Figure 32 APOBEC enrichment in L023

APOBEC enrichment odds ratio for early (blue) and late (red) mutations in L023.

3.6 Intratumour heterogeneity in copy number aberrations

With higher tumour purity and greater sequencing depth, copy number aberrations were explored in the TRACERx cohort only. Using TCGA LUAD and LUSC data, recurrent chromosomal segment gains and/or amplifications and losses were identified as described in Section 2.11. These were used to identify potential driver copy number aberrations in the TRACERx cohort. **Figure 33** and **Figure 34** demonstrate the regional distribution of potential driver copy number amplifications and/or gains in patients with LUAD and LUSC histological subtypes, respectively. Predicted driver genes within each chromosomal segment were identified and used to annotate each heat map. Intratumour spatial heterogeneity in copy number gains and/or amplifications was seen in all patients, except in patients LTX012 and LTX051 where only ubiquitous amplifications were seen. A gain and/or amplification of the chromosomal segment (chr7p11.2) containing the *EGFR* gene was identified as a ubiquitous event in all patients with LUAD (**Figure 33**). This was also identified as a heterogeneous event in a subset of patients with LUSC (LTX015, LTX019, LTX038 and LTX058). In addition to a gain and/or amplification, a mutation in the *EGFR* gene (category 1 to 3) was also identified in some of these patients; patient LTX001 with LUAD had a category 2 truncal *EGFR* G630E mutation, and patient LTX051 with LUAD had a category 1 truncal *EGFR* exon 21 L858R mutation. There were no *EGFR* mutations found amongst the patients with LUSC. In a cohort study of lung adenocarcinoma by Li and

colleagues, *EGFR* amplifications were found to often accompany *EGFR* mutations, whereby *EGFR* amplifications were seen in 15/29 (52%) *EGFR*-mutant tumours, but in only 5/31 (6%) non-mutant tumours (Li et al. 2008).

Figure 35 and **Figure 36** demonstrate the regional distribution of potential driver copy number losses in patients with LUAD and LUSC histological subtypes, respectively. Intratumour spatial heterogeneity was also seen in copy number losses in both LUAD and LUSC histological subtypes. Chromosomal segments containing known tumour suppressor genes, such as *FAT1, CDKN2A* (cyclin-dependent kinase inhibitor 2A)*, PTEN,* and *RB1* were found to be ubiquitously aberrant in some patients, and heterogeneously aberrant in others. For example, the chromosomal segment containing *FAT1* was ubiquitously lost in patients LTX015, LTX022, LTX028, LTX029, LTX033, and LTX036, and heterogeneously lost in patients LTX016, LTX019, and LTX038. The existence of heterogeneous gains and/or amplifications and losses across the cohort suggests branched tumour evolution of copy number aberrations, with ubiquitous and heterogeneous aberrations potentially involved in early and late tumourigenesis, retrospectively.

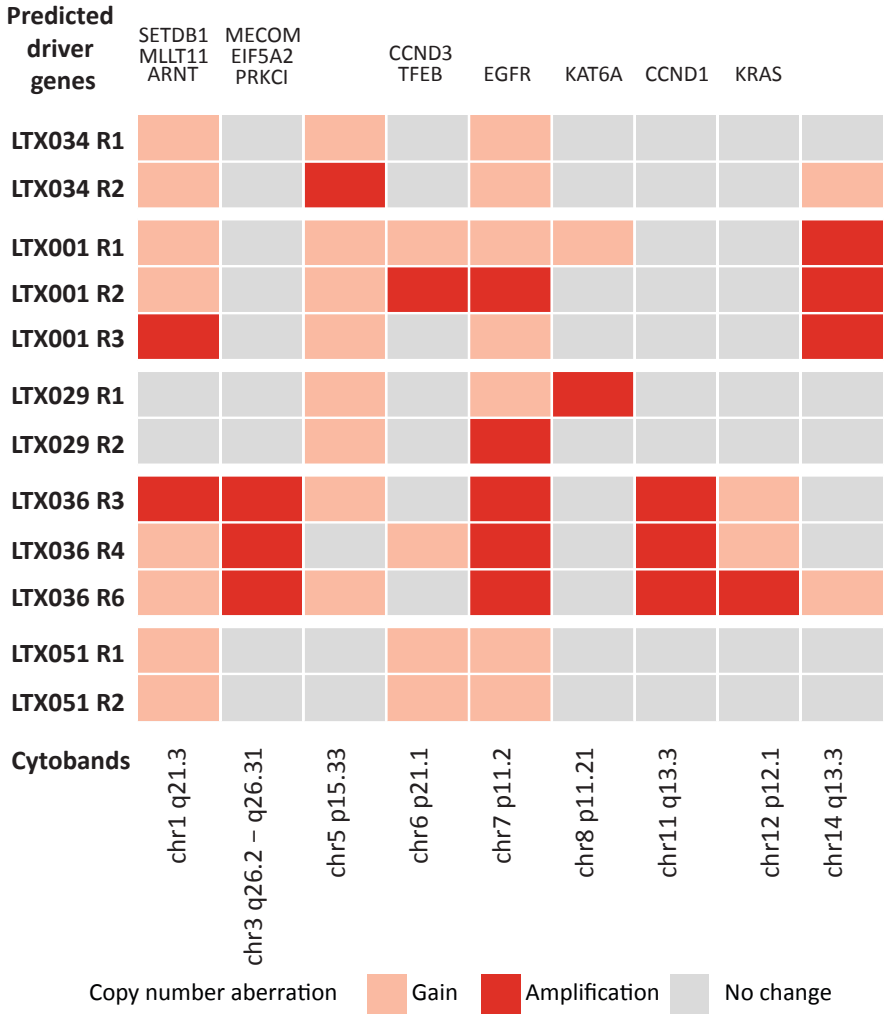


Figure 33 Copy number gains and amplifications in LUADs

Heat map representing the regional distribution of potential driver copy number gains and/or amplifications in the TRACERx LUAD cohort, based on recurrently amplified chromosomal segments identified in TCGA LUAD data. For each tumour region amplifications and gains were determined relative to the mean ploidy. Each cytoband range defines a chromosomal segment, and each segment is annotated by predicted driver genes contained within the segment.

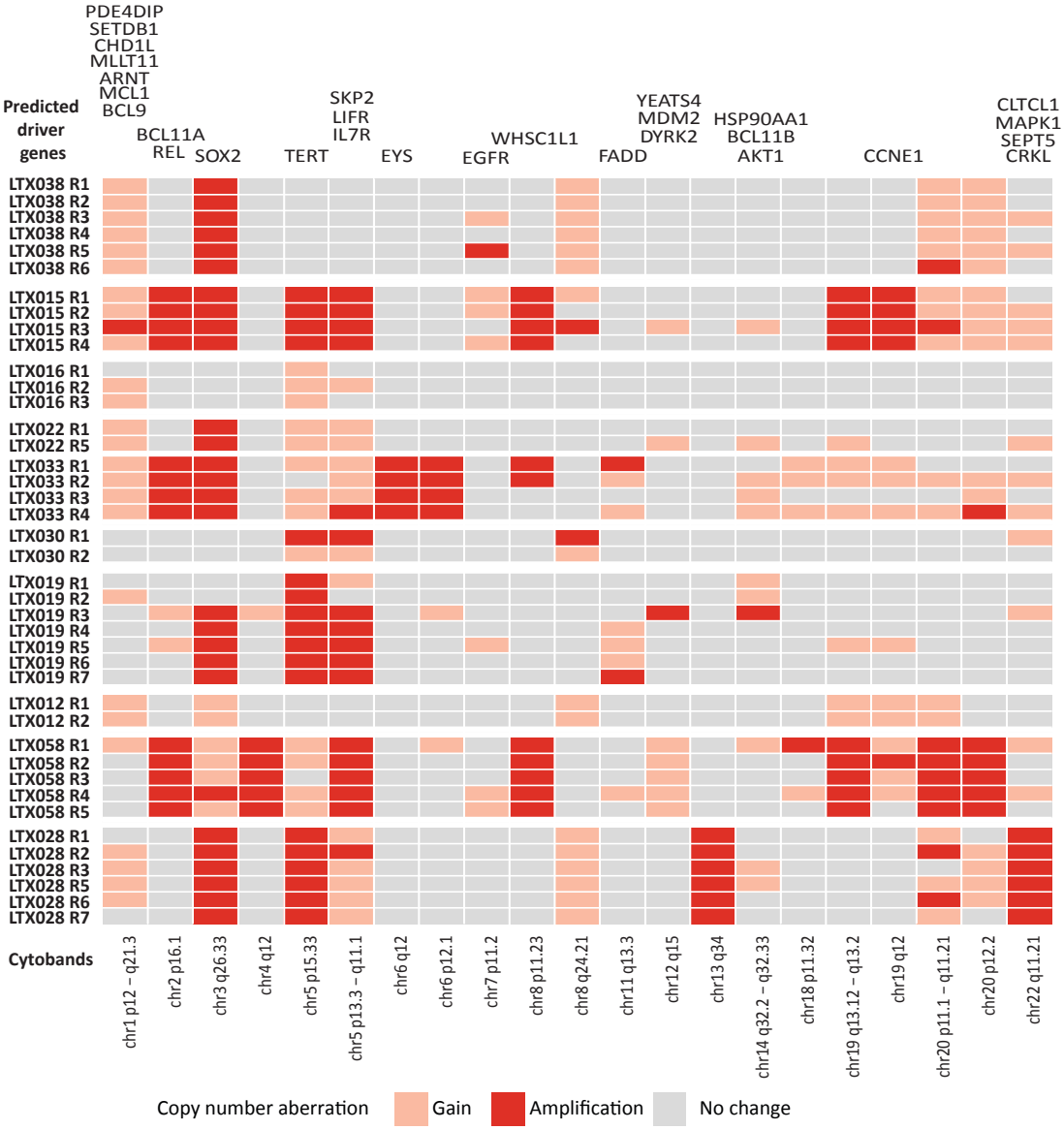


Figure 34 Copy number gains and amplifications in LUSCs

Heat map representing the regional distribution of potential driver copy number gains and/or amplifications in the TRACERx LUSC cohort, based on recurrently amplified chromosomal segments identified in TCGA LUSC data. For each tumour region amplifications and gains was determined relative to the mean ploidy. Each cytoband range defines a chromosomal segment, and each segment is annotated by predicted driver genes contained within the segment.

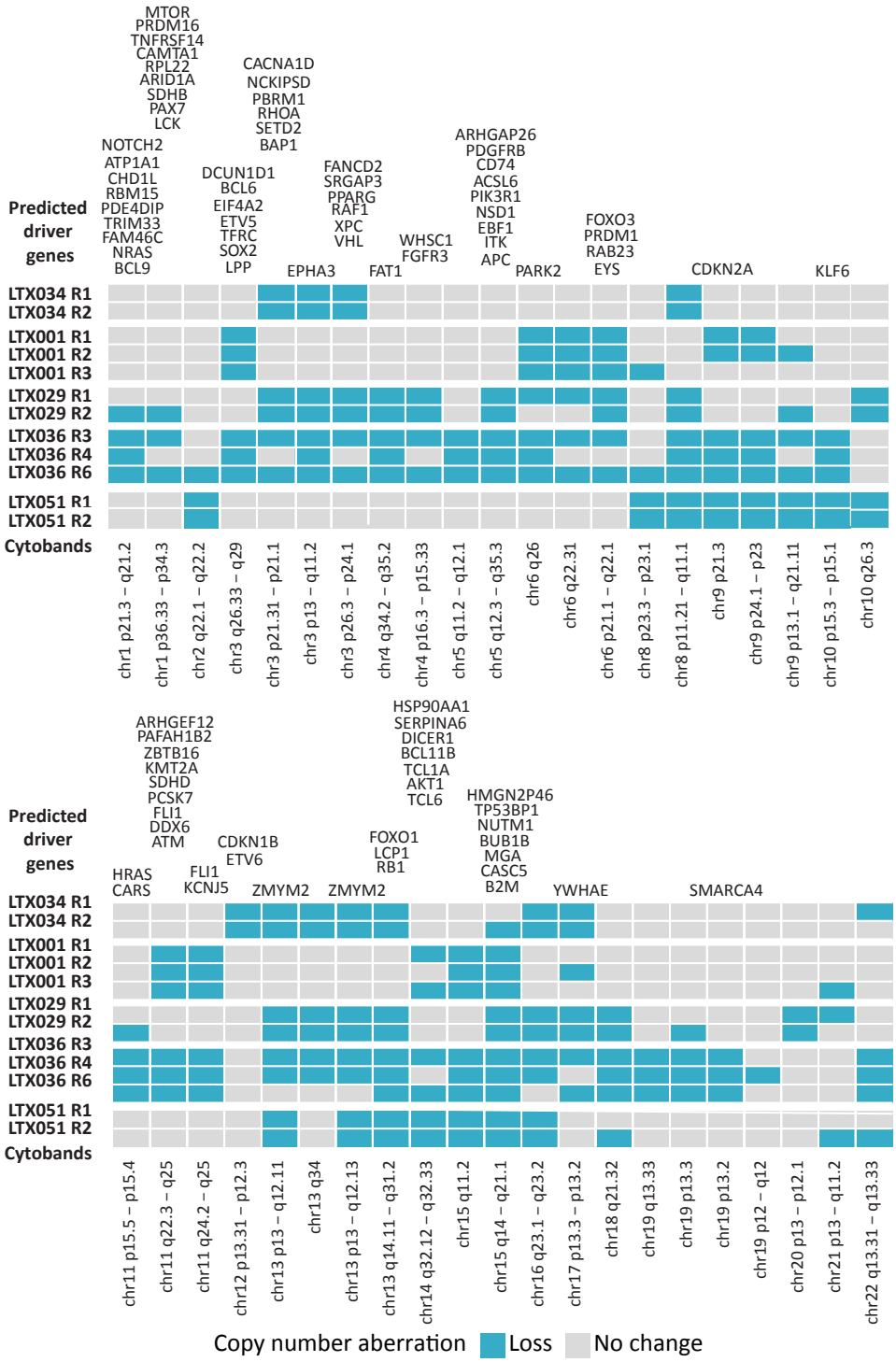


Figure 35 Copy number losses in LUADs

Heat map representing the regional distribution of potential driver copy number losses in the TRACERx LUAD cohort, based on recurrently deleted chromosomal segments identified in TCGA LUAD data. For each tumour region losses were determined relative to the mean ploidy. Each cytoband range defines a chromosomal segment, and each segment is annotated by predicted driver genes contained within the segment.

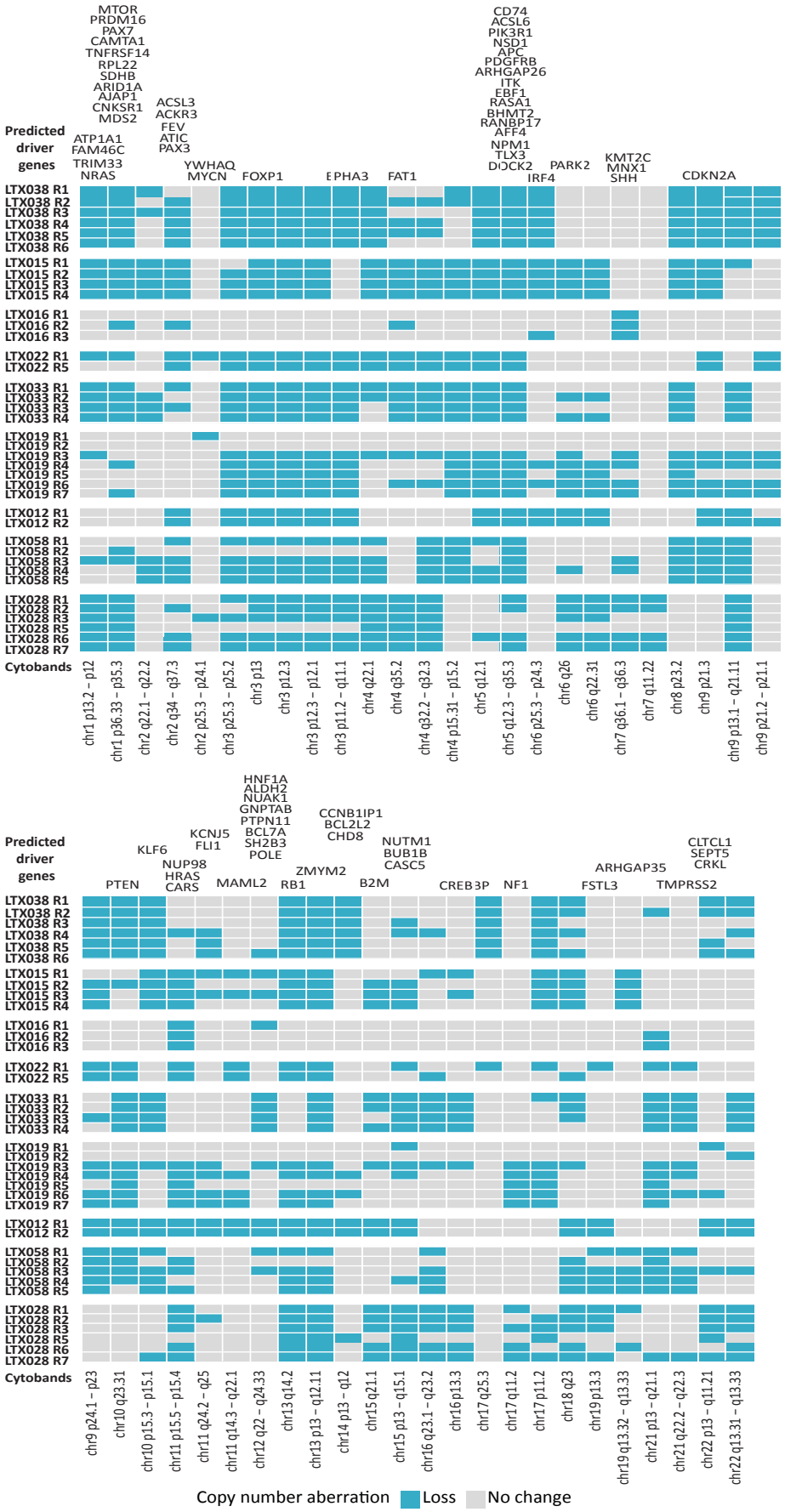

Predicted driver genes
MTOR
PRDM16
PAX7
CAMTA1
TNFRSF14
RPL22
SDHB
ARID1A
AJAP1
CNKSR1
MDS2
ATP1A1
FAM46C
TRIM33
NRAS
ACSL3
ACKR3
FEV
ATIC
PAX3
YWHAQ
MYCN
FOXP1
EPHA3
FAT1
CD74
ACSL6
PIK3R1
NSD1
APC
PDGFRB
ARHGAP26
ITK
EBF1
RASA1
BHMT2
RANBP17
AFF4
NPM1
TLX3
DOCK2
IRF4
PARK2
KMT2C
MNX1
SHH
CDKN2A
LTX038 R1
LTX038 R2
LTX038 R3
LTX038 R4
LTX038 R5
LTX038 R6
LTX015 R1
LTX015 R2
LTX015 R3
LTX015 R4
LTX016 R1
LTX016 R2
LTX016 R3
LTX022 R1
LTX022 R5
LTX033 R1
LTX033 R2
LTX033 R3
LTX033 R4
LTX019 R1
LTX019 R2
LTX019 R3
LTX019 R4
LTX019 R5
LTX019 R6
LTX019 R7
LTX012 R1
LTX012 R2
LTX058 R1
LTX058 R2
LTX058 R3
LTX058 R4
LTX058 R5
LTX028 R1
LTX028 R2
LTX028 R3
LTX028 R5
LTX028 R6
LTX028 R7
Cytobands
chr1 p13.2 – p12
chr1 p36.33 – p35.3
chr2 q22.1 – q22.2
chr2 q34 – q37.3
chr2 p25.3 – p24.1
chr3 p25.3 – p25.2
chr3 p13
chr3 p12.3
chr3 p12.3 – p12.1
chr3 p11.2 – q11.1
chr4 q22.1
chr4 q35.2
chr4 q32.2 – q32.3
chr4 p15.31 – p15.2
chr5 q12.1
chr5 q12.3 – q35.3
chr6 p25.3 – p24.3
chr6 q26
chr6 q22.31
chr7 q36.1 – q36.3
chr7 q11.22
chr8 p23.2
chr9 p21.3
chr9 p13.1 – q21.11
chr9 p21.2 – p21.1
Predicted driver genes
HNF1A
ALDH2
NUAK1
GNPTAB
PTPN11
BCL7A
SH2B3
POLE
KCNJ5
FLI1
KLF6
NUP98
HRAS
CARS
CCNB1IP1
BCL2L2
CHD8
NUTM1
BUB1B
CASC5
ZMYM2
CLTCL1
SEPT5
CRKL
ARHGAP35
PTEN
MAML2
RB1
B2M
CREBBP
NF1
FSTL3
TMPRSS2
LTX038 R1
LTX038 R2
LTX038 R3
LTX038 R4
LTX038 R5
LTX038 R6
LTX015 R1
LTX015 R2
LTX015 R3
LTX015 R4
LTX016 R1
LTX016 R2
LTX016 R3
LTX022 R1
LTX022 R5
LTX033 R1
LTX033 R2
LTX033 R3
LTX033 R4
LTX019 R1
LTX019 R2
LTX019 R3
LTX019 R4
LTX019 R5
LTX019 R6
LTX019 R7
LTX012 R1
LTX012 R2
LTX058 R1
LTX058 R2
LTX058 R3
LTX058 R4
LTX058 R5
LTX028 R1
LTX028 R2
LTX028 R3
LTX028 R5
LTX028 R6
LTX028 R7
Cytobands
chr9 p24.1 – p23
chr10 q23.31
chr10 p15.3 – p15.1
chr11 p15.5 – p15.4
chr11 q24.2 – q25
chr11 q14.3 – q22.1
chr12 q22 – q24.33
chr13 q14.2
chr13 p13 – q12.11
chr14 p13 – q12
chr15 q21.1
chr15 p13 – q15.1
chr16 q23.1 – q23.2
chr16 p13.3
chr17 q25.3
chr17 q11.2
chr17 p11.2
chr18 q23
chr19 p13.3
chr19 q13.32 – q13.33
chr21 p13 – q21.1
chr21 q22.2 – q22.3
chr22 p13 – q11.21
chr22 q13.31 – q13.33
Copy number aberration
Loss
No change

Figure 36 Copy number losses in LUSCs

Heat map representing the regional distribution of potential driver copy number losses in the TRACERx LUSC cohort, based on recurrently deleted chromosomal segments identified in TCGA LUSC data. For each tumour region losses and deletions were determined relative to the mean ploidy. Each cytoband range defines a chromosomal segment, and each segment is annotated by predicted driver genes contained within the segment.

3.7 Correlation between the intratumour heterogeneity index and clinical variables

The intratumour heterogeneity index was calculated for the primary tumours in the combined NSCLC pilot and TRACERx cohort (**Table 10**), with a median index of 0.22 (range 0.01 to 0.64). Using a two-sided Wilcoxon rank-sum test or a Spearman's rank correlation coefficient (Rs), the correlation between the index and the smoking status, histological subtype, tumour stage, tumour size (maximum diameter), lymph node involvement, and disease recurrence was assessed, as shown in **Figure 37** A-G. Patient L023 from the NSLCC pilot cohort, who had two tumours, was excluded since only one region had been sequenced from each tumour. Statistically significant correlations were not identified in this cohort, potentially due to the small number of patients. Such correlations are likely to be identified in larger patient cohorts, such as TRACERx, in which different measures of heterogeneity can be assessed with sufficient power for detection, in order to define an index that is truly representative of the underlying intratumour heterogeneity, with potential predictive and prognostic value.

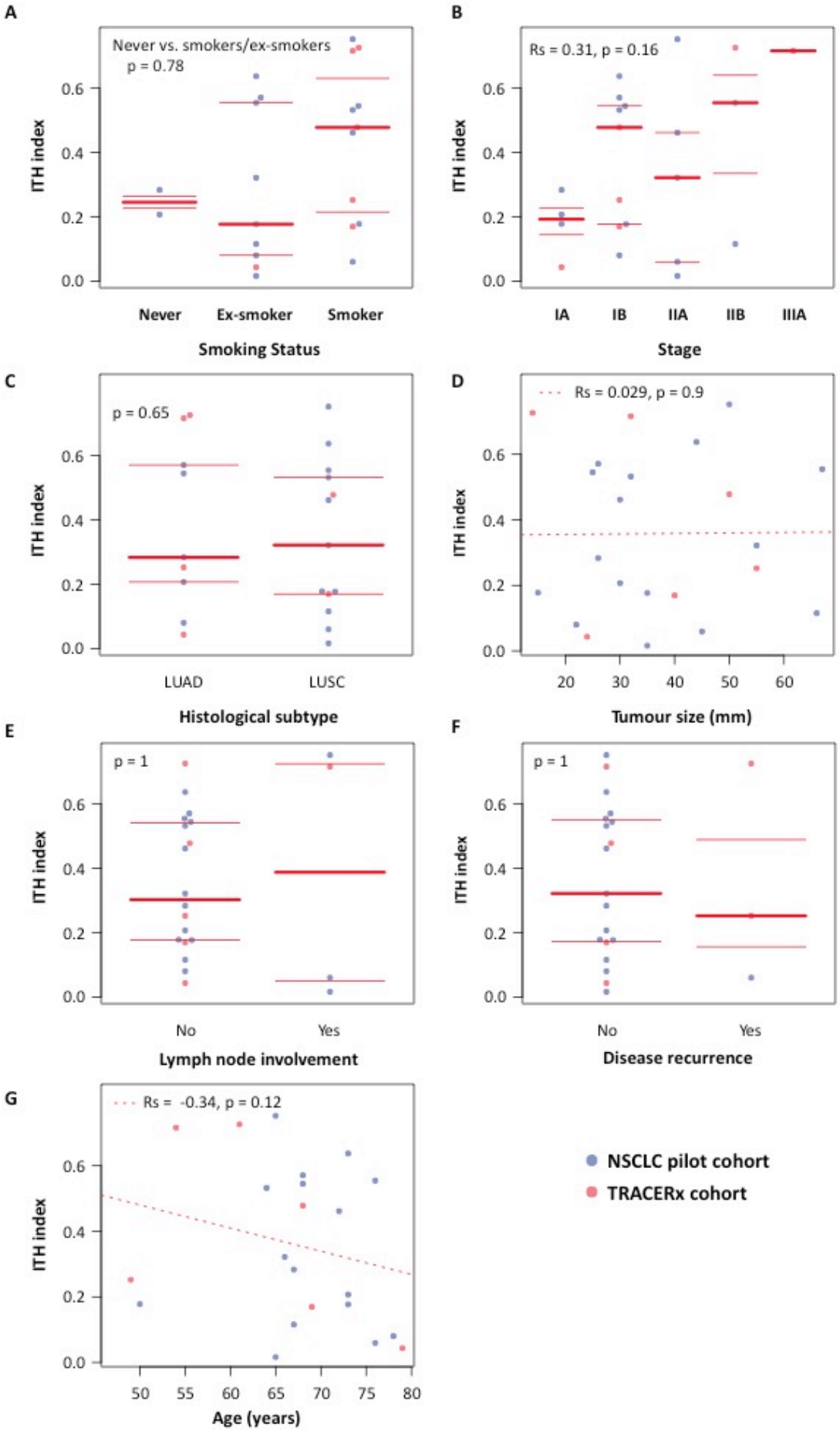

A
Never vs. smokers/ex-smokers
p = 0.78
ITH index
0.0
0.2
0.4
0.6
Never
Ex-smoker
Smoker
Smoking Status
B
Rs = 0.31, p = 0.16
ITH index
IA
IB
IIA
IIB
IIIA
Stage
C
p = 0.65
ITH index
LUAD
LUSC
Histological subtype
D
Rs = 0.029, p = 0.9
ITH index
20
30
40
50
60
Tumour size (mm)
E
p = 1
ITH index
No
Yes
Lymph node involvement
F
p = 1
ITH index
No
Yes
Disease recurrence
G
Rs = -0.34, p = 0.12
ITH index
50
55
60
65
70
75
80
Age (years)
NSCLC pilot cohort
TRACERx cohort

Figure 37 Intratumour heterogeneity index and clinical correlates

Intratumour heterogeneity (ITH) index and correlation with smoking status (A), tumour stage (B), tumour size (C), lymph node involvement (D), histological subtype (E), disease recurrence (F), and age (G). Thick red lines represent median values, and thin red lines represent the interquartile range. Abbreviations: Rs, Spearman's rank correlation coefficient; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma. Abbreviations: Rs; Spearman's rank correlation coefficient.

Patient	Age	Smoking status (pack-years)	Histo	Stage	Max diameter tumour size	Lymph node involvement	Recurrence (mths)	PW ITH
L011	49	Smoker	LUAD	IB	55	No	Yes	0.05
L012	69	Smoker	LUSC	IB	40	No	No	0.35
L013	68	Smoker	LUSC	IB	50	No	No	0.26
L015	68	Smoker	LUSC	IA	30	No	No	0.15
L016	65	Smoker	LUSC	IIIA	27	No	No	0.04
L017	61	Smoker	LUAD	IIB	14	No	Yes	0.15
L019	47	Never-smoker	LUAD	IIIA	25	Yes	Yes	0.41
L022	54	Smoker	LUAD	IIIA	32	Yes	No	0.28
L029	66	Ex-smoker	LUAD	IA	15	No	No	0.07
L030	79	Ex-smoker	LUAD	IA	24	No	No	0.14
LTX001	68	Ex-smoker	LUAD	IB	26	No	No	0.21
LTX012	65	Ex-smoker	LUSC	IIA	35	Yes	No	0.01
LTX015	64	Smoker	LUSC	IB	32	No	No	0.18
LTX016	67	Ex-smoker	LUSC	IIB	66	No	No	0.64
LTX019	65	Smoker	LUSC	IIA	50	Yes	No	0.46
LTX022	72	Smoker	LUSC	IIA	30	No	No	0.13
LTX028	66	Ex-smoker	LUSC	IIA	55	No	No	0.22
LTX029	78	Ex-smoker	LUAD	IB	22	No	No	0.27
LTX030	76	Smoker	LUSC	IIA	45	Yes	Yes	0.03
LTX031	50	Smoker	LUSC	IA	15	No	No	0.04
LTX033	73	Ex-smoker	LUSC	IB	35	No	No	0.11
LTX034	73	Never	LUAD	IA	30	No	No	0.32
LTX036	68	Smoker	LUAD	IB	25	No	No	0.23
LTX038	76	Ex-smoker	LUSC	IIB	67	No	No	0.45
LTX051	67	Never	LUAD	IA	26	No	No	0.27
LTX058	73	Ex-smoker	LUSC	IB	44	No	No	0.26

Table 10 Intratumour heterogeneity index and clinical variables

Intratumour heterogeneity index in the combined NSCLC pilot and TRACERx cohort. Abbreviations: Histo, histology; PW ITH, pairwise intratumour heterogeneity index.

3.8 Conclusions and discussion

Although work within our laboratory has previously demonstrated intratumour heterogeneity in NSCLC (de Bruin et al. 2014), the data presented here are based on a larger cohort of tumours, in particular squamous cell carcinoma tumours, which have been sequenced to a greater depth. In the combined NSCLC pilot and TRACERx cohort, 93 spatially distinct tumour regions from 27 patients were subjected to WES. These data demonstrated intratumour heterogeneity in SNV, indel and copy number aberrations. Furthermore, heterogeneity was seen in predicted driver mutations, and copy number aberrations involving predicted driver genes, with some aberrations occurring early (truncal) and some occurring late (branch) in tumour evolution.

Truncal, that is clonal in origin, mutations are those that are present in all of the cancer cells within a tumour. A mutation that is present in a small population of cancer cells, and therefore subclonal in origin, may be present in all tumour regions, but would not be truly truncal. One limitation of the data presented here is that a truncal mutation has been defined as a mutation that is present in every tumour region, suggesting that the degree of intratumour heterogeneity may have been underestimated in some of the NSCLC tumours. Furthermore, subclonal mutations, present in only a subset of cancer cells, are not necessarily later events, but may instead represent copy number driven intratumour heterogeneity. Thus, a mutation may have occurred prior to the emergence of the most recent common ancestor and yet still be heterogeneous within a tumour. As such, from a clinical perspective, it may be more meaningful to consider whether a mutation is clonal or subclonal given that this provides information on the number of cells harbouring a genetic aberration of interest at a given point in time, rather than in the past. In the TRACERx study, the definition of truncal and branch mutations will take into account the cancer cell fraction associated with each mutation, and with the incorporation of copy number data and deeper sequencing, a more accurate interpretation of the clonal architecture, as well as the true extent of intratumour heterogeneity in NSCLC, can be made.

Intratumour heterogeneity potentially poses significant challenges in the management of NSCLC patients with regards to the identification of predictive and prognostic biomarkers based on a single diagnostic biopsy. The significance of branch aberrations,

in terms of therapeutic response and clinical outcome, is yet to be determined in longitudinal genomic studies, such as TRACERx. However, with increasing evidence for intratumour heterogeneity in NSCLC, the current methods of tumour profiling based on single diagnostic, often historical, biopsies are unlikely to represent the true tumour genomic landscape of tumours. This in turn can lead to inadequately informed clinical decision-making, potentially compromising patient care. The insight into NSCLC evolution, and its complex genomic landscape, has led to the development and justification of the TRACERx study, which aims to take forward and further develop the analyses presented here, with the added benefit of using copy number data to confidently identify the clonality of detected mutations. In particular, TRACERx will explore the following:

- The prevalence of predicted driver mutations and copy number aberrations in multiple compared with single regions
- The prevalence of therapeutically targetable subclonal mutations
- The identification of new potential subclonal drivers of NSCLC
- The identification of significant correlations between indices of intratumour heterogeneity and clinical variables and outcome
- The identification of mutational signatures, and their evolution from diagnosis to relapse.

4 Results 2: Intratumour heterogeneity in lung adenocarcinoma *in situ* lesions

4.1 Introduction

Patients recruited into the TRACERx study either had a confirmed histological diagnosis of NSCLC based on a pre-surgical diagnostic biopsy, or were highly suspected to have such a diagnosis based on their clinical presentation and pre-operative imaging. Five patients recruited into the study were subsequently found to have a histological diagnosis of adenocarcinoma *in situ.* These patients were withdrawn from the study, but still underwent surgical resection of their presumed primary NSCLC tumours, which were subjected to multi-region WES using the same TRACERx bioinformatics pipeline. These cases were analysed as a separate substudy, and are discussed in this chapter.

Adenocarcinoma *in situ* (AIS) is recognised as a non-invasive form of glandular hyperplasia, with the potential to undergo malignant transformation to invasive adenocarcinoma. Little is known about the genetic aberrations and biological pathways involved in the early development of lung adenocarcinomas. In a cohort study by Izumchenko and colleagues (Izumchenko et al. 2014) targeted sequencing of cancer-related genes performed on AIS samples, identified mutations in the genes *KIT, KRAS, HRAS* (Harvey Rat Sarcoma Viral Oncogene Homolog), *IGF1R* (Insulin-Like Growth Factor 1 Receptor), *FGFR3* (Fibroblast Growth Factor Receptor 3)*, MET* and *TSC2* (Tuberous Sclerosis 2) in four out of five patients. Unsurprisingly, most of the lesions harboured mutations in genes associated with DNA repair and chromatin remodelling, suggesting that abnormalities in the DNA repair machinery may be associated with the early development and tumourigenesis of lung adenocarcinoma as a result of increased genomic instability. The study of pre-invasive lung lesions, such as AIS lesions, may give some insight into the mutational pathways involved in initiating and driving the progression of pre-invasive lesions to invasive tumours.

4.2 Baseline characteristics and histopathological variables

Table 11 shows the baseline characteristics of the patients from whom AIS lesions were collected and sequenced. All five patients were ex-smokers, the median age was

74-years (range 71-86 years), three were male and two were female. The median lesion size was 40mm (range 20-53 mm).

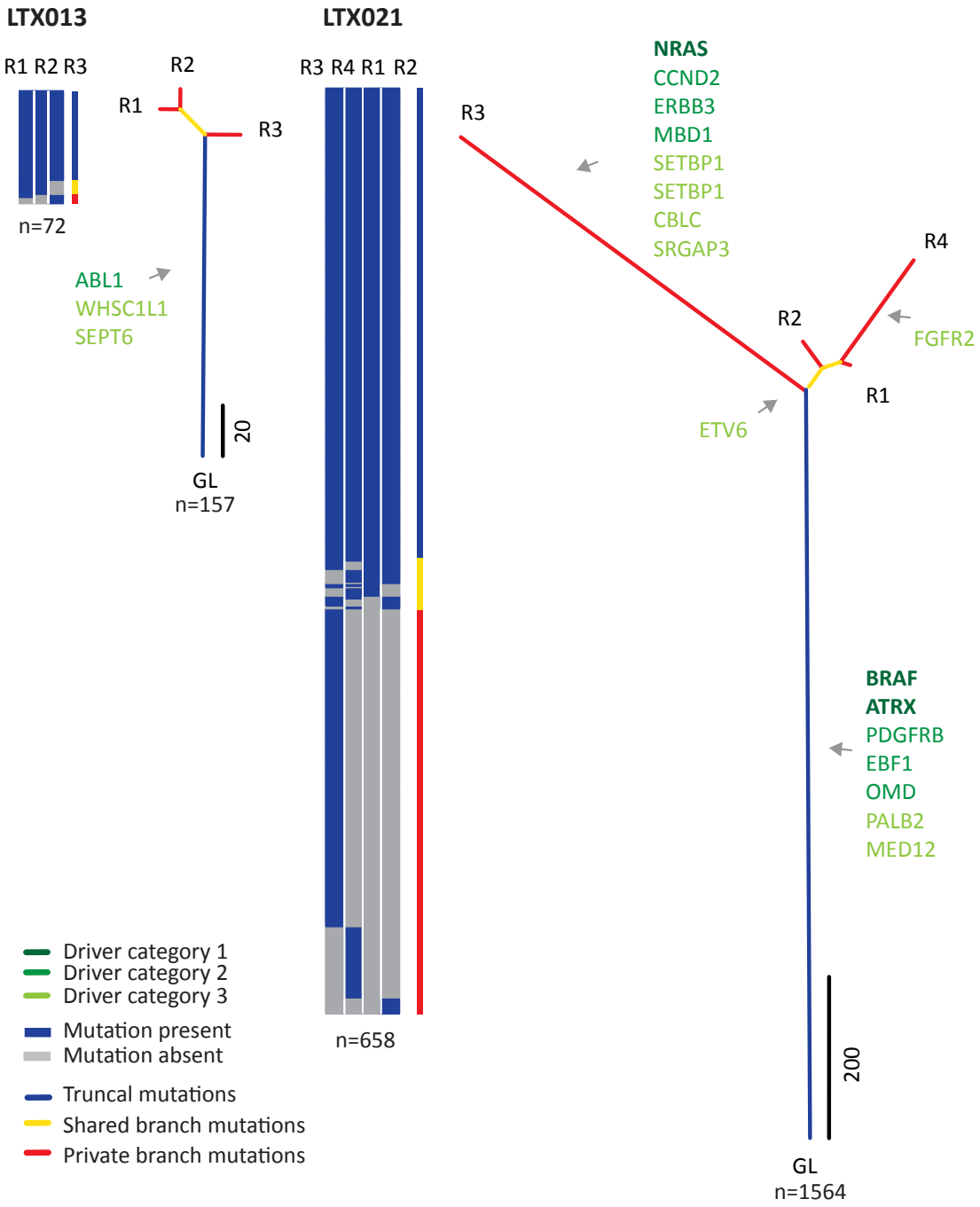
Patient	Age	Gender	Smoking status (pack-years)	Histology	Size on path (mm)
LTX013	86	M	Ex-Smoker (13)	Adenocarcinoma in situ	53
LTX021	74	F	Ex-Smoker (56)	Adenocarcinoma in situ	40
LTX041	76	M	Ex-Smoker (74)	Adenocarcinoma in situ	43
LTX049	68	F	Ex-Smoker (53)	Adenocarcinoma in situ	20
LTX055	71	M	Ex-Smoker (52)	Adenocarcinoma in situ	40

Table 11 Baseline characteristics of the AIS cohort

Baseline patient and lesion characteristics of the AIS cohort.

4.3 Regional distribution of mutations and phylogenetic trees

Filtered non-synonymous mutations were used to create heat maps representing the regional distribution of non-silent SNVs and indels, and phylogenetic trees were created using filtered non-synonymous and synonymous SNV and indel mutations (**Figure 38**). Both spatial and temporal heterogeneity was evident in the AIS lesions, with evidence for branched evolution. This suggested that similar to what was seen in the evolution of the NSCLC tumous, different populations of cells harbouring specific mutations existed in the pre-invasive lesions, and that they had undergone a degree of clonal evolution up until the point of surgical resection. Having previously shown that NSCLC tumours from patients who had never smoked often had shorter trunks compared to tumours from patients who were either current or ex-smokers, patient LTX013 had the shortest trunk, most likely related to his reduced exposure to tobacco smoke (13 pack-year history) compared with the other patients.


LTX013
R1 R2 R3
n=72
R2
R1
R3
ABL1
WHSC1L1
SEPT6
20
GL
n=157
LTX021
R3 R4 R1 R2
n=658
R3
NRAS
CCND2
ERBB3
MBD1
SETBP1
SETBP1
CBLC
SRGAP3
R4
R2
FGFR2
R1
ETV6
BRAF
ATRX
PDGFRB
EBF1
OMD
PALB2
MED12
200
GL
n=1564
Driver category 1
Driver category 2
Driver category 3
Mutation present
Mutation absent
Truncal mutations
Shared branch mutations
Private branch mutations

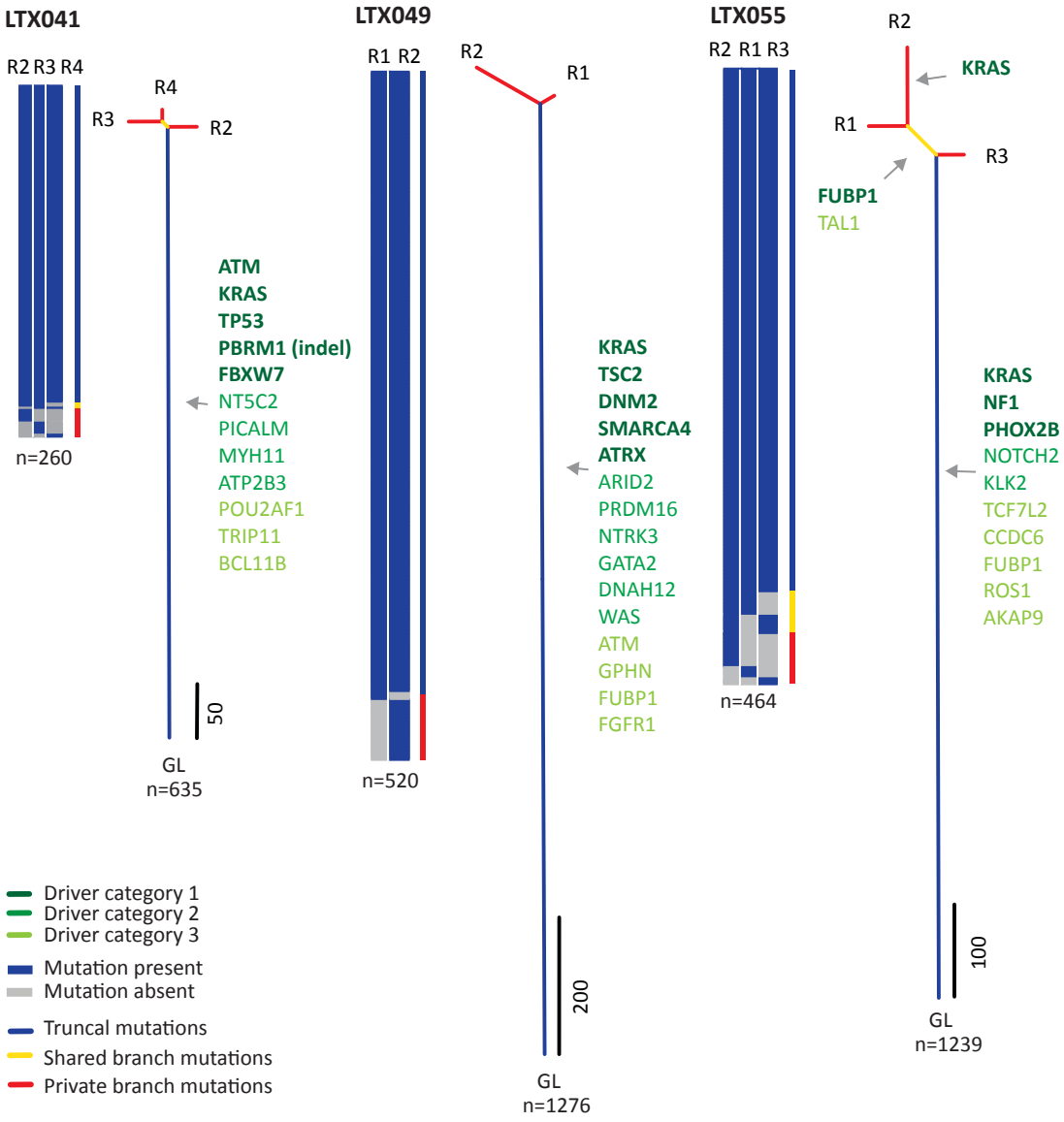


Figure 38 Heat maps and phylogenetic trees in AIS lesions

Heat maps representing the regional distribution of mutations, where the presence of a mutation is indicated by blue and its absence by grey. Alongside each heat map is a vertical bar, where blue represents mutations that are present in all regions (truncal), yellow represents mutations that are present in some, but not all, regions (shared, branch), and red represents mutations that are present in one region only (private, branch). Underneath each heat map is the total number of non-synonymous mutations. Phylogenetic trees annotated by genes representing predicted driver category 1 to 3 mutations, with arrows pointing towards the part of the tree on which they have been acquired. Ubiquitous mutations (present in all regions) are shown on the blue trunks of trees, shared mutations (present in some, but not all, regions) are shown on the yellow branches of trees, and private mutations (present in only one region) are shown on the red branches of trees. Underneath each tree is the total number of filtered non-synonymous and synonymous mutations.

4.4 Predicted driver mutations

As previously described, category 1 to 3 predicted driver mutations were identified in each lesion and used to annotate the phylogenetic trees shown in **Figure 38**. Both spatial and temporal heterogeneity of driver mutations was evident, with certain driver mutations occurring early (present on the trunks of the trees), and certain driver mutations occurring late (present on the branches of the trees) during the evolution of the pre-invasive lesions. Overall, there were 59 category 1 to 3 driver mutations, 17/59 (29%) were category 1, 19/59 (32%) were category 2, and 23/59 (39%) were category 3 predicted driver mutations. Amongst the category 1 high confidence driver mutations, 14/17 (82%) of them were truncal (predicted to be clonal in origin), and 3/17 (17%) of them were branch mutations (predicted to be subclonal in origin). Category 1 to 3 predicted driver mutations for each AIS lesion are listed in **Table 12**.

Patient	Gene	Mutation type	Driver category	Truncal/Branch
LTX013	ABL1	SNV	2	Truncal
	WHSC1L1	SNV	3	Truncal
	SEPT6	SNV	3	Branch
LTX021	BRAF	SNV	1	Truncal
	ATRX	SNV	1	Truncal
	NRAS	SNV	1	Branch
	PDGFRB	SNV	2	Truncal
	EBF1	SNV	2	Truncal
	OMD	SNV	2	Truncal
	CCND2	SNV	2	Branch
	ERBB3	SNV	2	Branch
	MBD1	SNV	2	Branch
	PALB2	SNV	3	Truncal
	MED12	SNV	3	Truncal
	ETV6	SNV	3	Branch
	SETBP1	SNV	3	Branch
	SETBP1	SNV	3	Branch
	CBLC	SNV	3	Branch
	SRGAP3	SNV	3	Branch
	FGFR2	SNV	3	Branch
LTX041	KRAS	SNV	1	Truncal
	TP53	SNV	1	Truncal
	PBRM1	Indel	1	Truncal
	FBXW7	SNV	1	Truncal

	NT5C2	SNV	2	Truncal
	PICALM	SNV	2	Truncal
	MYH11	SNV	2	Truncal
	ATP2B3	SNV	2	Truncal
	POU2AF1	SNV	3	Truncal
	TRIP11	SNV	3	Truncal
	BCL11B	SNV	3	Truncal
	KRAS	SNV	1	Truncal
	TSC2	SNV	1	Truncal
	DNM2	SNV	1	Truncal
	SMARCA4	SNV	1	Truncal
	ATRX	SNV	1	Truncal
	ARID2	SNV	2	Truncal
	PRDM16	SNV	2	Truncal
LTX049	NTRK3	SNV	2	Truncal
	GATA2	SNV	2	Truncal
	DNAH12	SNV	2	Truncal
	WAS	SNV	2	Truncal
	ATM	SNV	3	Truncal
	GPHN	SNV	3	Truncal
	FUBP1	SNV	3	Truncal
	FGFR1	SNV	3	Truncal
	FUBP1	SNV	1	Branch
	KRAS	SNV	1	Branch
	KRAS	SNV	1	Truncal
	NF1	SNV	1	Truncal
	PHOX2B	SNV	1	Truncal
	NOTCH2	SNV	2	Truncal
LTX055	KLK2	SNV	2	Truncal
	TAL1	SNV	3	Branch
	TCF7L2	SNV	3	Truncal
	CCDC6	SNV	3	Truncal
	FUBP1	SNV	3	Truncal
	ROS1	SNV	3	Truncal
	AKAP9	SNV	3	Truncal

Table 12 Category 1 to 3 predicted driver mutations in the AIS cohort

Predicted category 1 to 3 driver mutations identified in each AIS lesion. The gene name, type of variant (SNV or indel), and whether the mutation is truncal or branch is shown.

Category 1 driver mutations in the *KRAS* gene were identified in 3/5 patients. In patient LTX041 there was a truncal G12C mutation, in patient LTX049 there was a truncal G12V mutation, and in patient LTX055 there was both a truncal G12C and a branch G12V mutation. In 2/5 patients there was a category 1 truncal driver mutation in the *ATRX* gene (E851X in LTX021 and R445N in LTX049). The presence of these activating mutations in the *KRAS* oncogene, mostly as truncal events, suggests that *RAS*-dependent signalling may play a significant role in the early development of AIS lesions. In the case of patient LTX055, the presence of both truncal and branch driver mutations in the *KRAS* gene, suggests that such signalling may be involved in both early and late evolution of AIS lesions. Since G>C or G>T transversions of the guanine nucleotide residues located in the wild type *KRAS* codon 12 are known to be associated with smoking (Ahrendt et al. 2001), and are uncommon in NSCLC amongst non-smokers (Riely et al. 2008), it is not surprising that these mutations existed in these ex-smoker patients with significant numbers of smoking pack-years (**Table 11**).

Interestingly, in patient LTX021 there was a category 1 truncal driver mutation in *BRAF* (G466E, exon 11 kinase domain), as well as a category 1 branch driver mutation in *NRAS* (Q61K). This suggests that the *BRAF* mutation occurred early, and that the *NRAS* mutation occurred late, in the development of the AIS lesion. Both *BRAF* and *NRAS* are components of the MAPK/ERK pathway, which controls various cellular processes, including cellular proliferation, growth, migration and apoptosis (Dhillon et al. 2007). Although distinct mutations in the *BRAF* and *NRAS* genes have occurred in different regions of the lesion, they converge on the same signalling pathway, representing a case of convergent evolution. If this *in situ* lesion were to progress further and transform into an invasive tumour, a single diagnostic biopsy would most likely only identify the truncal *BRAF* mutation as a driver of disease. Applying a selection pressure, such as treatment with a BRAF inhibitor, could allow the subclone harbouring the *NRAS* mutation to dominate the tumour mass and continue to drive disease through MAPK/ERK pathway activation, potentially leading to drug resistance and disease progression. Inhibiting the MAPK/ERK pathway downstream of its *BRAF* and *NRAS* components, for example with MEK or ERK inhibition, could potentially result in better disease control. In melanoma cell lines harbouring a *BRAF* V600E and *NRAS* Q61K mutation, resistance to the BRAF inhibitor vemurafenib, but sensitivity to the MEK inhibitor selumetinib, has been previously shown (Atefi et al. 2011). Inhibition with the ERK1/2 inhibitor PLX7904 has also been shown to inhibit tumour growth in

melanoma cell lines harbouring both the *BRAF* V600E and *NRAS* Q61K mutations (D. B. Johnson et al. 2014; Le et al. 2013). Similarly in a *BRAF*-, *KRAS*- and *NRAS*-mutant xenograft models of melanoma, sensitivity and tumour response to the ERK1/2 inhibitor SCH772984 has also been shown (E. J. Morris et al. 2013). One potential treatment option in the context of both mutations and advanced invasive adenocarcinoma might be combination therapy with a BRAF and MEK inhibitor in melanoma, such as dabrafenib and trametinib, which has been shown to significantly improve PFS and OS (Long et al. 2015).

4.5 Intratumour heterogeneity in copy number aberrations

Copy number aberrations have been previously demonstrated in pre-invasive lung cancer lesions (Massion et al. 2009). Recurrent chromosomal segment gains and/or amplifications, and losses and/or deletions were identified in the AIS lesions as described in Section 2.11, and were used to identify potential driver copy number aberrations (**Figure 39** and **Figure 40**). Predicted driver genes within each chromosomal segment were identified and used to annotate each heat map. Intratumour spatial heterogeneity in copy number gains and/or amplifications was seen in all of the AIS lesions, apart from in patients LTX049, in whom only two regions had been sequenced (**Figure 39**). Similarly, intratumour spatial heterogeneity in copy number losses and/or deletions was seen in all of the AIS lesions (**Figure 40**). Overall, the existence of heterogeneous copy number aberrations in pre-invasive lung lesions suggests that branched evolution of copy number events can occur in the early stages of cancer development prior to the development of invasive disease.

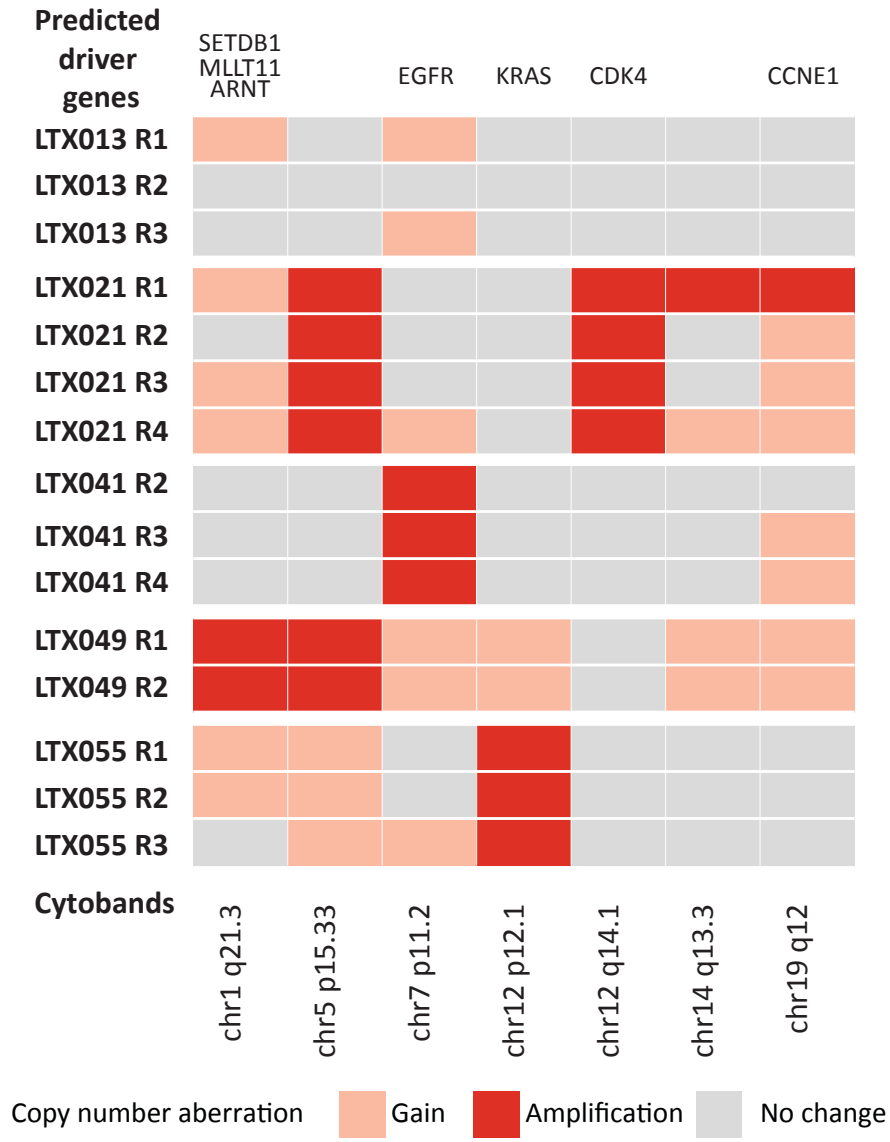


Figure 39 Copy number gains and/or amplifications in AIS lesions

Heat map representing the regional distribution of potential driver copy number gains and/or amplifications in the AIS lesiosn, based on recurrently amplified chromosomal segments identified in TCGA data. For each tumour region amplifications and gains were determined relative to the mean ploidy. Each cytoband range defines a chromosomal segment, and each segment is annotated by predicted driver genes contained within the segment.

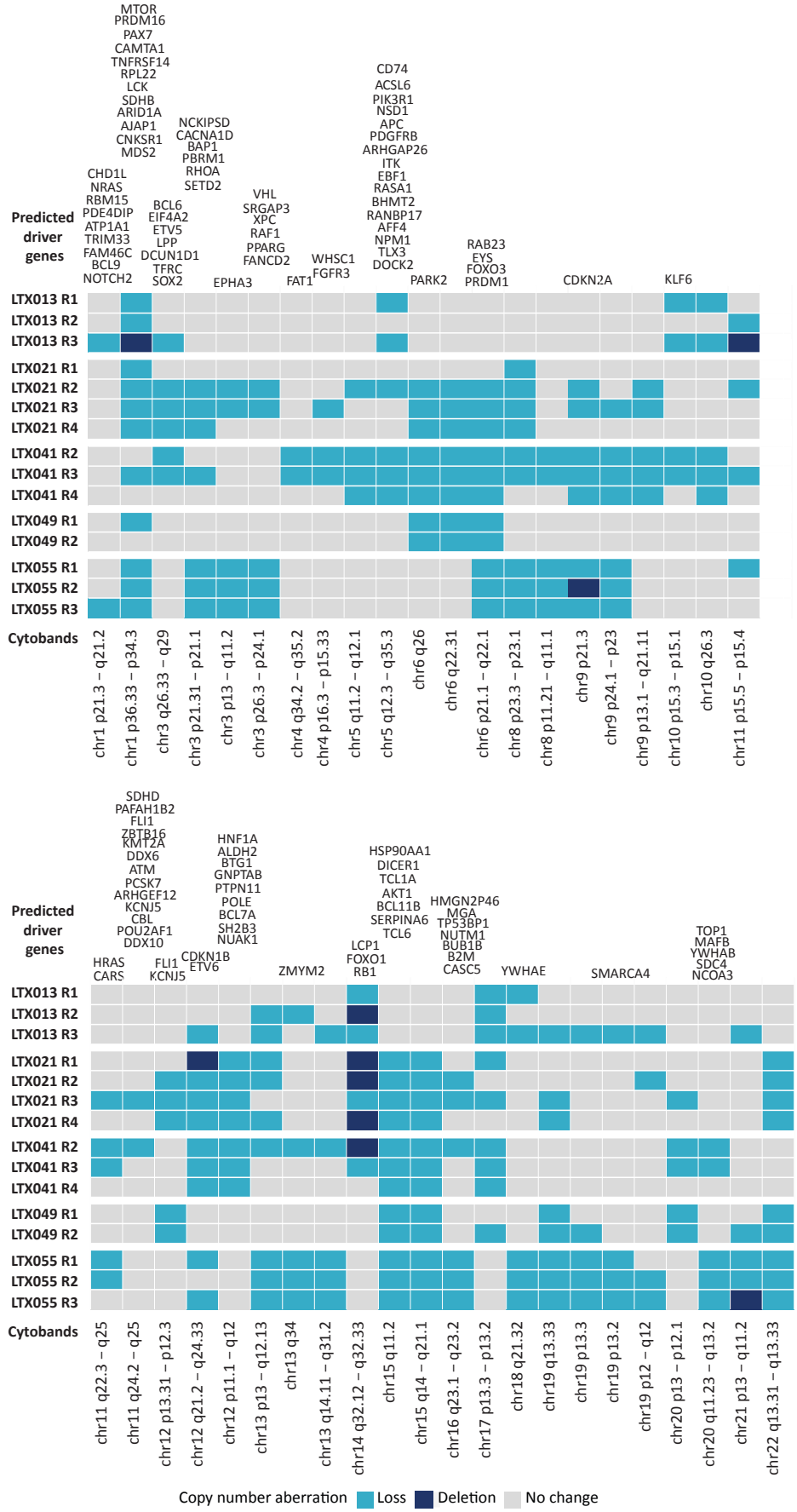

Predicted driver genes
MTOR
PRDM16
PAX7
CAMTA1
TNFRSF14
RPL22
LCK
SDHB
ARID1A
AJAP1
CNKSR1
MDS2
CHD1L
NRAS
RBM15
PDE4DIP
ATP1A1
TRIM33
FAM46C
BCL9
NOTCH2
NCKIPSD
CACNA1D
BAP1
PBRM1
RHOA
SETD2
BCL6
EIF4A2
ETV5
LPP
DCUN1D1
TFRC
SOX2
VHL
SRGAP3
XPC
RAF1
PPARG
FANCD2
EPHA3
FAT1
WHSC1
FGFR3
CD74
ACSL6
PIK3R1
NSD1
APC
PDGFRB
ARHGAP26
ITK
EBF1
RASA1
BHMT2
RANBP17
AFF4
NPM1
TLX3
DOCK2
PARK2
RAB23
EYS
FOXO3
PRDM1
CDKN2A
KLF6
LTX013 R1
LTX013 R2
LTX013 R3
LTX021 R1
LTX021 R2
LTX021 R3
LTX021 R4
LTX041 R2
LTX041 R3
LTX041 R4
LTX049 R1
LTX049 R2
LTX055 R1
LTX055 R2
LTX055 R3
Cytobands
chr1 p21.3 – q21.2
chr1 p36.33 – p34.3
chr3 q26.33 – q29
chr3 p21.31 – p21.1
chr3 p13 – q11.2
chr3 p26.3 – p24.1
chr4 q34.2 – q35.2
chr4 p16.3 – p15.33
chr5 q11.2 – q12.1
chr5 q12.3 – q35.3
chr6 q26
chr6 q22.31
chr6 p21.1 – q22.1
chr8 p23.3 – p23.1
chr8 p11.21 – q11.1
chr9 p21.3
chr9 p24.1 – p23
chr9 p13.1 – q21.11
chr10 p15.3 – p15.1
chr10 q26.3
chr11 p15.5 – p15.4
Predicted driver genes
SDHD
PAFAH1B2
FLI1
ZBTB16
KMT2A
DDX6
ATM
PCSK7
ARHGEF12
KCNJ5
CBL
POU2AF1
DDX10
HRAS
CARS
FLI1
KCNJ5
HNF1A
ALDH2
BTG1
GNPTAB
PTPN11
POLE
BCL7A
SH2B3
NUAK1
CDKN1B
ETV6
ZMYM2
HSP90AA1
DICER1
TCL1A
AKT1
BCL11B
SERPINA6
TCL6
LCP1
FOXO1
RB1
HMGN2P46
MGA
TP53BP1
NUTM1
BUB1B
B2M
CASC5
YWHAE
SMARCA4
TOP1
MAFB
YWHAB
SDC4
NCOA3
LTX013 R1
LTX013 R2
LTX013 R3
LTX021 R1
LTX021 R2
LTX021 R3
LTX021 R4
LTX041 R2
LTX041 R3
LTX041 R4
LTX049 R1
LTX049 R2
LTX055 R1
LTX055 R2
LTX055 R3
Cytobands
chr11 q22.3 – q25
chr11 q24.2 – q25
chr12 p13.31 – p12.3
chr12 q21.2 – q24.33
chr12 p11.1 – q12
chr13 p13 – q12.13
chr13 q34
chr13 q14.11 – q31.2
chr14 q32.12 – q32.33
chr15 q11.2
chr15 q14 – q21.1
chr16 q23.1 – q23.2
chr17 p13.3 – p13.2
chr18 q21.32
chr19 q13.33
chr19 p13.3
chr19 p13.2
chr19 p12 – q12
chr20 p13 – p12.1
chr20 q11.23 – q13.2
chr21 p13 – q11.2
chr22 q13.31 – q13.33
Copy number aberration
Loss
Deletion
No change

Figure 40 Copy number losses and/or deletions in AIS lesions

Heat map representing the regional distribution of potential driver copy number losses in the AIS lesions, based on recurrently deleted chromosomal segments identified in TCGA data. For each tumour region losses were determined relative to the mean ploidy. Each cytoband range defines a chromosomal segment, and each segment is annotated by predicted driver genes contained within the segment.

4.6 Conclusions and discussion

Whilst sequencing studies have given us some insight into the heterogeneous genomic landscape of tumours, and the genetic aberrations involved in tumourigenesis, our understanding of the early stages of cancer development remains limited. The somatic mutational burden, the dynamics of potentially competing populations of cells, and the effect of driver mutations, in pre-invasive solid tumour disease is yet to be determined. Since the development of cancer, and the evolution of tumours, is thought to rely on the accumulation of driver mutations in cancer-related genes (Vogelstein et al. 2013), the existence of acquired somatic mutations, including those in known driver genes, in pre-invasive lesions is not surprising. Furthermore, recent evidence has shown the existence of significant mutational burden, and high frequency of oncogenic driver mutations subject to positive clonal selection in physiologically normal skin cells (Martincorena et al. 2015).

The spatial and temporal heterogeneity of mutations and copy number aberrations seen in these AIS lesions suggests that prior to the potential development of invasive disease, these pre-invasive lesions have already undergone a degree of clonal evolution, such that they are composed of different populations of cells harbouring specific driver mutations. Although the identified drivers in these lesions may have not led to the development of invasive disease at the time of surgical resection, it is conceivable that under certain microenvironmental selection pressures, such as hypoxia or altered vasculature, these drivers may have the opportunity to play a role in driving clonal expansion, and therefore invasive transformation. Further investigation of pre-invasive lung lesions may reveal essential components of the pathways involved in the early stages of cancer development, and the combination of driver mutations necessary to aid the pre-invasive to invasive transition.

5 Results 3: Detection of genetic aberrations in cfDNA from patients with NSCLC

5.1 Introduction

Several studies have demonstrated the use of cfDNA in tracking tumour evolution and mutational burden in patients with cancer over time (Murtaza et al. 2014; Diehl et al. 2008; Diehl et al. 2005; Dawson et al. 2013; Diaz et al. 2012; Carreira et al. 2014; Siravegna et al. 2015). Circulating biomarkers have the potential to aid early detection and guide treatment initiation and stratification. Given the sampling bias associated with single tumour biopsies, and the difficulties in repeated tumour sampling, the use of cfDNA to identify therapeutically targetable mutations and the emergence of therapeutic resistance mechanisms, may have significant implications for the management of patients with cancer. The accurate identification of, and the distinction between, clonal and subclonal mutations in cfDNA can give some insight into the heterogeneous genomic landscape of tumours. Furthermore, the use of such circulating biomarkers to monitor tumour clonal dynamics has the potential to enable real-time monitoring of tumour evolution at clinically relevant stages of disease, such as disease progression and the development of resistance to therapy (Jamal-Hanjani, Quezada, et al. 2015). Further evidence supporting the use of circulating biomarkers to detect early stage cancers may inform cancer screening strategies, and the detection of tumors at earlier time points in their evolution. Such strategies could potentially aid early therapeutic intervention prior to the onset of intratumour heterogeneity.

The concordance between mutations detected in tumours and cfDNA in early stage disease remains unclear. Furthermore, exactly how representative cfDNA is of the underlying genomic landscape of tumours, and whether subclonal, as well as clonal, mutations can be successfully detected is yet to be determined. In this chapter, three different approaches were used to detect aberrations in cfDNA extracted from patients with early stage NSCLC in order to define an optimal approach to address these questions. Mutations were chosen based on multi-region WES and Ion AmpliSeq deep-sequencing validation data from a selection of patients in the NSCLC pilot cohort from whom plasma was collected prior to surgical resection of their primary tumours. Both ubiquitous truncal mutations, predicted to be clonal in origin, and heterogeneous branch mutations, predicted to be subclonal in origin, were chosen to represent the

heterogeneous landscape of each tumour, and to determine the sensitivity of the different approaches in detecting mutations with low variant allele frequencies (VAFs). Having previously defined reliable VAF thresholds to determine the presence or absence of a mutation in each tumour region (Section 2.5.1), a VAF threshold of 1% for SNVs and 2% for indels was used. Where Ion AmpliSeq data was not available, WES data was used. Germ line DNA was extracted from whole blood using the QIAamp DNA Blood Mini Kit (Qiagen). Plasma was extracted from approximately 10ml whole blood samples collected from patients prior to surgical resection of their primary NSCLC tumours (Section 2.3). The plasma extraction protocol was established in collaboration with Professor Jacqueline Shaw (University of Leicester), and is the recommended protocol by the ECMC and NCRI Biomarker and Imaging Clinical Studies Group (cfDNA consensus meeting, 2014). cfDNA was extracted from approximately 2ml plasma samples using the QIAamp Circulating Nucleic Acid kit (Qiagen).

In collaboration with Agena Bioscience (Michael Mosko and Anders Nygren, San Diego, CA, USA), the presence of mutations in cfDNA was analysed using multiplex PCR and Matrix Assisted Laser Desorption/Ionization-Time of Flight (MALDI-TOF) mass spectrometry, in collaboration with Illumina (Claire Fielding and Mark Ross, Cambridge, UK), multiplex PCR and targeted MiSeq sequencing was used, and in collaboration with Natera (Robert Pelham, San Carlos, CA, USA), multiplex PCR and targeted HiSeq sequencing was used.

5.2 Multiplex PCR and MALDI-TOF mass spectrometry

In collaboration with Agena Bioscience (Michael Mosko and Anders Nygren, San Diego, CA, USA), this approach involved determining the presence of mutations in cfDNA using a novel minor variant detection technique called UltraSEEK, which involves multiplex PCR followed by a mutation specific single base extension reaction and characterisation using MALDI-TOF mass spectrometry. Six patients (L003, L008, L011, L013, L017 and L019) from the NSCLC pilot cohort were analysed using this approach (**Table 13**). Patients L003 and L008 were from a previously published NSCLC pilot cohort (de Bruin et al. 2014), which has also been set-up prior to the TRACERx study. The UltraSEEK technique was used on the germ line DNA, DNA from each tumour region, and cfDNA. Mutations were detected using a median absolute deviation (MAD)-

based z-score, which was calculated for each assay using the median and the MAD values previously established with cfDNA samples from normal healthy volunteers (Section 2.12.2.6). This data served as the baseline for the mutation detection and was used for each data analysis. Samples that exceeded the user-defined assay z-score cut-off (default of 10) and met the peak quality criteria (adjustable minimum peak intensity and the call probability of 0.8 or better) were labelled as containing the mutation by the analysis software (Typer Software, Agena Bioscience).

Patient ID	Age	Gender	Histology	Stage	Vascular invasion	Pleural invasion	Smoking status (pack years)
L003	84	F	LUAD	IIIB	N	N	Never-Smoker
L008	75	M	LUAD	IIIA	Y	Y	Ex-Smoker (25)
L013	68	F	LUSC	IB	Y	Y	Smoker (50)
L017	61	F	LUAD	IIB	Y	N	Smoker (48)
L019	47	M	LUAD	IIIA	N	Y	Never-smoker
L011	49	F	LUAD	IB	N	N	Smoker (45)

Table 13 Clinical characteristic of the patient cohort.

Abbreviations: LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.

In patient L003, all selected mutations were identified in the tumour and lymph node DNA, with a regional distribution concordant with the sequencing data (**Table 14**). However, none of these mutations were identified in the cfDNA. **Figure 41** shows the average adjusted normalised intensity for each mutation in the germ line, tumour DNA and cfDNA compared to the wild type control DNA.

Gene (amino acid change)	Truncal/Branch	Ion AmpliSeq/WES VAF (%)			UltraSEEK detection			
		LN	R2	R4	LN	R2	R4	cfDNA
EGFR (L858R)	Truncal	20.8	6.1	68.8	✓	✓	✓	nd
MUC1 (H116D)	Branch	6.3	0	15.8	✓		✓	nd
BRAF (G30D)	Branch	19.0	0	0	✓			nd
MLL (S2834L)	Branch	0	0	9.5			✓	nd
JAK1 (E613K)	Branch	0	6.4	0		✓		nd
CTNNB1 (S37F)	Branch	0	7.1	0		✓		nd
FOXP1 (Y25X)	Branch	0	16.3	0		✓		nd
IL7R (P132S)	Branch	0	17.7	0		✓		nd

Table 14 Summary of UltraSEEK results for L003

The VAF for each selected mutation from the Ion AmpliSeq or WES data, and whether the mutation was detected in the tumour DNA and cfDNA is shown here. Genes highlighted in red represent category 1 to 3 predicted driver mutations. Abbreviations: nd, not detected.

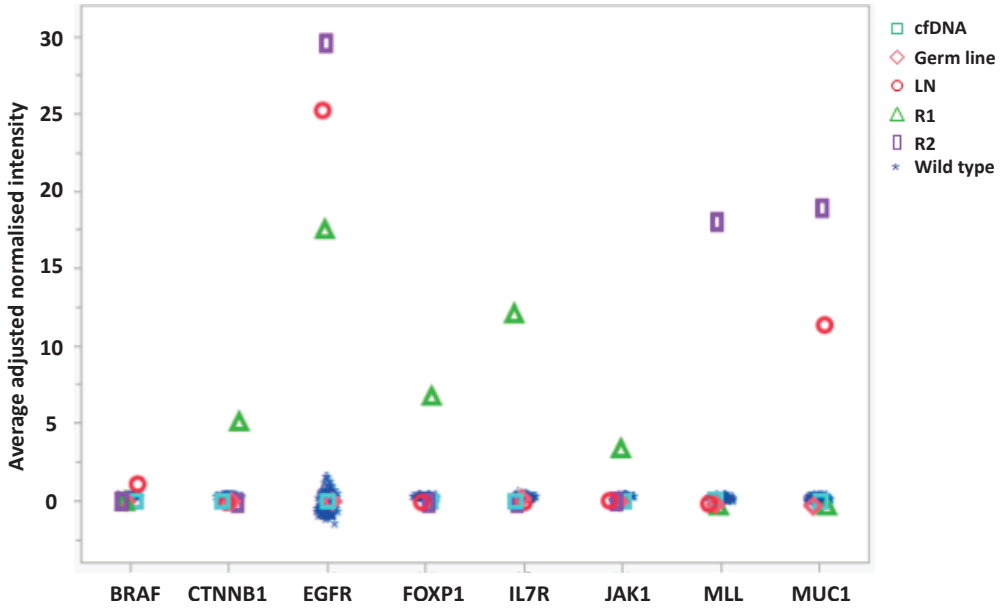


Figure 41 Distribution of detected mutations in L003

The average adjusted normalised intensity for each mutation (gene name on the x-axis) in the germ line, tumour, and wild type control DNA, and cfDNA.

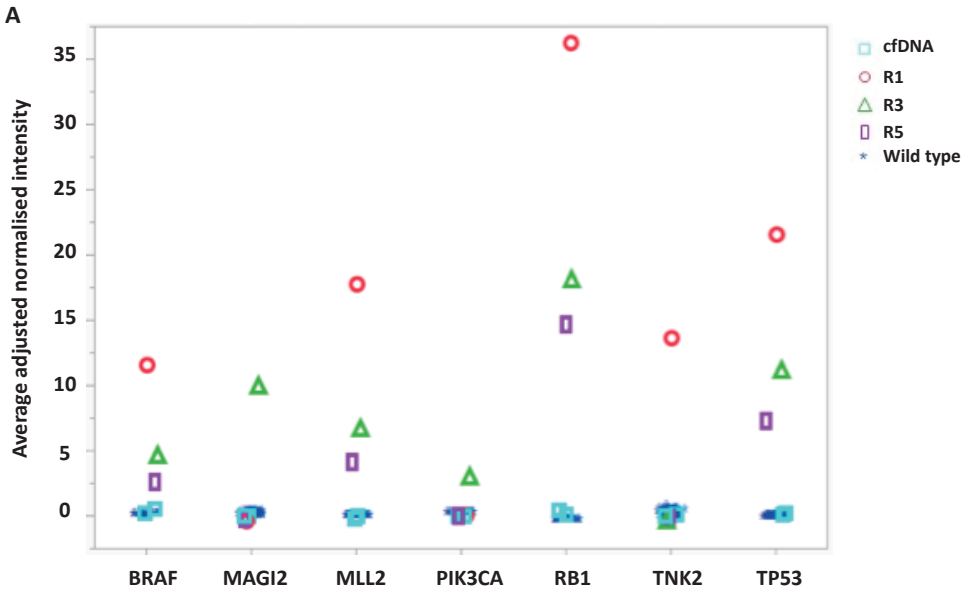
In patient L008, all the selected mutations were identified in the tumour DNA, with a regional distribution concordant with the sequencing data (**Table 15**). Four truncal

mutations were identified in the cfDNA; *BRAF* (G469A), *MLL2* (A5010T), *RB1* (5829X) and *TP53* (H193Y) (**Figure 42**A). cfDNA detection of these mutations was confirmed for *BRAF*, *RB1* and *TP53* in two replicate assays, and for *MLL2* in one replicate assay (**Figure 42**B).

Gene (amino acid change)	Truncal/Branch	Ion AmpliSeq/WES VAF (%)			UltraSEEK detection			
		R1	R3	R5	R1	R3	R5	cfDNA
BRAF (G469A)	Truncal	28.3	10.5	2.9	✓	✓	✓	✓
MLL2 (A5010T)	Truncal	41	10.5	6.5	✓	✓	✓	✓
RB1 (S829X)	Truncal	44.4	12.6	5.6	✓	✓	✓	✓
TP53 (H193Y)	Truncal	43	9.8	8.1	✓	✓	✓	✓
PIK3CA (E542K)	Branch	0	6.7	0		✓		nd
MAGI2 (D575N)	Branch	0	4.4	0		✓		nd
TNK2 (G6S)	Branch	6.6	0	0	✓			nd

Table 15 Summary of UltraSEEK results for L008

The VAF for each selected mutation from the Ion AmpliSeq or WES data, and whether the mutation was detected in the tumour DNA and cfDNA is shown here. Genes highlighted in red represent category 1 to 3 predicted driver mutations. Abbreviations: nd, not detected.



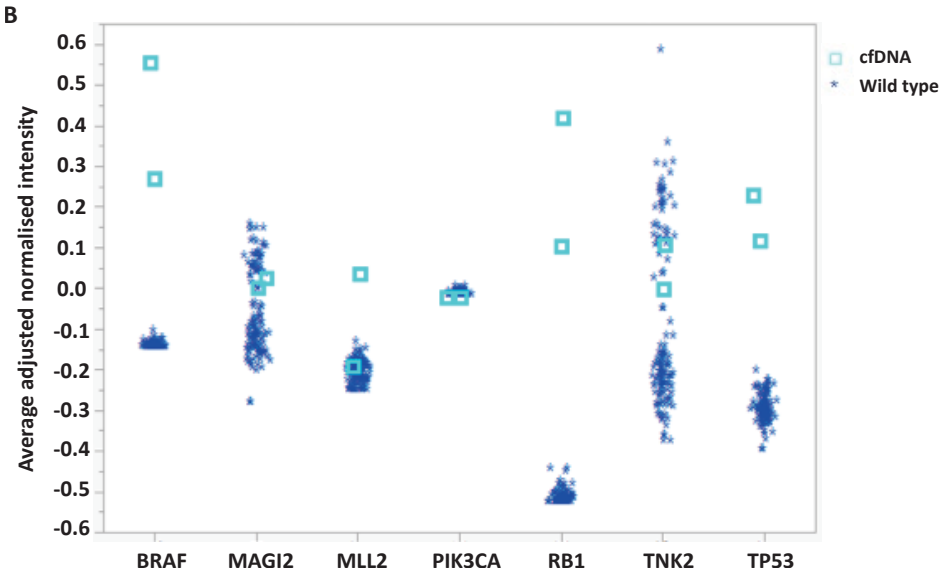


Figure 42 Distribution of mutations in L008

The average adjusted normalised intensity for each mutation (gene name on the x-axis) in the germ line, tumour, and wild type control DNA, and cfDNA (A), magnified portion of the y-axis with lower intensities showing the separation between wild type control and replicate cfDNA assays (B).

In patient L011, all the selected mutations were identified in the tumour DNA, with a regional distribution concordant with the sequencing data (**Table 16**). None of the selected mutations was identified in the cfDNA (**Figure 43**).

Gene (amino acid change)	Truncal/Branch	Ion AmpliSeq/WES VAF (%)			UltraSEEK detection			
		R1	R2	R3	R1	R2	R3	cfDNA
BRAF (V600E)	Truncal	19.3	31.2	15.8	✓	✓	✓	nd
TP53 (E166X)	Truncal	22.6	32	19.5	✓	✓	✓	nd
OR10Z1 (R270S)	Branch	6.8	0	0	✓			nd
LGALSL (D75Y)	Branch	9.2	0	0	✓			nd
CSMD3 (G2956R)	Branch	0	10.7	0		✓		nd
CHD2 (K1022M)	Branch	0.1	8.5	0		✓		nd
GPX6 (K106fs)	Branch	0.7	1	7.8	nd	nd	✓	nd
HSP90AA1	Branch	0	0	9.4			✓	nd

Table 16 Summary of UltraSEEK results for L011

The VAF for each selected mutation from the Ion AmpliSeq or WES data, and whether the mutation was detected in the tumour DNA and cfDNA is shown here. Genes highlighted in red represent category 1 to 3 predicted driver mutations. Abbreviations: fs, frameshift mutation; nd, not detected.

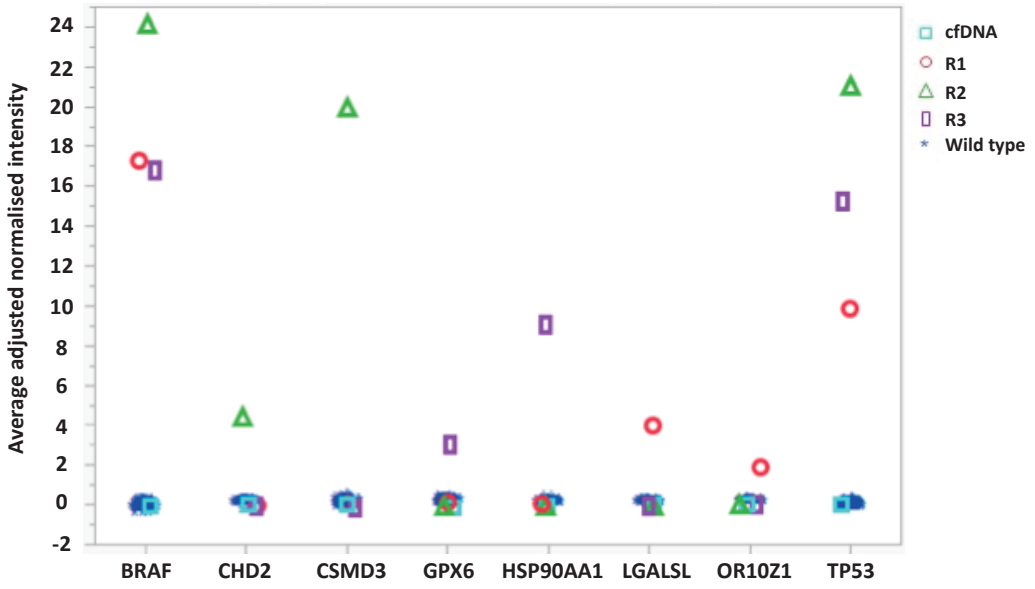


Figure 43 Distribution of mutations in L011

The average adjusted normalised intensity for each mutation (gene name on the x-axis) in the germ line DNA, tumour DNA from each respective region, wild type control DNA, and cfDNA.

In patient L013, all the selected mutations were identified in the tumour DNA, with a regional distribution concordant with the sequencing data. In addition, UltraSEEK detected the mutation in *JAK2* in tumour region R1, which was below the sequencing reliable detection threshold (VAF of 0.2%) (**Table 17**). One truncal mutation, *EGFR* (D761Y), was identified in the cfDNA (**Figure 44**).

Gene (amino acid change)	Truncal/Branch	Ion AmpliSeq/WES VAF (%)			UltraSEEK detection			
		R1	R2	R3	R1	R2	R3	cfDNA
EGFR (G719A)	Truncal	21.4	24.8	55.2	✓	✓	✓	nd
EGFR (D761Y)	Truncal	20.1	17.3	48	✓	✓	✓	✓
TP53 (P21S)	Truncal	7	4.3	15.6	✓	✓	✓	nd
MSH2 (Q510H)	Truncal	5	4	11	✓	✓	✓	nd
JAK2 (K33X)	Branch	(0.2)	11	7	✓	✓	✓	nd
MLL3 (K589R)	Branch	0	0	4.1			✓	nd
FBXO46 (R133W)	Branch	0	0	9.5			✓	nd

Table 17 Summary of UltraSEEK results for L013

The VAF for each selected mutation from the Ion AmpliSeq or WES data, and whether the mutation was detected in the tumour DNA and cfDNA is shown here. Genes highlighted in red represent category 1 to 3 predicted driver mutations. Abbreviations: nd, not detected.

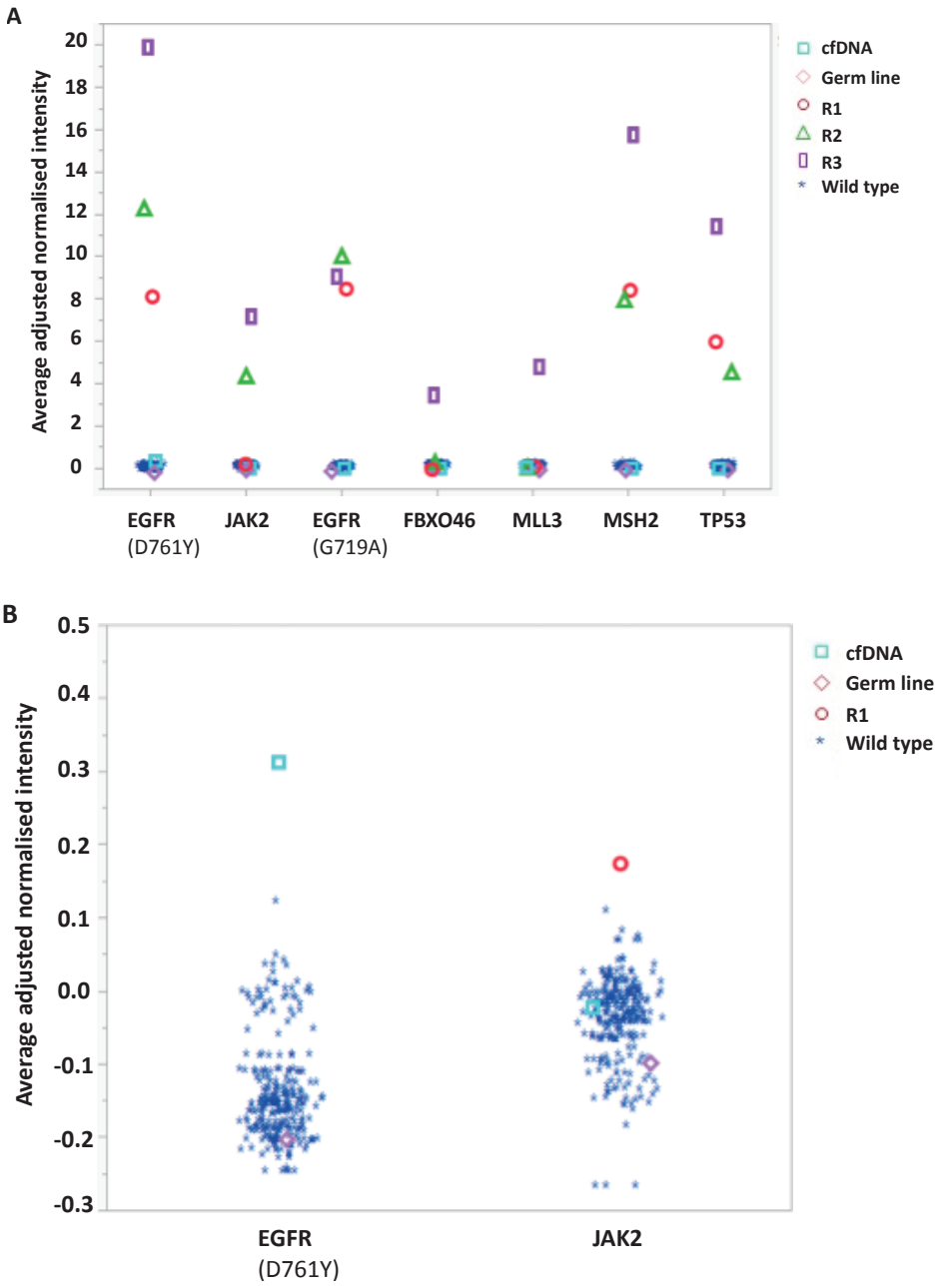


Figure 44 Distribution of mutations in L013

The average adjusted normalised intensity for each mutation (gene name on the x-axis) in the germ line, tumour, and wild type control DNA, and cfDNA (A), magnified portion of the y-axis with lower intensities showing the separation between wild type control and cfDNA assays (B).

In patient L017, all the selected mutations were identified in the tumour DNA, with a regional distribution concordant with the sequencing data. In addition, UltraSEEK detected the mutation in *TRIM67* in tumour region R1, which was below the reliable sequencing detection threshold (VAF of 0.21%) (**Table 18**). The mutation in *EIF2AKI*

was not assessed due to assay failure, and none of the selected mutations were identified in the cfDNA (**Figure 45**).

Gene (amino acid change)	Truncal/Branch	Ion AmpliSeq/WES VAF (%)			UltraSEEK detection			
		R1	R2	R3	R1	R2	R3	cfDNA
KRAS (G12V)	Truncal	20	22	0	✓	✓		nd
BRCA2 (S2156F)	Truncal	16	28	0	✓	✓		nd
TP53 (E155X)	Truncal	19.5	34.9	0	✓	✓		nd
NF1 (C1690Y)	Truncal	20	25	0	✓	✓		nd
PAX8 (P273L)	Branch	9	0	0	✓			nd
TRIP11 (E897Q)	Branch	6.6	0	0	✓			nd
TRIM67 (R298C)	Branch	(0.21)	66	0	✓	✓		nd
TP53 (splice variant)	NA	0	0	16			✓	nd
NF1 (R366X)	NA	0	0	20			✓	nd

Table 18 Summary of UltraSEEK results for L017

The VAF for each selected mutation from the Ion AmpliSeq or WES data, and whether the mutation was detected in the tumour DNA and cfDNA is shown here. Genes highlighted in red represent category 1 to 3 predicted driver mutations. Abbreviations: nd, not detected; NA, not applicable.

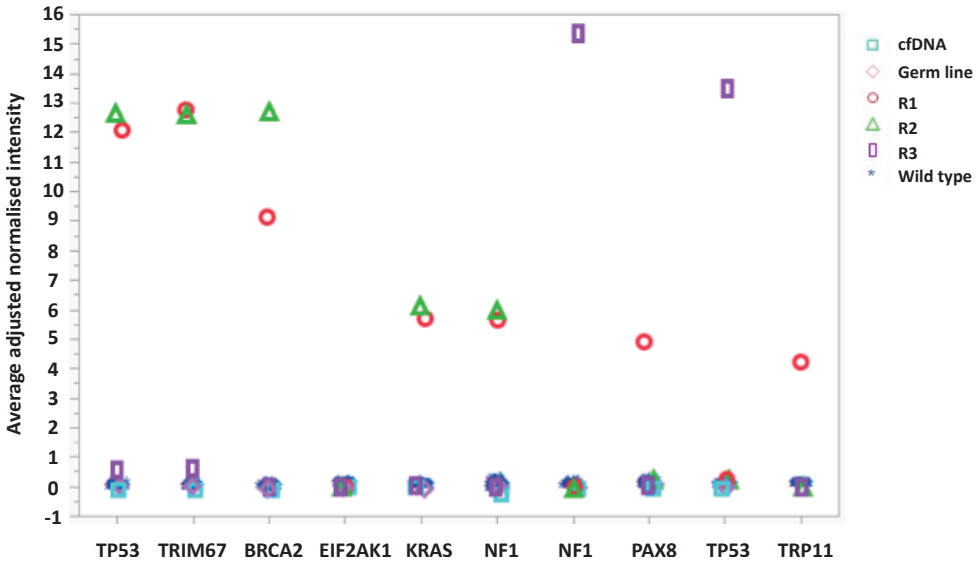


Figure 45 Distribution of mutations in L017

The average adjusted normalised intensity for each mutation (gene name on the x-axis) in the germ line, tumour, and wild type control DNA, and cfDNA.

In patient L019, the UltraSEEK approach identified all the selected mutations in the tumour DNA, with a regional distribution concordant with the sequencing data. The UltraSEEK approach was able to detect the mutation in *KNTC1* in R1, which was unexpected since this mutation was found to be absent in this region based on the WES data with 0/467 mutant reads (VAF = 0%). **Table 19** summarises the results for patient L019. The mutations in *ARID5A* and *MYCN,* and the EML4-ALK translocation were not assessed due to assay failure (**Figure 46**A). In addition, the mutation in *LEPRE1* was identified in R1 and LN3, despite a VAF of 0.8% and 0.1%, respectively (**Figure 46**B).

Gene (amino acid change)	**Truncal/Branch**	**Ion AmpliSeq/WES VAF (%)**				**UltraSEEK detection**				
		R1	**LN1**	**LN2**	**LN3**	**R1**	**LN1**	**LN2**	**LN3**	**cfDNA**
AKAP9 (T1741A)	Truncal	11	13	13	3	✓	✓	✓	✓	nd
XRRA1 (R100S)	Branch	0	10.3	19.2	3.3		✓	✓	✓	nd
KNTC1 (Q173R)	Branch	0	18	19	4	✓	✓	✓	✓	nd
LEPRE1 (D36Y)	Branch	(0.8)	1.7	1.4	(0.1)	✓	✓	✓	✓	nd
POU2AF1 (S89F)	Branch	4	0	0	0	✓				nd

Table 19 Summary of UltraSEEK results for L019

The VAF for each selected mutation from the Ion AmpliSeq or WES data, and whether the mutation was detected in the tumour DNA and cfDNA is shown here. Genes highlighted in red represent category 1 to 3 predicted driver mutations. Abbreviations: nd, not detected.

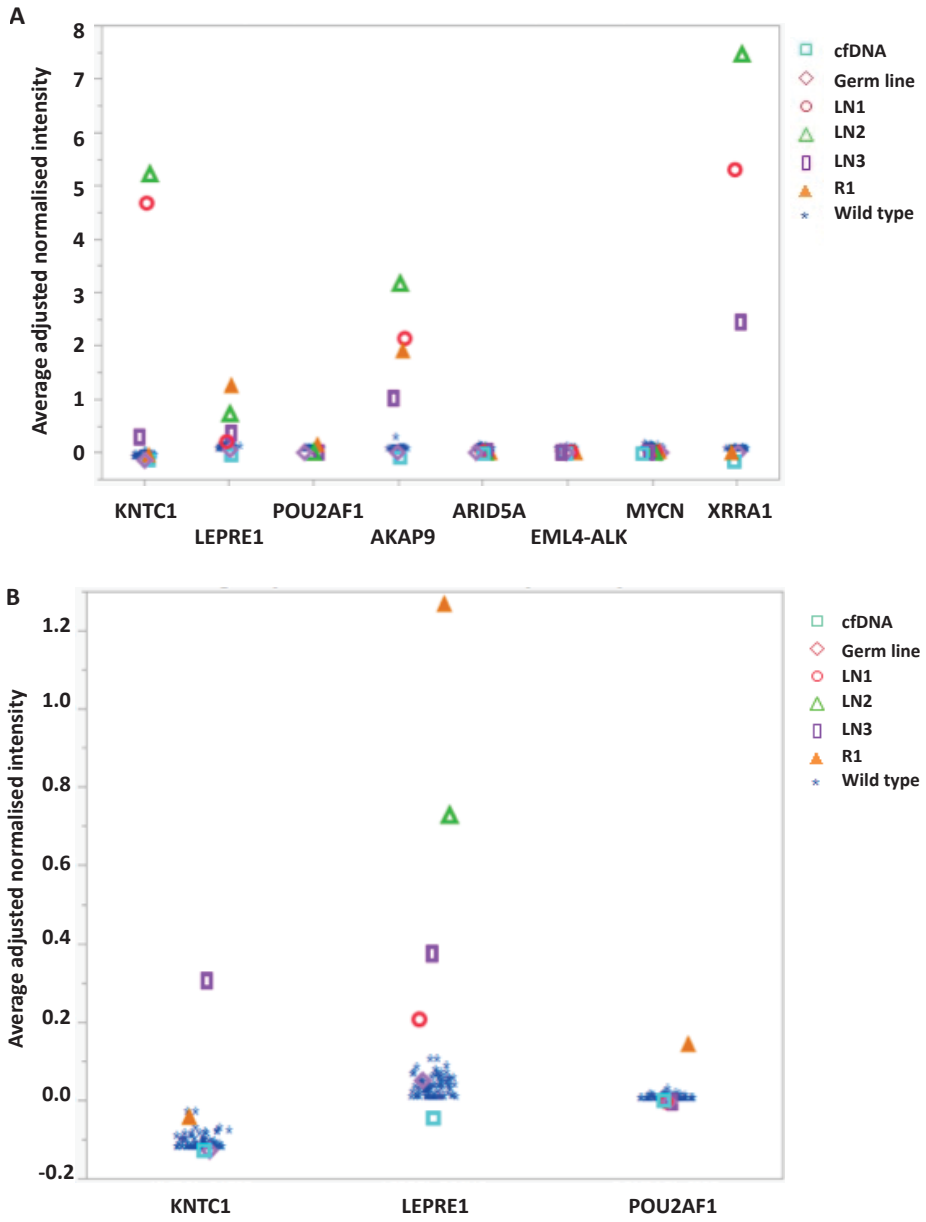


Figure 46 Distribution of mutations in L019

The average adjusted normalised intensity for each mutation (gene name on the x-axis) in the germ line, tumour, and wild type control DNA, and cfDNA (A), magnified portion of the y-axis with lower intensities showing the separation between wild type control and cfDNA assays (B).

5.2.1 Summary of the UltraSEEK technique

The UltraSEEK technique was successful in detecting 44/48 (92%) of the selected mutations in tumour DNA, providing validation for a targeted approach in tumour DNA. In general, the pattern of detection in the tumour DNA was in keeping with the regional distribution for each mutation based on the sequencing data using a VAF threshold for detection of 1%. There were 4 failed assays in which the predicted aberration was not

identified in tumour DNA using the UltraSEEK technique. Three of these were mutations that had failed Ion AmpliSeq amplification (*EIF2AK1, ARID5A* and *MYCN),* in which case the WES data was used instead to predict the presence or absence of the mutation in each tumour region. The mutations in *EIF2AK1* and *ARID5A* were both C>A mutations that were subsequently identified as sequencing artefacts associated with the Nextera library reparation kit (Illumina). A failure in primer design was thought to be the reason why the mutation in *MYCN* was not detected. Finally, a failure to detect the EML4-ALK translocation was also thought to be due to a failure in primer design. Therefore, this technique was able to detect 44/46 of the chosen variants in tumour and lymph node DNA. Although the UltraSEEK technique was able to detect mutations in tumour DNA with VAFs as low as 0.2% (*JAK2* in L013 and *TRIM67* in L017) and 0.1% (*LEPRE1* in L019), this was not consistently seen across all tumours, since several other mutations with similar or even higher VAFs were not detected. Only 5/44 (11%) of the identified mutations were detected in cfDNA. Four of these mutations (*BRAF* G469A, *MLL2* A5010T, *RB1* 5829X, and *TP53* H193Y) were detected in patient L008, who was a 75-year-old male with stage IIIA lung adenocarcinoma (LUAD) and evidence for both vascular and pleural invasion (**Table 13**). He was an ex-smoker with a 25 pack-year history, having stopped smoking 10-years ago. The remaining mutation (*EGFR* D761Y) was detected in patient L013, who was a 68-year-old female with stage IB lung squamous cell carcinoma (LUSC) and evidence for both vascular and pleural invasion. She was a current smoker with a 50 pack-year history. All five mutations were predicted to be truncal events based on the sequencing data, and were either category 1 or 2 predicted driver mutations. These mutations were amongst those with higher tumour VAFs compared to others (48% for *EGFR* in L013, 28.3% for *BRAF* in L008, 41% for *MLL2* in L008, 44.4% for *RB1* in L008, and 43% for *TP53* in L008). There was a statistically significant difference in the median tumour VAF between the mutations that were detected in cfDNA (20.3%, range 13.9 to 28.5%), and those that were not detected (9.3%, range 1.3 to 33.8%) (Mann-Whitney test, $p = 0.01207$), suggesting that mutations with higher tumour VAFs are more likely to be detected in cfDNA using this approach.

5.3 Multiplex PCR and targeted MiSeq sequencing

In collaboration with Illumina (Claire Fielding and Mark Ross, Cambridge, UK), this approach involved determining the presence of mutations in cfDNA using multiplex

PCR and targeted MiSeq sequencing in one patient (L019) from the NSCLC pilot cohort (**Table 20**).

Patient ID	Age	Gender	Histology	Stage	Vascular invasion	Pleural invasion	Smoking status (pack years)
L019	47	M	LUAD	IIIA	N	Y	Never-smoker

Table 20 Clinical characteristics of Patient L019

Abbreviations: LUAD, lung adenocarcinoma.

Targeted MiSeq sequencing was performed using germ line, tumour DNA and cfDNA. Five SNVs, 2 indels, and one EML4-ALK translocation were analysed in the cfDNA, with a mean sequencing coverage of 309,065x (range 9,495x - 701,653x) (**Table 21**). The number of mutant reads and VAFs for each variant are shown in **Figure 47** and **Figure 48**. The SNV mutation in *ARID5A* had low mutant reads and VAFs across all samples. This mutation was subsequently found to be a sequencing artefact associated with the Nextera XT library preparation kit (Illumina), and was therefore not taken forward. Variant mutant reads were detected in the germ line and control DNA samples, which was not unexpected at this level of ultra-deep sequencing. Furthermore, since germ line DNA was extracted from whole blood, mutant reads present in the germ line DNA may have been associated with contamination from circulating tumour DNA. As expected, the number of mutant reads (range 5 to 203) and the VAFs (range 0.015 to 0.094%) identified for each variant in the control DNA were significantly lower compared with the tumour and lymph node samples.

In order to determine whether variants were detected in the cfDNA, the number of mutant reads and the VAFs for each variant were compared with their corresponding control DNA results. In all variants, apart from the mutation in *ZNF175* and *XRRA1*, the number of mutant reads detected in the cfDNA was higher compared to the control DNA. This was most evident in the *DIS3L2* (truncal indel), *LEPRE1* (branch SNV), and the EML4-ALK translocation variants (**Figure 47**). Only the variants in *DIS3L3* and *LEPRE1* had VAFs that were higher in the cfDNA compared with the control DNA (**Figure 48**). These data suggest that the variants in *DIS3L3, LEPRE1*, and EML4-ALK may have been successfully detected in the cfDNA. However, with too few data points it is not possible to assign any statistical significance to this, and therefore the data is only suggestive. Patient L019 was also investigated using the UltraSEEK technique,

with three overlapping mutations between the two approaches (*XRRA1*, *POU2AF1* and *LEPRE1*). None of these mutations were detected in cfDNA using the UltraSEEK technique.

5.3.1 Summary

In summary, this approach detected mutations in the tumour DNA from patient L019, with a regional distribution in keeping with the sequencing data. Of the eight variants that were tested in cfDNA, one of the mutations (*ARID5A*) was disregarded since it was a sequencing artefact, and although the data suggest that the variants in *DIS3L3, LEPRE1*, and EML4-ALK may have been successfully detected in the cfDNA, this cannot be interpreted with any statistical confidence.

Gene			ZNF175	DIS3L2	XRRA1	LEPRE1	POU2AF1	MYCN	EML4 -ALK
	Germ line	WT reads	351983	334273	623397	274003	265457	1211	
		Mutant reads	55	65	251	208	54	1	18
		VAF (%)	0.016	0.019	0.040	0.076	0.020	0.083	
	R1	WT reads	331354	249215	556200	244527	253986	3945	
		Mutant reads	36280	24827	260	29106	11650	3	7003
		VAF (%)	9.9	9.1	0.047	10.6	4.4	0.076	
	LN1	WT reads	291810	249820	475613	266533	243945	2500	
		Mutant reads	43077	28155	48933	36819	220	1	8541
		VAF (%)	12.9	10.1	9.3	12.1	0.090	0.040	
	LN2	WT reads	267553	228526	428459	301177	223676	4037	
		Mutant reads	45946	39748	74027	55837	49	613	8215
		VAF (%)	14.7	14.8	14.7	15.6	0.022	13.2	
	LN3	WT reads	268721	264745	454662	358989	222236	4731	
		Mutant reads	12524	11038	12335	16924	44	21	1749
		VAF (%)	4.5	4.0	2.6	4.5	0.020	0.442	
	Control	WT reads	246309	204118	603413	151341	205466	9969	
		Mutant reads	36	38	203	142	50	5	37
		VAF (%)	0.015	0.019	0.034	0.094	0.024	0.050	
	cfDNA	WT reads	637046	585543	1769824	1319598	468821	40073	
		Mutant reads	3	190	156	806	68	25	277
		VAF (%)	0.001	0.033	0.010	0.198	0.019	0.035	

Table 21 Summary results for L019

Results of targeted MiSeq sequencing (performed by Illumina) of 8 variants in L019. Mutant and wild type reads for each variant are shown, with corresponding VAFs. Genes highlighted in red represent category 1 to 3 predicted driver events.

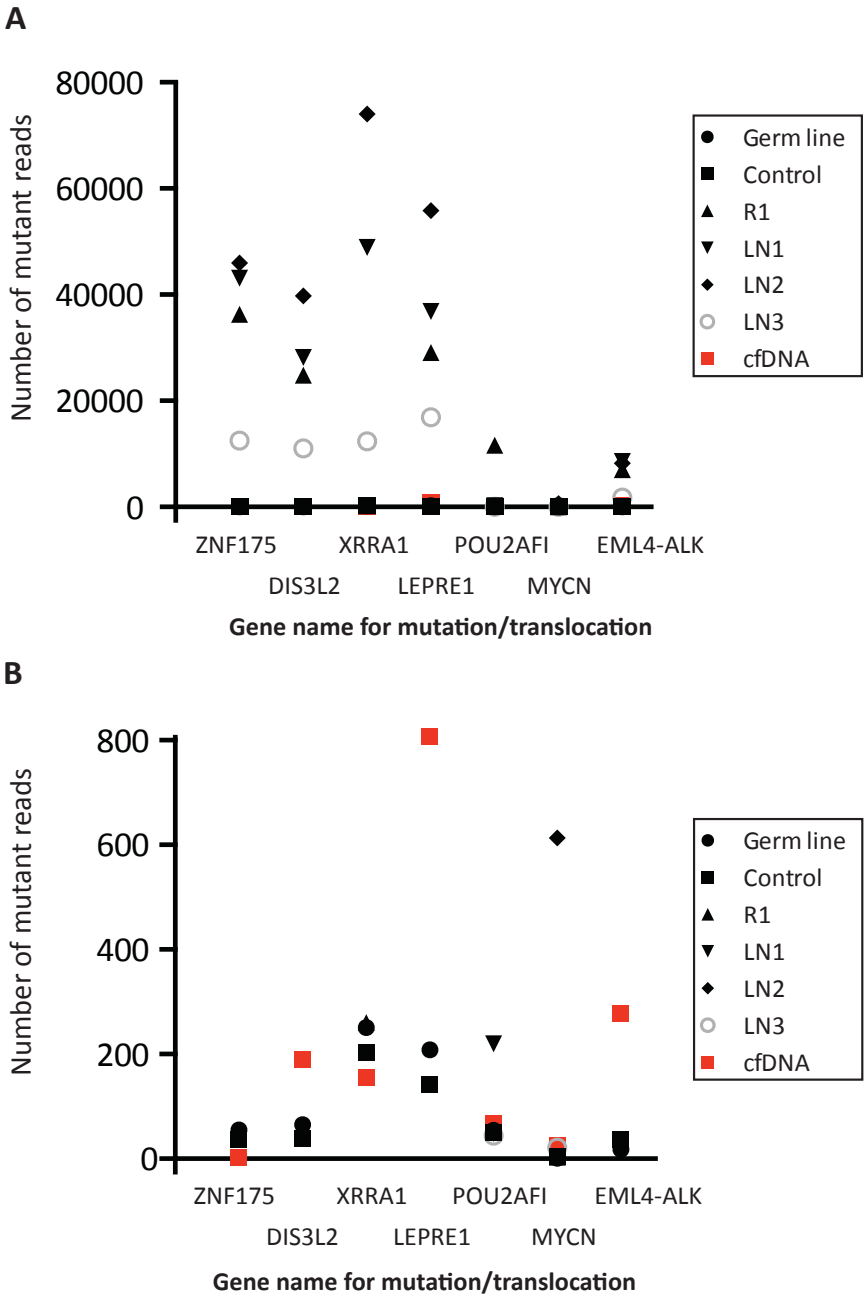


Figure 47 Number of mutant reads comparison between tumour, control, and cfDNA

The number of mutant reads for each mutation/translocation in germ line, control, tumour region R1, lymph node regions LN1 to LN3 DNA, and cfDNA in patient L019 (A), magnified y-axis scale demonstrating the detected variants with low number of mutant reads (B).

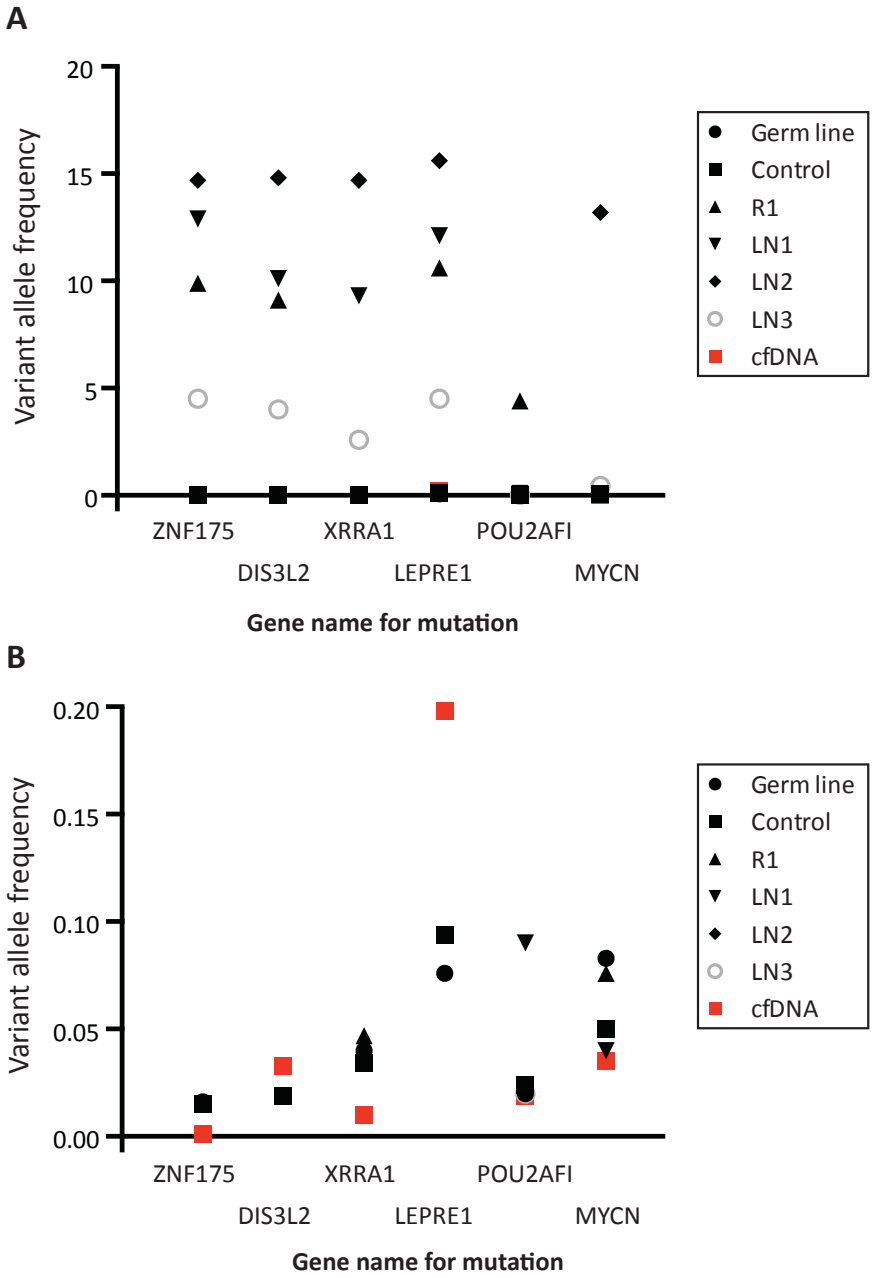


Figure 48 VAF comparison between tumour, control, and cfDNA

The variant allele frequency (VAF) for each mutation in germ line, control, tumour region R1, lymph node regions LN1 to LN3 DNA, and cfDNA (A), magnified y-axis scale demonstrating the detected variants with low number of mutant VAFs (B).

5.4 Multiplex PCR and targeted HiSeq sequencing

In collaboration with Natera (Robert Pelham, San Carlos, CA, USA), this approach involved determining the presence of mutations in cfDNA using multiplex PCR and targeted HiSeq sequencing. Four patients (L012, L013, L015 and L017) from the NSCLC pilot cohort were analysed using this approach (**Table 22**). Fifty SNV mutations were selected for analysis, including both ubiquitous truncal and heterogeneous branch mutations. This approach was used on the germ line DNA, DNA from each tumour region, and cfDNA for each patient, as well as cfDNA obtained from presumed-normal healthy volunteers representing negative controls for each mutation (n=48). Primers designed for each mutation were combined in a multiplex PCR approach, followed by targeted sequencing on the HiSeq 2500 system (Illumina). The median coverage per mutation target in cfDNA was 62,399x. One mutation (*CYFIP1* in patient L012) was disregarded, since it was subsequently found to be a sequencing artefact associated with the Nextera XT library preparation kit (Illumina). Of the remaining 49 SNV mutations, 12 assays failed to detect a mutation in the tumour DNA or cfDNA due to low coverage. A total of 37 mutations were taken forward.

Patient ID	Age	Gender	Histology	Stage	Vascular invasion	Pleural invasion	Smoking status (pack years)
L012	69	F	LUSC	IB	Y	Y	Smoker (40)
L013	68	F	LUSC	IB	Y	Y	Smoker (50)
L015	68	M	LUSC	IA	N	N	Smoker (100)
L017	61	F	LUAD	IIB	Y	N	Smoker (48)

Table 22 Clinical characteristics of the patient cohort

Abbreviations: LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.

In the germ line and tumour DNA there was good concordance between the VAFs identified by prior sequencing (WES and IonAmpliSeq), and those identified by the Natera approach, with a significant linear correlation between the two (coefficient of determination, $R^2 = 0.9144$, $p < 0.0001$) (**Figure 49**). The VAFs ranged from 0.1 to 65.7% with prior sequencing, and 0.15 to 60.8% with the Natera approach (**Table 23**).

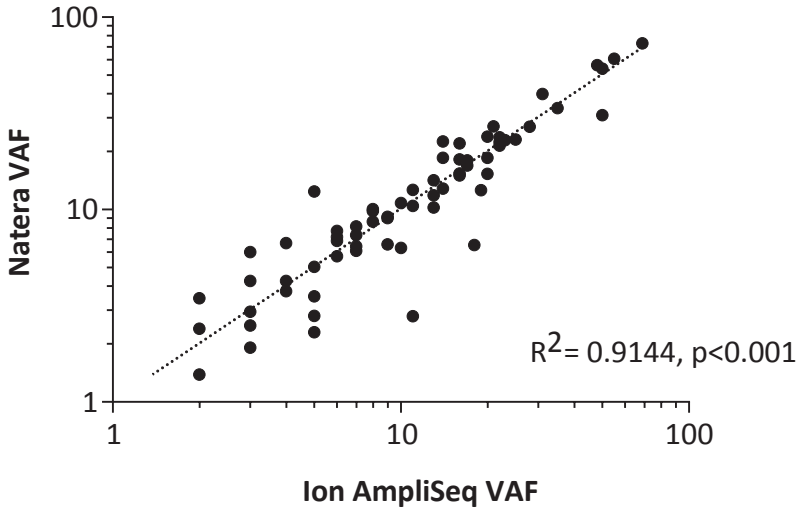


Figure 49 Concordance between VAFs

Concordance between individual mutation variant allele frequencies (VAFs) and mean VAF (vertical line) identified by prior sequencing (WES and Ion AmpliSeq) and the Natera approach by direct comparison (A), and by linear regression modeling (B).

Patient	Gene (amino acid change)	WES/Ion AmpliSeq (VAF %)			Natera (VAF %)			cfDNA (VAF %)	Truncal/ Branch
		R1	R2	R3	R1	R2	R3		
L012 LUSC Stage IB Smoker (40)	BRIP1 (R173C)	14.0	6.0	8.0	22.6	7.7	10.1	3.71	Truncal
	CARS (T51A)	22.0	11.0	16.0	22.4	10.5	18.2	5.02	Truncal
	CIC (E61X)	0.1	nd	7.0	nd	nd	8.2	1.71	Branch
	FAT1 (D924Y)	8.0	4.0	1.2	9.8	4.3	0.8	0.57	Branch
	KDM6A (R4079K)	10	5.0	0.8	6.3	3.5	nd	0.28	Branch
	MLLT4 (E478Q)	9.0	3.0	0.3	9.0	4.3	0.7	0.95	Branch
	NFE2L2 (E1142D)	69.0	31.0	50	73.1	39.9	53.9	23.25	Truncal
	RASA1 (E183X)	6.5	nd	0.7	7.4	nd	0.4	nd	Branch
	TP53 (R136H)	23.0	9.0	17.0	22.9	9.2	17.0	4.89	Truncal
	TP53 (Y181C)	22.0	8.0	16.0	21.5	8.7	22.1	5.77	Truncal
L013 LUSC Stage IB Smoker (50)	EGFR (G719A)	21.4	24.8	55.2	27.1	23.1	60.8	1.16	Truncal
	EGFR (D761Y)	20.1	17.3	48.0	23.9	18.0	56.3	1.09	Truncal
	HERC4 (N217K)	0.1	3.0	6.0	0.2	1.9	5.7	nd	Branch
	JAK2 (K33X)	0.2	11.1	6.8	0.3	2.8	6.1	nd	Branch
	MLL3 (K589R)	0.2	0.2	4.1	nd	nd	6.7	nd	Branch
	MSH2 (Q444H)	4.8	4.2	11.0	5.0	3.3	12.7	nd	Truncal
	MTOR (Q838E)	2.0	0.6	3.0	2.4	1.4	6.0	nd	Truncal
	PLCG2 (E525X)	5.1	1.7	10.5	2.8	1.7	6.9	nd	Truncal
	TP53 (P60S)	7.0	4.3	15.6	6.1	3.8	15.4	0.4	Truncal
L015 LUSC Stage IA Smoker (100)	ALK (P234R)	6.0	2.0	NS	7.2	1.4	NS	nd	Truncal
	GABRG1 (I279F)	13.0	0.8	NS	11.8	2.0	NS	nd	Truncal
	KDM6A (S539C)	18.0	6.0	NS	6.5	0.9	NS	0.17	Truncal
	MLL2 (E1186X)	5.0	0.3	NS	12.4	nd	NS	nd	Branch
	ROS1 (Y891C)	10.0	3.0	NS	10.8	2.9	NS	0.15	Truncal
	SLC39A4 (A546T)	14.0	nd	NS	18.6	nd	NS	nd	Branch
	TP53 (G199X)	19.0	3.0	NS	12.6	2.5	NS	nd	Truncal

	ZFHX4 (K3595N)	13.0	5.0	NS	10.3	2.3	NS	nd	Truncal
	ZMYM4 (E257K)	14.0	2.0	NS	12.8	3.5	NS	nd	Truncal
L017 LUAD Stage IIB Smoker (48)	BRCA2 (S2156F)	16.0	28.0	nd	15.0	27.0	nd	nd	Truncal
	KRAS (G12V)	13.0	21.6	nd	14.2	23.8	nd	nd	Truncal
	NF1 (C1711Y)	19.7	24.9	nd	18.6	23.1	nd	nd	Truncal
	NF1 (R366X)	nd	nd	20.0	nd	nd	17.8	0.37	NA
	PAX8 (P350L)	9.3	nd	nd	6.6	nd	nd	nd	Branch
	TP53 (E155X)	nd	nd	16.4	nd	nd	19.0	nd	NA
	TP53 (Splice variant)	19.5	34.9	nd	15.4	33.7	nd	nd	Truncal
	TRIM67 (R298C)	21.0	65.7	nd	18.4	31.0	nd	0.57	Truncal
	TRIP11 (E897Q)	6.6	nd	nd	6.4	nd	nd	nd	Branch

Table 23 List of selected mutations for the Natera approach

List of mutations and variant allele frequencies (VAFs) identified using WES/AmpliSeq sequencing and the Natera approach. VAFs are shown for each tumor region. Genes highlighted in red represent category 1 to 3 predicted driver mutations. Abbreviations: NA, not applicable; nd, not detected; NS, no sample; LUSC, lung squamous cell carcinoma; LUAD, lung adenocarcinoma.

In patient L012, from the 10 selected mutations, 10/10 were identified in the tumour DNA, and 9/10 were identified in the cfDNA. Four of the mutations detected in the cfDNA were branch mutations (*CIC, FAT1, KDM6A,* and *MLLT4)*. In patient L013, from the 9 selected mutations, 9/9 were identified in the tumour DNA, and 3/9 were identified in the cfDNA. The mutations identified in the cfDNA were all truncal mutations (*EGFR* G719A, *EGFR* D761Y, and *TP53* P60S). This patient was also investigated using the UltraSEEK technique, and of the 6 overlapping selected mutations, 3 were detected in cfDNA using the Natera approach compared with only 1 (*EGFR* D761Y) using the UltraSEEK technique. In patient L015, from the 9 selected mutations, 9/9 were identified in the tumour DNA, and 2/9 were identified in the cfDNA. The mutations identified in the cfDNA were truncal mutations (*KDM6A* and *ROS1*). In patient L017, from the 9 selected mutations, 9/9 were identified in the tumour DNA, and 2/9 were identified in the cfDNA (*NF1* and *TRIM67*). This patient presented with two tumours at diagnosis; regions R1 and R2 came from one tumour, and region R3 came from

another tumour. The two mutations detected in the cfDNA were from separate tumours. The mutation in *TRIM67* was truncal, and the status of the other mutation (*NF1*) could not be determined since it was identified in the second tumour from which only one region (R3) had been collected. This patient was also investigated using the UltraSEEK technique, and of the 9 overlapping selected mutations, 2 were detected in cfDNA using the Natera approach compared with none using the UltraSEEK technique. The median cfDNA VAF for truncal mutations was 1.16% (range 0.15 to 23.25%), and for branch mutations it was 0.76% (range, 0.28 to 1.71%) (**Table 23**). The number of branch mutations was too small for a statistical comparison between the two groups. Although the median tumour VAF for mutations that were detected in cfDNA was higher compared to those that were not detected (14.92%, range 3.27 to 55.63 vs. 6.9%, range 2.6 to 24.55%), this was not statistically significant (Mann-Whitney test, $p = 0.1584$). To confidently call the mutations detected in cfDNA, and to estimate the background noise for each assay, the selected 37 mutations were also analysed in cfDNA from 48 presumed-normal healthy volunteers (**Figure 50**). The mean cfDNA VAF for the 37 mutations in the healthy volunteer cohort ranged from 0.001 to 0.343% (standard deviation range 0.001 to 0.036%). These samples represented normal baseline controls with background VAFs for comparison with mutation VAFs detected in cfDNA from patients. Since there was a normal distribution for VAF in the healthy volunteer cohort, for each mutation, a z-score was calculated using the mean VAF and standard deviation in this cohort of samples (negative controls), and the corresponding VAF in the patient samples. If the z-score was > 10, the mutation was considered to be present. None of the 16 mutations detected in cfDNA from the patient cohort were detected in cfDNA from the healthy volunteer cohort ($p < 0.00001$ for each mutation).

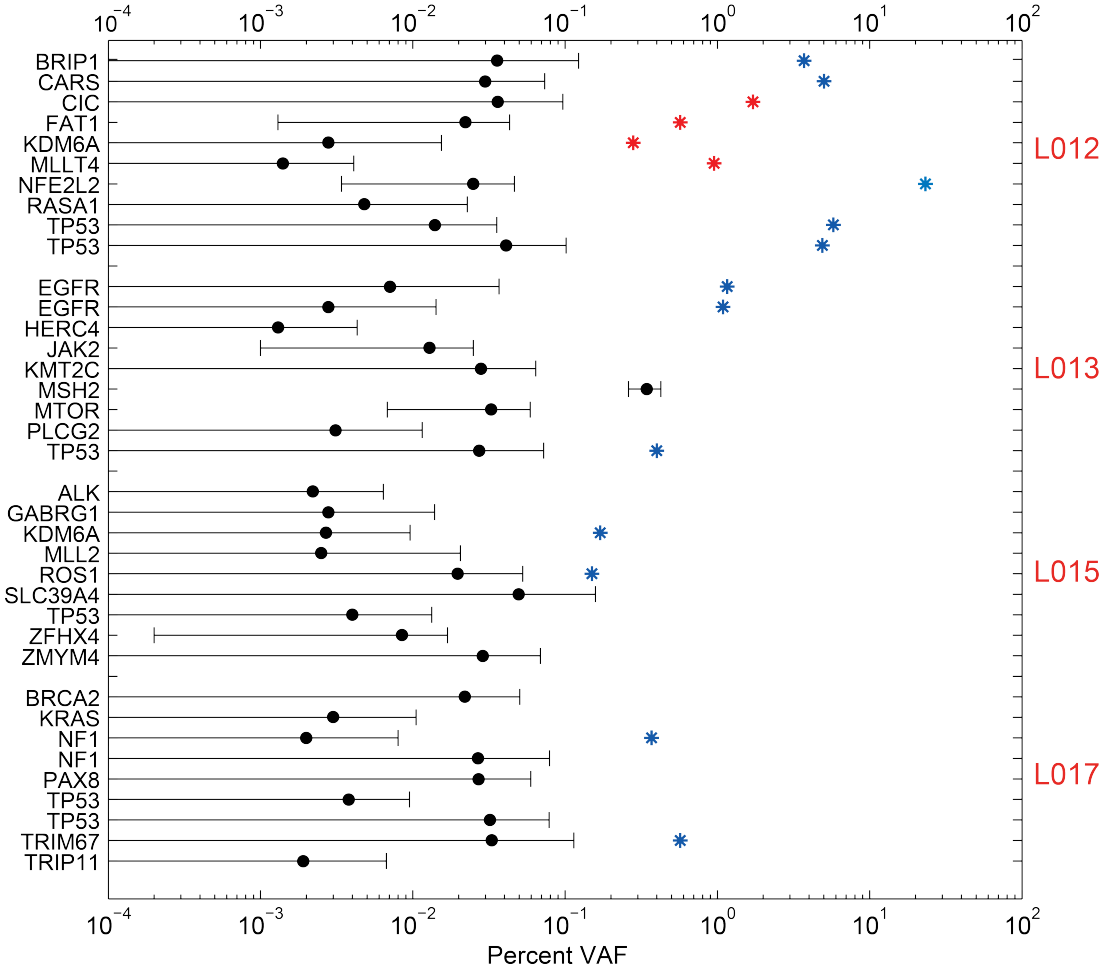


Figure 50 Detection of the mutations in cfDNA

Detection of truncal and branch mutations in cfDNA from presumed-normal healthy volunteers (n=48) and patients L012, L013, L015 and L017. The mean VAF ± 3 standard deviations (SD) for the negative control cfDNA samples is shown in black, and the VAF for the 16 mutations detected in the patient cfDNA samples is shown in blue for the truncal mutations, and in red for the branch mutations ($p<0.00001$ for all 16 mutations).

5.4.1 Summary

In summary, this approach was able to detect all 37 selected mutations in tumour DNA, providing validation for the use of this technique in tumour profiling. However, less than half (16/37, 43%) of these mutations were detected in cfDNA. Four out of 16 (25%) of these were branch mutations, and 15/16 (94%) of these were predicted to be clinically relevant driver mutations. Overall, the 37 mutations consisted of 23 truncal mutations, 11/23 (48%) of which were detected in cfDNA, and 12 branch mutations, 4/12 (33%) of which were detected in cfDNA. Two out of the 37 mutations were from tumour region R3 in patient L017, for which truncal or branch status could not be determined.

5.5 Conclusions and discussion

The main aim of the work presented here was to determine how representative cfDNA was of the mutational burden in early stage primary NSCLC. In particular, the aim was to determine whether both truncal and branch mutations could be detected, such that the heterogeneous genomic landscape of tumours, and therefore a measure of intratumour heterogeneity, could be interpreted using cfDNA. In this chapter, three different approaches were used to detect a selection of mutations in cfDNA from a cohort of early stage NSCLC patients. All three approaches involved using mutation panels customised for each patient's tumour based on prior deep sequencing. A multiplex PCR (mPCR) approach was combined with either mass spectrometry or next-generation sequencing (MiSeq or HiSeq). The UltraSEEK technique (mPCR/mass spectrometry) detected 5/44 (11.4%) SNVs (all truncal) in cfDNA, and the Natera approach (mPCR/HiSeq) detected 16/37 (43.2%) SNVs in cfDNA (4 branch and 2 truncal). The Natera approach detected the greatest proportion of selected mutations in cfDNA. In patients L013 and L017, both the UltraSEEK and Natera approaches were used, with the Natera approach detecting a greater number of mutations in comparison for both tumour DNA and cfDNA. Although it appeared that the Illumina approach (mPCR/MiSeq) was able to detect an SNV, an indel, and a translocation variant in cfDNA, this was based on single assays and overall there were very few selected mutations for testing, such that a measure of statistical significance could not be assigned. **Table 24** summarises the results for all three approaches.

A correlation between tumour burden or stage and the ability to detect genetic aberrations in circulating biomarkers, has been shown in several tumour types, including colorectal (Diehl et al. 2008), breast (Board et al. 2010; Dawson et al. 2013), ovarian (Forshew et al. 2012; Kuhlmann et al. 2012), hepatocellular (Chan et al. 2013), pancreatic (Yamada et al. 1998), and NSCLC (Nygaard et al. 2013). In the overall cohort of eight patients analysed here, there was no significant correlation identified between cfDNA mutation detection and stage ($p = 1$), lymph node involvement ($p = 1$), vascular invasion ($p = 0.4286$), pleural invasion (0.4286), or smoking status (0.4643). This is likely to have been due to the small sample size.

The data presented here are not without limitations. The use of a relatively small amount (approximately 1ml) of plasma available for analysis may have influenced the ability to detect mutations in cfDNA. Increased plasma volume and further assay

optimisation, could considerably improve the analytical sensitivity of these approaches, and potentially allow the detection of mutations at lower VAFs, such as heterogeneous branch mutations. The incomplete concordance between mutations detected in tumour DNA and cfDNA may also have been a reflection of biological factors that can influence the release of tumour DNA into circulation, such as the extent of tumour vascularisation, rate of tumour growth and cellular apoptosis (Diaz & Bardelli 2014). In some cases, branch mutations that were predicted to be absent from some tumour regions based on prior Ion AmpliSeq and/or WES, were identified in these regions using the cfDNA approaches discussed here. Although the use of different sequencing platforms and greater sequencing coverage with the Illumina and Natera approaches may have increased the ability of these techniques to identify variants with low VAFs, any discrepancies between these approaches and prior sequencing was mostly due to the fact that with prior sequencing VAF thresholds were used to predict the presence or absence of a mutation in each tumour region. These thresholds were derived based on the multi-region WES and Ion AmpliSeq validation sequencing performed in the NSCLC pilot cohort, but if these thresholds were not used the concordance between prior sequencing and the MiSeq/HiSeq cfDNA approached would be 100%. In addition, the tumours presented here did not have copy number data, making it difficult to confidently rely on the VAFs obtained from variant calling alone, and therefore potentially affecting the comparison between VAFs obtained in cfDNA from NSCLC patients and from normal healthy volunteers in the Natera approache. There was one mutation in *KNTC1* in tumour region R1 for patient L019, that was predicted to be absent with prior WES based on 0/467 mutant reads (VAF 0%) in this region, but was identified using the UltraSEEK technique. This approach involved the use of mass spectrometry, with little difference between the average adjusted normalised intensities for the mutation in region R1 and the wild type control, making it difficult to confidently interpret this result.

Current assessments of treatment response and disease progression rely on clinical examination and the presence of measurable disease on radiological imaging. The detection of existing or new drivers of disease in plasma may precede such assessments, and therefore aid the early detection of disease recurrence and emerging drug resistance. This may in turn guide early therapeutic intervention, or the modification of ongoing treatment strategies. In addition, determining the clonality of these driver mutations may help identify the overriding drivers of disease at any given

time point during the disease course. Formal validation is required in clinical studies, such as TRACERx, where the use of circulating biomarkers is tested as part of the study endpoints to truly assess its utility in clinical practice and predicting survival outcome (Crowley et al. 2013). As a result of the sequencing artefacts and failed assays discussed here that were found to be associated with the Nextera XT library preparation kit (Illumina), this kit was not taken forward for use in the sequencing of tumours from the TRACERx lung study.

Pt	Age	G	Smoking status (pack-years)	Histo	Stage	Tumour size (max diameter, mm)	Lymph node involve-ment	Mutation detected in cfDNA	T/B	Detected by
L008	75	M	Ex-smoker (25)	LUAD	IIIA	15 and 24	Yes	BRAF (G469A)	T	UltraSEEK
								MLL2 (A5010T)	T	UltraSEEK
								RB1 (S829X)	T	UltraSEEK
								TP53 (H193Y)	T	UltraSEEK
L012	69	F	Smoker (40)	LUSC	IB	40	No	BRIP1 (R173C)	T	Natera
								CARS (T51A)	T	Natera
								CIC (E61X)	B	Natera
								FAT1 (D924Y)	B	Natera
								KDM6A (R4079K)	B	Natera
								MLLT4 (E478Q)	B	Natera
								NFE2L2 (E1142D)	T	Natera
								TP53 (R136H)	T	Natera
								TP53 (Y181C)	T	Natera
L013	68	F	Smoker (50)	LUSC	IB	50	No	EGFR (D761Y)	T	UltraSEEK /Natera
								EGFR (G719A)	T	Natera
								TP53 (P21S)	T	Natera
L015	68	F	Smoker (100)	LUSC	IA	30	No	KDM6A (S539C)	T	Natera
								ROS1 (Y891C)	T	Natera
L017	61	F	Smoker (48)	LUAD	IIB	14 and 20	No	NF1 (R366X)	T	Natera
								TRIM67 (R298C)	T	Natera

Table 24 Summary results for the mutations detected in cfDNA

Clinical variables for patients in which cfDNA mutations were detected using the UltraSEEK, Illumina, and Natera approaches. Genes highlighted in red represent category 1 to 3 predicted driver mutations. Abbreviations: Pt, patient; G, gender; M, male; F, female; Histo, histology; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; T, truncal; B, branch.

6 Results 4: CIN in the TransTACT cohort

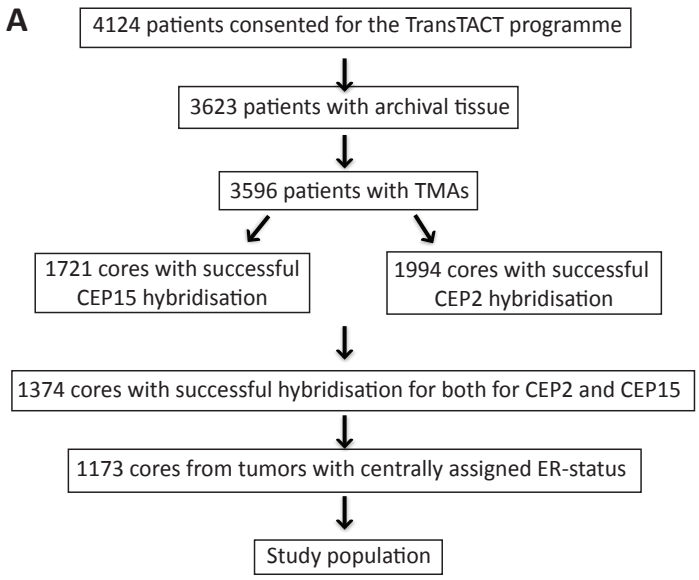
6.1 Introduction

Work from our laboratory has previously demonstrated a complex relationship between CIN and clinical outcome in ER-negative breast cancer (Roylance et al. 2011; Birkbak et al. 2011). In a discovery cohort of 246 patients with primary breast cancer, ER-negative breast cancers were found to have increased CIN and clonal heterogeneity relative to ER-positive breast cancers. Whilst there was a linear relationship between increasing CIN and worsening prognosis in ER-positive tumours, in ER-negative tumours extreme CIN was paradoxically associated with improved prognosis (Roylance et al. 2011; Birkbak et al. 2011). In this chapter the results of the largest validation cohort study are presented assessing the relationship between CIN and outcome, using the same established dual centromeric FISH assay with pre-defined CIN thresholds as the discovery cohort study (Jamal-Hanjani, A'Hern, et al. 2015). This cohort of patients was chosen from the TACT trial (Ellis et al. 2009), where detailed clinical and follow-up data were available for each patient. The CIN status of each tumour was assessed in relation to clinical and histopathological variables, treatment response and clinical outcome. As previously discussed (Section 2.13.2) TMA slides containing tumour cores for each patient were used to identify discrete hybridisation signals for chromosomes 2 and 15. Using previously established methods (Roylance et al. 2011), a CIN score was derived by counting the numbers of centromeres for chromosomes 2 and 15 in 40 nuclei per core. The mean (centromere counts for chromosomes 2 and 15 combined) percentage of cells deviating from the modal centromere number was used to define 4 CIN score groups (Modal Centromere Deviation groups 1 to 4: MCD1, 0%–15%; MCD2, 15%–30%; MCD3 30%–45%; and MCD4, >45%). Roger A'hern at the Institute of Cancer Research Clinical Trials and Statistics Unit performed the statistical analyses presented here.

6.2 The TransTACT TMA cohort

Thirty TMAs were assessed, with each TMA containing an average of 100 tumour cores. All TMA slides were hybridised using centromeric probes for chromosome 2 (CEP2) and 15 (CEP15). For some TMA slides hybridisation was not successful, and in some cases tumour cores had fallen off the slides. Hybridisation was successful in 1994 cores for CEP2, and in 1721 cores for CEP15. 1374 cores with successful

hybridisation for both CEP2 and CEP15 were selected, and those cores with full clinical and histopathological annotation and centrally assigned ER-receptor status were further selected to form the study population of 1173 patients (**Figure 51**A). The distribution of the percentage of nuclei deviating from the modal centromere count, and the thresholds used in the allocation of MCD groups is shown in **Figure 51**B. Representative images of nuclei from tumour cores allocated to different MCD groups are shown in **Figure 52**. No statistically significant differences were observed between the clinical and pathological characteristics of patients in the overall TACT trial compared with the TransTACT study population analysed here (**Table 25**).



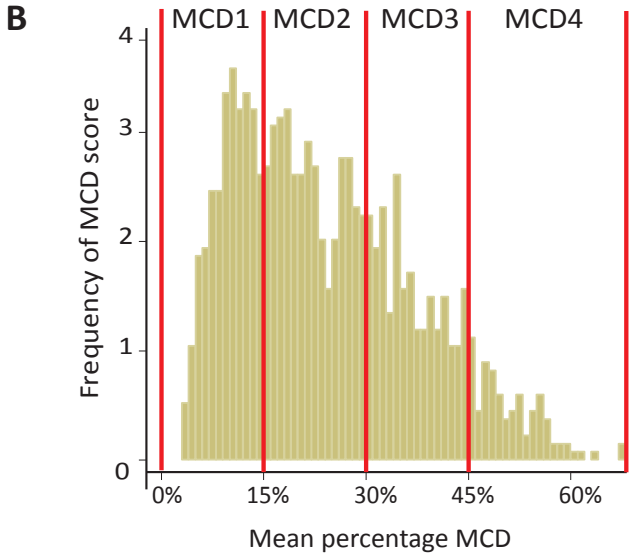


Figure 51 TransTACT study population and allocation of MCD groups

CONSORT diagram outlining the selected study population from the TACT study (A), and histogram distribution of the percentage of nuclei deviating from the modal centromere signals and allocation of MCD1 to MCD4 groups (B) (Jamal-Hanjani, A'Hern, et al. 2015).

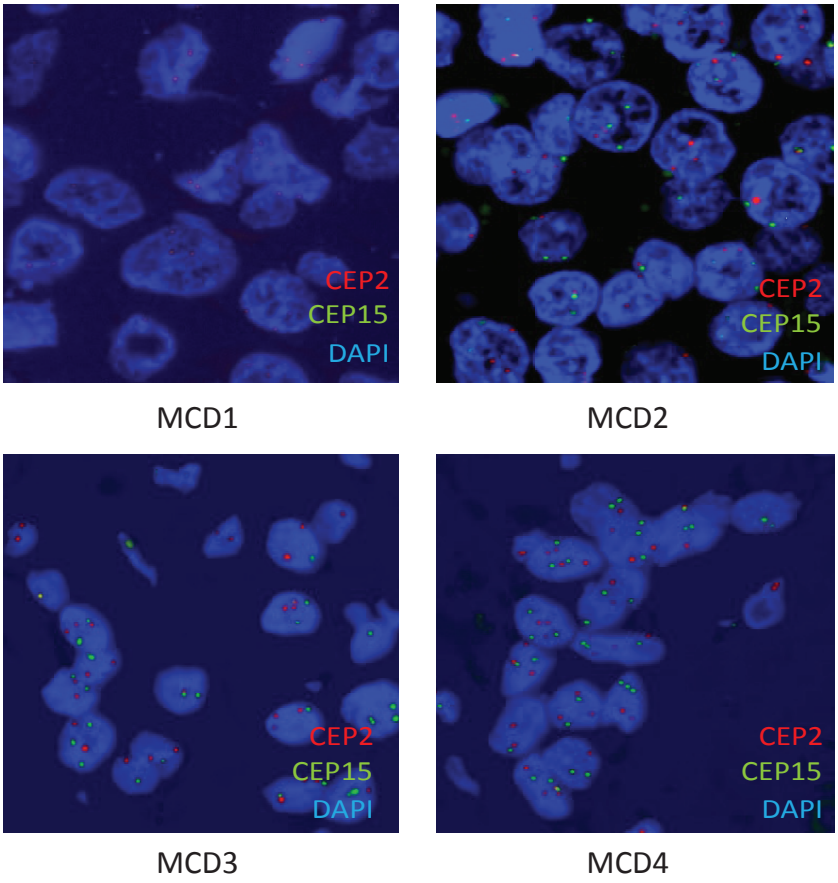


Figure 52 Dual centromeric FISH

Representative FISH images of MCD1 to MCD4 tumours with centromeric probes for chromosomes 2 and 15 labelled in red and green respectively, and DNA stained with DAPI, 40,6-diamidino-2-phenylindole (Jamal-Hanjani, A'Hern, et al. 2015).

	Study cohort (n=1173)	Whole cohort (n=4162)
	n (%)	n (%)
Median age in years (range)	48 (25-88)	49 (24-88)
Positive lymph nodes		
0	242 (20.6)	835 (20.1)
1-3	524 (44.7)	1839 (44.2)
4+	407 (34.7)	1488 (35.8)
Tumour size (cm)		
0-2	771 (65.7)	2657 (63.9)
2-5	311 (26.5)	1108 (26.6)
>5	91 (7.8)	392 (9.4)
Grade		
I	53 (4.5)	229 (5.5)
II	378 (32.2)	1536 (36.9)
III	742 (63.3)	2382 (57.2)
ER status		
Negative	429 (40.4)	1041 (36)
Positive	744 (59.6)	1851 (64)
PR status		
Negative	579 (49.4)	1582 (50.2)
Positive	594 (50.6)	1572 (49.8)
HER2 status		
Negative	886 (75.5)	2724 (76.4)
Positive	287 (24.5)	841 (23.6)
Age ≥ 40		
<40	216 (18.4)	718 (17.3)
≥40	957 (81.6)	3444 (82.8)
Treatment		
Control	588 (50.1)	2089 (50.2)
FEC-T	585 (49.9)	2073 (49.8)

Table 25 The TACT trial and TransTACT cohorts

Patient and tumour characteristics in the TACT trial and TransTACT patient cohorts (Jamal-Hanjani, A'Hern, et al. 2015).

6.3 Patient cohort characteristics and histopathological variables

The distribution of histological grade, tumour size, lymph node involvement and ER- and HER2-receptor status in the study population across MCD groups is shown in **Table 26**. ER and PR-positive status was negatively correlated with increasing CIN (**Table 26**; Spearman's rank correlation coefficient (Rs) = -0.09, p = 0.003 and -0.16, p < 0.001. HER2-positive status and increasing histological grade was positively

correlated with increasing CIN (**Table 26**; Rs = 0.16, p < 0.01 and 0.15, p < 0.01, respectively). In ER-positive tumours alone, there was a weak but significant positive correlation between increasing CIN and histologic grade (Rs = 0.14, p < 0.001, n (number of patients) = 744), which was not seen in ER-negative tumours (Rs = 0.06, p = 0.22, n = 429). Work from our laboratory has previously shown that CIN is associated with a younger age at diagnosis in ER-negative breast cancer (Endesfelder et al. 2011), however this was not demonstrated in the study population (**Table 26**). **Figure 53**A shows the relationship between histopathological parameters ER and HER2 status with MCD group, and **Figure 53**B shows the relationship between tumour size, lymph node status, and grade with MCD group.

	Study cohort	MCD1	MCD2	MCD3	MCD4	Rs	p
	n	n (%)	n (%)	n (%)	n (%)		
		354 (30.2)	447 (38.1)	283 (24.1)	89 (7.7)		
Median age in yrs (range)	49 (24-88)	48 (25-88)	49 (28-67)	49 (25-70)	49 (27-75)	0.03	0.34
<40	216	66 (30.6)	77 (35.7)	58 (26.9)	15 (6.9)		
≥ 40	957	288 (30.1)	370 (38.7)	225 (23.5)	74 (7.7)	0	0.80
Positive lymph nodes							
0	242	61 (25.2)	105 (43.4)	59 (24.4)	17 (7)		
1-3	524	181 (34.5)	191 (36.5)	114 (21.8)	38 (7.3)		
4+	407	112 (27.5)	151 (37.1)	110 (27)	34 (8.4)	0.03	0.38
Tumour size (cm)							
0-2	311	103 (33.1)	124 (39.9)	63 (20.3)	21 (6.8)		
2-5	771	224 (29.1)	294 (38.1)	195 (25.3)	58 (7.5)		
>5	91	27 (29.7)	29 (31.9)	25 (27.5)	10 (11)	0.05	0.07
Grade							
I	53	28 (52.8)	16 (30.2)	4 (7.6)	5 (9.4)		
II	378	137 (36.2)	144 (38.1)	76 (20.1)	21 (5.6)		
III	742	189 (25.5)	287 (38.7)	203 (27.4)	63 (8.5)	0.15*	<0.001
ER status							
Negative	429	93 (21.7)	179 (41.7)	121 (28.2)	36 (8.4)		
Positive	744	261 (35.1)	268 (36)	162 (21.8)	53 (7.1)	-0.09*	0.003
PR status							
Negative	579	134 (23.1)	232 (40.1)	160 (27.6)	53 (9.2)		
Positive	594	220 (37)	215 (36.2)	123 (20.7)	36 (6.1)	-0.16*	<0.001
HER2 status							
Negative	886	305 (34.4)	326 (36.8)	193 (21.8)	62 (7)		
Positive	287	49 (17.1)	121 (42.2)	90 (31.4)	27 (9.4)	0.16*	<0.001
Treatment							
Control	588	169 (28.7)	241 (41)	136 (23.1)	42 (7.1)		
FEC-T	585	185 (31.6)	206 (35.2)	147 (25.1)	47 (8)	N/A	

Table 26 Clinical and histopathological variables according to MCD group

Distribution of clinical and histopathological variables according to MCD group. Percentages represent the proportion of patients in each variable category. Abbreviations: ER, oestrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; Rs, Spearman's rank correlation coefficient; and n, number of patients. * Statistically significant association between a variable and increasing MCD group (Jamal-Hanjani, A'Hern, et al. 2015).

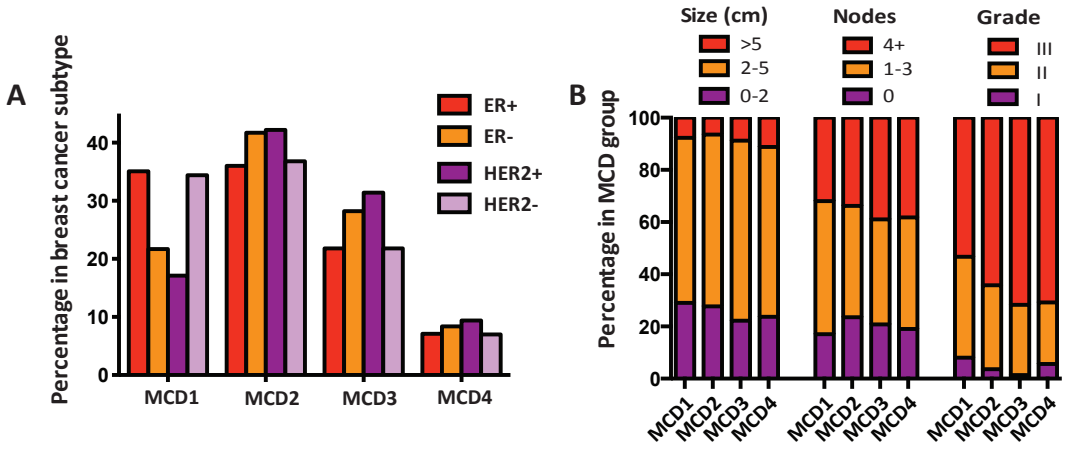


Figure 53 Histopathological parameters and MCD group

Distribution of ER and HER status across MCD groups (A), and relationship between tumour size, number of lymph nodes involved, and tumour grade across MCD groups (Jamal-Hanjani, A'Hern, et al. 2015).

6.4 Relationship between clonal heterogeneity and MCD group

The Shannon Diversity Index (SDI) integrates both the number and abundance of clones within a tumour, and is therefore a measure of clonal heterogeneity (Maley et al. 2006). Consistent with previous findings (Roylance et al. 2011), there was a highly significant correlation between MCD group and SDI for both chromosomes 2 and 15 (**Figure 54**, Rs = 0.95 and 0.87, respectively). Tumours in the MCD4 group had the highest SDI, and therefore the greatest degree of clonal heterogeneity, which is keeping with CIN as a driver of heterogeneity (**Figure 55**). In a univariate analysis there was no statistically significant association found between SDI and DFS with either chromosome 2 (Rs = 0.76, 0.38-1.51; p = 0.43) or chromosome 15 (Rs = 0.71, 0.38-1.31; p = 0.27) in ER-negative cancers. This was also the case in ER-positive cancers. Overall, there was a trend in ER-negative cancers toward improved outcome with increasing SDI, but this was not statistically significant on univariate analysis. A similar association or trend was not found in ER-positive cancers.

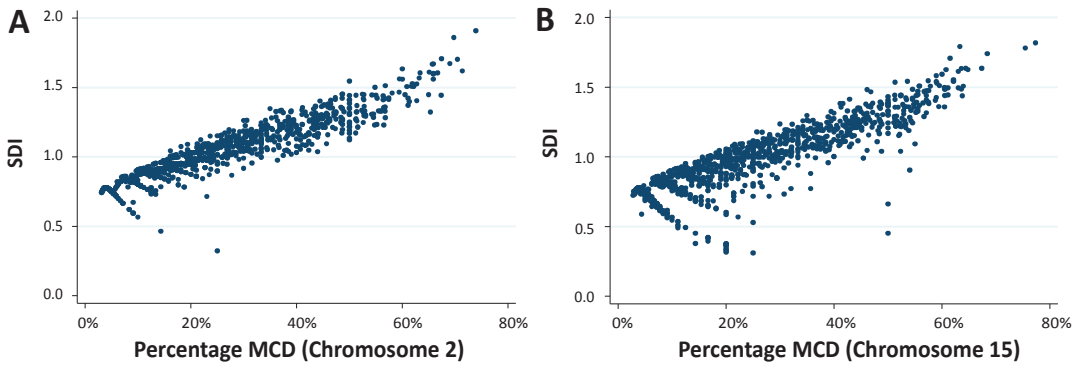


Figure 54 SDI and MCD group

Correlation between the Shannon Diversity Index (SDI) and MCD group for chromosome 2 (A) and chromosome 15 (B).

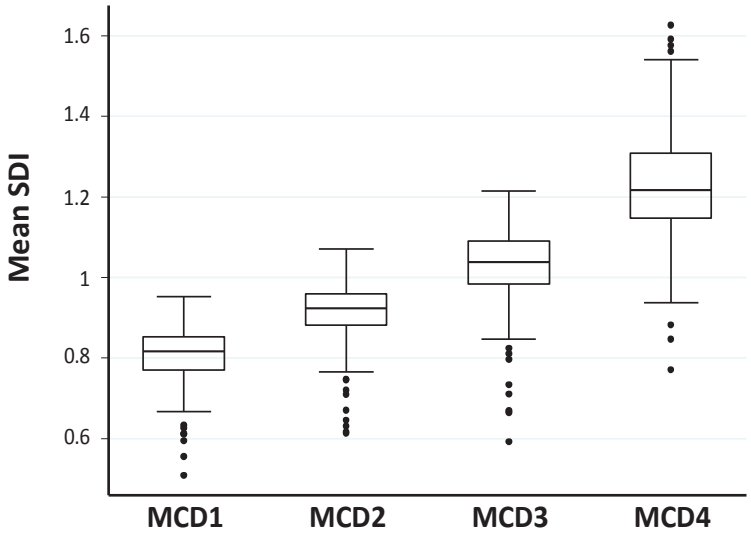


Figure 55 Median SDI and MCD group

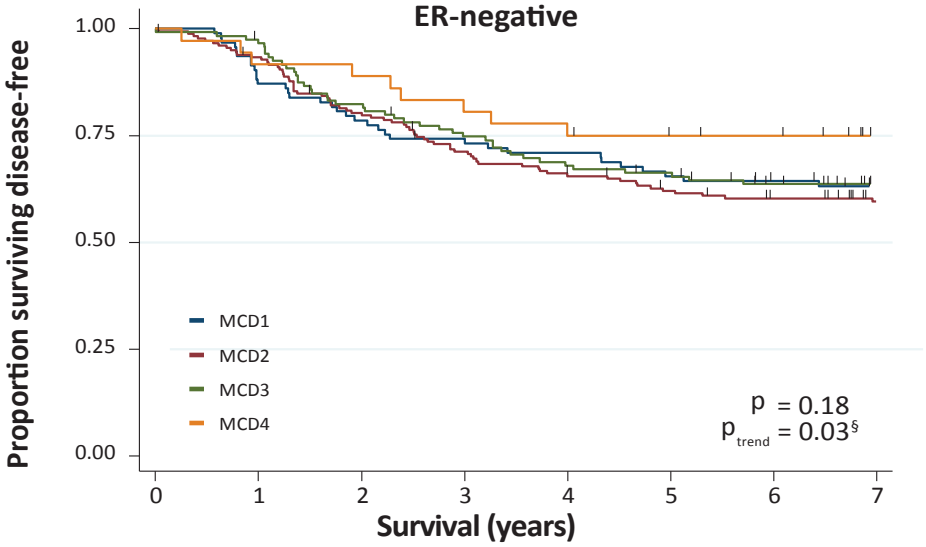
Correlation between the median SDI for both chromosome 2 and 15 and MCD group, with the MCD4 group having the greatest median SDI, and therefore a greater degree of clonal heterogeneity (Jamal-Hanjani, A'Hern, et al. 2015).

6.5 Extreme CIN and clinical outcome in ER-negative breast cancers

In our previous discovery cohort, extreme CIN was associated with an improved prognosis in patients with ER-negative breast cancer (Roylance et al. 2011). In the

TransTACT cohort with a median follow-up of 91 months, a univariate analysis of DFS showed that patients with ER-negative breast cancer in the MCD4 group with extreme CIN also had an improved outcome compared with the other MCD groups, but this was not statistically significant (hazard ratio, HR = 0.61, 95% CI 0.29-1.26; p = 0.18) (**Figure 56**A). In contrast, patients with ER-positive cancer in the MCD4 group had worse outcome, although this was again not statistically significant (HR = 1.28, 95% CI 0.74-2.21; p = 0.38). Improved outcome in the MCD4 group was also seen in ER-negative/HER2-negative cancers (0.41, 0.14-1.17; p = 0.1) (**Figure 56**B). Using breast cancer-specific death as a measure of outcome, a similar pattern was seen in ER-negative cancers (0.67, 0.29-1.54, p = 0.35) (**Figure 56**C). A multivariate analysis of DFS according to ER status showed that in ER-positive cancers, positive nodal status ($p < 0.001$), tumour size 2-5cm (p = 0.002), and age range 40-49 years ($p < 0.001$) were significant determinants of DFS. In ER-negative cancers, positive nodal status ($p < 0.001$), tumour size > 5cm (p = 0.001), and extreme CIN (p trend = 0.03) were significant determinants of DFS. In ER-negative/HER2-negative cancers, positive nodal status ($p < 0.001$), being in the MCD3 group (p = 0.039), and extreme CIN (p trend = 0.007) were significant determinants of DFS (**Table 28**). Overall, improved outcome was seen with increasing CIN in ER-negative (p trend = 0.03) and ER-negative/HER2-negative cancers (p trend = 0.007). This further substantiates the previously demonstrated paradoxical relationship between extreme CIN and improved outcome in ER-negative cancers (Roylance et al. 2011), and the potential value in using tumour CIN status as a prognostic biomarker.

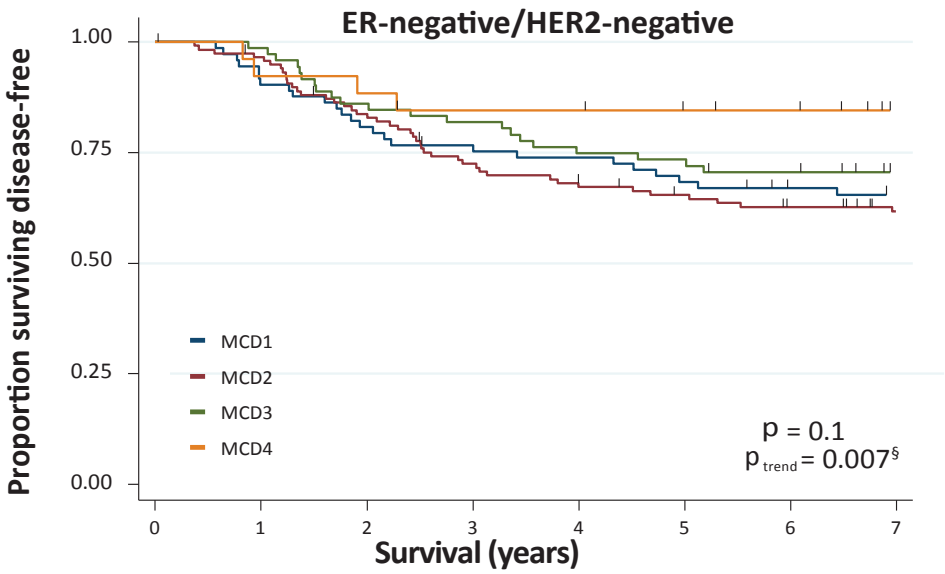
A



Number at risk

MCD1	93	81	73	68	65	60	56	52
MCD2	179	166	142	125	115	106	100	89
MCD3	121	115	98	88	80	77	70	63
MCD4	36	33	32	29	28	25	24	18

B



Number at risk

MCD1	73	66	59	55	53	49	45	43
MCD2	118	113	97	83	77	72	67	59
MCD3	73	71	62	58	53	52	49	44
MCD4	26	24	23	22	22	20	19	14

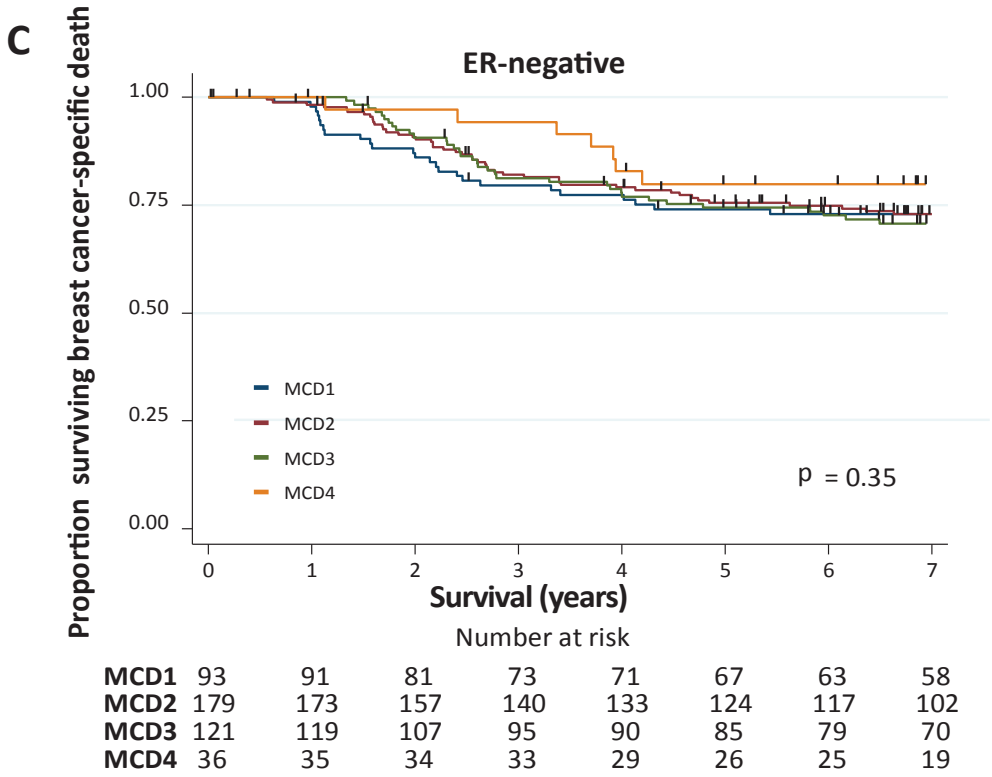


Figure 56 Clinical outcome and MCD group

Kaplan-Meier survival curves for clinical outcome in ER-negative and ER-negative/HER2-negatve breast cancers measured by DFS (A-B), and breast cancer-specific death in ER-negative cancers (C) (Jamal-Hanjani, A'Hern, et al. 2015). [§] Multivariate analysis p values (Jamal-Hanjani, A'Hern, et al. 2015).

		ER-positive (n=744)		ER-negative (n=429)	
		HR (95%CI)	**p**	**HR (95%CI)**	**p**
Nodal Status	Pos. vs Neg.	1.71 (1.45-2.01)	<0.001*	1.85 (1.58-2.16)	<0.001*
Tumour Size	0-2cm	1		1	
	2-5cm	0.56 (0.39-0.8)	0.002*	0.68 (0.45-1.02)	0.07
	>5cm	1.13 (0.73-1.76)	0.58	2.2 (1.37-3.52)	0.001*
Age	Up to 39	1		1	
	40-49	0.41 (0.28-0.61)	<0.001*	0.95 (0.62-1.46)	0.83
	50-59	0.56 (0.39-0.81)	0.002*	0.96 (0.63-1.47)	0.85
	>60	0.7 (0.42-1.16)	0.17	0.64 (0.32-1.29)	0.21
Grade	I	1		-	
	II	0.67 (0.38-1.17)	0.16	1	
	III	1.16 (0.67-2.02)	0.59	1.28 (0.83-1.96)	0.27
Treated with taxane		1.08 (0.83-1.41)	0.57	0.93 (0.68-1.28)	0.66
HER2 status	Pos. vs Neg.	0.76 (0.54-1.08)	0.13	1.23 (0.89-1.7)	0.21
PR status	Pos. vs Neg.	0.77 (0.57-1.05)	0.1	0.62 (0.34-1.13)	0.12
MCD group	1	1		1	
	2	1.19 (0.86-1.64)	0.3	0.98 (0.65-1.48)	0.94
	3	1.06 (0.73-1.55)	0.75	0.67 (0.42-1.08)	0.1
	4	1.06 (0.61-1.84)	0.85	0.59 (0.28-1.24)	0.17
			0.80(p trend)		0.03(p trend)*

Table 27 Multivariate analysis of DFS according to ER status

Multivariate analysis of DFS in ER-negative and ER-positive breast cancers to determine significant determinants of outcome (Jamal-Hanjani, A'Hern, et al. 2015). * Statistically significant associations.

		ER-positive/ HER2-negative (n=596)		ER-negative/ HER2-negative (n=429)	
		HR (95%CI)	**p**	**HR (95%CI)**	**p**
Nodal Status	Pos. vs Neg.	1.82 (1.52-2.19)	<0.001*	1.87 (1.5-2.33)	<0.001*
Tumour Size	0-2cm	1		1	
	2-5cm	0.48 (0.32-0.72)	<0.001*	0.76 (0.45-1.26)	0.28
	>5cm	0.92 (0.55-1.57)	0.77	1.84 (0.93-3.64)	0.08
Age	Up to 39	1		1	
	40-49	0.34 (0.22-0.52)	<0.001*	1.46 (0.8-2.64)	0.22
	50-59	0.47 (0.31-0.7)	<0.001*	1.43 (0.79-2.59)	0.24
	>60	0.6 (0.35-1.03)	0.07	1.06 (0.44-2.55)	0.9
Grade	I	1		-	
	II	0.7 (0.39-1.26)	0.23	1	
	III	1.15 (0.65-2.05)	0.63	1.32 (0.75-2.33)	0.34
Treated with taxane		1.09 (0.81-1.48)	0.56	0.97 (0.64-1.47)	0.89
PR status	Pos. vs Neg.	0.68 (0.48-0.96)	0.031*	0.56 (0.24-1.31)	0.18
MCD group	1	1		1	
	2	1.17 (0.82-1.68)	0.39	0.99 (0.6-1.65)	0.98
	3	0.94 (0.62-1.45)	0.79	0.51 (0.27-0.97)	0.039*
	4	1.17 (0.63-2.16)	0.61	0.37 (0.13-1.06)	0.06
			0.91(p trend)		0.007(p trend)*

Table 28 Multivariate analysis of DFS according to ER and HER2 status

Multivariate analysis of DFS in ER-positive/HER2-negative and ER-negative/HER2 negative breast cancers to determine significant determinants of outcome (Jamal-Hanjani, A'Hern, et al. 2015). * Statistically significant associations.

6.6 Taxane response and MCD group

Work from our laboratory has previously demonstrated an association between CIN and taxane resistance (Swanton et al. 2007; Swanton et al. 2009; Lee et al. 2011). In order to determine whether MCD group had any influence on response to taxane therapy in the TransTACT cohort, MCD1/MCD2 and MCD3/MCD4 groups were combined (due to small patient numbers), and the DFS in patients in the anthracycline/taxane-based treatment arm were compared to the DFS in patients in the anthracycline-based only treatment arm. In ER-negative cancers, the HR was 1.05 (95% CI 0.72-1.53; p = 0.80) in the MCD1/MCD2 groups vs. 0.89 (95% CI 0.51-1.53; p = 0.66) in the MCD3/MCD4 groups, demonstrating no difference in DFS between the two treatment groups (**Figure 57**A). Similarly, there was no significant correlation found

in the ER-negative/HER2-negative cancers (**Figure 57**B). Overall there was no evidence that DFS in the anthracycline/taxane treated group was dependent on MCD group, but it is important to note that with such small patient numbers this analysis was insufficiently powered to detect an interaction, and therefore a difference in outcome between the two treatment arms.

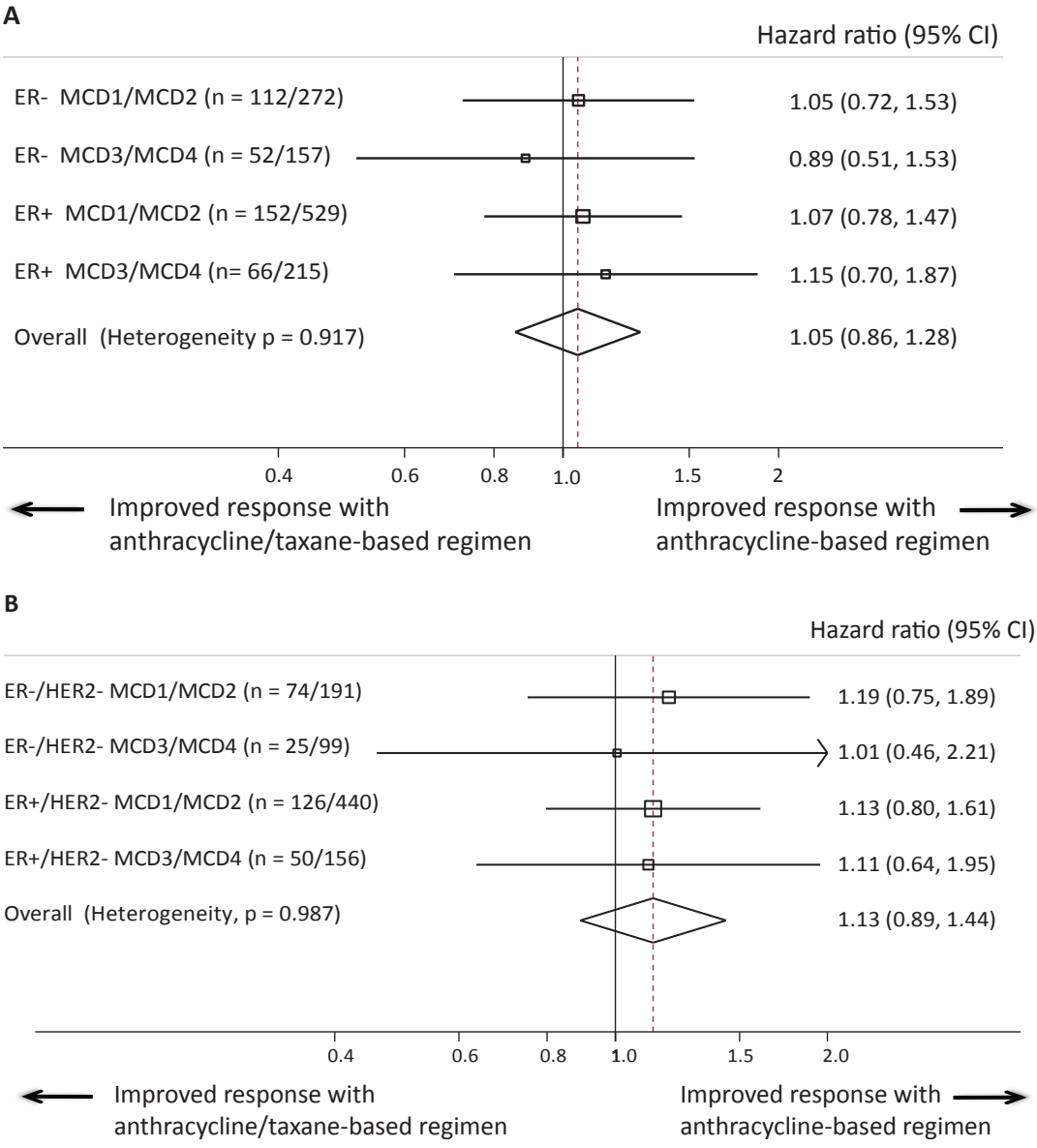


Figure 57 Taxane response and MCD group

Forest plots for DFS in patients in the anthracycline/taxane-based treatment arm compared to those in the anthracycline-based only treatment arm according to MCD group, and ER (A) and HER2 (B) status. n = number of patients in each respective MCD group (Jamal-Hanjani, A'Hern, et al. 2015).

6.7 Conclusions and discussion

In the initial discovery cohort study of 246 patients, extreme CIN was associated with improved long-term survival in ER-negative cancers (HR = 0.0827, 95% CI 0.0097 - 0.7066, p = 0.0228) (Roylance et al. 2011). This trend has been further validated in the TransTACT cohort, representing 1173 patients, using the same techniques and thresholds for analyses to demonstrate a statistically significant trend in improved prognosis with increasing CIN, as measured by MCD group (Jamal-Hanjani, A'Hern, et al. 2015). In order to optimise breast cancer treatment and management there is a need to identify and develop improved methods to predict prognosis, treatment benefit and risk of early relapse. The development of gene expression-based signatures has aided the assessment of outcome in certain breast cancer subtypes, in particular ER-positive tumours (Kim & Paik 2010; Pusztai 2009). However, there remains a need to supplement existing, and identify new, prognostic signatures in ER-negative breast tumours. In view of the potential prognostic value of tumour CIN status, the development of reproducible and efficient techniques to assess CIN status in tumour samples may be of significant benefit. Using defined thresholds for CIN, such as MCD group, may help distinguish between good and bad prognostic groups, supporting both treatment and risk stratification in breast cancer patients. In addition, stratification of ER-negative breast cancer patients by MCD group may help identify a cohort of patients with improved outcome, which may be important to take into account when assessing therapeutic response and clinical outcome, especially in the context of clinical trials. Overall, the data presented here, from a large validation cohort study of patients with breast cancer, has further substantiated the paradoxical relationship previously demonstrated between extreme CIN and outcome in ER-negative breast cancers (Roylance et al. 2011), and the potential role for tumour CIN status as a prognostic biomarker.

7 Discussion

7.1 Intratumour heterogeneity in NSCLC

In this thesis, the spatial and temporal heterogeneity of NSCLC were explored in a cohort of 27 patients using multi-region WES of fresh frozen tissue collected from surgically resected primary NSCLC tumours. In keeping with previously published data from our laboratory (de Bruin et al. 2014), both spatial and temporal heterogeneity was seen in both LUAD and LUSC histological subtypes of NSCLC. This heterogeneity was seen for SNV and indel mutations, copy number aberrations, and potential mutational signatures, such as those associated with tobacco smoke exposure and APOBEC cytidine deaminase activity. These data support the premise of branched tumour evolution in NSCLC, with some aberrations present in all regions of a tumour (likely to be clonal in origin), occurring early in tumour evolution, and some aberrations being present in some, but not all, regions of a tumour (likely to be subclonal in origin), occurring late in tumour evolution.

The importance of intratumour heterogeneity is increasingly recognised as a driver of tumour progression, drug resistance and treatment failure in solid tumours (Shah et al. 2012; Campbell et al. 2010; Lee et al. 2011; Navin et al. 2011; Wu et al. 2012; Szerlip et al. 2012; Yap et al. 2012). The presence of subclonal driver mutations may prove to present a significant challenge, in terms of biomarker development and drug target discovery efforts, and contribute to drug resistance and poor survival outcome (Anderson et al. 2011; Navin et al. 2011; Navin et al. 2010). Despite the impressive developments of international large-scale sequencing consortia, the spatial separation of tumour subclones, the changing nature of tumour clonal architecture over time, and the impact of such diversity upon clinical outcome are yet to be determined systematically and longitudinally. The lung TRACERx study aims to integrate genomic data with detailed clinical annotation and outcome, in order to decipher the heterogeneity of the cancer genome and mutational pathways involved in NSCLC pathogenesis. It aims to develop clinically meaningful measures of intratumour heterogeneity to guide patient management and treatment stratification (Merlo et al. 2010), and to prospectively define thresholds of tumour heterogeneity for clinical risk stratification. With increasing awareness for the need to obtain tissue and genetically profile cancers in order to stratify treatment, the concept of longitudinal tissue collection and analysis has become more acceptable in oncological practice. In following cancers

from diagnosis to relapse, tracking the evolutionary trajectories of tumours in relation to therapeutic interventions, and determining the impact of clonal heterogeneity on clinical outcomes, TRACERx may also serve as a model applicable to other cancer types.

7.2 Heterogeneity in lung adenocarcinoma *in situ* lesions

Patients recruited into the TRACERx study either had a confirmed histological diagnosis of NSCLC, or were highly suspected to have such a diagnosis. Five patients recruited into the study were subsequently withdrawn after being diagnosed with adenocarcinoma *in situ* (AIS). These pre-invasive lesions were still subjected to multi-region WES, using the same TRACERx bioinformatics pipeline, and were analysed separately. Whilst it is known that pre-invasive lung lesions can harbour some somatic mutations (Izumchenko et al. 2014) and copy number aberrations (Massion et al. 2009), whether clonal diversity and clonal evolution are prominent features of such lesions is yet to be determined. Similar to the NSCLC tumours described in this thesis, both temporal and spatial heterogeneity of mutations and copy number aberrations was evident in the AIS lesions analysed. Certain driver mutations appeared to be involved in early, and some appeared to be involved in late, clonal evolution of these lesions. In particular, there was an interesting case of convergent evolution where in one lesion there was a truncal mutation in the *BRAF* gene as well as a branch mutation in the *NRAS* gene, suggesting involvement of the MAPK/ERK signalling pathway throughout clonal evolution of the lesion up until the point of surgical resection. These data would support the hypothesis that prior to the development of invasive lung adenocarcinoma, pre-invasive lung lesions possess a heterogeneous clonal landscape, in which certain somatic mutations can give rise to a clonal selective advantage. Further insight into the clonal dynamics and evolution of these pre-invasive lesions may help improve our understanding of the earliest stages of lung cancer development, and the biological processes involved in the transition from pre-invasive to invasive disease.

7.3 Metastatic disease and defining the origins of the lethal subclone(s)

Clonal diversity between primary and metastatic tumours in the same patient has been demonstrated in different tumour types, including lung (Paik et al. 2015), breast (Shah et al. 2009), pancreatic (Campbell et al. 2010; Yachida et al. 2010), colorectal (Kogita et al. 2015; Tan et al. 2015), gastric (Nadauld et al. 2014), prostate (Haffner et al.

2013; Hong et al. 2015; Gundem et al. 2015) and medulloblastoma (Wu et al. 2012) tumours. The dynamics and patterns of clonal composition can indicate the tumour evolutionary paths that underlie tumour progression, including potential mechanisms involved in therapeutic resistance. Using WGS and WES, Juric and colleagues sequenced the primary and metastatic tumours in a breast cancer patient with a known activating *PIK3CA* mutation who had developed drug resistance whilst on treatment with a PIK3CA inhibitor (Juric et al. 2015). All the metastatic tumours were found to have a common copy number loss of *PTEN*, and those metastatic tumours that were identified as drug resistant based on imaging, were found to have acquired additional genetic aberrations in *PTEN* resulting in the loss of its expression, including SNVs, indels and copy number aberrations. This demonstrated parallel evolution in separate metastatic tumours with different aberrations in *PTEN* leading to a convergent phenotype resistant to PIK3CA inhibition.

In order to investigate the evolving constellation of genetic aberrations involved in the metastatic process, and the patterns of intratumour heterogeneity that can vary between primary and metastatic tumours, there is a need for longitudinal tissue collection from multiple sites of disease from living patients. In general, this relies on tissue that is surplus to diagnostic and clinical requirements. There are several challenges in obtaining tissue from patients with advanced cancer including the increased biopsy risk associated with multiple site sampling, the technical difficulty in accessing sites of metastatic disease and the inevitable time, travel and costs incurred by patients who are often unwell with a poor performance status. Most patients with widely disseminated disease do not undergo tumour resection, and therefore the ability to perform multi-region sequencing of surgically resected tumours is limited. Obtaining a limited number of core biopsies at different disease time points is a possibility in this group of patients, but may not be representative of the entire tumour.

In this thesis, the primary and metastatic NSCLC tumours of two patients who subsequently presented with disease recurrence were subjected to multi-region WES and analysed. In both cases, a region within the primary tumour from which the metastasis had originated was identified, and both spatial and temporal intratumour heterogeneity was evident in the primary and metastatic tumours. Patient L011 had a truncal *BRAF* V600E mutation, suggesting that this driver mutation was an early event in the evolution of her disease and that treatment with a BRAF or MEK inhibitor may

have been associated with an effective tumour response. Prior knowledge of the genomic profile of her primary tumour may have pre-empted early therapeutic intervention with such targeted drugs at the point of metastatic disease. Patient L017 had a truncal *KRAS* G12V mutation, suggesting that treatment in the future with an EGFR TKI may have resulted in decreased tumour response. These examples demonstrate how early genomic profiling of tumours has the potential to guide therapeutic intervention and to inform tumour response. Furthermore, the analysis of the clonal composition of paired primary-metastases tumours also has the potential to give some insight into the metastatic process and the origin of the lethal subclone(s). The general lack of paired primary and metastatic tumour samples in patients has meant that cancer research has been restricted to either established tumour cell lines or archival material (Rubin et al. 2000) limiting the systematic study of the metastatic process (Embuscado et al. 2005). Access to post-mortem tissue allows multi-region sampling within larger tumours, as well as sampling of all sites of metastatic disease, in order to define the relationship between metastatic tumours and the primary tumour.

7.3.1 The PEACE study

There are few studies in which post-mortem tissue sampling is performed immediately after death, including studies in breast cancer (Juric et al. 2015), prostate cancer (Rubin et al. 2000; Liu et al. 2009), pancreatic cancer (Embuscado et al. 2005), uveal melanoma (Borthwick et al. 2011), and idiopathic pulmonary fibrosis (Lindell et al. 2006). None of these studies have explored the clonal composition of tumours, and in particular, the genetic aberrations and biological pathways involved in the metastatic process. The PEACE (Posthumous tissuE donAtion in CancEr, 13/LO/0972, UKCRN ID 18422) study has been set-up to address this, and to support cancer research projects that would benefit from access to tumour tissue. The PEACE study is a multi-centre prospective observational study intended to facilitate tissue donation, in metastatic cancer, from multiple tumour sites in the post-mortem setting. This study also involves the collection of blood samples for germ line DNA, cfDNA and CTCs. The PEACE study will take advantage of the fresh tumour tissue sampling and processing protocols, and bioinformatics pipelines designed to accurately assess the mutational and copy number profiles of tumours, already developed and optimised within our laboratory.

The PEACE study aims to track the clonal evolution of primary to metastatic tumours, and to therefore establish a model for tumour progression and the metastatic process. By mapping the evolution of tumours from the primary to different metastatic sites, the study aims to reveal the lethal subclone(s) harbouring candidate driver mutations resulting in disease progression, therapeutic resistance, and subsequent overwhelming disease burden, which may have led to eventual death.

The analysis of cfDNA and CTCs in the metastatic setting will also be used to determine the extent to which such circulating biomarkers can represent the underlying genomic landscape of primary (if still *in situ*) and metastatic tumours. The study allows for access to archival tissue samples in cases where the primary tumour in no longer *in situ*. Patients who develop metastatic NSCLC in the lung TRACERx study will be approached to enter the PEACE study. This would truly represent longitudinal tissue and blood sampling from diagnosis to relapse to death, allowing the assessment of tumour evolution and tumour biology throughout the disease course. Patients for the PEACE study will be selected using the following eligibility criteria:

Inclusion criteria:

- Age 18 years or older
- Confirmed diagnosis of a solid malignancy with metastatic disease
- Oral and written informed consent from patient to enter the study and to undergo post-mortem tissue sampling

Exclusion criteria:

- Medical or psychiatric condition that would preclude informed consent
- History of high-risk infections (e.g. HIV-positive, hepatitis C, tuberculosis and Creutzfeldt-Jacob disease).

I wrote and developed the protocol for this study under the guidance of my supervisor during the production of this thesis, with the input of clinicians, pathologists, palliative care teams, medical lawyers, ethicists, patient advocate groups, and hospital mortuary staff. This study is currently open to recruitment in University College London Hospital, The Royal Marsden Hospital, Guy's and St Thomas' Hospital, and the Royal Leicester Infirmary Hospital. This study is currently being set-up across the lung TRACERx study sites, as well as other national sites with the intention of recruiting patients with metastatic disease of different solid tumour types.

7.4 Implications of intratumour heterogeneity for precision medicine

Given the evidence for spatial intratumour heterogeneity, using a single diagnostic tumour biopsy, representing a snap shot of a tumour in time, may limit the ability to identify and qualify effective biomarkers for clinical use (Jamal-Hanjani, Quezada, et al. 2015). Although taking multiple tumour biopsies to determine the true clonal composition of tumours is not a simple or practical solution in daily clinical practice, it is an effective approach in order to better understand tumour evolution during the disease course in the context of clinical studies. Longitudinal studies, such as TRACERx, involving the analysis of both tissue and liquid biopsies may help develop predictive tools or cancer evolutionary 'rule books' allowing us to predict likely beneficial therapeutic interventions, mechanisms of drug resistance, and eventual disease progression based on the wealth of data and experience gained from tracking the clonal evolution of tumours at different disease time points (Turajlic et al. 2015). Such studies may also allow the characterisation of patients who respond to specific therapeutic interventions exceptionally or poorly, which may be of benefit in drawing parallels with other patients with similar genetic or phenotypic patterns of disease.

Therapeutic studies aimed at delivering precision medicine may have a greater likelihood of determining the true impact of intratumour heterogeneity on clinical outcome if potential biomarkers of disease are identified and validated in the context of tumour spatial heterogeneity. Furthermore, drug combination designs may need to consider the impact of different potential drivers of disease during tumour evolution, and the existence of subclonal drivers of disease at any given time point, on drug efficacy and resistance (Zhao et al. 2014). The relevance of driver clonality and treatment response is yet to be determined in prospective clinical studies. Therapeutic targeting of actionable aberrations in dominant compared to minor subclones may not lead to the same therapeutic outcome. Therapeutic studies such as the DARWIN (Deciphering Anti-tumour Response With INtratumour Heterogeneity, ClinicalTrials.gov number NCT02183883) clinical trials programme, may help determine how driver clonally can influence therapeutic response (Jamal-Hanjani et al. 2014; Hiley et al. 2014).

Precision medicine has the potential to improve clinical outcome and to reduce drug-associated toxicity, but it is not without its challenges and significant cost implications. These include access to sequencing equipment, the development of clinical trials employing adequate tumour profiling to identify actionable aberrations, the validation of predictive and prognostic biomarkers, and the availability of targeted drugs (Jamal-Hanjani, Quezada, et al. 2015). Infrastructure supporting adequate computing resources and approved sequencing technologies is necessary to help clinical practice move toward the implementation of genomics and detailed molecular characterisation of a patient's tumour in order to enable tailored therapies and attempt to improve patient outcome. Furthermore, clinicians will need to adapt to the challenges in analysing the wealth of data that genomics studies can create, and to gain experience in the communication of increased risk or false-positive and false-negative results (May et al. 2014). The ultimate aim of such studies should be to develop economically feasible predictive and prognostic biomarkers to aid clinical decision-making and improve risk stratification. This way the medical and scientific community can work towards significant, as opposed to marginal, improvements in outcome (Fojo et al. 2014).

7.5 Representation of the tumour landscape in cfDNA

Given the increasing evidence for intratumour heterogeneity, sampling tumour tissue at a single time point to identify predictive and prognostic biomarkers has significant limitations, in terms of sampling bias of spatially heterogeneous tumours, and the lack of information regarding temporal heterogeneity. Circulating biomarkers may be an effective non-invasive alternative to longitudinal tissue sampling, and have the potential to track the genomic landscape of tumours, and help monitor tumour burden, minimal residual disease, and detect the emergence of drug resistance (Diaz & Bardelli 2014; Bidard et al. 2013). In addition, further evidence supporting the use of circulating biomarkers in early stage disease may also inform cancer screening strategies, and could potentially aid early therapeutic intervention. Furthermore, with the ability to detect copy number aberrations in circulating biomarkers, regions of chromosomal instability could be assessed to determine the relevance of CIN as a predictive and/or prognostic biomarker (Schütz et al. 2015).

The analysis of cfDNA in cancer patients may further complement tumour sequencing data, and potentially identify additional genetic aberrations that have not been detected by primary or metastatic tumour sequencing depending on sequencing depth. In order for this to be true, robust and sensitive enough techniques need to be employed for the detection of both clonal and subclonal genetic aberrations. In this thesis, three approaches were used to identify genetic aberrations in cfDNA from patients with NSCLC prior to surgical resection of their primary tumours. Tumour-specific SNV and indel mutations, as well as a single translocation, were chosen based on WES and Ion AmpliSeq sequencing data. The main aim was to determine how representative cfDNA was of the mutational burden in early stage primary NSCLC, and in particular, to determine whether both truncal and branch mutations could be detected, such that the heterogeneous genomic landscape of tumours, and therefore a measure of intratumour heterogeneity, could be interpreted using cfDNA. The approaches involved combining multiplex PCR with mass spectrometry (in collaboration with Agena Bioscience, the UltraSEEK technique), MiSeq sequencing (in collaboration with Illumina), or HiSeq sequencing (in collaboration with Natera). The UltraSEEK technique was able to detect 5/44 (11.4%) SNVs (all truncal) in cfDNA, and the Natera approach was able to detect 16/37 (43.2%) SNVs in cfDNA (4 branch and 2 truncal). The Illumina approach appeared to detect an SNV, an indel, and a translocation variant in cfDNA, but this was based on single assays and very few selected mutations overall, such that a measure of statistical significance could not be assigned. The Natera approach detected the greatest proportion of selected mutations, and provides some early evidence for the detection of branch mutations in cfDNA. Increased plasma volume and further assay optimisation, could considerably improve the analytical sensitivity of these approaches, and potentially allow the detection of mutations at lower variant allele frequencies (VAFs), which is often the case with branch mutations. A cfDNA substudy has been set-up within TRACERx in collaboration with Natera, involving the longitudinal sampling of 100 patients from diagnosis to relapse in order to further explore the utility of cfDNA in tracking tumour evolution using customised SNV and CNV panels based on the sequencing data, as well as lung cancer specific SNV and CNV panels.

Whilst cfDNA may have great potential, there are still obstacles to overcome before its use can be implemented in standard clinical practice. Standardisation of the methods and techniques employed is required, including the collection of blood samples, the methods used to extract plasma and subsequently cfDNA from whole blood, the

quantification of cfDNA, and the techniques used to identify and quantify tumour-associated genetic aberrations. As previously discussed, several NGS techniques exist to detect genetic aberrations in cfDNA. The cost and logistics of developing these techniques for use in the clinic is not insignificant. Furthermore, the wealth of sequencing data that can be produced from such sequencing strategies will require substantial bioinformatic and computing resource infrastructure for data analysis and interpretation (Crowley et al. 2013). In addition, since cfDNA may represent cancer genomes derived from multiple metastatic sites within the body, the interpretation of data may be confounded by intratumour heterogeneity (Pantel et al. 2013). The method by which genetic aberrations are selected is also an important consideration. Customised panels specific to each patient's tumour rely on screening of tumour biopsies so that the existing mutational profile is known on a patient-by-patient basis, take longer to design, and involve greater costs. Alternatively, generic cancer panels that are able to detect hundreds of somatic mutations, known to be commonly mutated in cancer, can be more cost-effective and would require less optimisation. The choice between the two approaches will depend on the question that is being asked. For example, if the intention is to track the clonal evolution of tumours and to monitor the rise and fall in the frequency of both known and rare drivers of disease, a customised approach would be appropriate, and if the intention is to identify a therapeutically targetable genetic aberration, a generic screening panel would be appropriate.

Which technique is the most optimal in the field of circulating biomarkers, exactly how representative circulating biomarkers are of the underlying tumour landscape, and how sensitive they are in identifying clonal and subclonal alterations, is yet to be determined (Diaz & Bardelli 2014). Ultimately, formal validation in clinical studies, such as TRACERx, where the use of circulating biomarkers is tested as part of a study endpoint, is necessary to assess the use of such biomarkers in clinical practice (Crowley et al. 2013). Both the lung and breast TRACERx studies involve serial sampling of circulating biomarkers so that their assessment can be correlated with clinically relevant time points, such as disease progression and the development of drug resistance, as well as clinical and radiological examinations.

7.6 CIN as a prognostic biomarker in ER-negative breast cancer

Increased CIN has been shown to be associated with poor prognosis in several tumour types (Carter et al. 2006; Walther et al. 2008; Habermann et al. 2009; Kronenwett et al. 2006; Bakhoum et al. 2011; Sheffer et al. 2009; M'kacher et al. 2010), suggesting that greater genomic instability may promote cancer cell growth and therefore tumour progression. However, both aneuploidy and CIN have also been shown to be disadvantageous for cancer cell survival and fitness (Ganem et al. 2009; Sotillo et al. 2010; Pavelka et al. 2010; Kops et al. 2004; Schmidt et al. 2005; Janssen et al. 2009), such that extreme levels of genomic instability are associated with improved clinical outcome (Birkbak et al. 2011; Roylance et al. 2011; Baumbusch et al. 2013). In this thesis, data from a large validation cohort of 1173 patients are presented supporting the previously shown significant association between CIN and improved outcome in ER-negative breast cancers. Both aneuploidy and CIN are features of malignant tumours, which may represent potential targetable phenotypes for cancer treatment (Roschke et al. 2003; Roschke et al. 2005; Roschke & Kirsch 2005). Given the clinical relevance of CIN in cancer, as a potential prognostic biomarker and therapeutic target, the identification of tumour CIN status may have significant exploitable benefits. Understanding the mechanisms contributing to numerical and structural CIN, and developing improved techniques to identify tumour CIN status, may help guide therapeutic intervention, and therefore improve outcome in cancer, in particular, patients with ER-negative breast cancers.

7.6.1 The breast TRACERx study

Intratumour heterogeneity is not simply manifested at the somatic mutational level, but also at the chromosomal level, in the form of structural and numerical chromosomal instability. Intratumour heterogeneity of copy number aberrations has been described in several studies, in which some copy number gains or losses can be clonal, occurring early in tumour evolution, and some can be subclonal, occurring late in tumour evolution (Szerlip et al. 2012; Snuderl et al. 2011; Sottoriva et al. 2013; Campbell et al. 2010; Yachida et al. 2010; Nik-Zainal et al. 2012; Gerlinger et al. 2014; Shah et al. 2012). Whilst these studies may shed some light on tumour heterogeneity in terms of numerical CIN status of a tumour, our understanding of both spatial and temporal heterogeneity in CIN is limited. The breast TRACERx study is a prospective

observational clinical study, which aims to address this, using different techniques for measuring CIN, including direct centroemric FISH and DNA ploidy analyses, on multi-region samples within the same tumour. Patients with breast cancer suitable for neadjuvant or primary chemotherapy followed by surgery, in the case of resected primary tumours, will be eligible for the study, which aims to recruit 546 patients with a total 10-year follow-up period. The study involves the collection of multi-region tumour tissue at diagnosis, during neoadjuvant or primary chemotherapy, at surgery, and at relapse. Longitudinal blood sampling, similar to the TRACERx lung study, for cfDNA will also be analysed during follow-up. The study aims to define the impact of intratumour heterogeneity on the response to treatment (by assessing the pathological complete response (pCR) rate), and on clinical outcome in terms of DFS and OS. Similar to the lung study, it aims to also understand the molecular drivers of tumour progression and drug resistance. The sequential analysis of tumour through the disease course will determine whether exposure to anthracycline- and taxane-containing regimens, as well as trastuzumab in HER2-positive patients, can select for intratumour heterogeneity. Patients will be selected using the following eligibility criteria:

Inclusion criteria:

- Written Informed consent
- Patients ≥ 18-years of age, with histological confirmation of invasive breast cancer with known receptor status suitable for neoadjuvant or primary chemotherapy
- Histological subtypes to include: with either TNBC (HER2-negative/ER-negative/PR-negative), HER2-positive (ER-positive or ER-negative), or ER-positive (HER2-negative) breast cancer
- Neoadjuvant therapy, surgery, and any subsequent adjuvant therapy and follow-up to be in accordance with NICE and local guidelines
- Stage T2-4 tumours (including inflammatory cancers), with any nodal status, can be included
- Multifocal cancer (as long as it is of the same subtype as the index cancer) can be included

Exclusion criteria:

- Confirmed metastatic disease at initial presentation

- Any other current malignancy, or malignancy diagnosed or relapsed within the last 5-years (other than non-melanomatous skin cancer and *in situ* cervical cancer)
- Any concomitant medical or psychiatric problems which in the opinion of the investigator would prevent completion of treatment or follow-up
- Lack of adequate tissue

The objectives of the study are outlined below:

Primary objectives:

- To determine the relationship between intratumour heterogeneity and the pCR rate achieved
- To determine the relationship between intratumour heterogeneity and clinical outcome, in terms of DFS and OS
- To determine the relationship between intratumour heterogeneity and the different histological subtypes of breast cancer

Secondary objectives:

- To identify potential drivers of genomic instability, metastatic progression and drug resistance
- To improve our understanding of the different breast cancer molecular subtypes and the extent of intratumour heterogeneity
- To validate key driver mutations to help guide stratification of breast cancer treatment and clinical study inclusion
- To establish whether cfDNA analysis can predict early relapse and track tumour clonal evolution through the disease course

By tracking the somatic mutational heterogeneity, and chromosomal structural and numerical instability, in breast cancer disease using established methods (Jamal-Hanjani et al. 2014; Roylance et al. 2011), this study aims to determine the clinical relevance of intratumour heterogeneity, in term of somatic mutations, copy number aberrations, and CIN, in breast cancer pathogenesis, recurrence and metastatic disease.

I wrote and developed the protocol for this study under the guidance of my supervisor during the production of this thesis, with input from Dr. Rebecca Roylance (Barts Cancer Institute, Queen Mary University of London). This study is currently awaiting R&D approval from University College London Hospital.

7.7 Conclusion

The data presented in this thesis demonstrate the existence of a complex and heterogeneous genomic landscape in NSCLC. Furthermore, spatial and temporal heterogeneity has been demonstrated in pre-invasive lung adenocarcinoma *in situ* lesions, suggesting that heterogeneity in mutations and copy number aberrations may be a feature of the early stages of lung cancer development. Intratumour heterogeneity may have significant clinical implications for both cancer diagnosis and therapeutic intervention. The impact of such heterogeneity on therapeutic response and clinical outcome is yet to be determined. Taking multiple tissue biopsies and serial sampling of tumours to track progression, and to identify new targetable drivers of disease is practically challenging, and is currently not part of standard clinical practice. An alternative non-invasive approach may be the use of circulating biomarkers, such as cfDNA. In this thesis different approaches were taken to determine the use of cfDNA in early stage NSCLC to represent the underlying tumour landscape, consisting of both truncal and branch mutations. Although the data suggest that some branch mutations can be detected in cfDNA, the identification of low frequency branch mutations in cfDNA remains a difficult task, and studies are required to further develop effective strategies for mutation detection.

Chromosomal instability (CIN), an initiator of intratumour heterogeneity, has been shown to be associated with poor prognosis in several tumour types (Carter et al. 2006; Habermann et al. 2009; Walther et al. 2008). In this thesis data are presented from the largest breast cancer cohort study in which the association between CIN and clinical outcome is investigated. Having previously demonstrated a paradoxical relationship between extreme levels of CIN and improved long-term survival in ER-negative breast cancers (Roylance et al. 2011; Birkbak et al. 2011), this relationship was further substantiated in this large prospective validation cohort study (Jamal-Hanjani, A'Hern, et al. 2015). The association between extreme CIN and improved outcome may be related to increased tumour aneuploidy resulting in decreased cellular fitness. This suggests that in certain tumour types a threshold for CIN and intratumour heterogeneity may exist above which tumour growth and progression is limited, analogous to a mutational burden meltdown. Aside from the potential role of CIN as a prognostic biomarker in ER-negative cancers, the association between CIN and improved outcome may provide therapeutic opportunities whereby patients with

specific tumour types may benefit from certain drugs, such as DNA damaging agents, as a result of increased tumour CIN levels.

Genomic studies involving longitudinal tissue and blood sampling, with the integration of genomic data with detailed clinical and phenotypic annotation of samples, have the potential to give us great insights into tumour biology. The realisation of precision medicine in the genomic era will rely on the collaborative efforts of multi-disciplinary teams, and the study of tumour evolution will require a multi-disciplinary approach combining studies in tumour microenvironment, the immune system, genetics, epigenetics, transcriptomics, and functional screening. Clinical studies, such as TRACERx and PEACE, will allow us to gain a deeper insight into the true extent of intratumour heterogeneity and CIN, how these evolve during the disease course, and their predictive and prognostic relevance.

8 Appendix

8.1 Appendix 1: List of papers and abstracts published during this PhD

8.1.1 Primary research articles

Detection of Ubiquitous and Heterogeneous Mutations in Cell-Free DNA from Patients with Early-Stage Non–Small-Cell Lung Cancer
Jamal-Hanjani M, et al. (*Submitted to Annals of Onoclogy*)

Clonal lung cancer neoantigens elicit T cell immunoreactivity and anti-PD-1 response
Nicholas McGranahan, Andrew J.S. Furness, Rachel Rosenthal, Sofie Ramskov, Rikke Lyngaa, Sunil Kumar Saini, Mariam Jamal-Hanjani et al. (*Submitted to Science*)

TumorTracer: A method to identify the tissue of origin from the somatic mutations of a tumor specimen
Andrea Marion Marquard, Nicolai Juul Birkbak, Cecilia Engel Thomas, Francesco Favero, Marcin Krzystanek, Celine Lefebvre, Charles Ferté, Mariam Jamal-Hanjani et al. BMC Med Genomics. 2015 Oct 1;8:58. doi: 10.1186/s12920-015-0130-0.

Extreme chromosomal instability forecasts improved outcome in ER-negative breast cancer: a prospective validation cohort study from the TACT trial
Jamal-Hanjani M, et al. Ann Oncol. 2015 Jul;26(7):1340-6.

Spatial and temporal diversity in genomic instability processes defines lung cancer evolution
de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L*, Jamal-Hanjani M*, et al. Science. 2014 Oct 10;346(6206):251-6.

Expression of regulators of mitotic fidelity are associated with intercellular heterogeneity and chromosomal instability in primary breast cancer
Roylance R*, Endesfelder D*, Jamal-Hanjani M*, et al. Breast Cancer Res Treat. 2014 Nov;148(1):221-9.

** Joint authors*

8.1.2 Review articles

The value of patient and public involvement intrial design and development
S. Gasson, J. Bliss, M. Jamal-Hanjani et al. Clin Oncol (R Coll Radiol). 2015 Dec 27(12):747-9.

Lung cancer
Jamal-Hanjani and Siow-Ming Lee. Future Medicine, Lung Cancer Manag. Lung Cancer Manag. 2015 4(3), 117–123.

Translational Implications of Tumor Heterogeneity
Jamal-Hanjani M, Quezada SA, Larkin J, Swanton C. Clin Cancer Res. 2015 Mar

15;21(6):1258-1266.

Tracking genomic cancer evolution for precision medicine: the lung TRACERx study
Jamal-Hanjani M, et al. PLoS Biol. 2014 Jul 8;12(7):e1001906.

Implications of intratumour heterogeneity for treatment stratification
Crockford A*, Jamal-Hanjani M*, Hicks J, Swanton C. J Pathol. 2014 Jan;232(2):264-73.

Tumour heterogeneity and immune-modulation
Jamal-Hanjani M, Thanopoulou E, Peggs KS, Quezada SA, Swanton C. Curr Opin Pharmacol. 2013 Aug;13(4):497-503.

Taxane benefit in breast cancer-a role for grade and chromosomal stability
A'Hern RP*, Jamal-Hanjani M*, Szász AM, Johnston SR, Reis-Filho JS, Roylance R, Swanton C. Nat Rev Clin Oncol. 2013 Jun;10(6):357-64.

** Joint authors*

8.1.3 Abstracts and presentations

The TRACERx Study
Cancer Research UK Lung Cancer Centre of Excellence Conference 2015 (poster presentation).

Intratumour Heterogeneity and Clonal Evolution in Non-Small Cell Lung Cancer
Cancer Research UK Lung Cancer Centre of Excellence Summer Conference, July 2015 (oral presentation).

Association between intratumour heterogeneity and genomic markers of chromosomal instability and chemotherapy response in the prospective lung cancer clinical trial TRACERx
The British Association for Cancer Research, June 2015 (poster presentation).

Intratumor heterogeneity and the detection of potential driver mutations in cfDNA in a NSCLC cohort using UltraSEEK™
The American Association for Cancer Research, April 2015 (poster presentation).

The TRACERx Study
The National Cancer Research Institute Conference, November 2014 (poster presentation, NCRI Prize Award).

Evolution of the genomic landscape in non-small cell lung cancer
The European Society of Medical Oncology, September 2014 (proffered paper - oral presentation).

Intratumour Heterogeneity and Clonal Evolution in Non-Small Cell Lung Cancer
The Royal Society of Medicine, Oncology Section Sylvia Lawler Prize Meeting, September 2014 (poster presentation).

8.2 Appendix 2: The lung TRACERx study protocol summary

Title	TRAcking non-small cell lung Cancer Evolution through therapy (Rx)
Short Title/acronym	TRACERx
Sponsor name & reference	UCL - UCL/12/0279
Funder name	Cancer Research UK Rosetrees Trust Academy for Medical Sciences UCL Biomedical Research Centre
Clinicaltrials.gov no	NCT01888601
Design	A prospective observational cohort study of patients with non-small cell lung cancer (NSCLC), in which translational research is the fundamental aspect of the study.
Primary objectives	• Define the relationship between intratumour heterogeneity and clinical outcome following surgery and adjuvant therapy (including relationships between intratumour heterogeneity and clinical disease stage and histological subtypes of NSCLC). • Establish the impact of adjuvant platinum-containing regimens upon intratumour heterogeneity in relapsed disease.
Key secondary objectives	• Development and validation of intratumour heterogeneity (ITH) ratio index (I_{TB}) as a prognostic and predictive biomarker in relation to DFS and OS.
Primary endpoints	• Intratumour heterogeneity quantified by the ratio index I_{TB} • Disease-free survival • Overall survival
Target accrual	842 patients, of which 270 are expected to have a first recurrence and agree to provide a biopsy of the site of local recurrence/metastases
Inclusion criteria	• Written Informed consent • Patients ≥18 years of age, with early stage I-IIIA disease who are eligible for primary surgery • Histopathologically confirmed NSCLC, or a strong suspicion of cancer on lung imaging necessitating surgery (e.g. diagnosis determined from frozen section in theatre) • Primary surgery in keeping with NICE guidelines planned (see section 9.3) • Agreement to be followed up in a specialist centre • Performance status 0 or 1
Exclusion criteria	• Any other current malignancy or malignancy diagnosed or relapsed within the past 5 years (other than non-melanomatous skin cancer and in situ cervical cancer) • Psychological condition that would preclude informed consent

	• Treatment with neo-adjuvant chemotherapy deemed necessary • Adjuvant chemotherapy regimen other than platinum-based chemotherapy (if a patient is deemed suitable for adjuvant chemotherapy) • Known Human Immunodeficiency Virus (HIV), Hepatitis B Virus (HBV), Hepatitis C Virus (HCV) or syphilis infection. • Sufficient tissue is unlikely to be obtained for the study based on pre-operative imaging Patients found to have in situ lesions rather than invasive cancer following surgery, such as adenocarcinoma in situ or minimally invasive lesions will be withdrawn. However, the surgical tissue already collected will be sent to the central laboratory, but these patients will not be followed-up in the study or required to provide any further blood samples. If these patients subsequently develop invasive cancer, the date of diagnosis and the tumour histology will be reported on the electronic data capture system.
Planned number of sites	6 recruiting sites: UCLH, University Hospital of South Manchester, Heart of England, University Hospitals of Leicester, Cardiff & Vale and NHS Grampian.
Treatment summary	All recruited patients will be suitable for primary surgery, in accordance with NICE guidelines. Further treatments (e.g. chemotherapy) would be given according to standard of care. No treatments are specified as part of this observational study.
Collection of tissue samples	Baseline: • From the primary tumour and normal tissue from the resected specimen, surplus to diagnostic and pathological requirements. After first confirmed recurrence: • Biopsy (after consent) of local-regional and/or metastatic sites Following progression, after the first recurrence: • Biopsy (after consent) of local-regional and/or metastatic sites
Collection of blood samples	Baseline (before surgery): • 1 x 10mL for germ line DNA • 4 x 10mL for cfDNA • 2 x 10mL for cfDNA high sensitivity assay for approximately 100 patients initially • 2 x 10mL CellSave tube, for circulating tumour cells (pulmonary blood); for approximately 200 patients • 1 x 10mL for immunology • 1 x 40mL for immunology (only in 150 patients with larger tumours, >7cm after allowing for factors such as necrosis) • Leicester only – additional 1 x 10mL cfDNA • UCLH only – additional 1 x 10mL CellSave tube for 30 patients

	• Manchester only – additional 1 x 10mL CellSave tube. This may be omitted in patients who provide the 40mL immunology sample. <u>48 hours after surgery:</u> • 2 x 10mL for cfDNA high sensitivity assay for approximately 100 patients initially <u>Follow-up after surgery:</u> 5 years of follow up in the adjuvant setting; every 3 months in years 1-2, then every 6 months in years 3-5. For patients who receive adjuvant chemotherapy, the first post-surgery sample should be collected just before cycle 1 treatment and the next sample should be collected 6 months from the day of surgery • 2 x 10mL for cfDNA • 2 x 10mL for cfDNA high sensitivity assay for approximately 100 patients initially (for up to 3 years of follow-up only) • 2 x 10mL for cfDNA for patients identified with an EGFR or BRAF mutation • Leicester only – additional 1 x 10mL cfDNA <u>At first confirmed recurrence</u>: • 4 x 10mL for cfDNA • 2 x 10mL for cfDNA high sensitivity assay for approximately 100 patients initially (only if first recurrence is within 3 years of surgery) • 2 x 10mL for cfDNA for patients identified with an EGFR or BRAF mutation • 1 x 10mL CellSave tube (circulating tumour cells) • 1 x 10mL for immunology • 1 x 40mL (only in the patients with larger tumours who provided a 40mL blood sample for immunology before surgery) • Leicester only – additional 1 x 10mL cfDNA • UCLH only – additional 1 x 10mL CellSave tube for those of the 30 patient cohort whose disease recurs • UCLH only – additional 1 x 10ml CellSave tube for 10 patients with EGFR, ALK, ROS, RET, KRAS or PD-L1 mutation <u>Follow-up after first recurrence:</u> At first CT on treatment for first recurrence • 2 x 10mL for cfDNA • 2 x 10mL for cfDNA for patients identified with an EGFR or BRAF mutation • 1 x 10mL CellSave tube • Leicester only – additional 1 x 10mL cfDNA • UCLH only – additional 1 x 10mL CellSave tube for those of the 30 patient cohort whose disease recurs • UCLH only – additional 1 x 10ml CellSave tube for the 10 patient cohort with EGFR, ALK, ROS, RET, KRAS or PD-L1 mutation

	<u>At progression, after the first recurrence</u>: • 2 x 10mL for cfDNA • 2 x 10mL for cfDNA for patients identified with an EGFR or BRAF mutation • 1 x 10mL Cellsave tube • Leicester only – additional 1 x 10mL cfDNA • UCLH only – additional 1 x 10mL CellSave tube for those of the 30 patient cohort whose disease progresses • UCLH only – additional 1 x 10ml CellSave tube for the 10 patient cohort with EGFR, ALK, ROS, RET, KRAS or PD-L1 mutation <u>Following progression</u>: At up to 4 more timepoints, for example at the first CT scan whilst on treatment and at subsequent progressions • 2 x 10mL for cfDNA • 2 x 10mL for cfDNA for patients identified with an EGFR or BRAF mutation • Leicester only – additional 1 x 10mL cfDNA <u>Completion of all treatment</u>: • 4 x 10mL for cfDNA • 2 x 10mL for cfDNA for patients identified with an EGFR or BRAF mutation • 1 x 10mL Cellsave tube • Leicester only – additional 1 x 10mL cfDNA • UCLH only – additional 1 x 10mL CellSave tube where applicable to the 30 patients within the cohort • UCLH only – additional 1 x 10ml CellSave tube for the 10 patient cohort with EGFR, ALK, ROS, RET, KRAS or PD-L1 mutation
Translational research	• Multi-region tissue sampling (including DNA and RNA sequencing and immunological analyses) • Blood samples for germ line DNA • Blood samples for circulating free tumour DNA (cfDNA) • Blood samples for circulating tumour cells • Blood sample for immunology
Duration of patient follow-up	5-years

8.3 Appendix 3: Number of regions sequenced and sequencing depth in the NSCLC and TRACERx cohort

Table 29 shows the number of regions sequenced, and the mean and median sequencing depth for WES and/or Ion AmpliSeq sequencing for each sample. Where a recurrence sample was sequenced in addition to the primary tumour (L011 and L017), the mean sequencing depth is also shown.

Patient	Region	Mean depth of WES	Median depth of WES	Mean depth of Ion AmpliSeq
L011	GL	99	78	1442
	R1	92	70	1542
	R2	95	73	1181
	R3	92	70	1695
	M1	117	105	
	M2	118	106	
	M3	123	112	
	M4	120	110	
L012	GL	279	130	1912
	R1	183	85	2148
	R2	237	97	2075
	R3	217	92	1291
L013	GL	124	116	868
	R1	107	100	844
	R2	114	104	902
	R3	102	90	821
L015	GL	411	380	415
	R1	335	302	446
	R2	373	339	459
L016	GL	415	383	2660
	R2	335	317	2516
	R3	250	231	2228
L017	GL	133	123	1415
	R1	156	140	1433
	R2	176	162	1284
	R3	162	149	1656
	M1	165	142	
L019	GL	232	130	1028
	R1	145	75	1098
	LN1	161	94	995
	LN2	307	160	1326
	LN3	172	82	761
L022	GL	71	65	

	R1	233	210	
	R2	124	110	
	LN1	238	215	
	GL	143	133	
L023	R1	154	135	
	R2	212	182	
	GL	98	686	688
L029	R1	145	785	781
	R2	149	1168	1180
	GL	102	1931	1914
L030	R1	166	2046	1967
	R2	170	2277	2221
	GL	478	431	
LTX001	R1	384	346	
	R2	432	383	
	R3	534	478	
	GL	420	381	
LTX012	R1	424	371	
	R2	436	380	
	GL	330	299	
	R1	355	310	
LTX015	R2	405	349	
	R3	366	325	
	R4	418	365	
	GL	384	351	
	R1	411	372	
	R2	502	456	
LTX016	R3	450	408	
	R4	432	395	
	R5	400	362	
	GL	436	400	
	R3	465	422	
	R4	422	379	
LTX019	R5	473	423	
	R6	495	449	
	R7	415	377	
	GL	381	350	
LTX022	R1	431	393	
	R5	280	253	
	R6	481	438	
	GL	414	370	
LTX028	R1	510	445	
	R2	478	428	
	R3	633	500	

	R5	433	387
	R6	496	439
	R7	527	459
	GL	393	358
LTX029	R1	462	422
	R2	518	467
	GL	415	374
LTX030	R1	352	325
	R2	561	500
	GL	460	419
LTX031	R1	444	401
	R2	461	412
	GL	302	271
	R1	376	339
LTX033	R2	328	286
	R3	390	338
	R4	360	319
	GL	427	389
LTX034	R1	352	317
	R2	440	395
	GL	353	314
	R1	478	431
	R2	497	447
LTX036	R3	469	414
	R4	518	457
	R6	442	394
	GL	282	252
	R1	457	390
	R2	411	356
	R3	404	355
LTX038	R4	434	384
	R5	389	341
	R6	364	315
	R7	298	262
	GL	461	416
LTX051	R1	412	371
	R2	391	350
	GL	397	359
	R1	406	352
	R2	343	299
LTX058	R3	370	315
	R4	355	311
	R5	368	326

Table 29 Number of regions sequenced and sequencing depth

Number of regions sequenced, mean and median depth of WES, and mean depth of Ion AmpliSeq sequencing for each tumour region in the NSCLC and TRACERx cohorts.

8.4 Appendix 4: Driver category 1 to 3 mutations in the NSCLC and TRACERx cohort

Table 30 shows the genes in which driver mutations were predicted in the combined NSCLC and TRACERx cohort. The gene name, mutation type (SNV or indel), driver category, which region of the tumour the mutation was present in, and whether the mutation was on the trunk or branch of the evolutionary tree is included.

Patient	Gene	Mutation type	Driver category	Truncal/Branch
L011	BRAF	SNV	1	Truncal
	TP53	SNV	1	Truncal
	BCLAF1	SNV	2	Branch
	FAT1	Indel	2	Truncal
	FAT1	SNV	2	Truncal
	FAT1	SNV	2	Truncal
	MTOR	Indel	2	Truncal
	AKT1	SNV	3	Truncal
	CACNA1D	SNV	3	Branch
	CDC73	SNV	3	Truncal
	CLTCL1	SNV	3	Truncal
	HOOK3	SNV	3	Truncal
	JAK1	SNV	3	Truncal
	MYCN	SNV	3	Truncal
	MYH9	SNV	3	Truncal
	PIM1	SNV	3	Truncal
	XPC	SNV	3	Truncal
L012	BRIP1	SNV	1	Truncal
	CDKN2A	SNV	1	Truncal
	CIC	SNV	1	Branch
	KDM6A	SNV	1	Branch
	TP53	SNV	1	Truncal
	TP53	SNV	1	Truncal
	CARS	SNV	2	Truncal
	NFE2L2	SNV	2	Truncal
	RASA1	SNV	2	Branch
	FANCA	SNV	3	Branch
	FAT1	SNV	3	Branch
	MLLT4	SNV	3	Branch
	TLX1	SNV	3	Branch
L013	EGFR	SNV	1	Truncal
	EGFR	SNV	1	Truncal

	MSH2	SNV	1	Branch
	MLL3	SNV	2	Branch
	MTOR	SNV	2	Branch
	PLCG2	SNV	2	Branch
	TP53	SNV	2	Truncal
	DNAH12	SNV	3	Truncal
	JAK2	SNV	3	Branch
	KIT	SNV	3	Truncal
	MED12	SNV	3	Truncal
	SMO	SNV	3	Truncal
L015	KDM6A	SNV	1	Truncal
	KMT2D	SNV	1	Branch
	TP53	SNV	1	Truncal
	ETV6	SNV	2	Truncal
	ALK	SNV	3	Truncal
	GOLGA5	SNV	3	Truncal
	ROS1	SNV	3	Truncal
	TFRC	SNV	3	Truncal
L016	BRCA1	SNV	1	Branch
	TP53	SNV	1	Truncal
	GAS7	SNV	2	Truncal
	NOTCH1	SNV	2	Truncal
	FLT3	SNV	3	Truncal
L017	ARID1A	SNV	1	Truncal
	ATRX	SNV	1	Truncal
	ATRX	SNV	1	Truncal
	ATRX	SNV	1	Truncal
	ATRX	SNV	1	Truncal
	BCOR	SNV	1	Branch
	FBXW7	SNV	1	Truncal
	KRAS	SNV	1	Truncal
	NF1	SNV	1	Truncal
	PMS1	SNV	1	Truncal
	TP53	SNV	1	Truncal
	TP53	SNV	1	Truncal
	WT1	SNV	1	Truncal
	ALK	SNV	2	Truncal
	ALK	SNV	2	Truncal
	AMER1	SNV	2	Truncal
	AMER1	SNV	2	Truncal
	ATP2B3	SNV	2	Truncal
	BRCA2	SNV	2	Truncal
	CCND1	SNV	2	Truncal
	CHD4	SNV	2	Truncal

	ETV5	SNV	2	Truncal
	FOXP1	SNV	2	Truncal
	GNPTAB	SNV	2	Truncal
	GPC3	SNV	2	Truncal
	KEAP1	SNV	2	Truncal
	MGA	SNV	2	Truncal
	NCOA1	SNV	2	Truncal
	NF1	SNV	2	Truncal
	POLE	SNV	2	Truncal
	RBM10	SNV	2	Truncal
	STAG2	SNV	2	Truncal
	ASPSCR1	SNV	3	Truncal
	ATP2B3	SNV	3	Truncal
	BRAF	SNV	3	Truncal
	BRD4	SNV	3	Truncal
	CDC73	SNV	3	Truncal
	CNKSR1	SNV	3	Truncal
	FAT1	SNV	3	Truncal
	FGFR1OP	SNV	3	Truncal
	HIST1H3B	SNV	3	Branch
	HOOK3	SNV	3	Truncal
	HOXD13	SNV	3	Branch
	KDM5A	SNV	3	Truncal
	MTOR	SNV	3	Truncal
	NOTCH1	SNV	3	Truncal
	NTRK1	SNV	3	Truncal
	PAX8	SNV	3	Branch
	PRDM16	SNV	3	Truncal
	RANBP17	SNV	3	Branch
	TPR	SNV	3	Branch
	TRAF7	SNV	3	Truncal
	TRIP11	SNV	3	Branch
	WAS	SNV	3	Truncal
	WWTR1	SNV	3	Truncal
	ZRSR2	SNV	3	Truncal
L019	EML4-ALK	Translocation	1	Truncal
	AKAP9	SNV	3	Truncal
	MYCN	SNV	3	Branch
	POU2AF1	SNV	3	Branch
L022	STAG2	SNV	1	Truncal
	TP53	SNV	1	Truncal
	WT1	SNV	1	Truncal
	FAT1	SNV	2	Truncal
	MGA	SNV	2	Truncal

	SETBP1	SNV	2	Truncal
	ABL2	SNV	3	Branch
	CALR	SNV	3	Branch
	CNBD1	SNV	3	Truncal
	JAK1	SNV	3	Truncal
L023	CASP8	SNV	1	Branch
	FANCA	SNV	1	Branch
	KRAS	SNV	1	Truncal
	WT1	SNV	1	Truncal
	GOLGA5	SNV	2	Truncal
	PIK3CA	SNV	2	Truncal
	ARID1A	SNV	3	Branch
	CCNB1IP1	SNV	3	Branch
	GMPS	SNV	3	Branch
	GNAS	SNV	3	Branch
	HLF	SNV	3	Truncal
	HNRNPA2B1	SNV	3	Branch
	MAP2K1	SNV	3	Branch
	PIM1	SNV	3	Branch
	SETBP1	SNV	3	Branch
	SMO	SNV	3	Truncal
	TET2	SNV	3	Branch
L029	KRAS	SNV	1	Truncal
	MSH6	SNV	1	Truncal
	TP53	SNV	1	Truncal
	AMER1	SNV	2	Truncal
	CHD8	SNV	2	Truncal
	KIT	SNV	2	Truncal
	PICALM	SNV	2	Truncal
L030	EGFR	Indel	1	Truncal
	ATM	Indel	2	Truncal
LTX001	ATM	SNV	1	Truncal
	ATM	SNV	1	Truncal
	CIC	SNV	1	Branch
	KRAS	SNV	1	Truncal
	EGFR	SNV	2	Truncal
	HERPUD1	SNV	2	Branch
	MBD1	SNV	2	Truncal
	MGA	SNV	2	Branch
	EBF1	SNV	3	Truncal
	ELF4	SNV	3	Truncal
	JAK2	SNV	3	Branch
	PDE4DIP	SNV	3	Branch
	RANBP17	SNV	3	Truncal

	SUZ12	SNV	3	Branch
	SYK	SNV	3	Truncal
	TPR	SNV	3	Truncal
LTX012	CDKN2A	Indel	1	Truncal
	NRAS	SNV	1	Truncal
	BCL11B	SNV	2	Truncal
	DDX3X	SNV	2	Truncal
	EBF1	SNV	2	Truncal
	KDR	SNV	2	Truncal
	MTOR	SNV	2	Truncal
	MTOR	SNV	2	Truncal
	PAX5	SNV	2	Truncal
	PLCG2	SNV	2	Truncal
	SOCS1	SNV	2	Truncal
	CDH11	SNV	3	Truncal
	GAS7	SNV	3	Truncal
	NSD1	SNV	3	Truncal
	PIK3R1	Indel	3	Truncal
	PRDM16	SNV	3	Truncal
	PTPRC	SNV	3	Truncal
	SLC45A3	SNV	3	Truncal
	SRGAP3	SNV	3	Branch
LTX015	ATM	SNV	1	Branch
	TP53	SNV	1	Truncal
	ZRSR2	SNV	1	Truncal
	SEPT12	SNV	2	Truncal
	EIF4A2	SNV	2	Truncal
	PIK3CA	SNV	2	Truncal
	RB1	SNV	2	Branch
	RPS2	SNV	2	Truncal
	SERPINB13	SNV	2	Truncal
	CD274	SNV	3	Branch
	CLTCL1	SNV	3	Branch
	CREB3L1	SNV	3	Truncal
	CSF3R	SNV	3	Truncal
	HOOK3	SNV	3	Branch
	IRF4	SNV	3	Truncal
	KMT2D	SNV	3	Truncal
	MGA	SNV	3	Truncal
	MPL	SNV	3	Branch
	NUP214	SNV	3	Truncal
	ODAM	SNV	3	Branch
LTX016	PIK3CA	SNV	1	Truncal
	ATM	SNV	3	Branch

	MED12	SNV	3	Branch
	MYH9	SNV	3	Branch
	PIK3CA	SNV	3	Branch
	PRF1	SNV	3	Branch
LTX019	CDKN2A	Indel	1	Truncal
	CDKN2A	SNV	1	Truncal
	CYLD	SNV	1	Truncal
	DAXX	SNV	1	Branch
	EP300	SNV	1	Branch
	FBXW7	SNV	1	Branch
	MSH2	SNV	1	Truncal
	NF1	SNV	1	Branch
	PIK3CA	SNV	1	Truncal
	TP53	SNV	1	Truncal
	EZH1	SNV	2	Truncal
	GNPTAB	SNV	2	Branch
	HOOK3	SNV	2	Truncal
	CDKN2A	SNV	3	Truncal
	EGFR	SNV	3	Branch
LTX022	CYLD	SNV	1	Branch
	FANCC	SNV	1	Branch
	TP53	SNV	1	Truncal
	BCLAF1	SNV	2	Truncal
	BCLAF1	SNV	2	Truncal
	CNBD1	SNV	2	Truncal
	ERCC4	SNV	2	Branch
	FAT1	SNV	2	Truncal
	MBD1	SNV	2	Truncal
	MBD1	SNV	2	Truncal
	NSD1	SNV	2	Branch
	ZNF521	SNV	2	Truncal
	CREBBP	SNV	3	Truncal
	DDX10	SNV	3	Branch
	GNA11	SNV	3	Truncal
	LCK	SNV	3	Truncal
	NOTCH1	SNV	3	Truncal
	SETBP1	SNV	3	Truncal
LTX028	TP53	Indel	1	Truncal
	CIITA	SNV	2	Truncal
	IL21R	SNV	2	Truncal
	MLLT10	SNV	2	Truncal
	MLLT10	SNV	2	Truncal
	PIK3CA	SNV	2	Truncal
	ROS1	SNV	2	Truncal

	TRIP11	SNV	2	Truncal
	ZNF521	SNV	2	Truncal
	ACSL3	SNV	3	Truncal
	ETV1	SNV	3	Truncal
	HIST1H3B	SNV	3	Branch
	MUC1	SNV	3	Truncal
	TAL1	SNV	3	Truncal
LTX029	APC	Indel	1	Truncal
	PIK3CA	SNV	1	Branch
	TP53	SNV	1	Truncal
	ERBB2	SNV	2	Truncal
LTX030	CBLB	SNV	1	Truncal
	PIK3CA	SNV	1	Truncal
	TP53	SNV	1	Truncal
	RASA1	SNV	2	Truncal
	SMAD4	SNV	2	Truncal
	TRIM33	SNV	2	Truncal
	TCF3	SNV	3	Branch
LTX031	CDKN2A	SNV	1	Truncal
	SMARCA4	SNV	1	Truncal
	TP53	SNV	1	Truncal
	KTN1	SNV	2	Truncal
	MTOR	SNV	2	Truncal
	NTRK1	SNV	2	Truncal
	FAT1	SNV	3	Truncal
	GNAS	SNV	3	Truncal
	MLLT10	SNV	3	Truncal
	MYB	SNV	3	Truncal
	TOP1	SNV	3	Truncal
LTX033	TP53	SNV	1	Truncal
	ERBB3	SNV	2	Truncal
	KEAP1	SNV	2	Truncal
	WAS	SNV	2	Branch
	ASXL1	SNV	3	Truncal
	BRIP1	SNV	3	Truncal
	HOXD11	SNV	3	Truncal
	MTOR	SNV	3	Truncal
	NSD1	SNV	3	Truncal
	ZNF521	SNV	3	Truncal
LTX034	BAP1	SNV	1	Truncal
	RB1	SNV	1	Branch
	TP53	SNV	1	Branch
	EZH2	SNV	2	Truncal
	MED12	SNV	3	Truncal

LTX036	ATM	SNV	1	Branch
	STK11	SNV	1	Truncal
	TP53	SNV	1	Truncal
	MYC	SNV	2	Branch
	PAX7	SNV	2	Branch
	POLE	SNV	2	Truncal
	ZNF521	SNV	2	Truncal
	ATP2B3	SNV	3	Branch
	BCL2	SNV	3	Branch
	FLI1	SNV	3	Branch
	IL7R	SNV	3	Truncal
	JAK3	SNV	3	Truncal
	MAP3K1	SNV	3	Branch
	MYH9	Indel	3	Branch
	RALGDS	SNV	3	Branch
	USP6	SNV	3	Truncal
LTX038	ARID1A	SNV	1	Branch
	CDKN2A	SNV	1	Truncal
	MLH1	SNV	1	Branch
	NFE2L2	SNV	1	Truncal
	NIN	SNV	1	Truncal
	PIK3CA	SNV	1	Branch
	TP53	SNV	1	Truncal
	SEPT12	SNV	2	Branch
	ATRX	SNV	2	Branch
	CCND1	SNV	2	Branch
	FAM166A	SNV	2	Branch
	JAK1	SNV	2	Branch
	KDM5A	SNV	2	Branch
	KDR	SNV	2	Branch
	NOTCH1	SNV	2	Branch
	NUP214	SNV	2	Truncal
	PTPRC	SNV	2	Branch
	SMC3	SNV	2	Branch
	STX2	SNV	2	Branch
	ACVR1B	SNV	3	Branch
	ARHGEF12	SNV	3	Branch
	CACNA1D	SNV	3	Truncal
	CHD8	SNV	3	Truncal
	CRTC3	SNV	3	Branch
	FANCF	SNV	3	Branch
	FAT1	SNV	3	Branch
	GNAS	SNV	3	Truncal
	KCNJ5	SNV	3	Truncal

	KIAA1549	SNV	3	Branch
	MSN	SNV	3	Branch
	NCOA1	SNV	3	Branch
	RBM10	SNV	3	Truncal
LTX051	EGFR	SNV	1	Truncal
	TP53	SNV	1	Truncal
	EGFR	SNV	2	Truncal
	INTS12	Indel	2	Truncal
	NSD1	SNV	2	Truncal
	SERPINB13	SNV	2	Branch
LTX058	AMER1	SNV	1	Branch
	KRAS	SNV	1	Truncal
	TP53	SNV	1	Truncal
	BCLAF1	SNV	2	Branch
	FAT1	SNV	2	Truncal
	LIFR	SNV	2	Branch
	MYH11	SNV	2	Truncal
	NTRK3	SNV	2	Truncal
	PAX5	SNV	2	Truncal
	ARHGEF12	SNV	3	Branch
	FCGR2B	SNV	3	Branch
	JAK2	SNV	3	Branch
	MN1	SNV	3	Truncal
	NFKB2	SNV	3	Branch
	ROS1	SNV	3	Truncal
	TRIP11	SNV	3	Truncal

Table 30 Driver category 1 to 3 mutations in the NSCLC pilot and TRACERx cohorts

Predicted category 1 to 3 driver mutations identified in each tumour for the NSCLC and TRACERx cohorts. The gene name, type of variant (SNV or indel), and whether the mutation is truncal or branch is shown.

8.5 Appendix 5: Number of regions sequenced and sequencing depth in the lung adenocarcinoma *in situ* lesions

Table 31 shows the number of regions sequenced, and the mean and median sequencing depth for WES for each sample from the adenocarcinoma *in situ* lesions.

Patient	Region	Mean depth of WES	Median depth of WES
LTX013	GL	363.5	331
	R1	387.3	351
	R2	386.6	348
	R3	369.6	348
LTX021	GL	534	483
	R1	444.7	402
	R2	515.3	452
	R3	482.5	438
	R4	441.2	396
LTX041	GL	339.6	308
	R2	328	295
	R3	451.4	403
	R4	564.9	500
LTX049	GL	441.5	406
	R1	342.8	309
	R2	480.3	423
LTX055	GL	547.5	500
	R1	573.7	500
	R2	381.1	341
	R3	370	331

Table 31 Number of regions sequence and sequencing depth

Number of regions sequenced, mean and median depth of WES for each tumour region in the AIS tumours.

Reference List

A'Hern, R.P. et al., 2013. Taxane benefit in breast cancer—a role for grade and chromosomal stability. *Nature Publishing Group*, 10(6), pp.357–364.

Abrahamsson, P.-A., 2010. Potential benefits of intermittent androgen suppression therapy in the treatment of prostate cancer: a systematic review of the literature. *European urology*, 57(1), pp.49–59.

Abramyuk, A. et al., 2010. Comment on "Developing DCE-CT to quantify intra-tumor heterogeneity in breast tumors with differing angiogenic phenotype." *IEEE Trans Med Imaging*, 29(4), pp.1088–9– author reply 1089–92.

Adzhubei, I.A. et al., 2010. A method and server for predicting damaging missense mutations. *Nature Methods*, 7(4), pp.248–249.

Ahrendt, S.A. et al., 2001. Cigarette smoking is strongly associated with mutation of the K-ras gene in patients with primary adenocarcinoma of the lung. *Cancer*, 92(6), pp.1525–1530.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., et al., 2013. Signatures of mutational processes in human cancer. *Nature*, 500(7463), pp.415–421.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Campbell, P.J., et al., 2013. Deciphering signatures of mutational processes operative in human cancer. *CellReports*, 3(1), pp.246–259.

Anand, S., Penrhyn-Lowe, S. & Venkitaraman, A.R., 2003. AURORA-A amplification overrides the mitotic spindle assembly checkpoint, inducing resistance to Taxol. *Cancer Cell*, 3(1), pp.51–62.

Anderson, K., Lutz, C., van Delft, F.W., Bateman, C.M., Guo, Y., Colman, S.M., Kempski, H., Moorman, A.V., Titley, I., Swansbury, J., Kearney, L., Enver, T. & Greaves, M., 2011. Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature*, 469(7330), pp.356–361.

Anglesio, M.S. et al., 2013. Molecular characterization of mucinous ovarian tumours supports a stratified treatment approach with HER2 targeting in 19% of carcinomas. *The Journal of Pathology*, 229(1), pp.111–120.

Antonarakis, E.S. et al., 2014. AR-V7 and Resistance to Enzalutamide and Abiraterone in Prostate Cancer. *New England Journal of Medicine*, p.140903140143003.

Atefi, M. et al., 2011. Reversing melanoma cross-resistance to BRAF and MEK inhibitors by co-targeting the AKT/mTOR pathway. J. M. Brandner, ed. *PLoS ONE*, 6(12), p.e28973.

Bailey, S.M. & Murnane, J.P., 2006. Telomeres, chromosome instability and cancer. *Nucleic Acids Research*, 34(8), pp.2408–2417.

Bakhoum, S.F. et al., 2011. Chromosomal instability substantiates poor prognosis in patients with diffuse large B-cell lymphoma. *Clinical Cancer Research*, 17(24), pp.7704–7711.

Balak, M.N. et al., 2006. Novel D761Y and common secondary T790M mutations in epidermal growth factor receptor-mutant lung adenocarcinomas with acquired resistance to kinase inhibitors. *Clinical Cancer Research*, 12(21), pp.6494–6501.

Baldus, S.E. et al., 2010. Prevalence and heterogeneity of KRAS, BRAF, and PIK3CA mutations in primary colorectal adenocarcinomas and their corresponding metastases. *Clinical Cancer Research*, 16(3), pp.790–799.

Barber, T.D. et al., 2008. Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. *Proceedings of the National Academy of Sciences of the United States of America*, 105(9), pp.3443–3448.

Bashashati, A. et al., 2013. Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *The Journal of Pathology*, 231(1), pp.21–34.

Baumbusch, L.O. et al., 2013. High Levels of Genomic Aberrations in Serous Ovarian Cancers Are Associated with Better Survival A. J. R. Bishop, ed. *PLoS ONE*, 8(1), pp.e54356–8.

Bean, J. et al., 2008. Acquired resistance to epidermal growth factor receptor kinase inhibitors associated with a novel T854A mutation in a patient with EGFR-mutant lung adenocarcinoma. *Clinical Cancer Research*, 14(22), pp.7519–7525.

Bergers, E., van Diest, P.J. & Baak, J.P., 1996. Tumour heterogeneity of DNA cell cycle variables in breast cancer measured by flow cytometry. *J Clin Pathol*, 49(11), pp.931–937.

Bettegowda, C. et al., 2014. Detection of Circulating Tumor DNA in Early- and Late-Stage

Human Malignancies. *Science Translational Medicine*, 6(224), pp.224ra24–224ra24.
Bidard, F.-C., Weigelt, B. & Reis-Filho, J.S., 2013. Going with the flow: from circulating tumor cells to DNA. *Science Translational Medicine*, 5(207), pp.207ps14–207ps14.
Birkbak, N.J. et al., 2011. Paradoxical relationship between chromosomal instability and survival outcome in cancer. *Cancer Research*, 71(10), pp.3447–3452.
Bloom, H.J. & Richardson W.W., 1957. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *British Journal of Cancer*, 11(3), pp.359–377.
Board, R.E. et al., 2010. Detection of PIK3CA mutations in circulating free DNA in patients with breast cancer. *Breast Cancer Research and Treatment*, 120(2), pp.461–467.
Boeckx, N. et al., 2014. Anti-EGFR Resistance in Colorectal Cancer: Current Knowledge and Future Perspectives. *Current Colorectal Cancer Reports*, pp.1–15.
Bolli, N. et al., 2014. ncomms3997. *Nature Communications*, 5, pp.1–13.
Borthwick, N.J. et al., 2011. The biology of micrometastases from uveal melanoma. *Journal of Clinical Pathology*, 64(8), pp.666–671.
Bouchet, B.P. et al., 2007. Paclitaxel resistance in untransformed human mammary epithelial cells is associated with an aneuploidy-prone phenotype. *British Journal of Cancer*, 97(9), pp.1218–1224.
Bower, K. et al., 2012. Loss of wild-type ATRX expression in somatic cell hybrids segregates with activation of Alternative Lengthening of Telomeres. S. Ahmed, ed. *PLoS ONE*, 7(11), p.e50062.
Brevet, M. et al., 2011. Detection of EGFR mutations in plasma DNA from lung cancer patients by mass spectrometry genotyping is predictive of tumor EGFR status and response to EGFR inhibitors. *Lung Cancer*, 73(1), pp.96–102.
Brose, M.S. et al., 2002. BRAF and RAS mutations in human lung cancer and melanoma. *Cancer Research*, 62(23), pp.6997–7000.
Bryan, T.M. et al., 1997. Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nature Medicine*, 3(11), pp.1271–1274.
Burns, M.B., Temiz, N.A. & Harris, R.S., 2013. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nature Genetics*, 45(9), pp.977–983.
Burrell, R.A. & Swanton, C., 2014. Tumour heterogeneity and the evolution of polyclonal drug resistance. *Molecular Oncology*, pp.1–17.
Burrell, R.A. et al., 2013. Replication stress links structural and numerical cancer chromosomal instability. *Nature*, 494(7438), pp.492–496.
Cahill, D.P. et al., 1999. Genetic instability and darwinian selection in tumours. *Trends in cell biology*, 9(12), pp.M57–60.
Cahill, D.P. et al., 1998. Mutations of mitotic checkpoint genes in human cancers. *Nature*, 392(6673), pp.300–303.
Campbell, P.J. et al., 2010. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature*, 467(7319), pp.1109–1113.
Cardarella, S. et al., 2013. Clinical, pathologic, and biologic features associated with BRAF mutations in non-small cell lung cancer. *Clinical Cancer Research*, 19(16), pp.4532–4540.
Carreira, S. et al., 2014. Tumor clone dynamics in lethal prostate cancer. *Science Translational Medicine*, 6(254), pp.254ra125–254ra125.
Carter, S.L. et al., 2006. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nature Genetics*, 38(9), pp.1043–1048.
Castells, A. et al., 1999. K-ras mutations in DNA extracted from the plasma of patients with pancreatic carcinoma: diagnostic utility and prognostic significance. *Journal of Clinical Oncology*, 17(2), pp.578–584.
Chan, K.C.A. et al., 2013. Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. *Clinical Chemistry*, 59(1), pp.211–224.
Chin, S.F. et al., 2007. High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biology*, 8(10), p.R215.
Choi, C.-M. et al., 2009. Chromosomal instability is a risk factor for poor prognosis of adenocarcinoma of the lung: Fluorescence in situ hybridization analysis of paraffin-embedded tissue from Korean patients. *Lung Cancer*, 64(1), pp.66–70.

Cooper, C.S. et al., 2015. Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nature Genetics*, 47(4), pp.367–372.

Cottu, P.H. et al., 2008. Intratumoral heterogeneity of HER2/neu expression and its consequences for the management of advanced breast cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, 19(3), pp.595–597.

Crasta, K. et al., 2012. DNA breaks and chromosome pulverization from errors in mitosis. *Nature*, 482(7383), pp.53–58.

Crowley, E. et al., 2013. Liquid biopsy: monitoring cancer-genetics in the blood. *Nature Reviews Clinical Oncology*, 10(8), pp.472–484.

Darzynkiewicz, Z., Halicka, H.D. & Zhao, H., 2010. Analysis of cellular DNA content by flow and laser scanning cytometry. *Adv Exp Med Biol*, 676, pp.137–147.

Davila, E. & Amazon, K., 2010. The Clinical Importance of the Heterogeneity of HER2 neu. *Case reports in oncology*, 3(2), pp.268–271.

Dawson, S.-J. et al., 2013. Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer. *New England Journal of Medicine*, 368(13), pp.1199–1209.

de Bruin, E.C. et al., 2014. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science*, 346(6206), pp.251–256.

Dereli-Öz, A., Versini, G. & Halazonetis, T.D., 2011. Studies of genomic copy number changes in human cancers reveal signatures of DNA replication stress. *Molecular Oncology*, 5(4), pp.308–314.

Dhillon, A.S. et al., 2007. MAP kinase signalling pathways in cancer. *Oncogene*, 26(22), pp.3279–3290.

Diaz, L.A. & Bardelli, A., 2014. Liquid biopsies: genotyping circulating tumor DNA. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 32(6), pp.579–586.

Diaz, L.A. et al., 2012. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature*, 486(7404), pp.537–540.

Diaz-Rodríguez, E. et al., 2008. Hec1 overexpression hyperactivates the mitotic checkpoint and induces tumor formation in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, 105(43), pp.16719–16724.

Dick, J.E., 2008. Stem cell concepts renew cancer research. *Blood*, 112(13), pp.4793–4807.

Diehl, F. et al., 2008. Circulating mutant DNA to assess tumor dynamics. *Nature Medicine*, 14(9), pp.985–990.

Diehl, F. et al., 2005. Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proceedings of the National Academy of Sciences*, 102(45), pp.16368–16373.

Ding, L. et al., 2008. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature*, 455(7216), pp.1069–1075.

Dressman, D. et al., 2003. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proceedings of the National Academy of Sciences*, 100(15), pp.8817–8822.

Duesberg, P., Stindl, R. & Hehlmann, R., 2000. Explaining the high mutation rates of cancer cells to drug and multidrug resistance by chromosome reassortments that are catalyzed by aneuploidy. *Proceedings of the National Academy of Sciences*, 97(26), pp.14295–14300.

Ellis, P. et al., 2009. Sequential docetaxel as adjuvant chemotherapy for early breast cancer (TACT): an open-label, phase III, randomised controlled trial. *The Lancet*, 373(9676), pp.1681–1692.

Embuscado, E.E. et al., 2005. Immortalizing the complexity of cancer metastasis: genetic features of lethal metastatic pancreatic cancer obtained from rapid autopsy. *Cancer biology & therapy*, 4(5), pp.548–554.

Endesfelder, D. et al., 2011. A breast cancer meta-analysis of two expression measures of chromosomal instability reveals a relationship with younger age at diagnosis and high risk histopathological variables. *Oncotarget*, 2(7), pp.529–537.

Endesfelder, D. et al., 2014. Chromosomal Instability Selects Gene Copy-Number Variants Encoding Core Regulators of Proliferation in ER+ Breast Cancer. *Cancer Research*, 74(17), pp.4853–4863.

Engelman, J.A. & Settleman, J., 2008. Acquired resistance to tyrosine kinase inhibitors during

cancer therapy. *Current Opinion in Genetics & Development*, 18(1), pp.73–79.
Engelman, J.A. et al., 2006. Allelic dilution obscures detection of a biologically significant resistance mutation in EGFR-amplified lung cancer. *Journal of Clinical Investigation*, 116(10), pp.2695–2706.
Farabegoli, F. et al., 2001. Clone heterogeneity in diploid and aneuploid breast carcinomas as detected by FISH. *Cytometry*, 46(1), pp.50–56.
Favero, F. et al., 2015. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, 26(1), pp.64–70.
Fiegl, M. et al., 2000. Malignant cell detection by fluorescence in situ hybridization (FISH) in effusions from patients with carcinoma. *Human pathology*, 31(4), pp.448–455.
Flynn, A. et al., 2015. The genomic landscape of phaeochromocytoma. *The Journal of Pathology*, 236(1), pp.78–89.
Fojo, T., Mailankody, S. & Lo, A., 2014. Unintended consequences of expensive cancer therapeutics-the pursuit of marginal indications and a me-too mentality that stifles innovation and creativity: the john conley lecture. *JAMA otolaryngology-- head & neck surgery*, 140(12), pp.1225–1236.
Forbes, S.A. et al., 2011. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Research*, 39(Database issue), pp.D945–50.
Forshew, T. et al., 2012. Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA. *Science Translational Medicine*, 4(136), pp.136ra68–136ra68.
Frattini, M. et al., 2008. Quantitative and qualitative characterization of plasma DNA identifies primary and recurrent colorectal cancer. *Cancer Letters*, 263(2), pp.170–181.
Futreal, P.A. et al., 2004. A census of human cancer genes. *Nature Reviews Cancer*, 4(3), pp.177–183.
Gallia, G.L. et al., 2006. PIK3CA gene mutations in pediatric and adult glioblastoma multiforme. *Molecular cancer research : MCR*, 4(10), pp.709–714.
Gancberg, D. et al., 2002. Comparison of HER-2 status between primary breast cancer and corresponding distant metastatic sites. *Annals of Oncology*, 13(7), pp.1036–1043.
Gandhi, J. et al., 2009. Alterations in genes of the EGFR signaling pathway and their relationship to EGFR tyrosine kinase inhibitor sensitivity in lung cancer cell lines. A. Lewin, ed. *PLoS ONE*, 4(2), p.e4576.
Ganem, N.J., Godinho, S.A. & Pellman, D., 2009. A mechanism linking extra centrosomes to chromosomal instability. *Nature*, 460(7252), pp.278–282.
Garassino, M.C. et al., 2011. Different types of K-Ras mutations could affect drug sensitivity and tumour behaviour in non-small-cell lung cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, 22(1), pp.235–237.
Gasch, C. et al., 2013. Heterogeneity of epidermal growth factor receptor status and mutations of KRAS/PIK3CA in circulating tumor cells of patients with colorectal cancer. *Clinical Chemistry*, 59(1), pp.252–260.
Gatenby, R.A. et al., 2009. Adaptive Therapy. *Cancer Research*, 69(11), pp.4894–4903.
Gautschi, O. et al., 2012. A patient with BRAF V600E lung adenocarcinoma responding to vemurafenib. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*, 7(10), pp.e23–4.
Geigl, J.B. et al., 2008. Defining “chromosomal instability.” *Trends in Genetics*, 24(2), pp.64–69.
Gerlinger, M. & Swanton, C., 2010. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *British Journal of Cancer*, 103(8), pp.1139–1143.
Gerlinger, M. et al., 2014. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nature Genetics*, 46(3), pp.225–233.
Gerlinger, M. et al., 2012. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *New England Journal of Medicine*, 366(10), pp.883–892.
Ghersi, D., Wilcken, N. & Simes, R.J., 2005. A systematic review of taxane-containing regimens for metastatic breast cancer. *British Journal of Cancer*, 93(3), pp.293–301.
Gillies, R.J., Verduzco, D. & Gatenby, R.A., 2012. Evolutionary dynamics of carcinogenesis and why targeted therapy does not work. *Nature Reviews Cancer*, 12(7), pp.487–493.
Gisselsson, D. et al., 2000. Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. *Proceedings of the National Academy of Sciences*, 97(10),

pp.5357–5362.
Gisselsson, D. et al., 2010. Generation of trisomies in cancer cells by multipolar mitosis and incomplete cytokinesis. *Proceedings of the National Academy of Sciences of the United States of America*, 107(47), pp.20489–20493.
Gisselsson, D. et al., 2001. Telomere dysfunction triggers extensive DNA fragmentation and evolution of complex chromosome abnormalities in human malignant tumors. *Proceedings of the National Academy of Sciences*, 98(22), pp.12683–12688.
Gomez-Roca, C. et al., 2009. Differential expression of biomarkers in primary non-small cell lung cancer and metastatic sites. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*, 4(10), pp.1212–1220.
Goswami, R.S. et al., 2015. Hotspot mutation panel testing reveals clonal evolution in a study of 265 paired primary and metastatic tumors. *Clinical Cancer Research*, 21(11), pp.2644–2651.
Govindan, R. et al., 2012. Genomic Landscape of Non-Small Cell Lung Cancer in Smokers and Never-Smokers. *Cell*, 150(6), pp.1121–1134.
Grabsch, H. et al., 2003. Overexpression of the mitotic checkpoint genes BUB1, BUBR1, and BUB3 in gastric cancer--association with tumour cell proliferation. *The Journal of Pathology*, 200(1), pp.16–22.
Grant, D.S. et al., 2003. Comparison of antiangiogenic activities using paclitaxel (taxol) and docetaxel (taxotere). *International Journal of Cancer*, 104(1), pp.121–129.
Greaves, M. & Maley, C.C., 2012. Clonal evolution in cancer. *Nature*, 481(7381), pp.306–313.
Gregan, J. et al., 2011. Merotelic kinetochore attachment: causes and effects. *Trends in cell biology*, 21(6), pp.374–381.
Gundem, G. et al., 2015. The evolutionary history of lethal metastatic prostate cancer. *Nature*, 520(7547), pp.353–357.
Habermann, J.K. et al., 2009. The gene expression signature of genomic instability in breast cancer is an independent predictor of clinical outcome. *International Journal of Cancer*, 124(7), pp.1552–1564.
Haffner, M.C. et al., 2013. Tracking the clonal origin of lethal prostate cancer. *Journal of Clinical Investigation*, 123(11), pp.4918–4922.
Hamana, K. et al., 2005. Monitoring of circulating tumour-associated DNA as a prognostic tool for oral squamous cell carcinoma. *British Journal of Cancer*, 92(12), pp.2181–2184.
Hammerman, P.S. et al., 2012. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*, 489(7417), pp.519–525.
Han, S.-W. et al., 2005. Predictive and prognostic impact of epidermal growth factor receptor mutation in non-small-cell lung cancer patients treated with gefitinib. *Journal of Clinical Oncology*, 23(11), pp.2493–2501.
Hata, A., Katakami, N., Fujita, S., et al., 2013. Panitumumab rechallenge in chemorefractory patients with metastatic colorectal cancer. *Journal of gastrointestinal cancer*, 44(4), pp.456–459.
Hata, A., Katakami, N., Kaji, R., et al., 2013. Does T790M disappear? Successful gefitinib rechallenge after T790M disappearance in a patient with EGFR-mutant non-small-cell lung cancer. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*, 8(3), pp.e27–9.
Hayes, D.F. et al., 2007. HER2 and response to paclitaxel in node-positive breast cancer. *New England Journal of Medicine*, 357(15), pp.1496–1506.
Heaphy, C.M. et al., 2011. Altered telomeres in tumors with ATRX and DAXX mutations. *Science*, 333(6041), pp.425–425.
Hegi, M.E. et al., 2011. Pathway analysis of glioblastoma tissue after preoperative treatment with the EGFR tyrosine kinase inhibitor gefitinib--a phase II trial. *Molecular cancer therapeutics*, 10(6), pp.1102–1112.
Heitzer, E., Auer, M., et al., 2013. Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing. *Cancer Research*, 73(10), pp.2965–2975.
Heitzer, E., Ulz, P., et al., 2013. Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. *Genome medicine*, 5(4), p.30.
Heppner, G.H. & Miller, B.E., 1983. Tumor heterogeneity: biological implications and therapeutic

consequences. *Cancer and Metastasis Reviews*, 2(1), pp.5–23.
Higgins, M.J. et al., 2012. Detection of tumor PIK3CA status in metastatic breast cancer using peripheral blood. *Clinical Cancer Research*, 18(12), pp.3462–3469.
Hiley, C. et al., 2014. Deciphering intratumor heterogeneity and temporal acquisition of driver events to refine precision medicine. *Genome Biology*, 15(8), p.453.
Hodgkinson, C.L. et al., 2014. Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. *Nature Medicine*.
Hong, M.K.H. et al., 2015. Tracking the origins and drivers of subclonal metastatic expansion in prostate cancer. *Nature Communications*, 6, p.6605.
Imielinski, M. et al., 2012. Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing. *Cell*, 150(6), pp.1107–1120.
Inaba, S. et al., 2005. Synuclein gamma inhibits the mitotic checkpoint function and promotes chromosomal instability of breast cancer cells. *Breast Cancer Research and Treatment*, 94(1), pp.25–35.
Jahr, S. et al., 2001. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Research*, 61(4), pp.1659–1665.
Jamal-Hanjani, M. et al., 2014. Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. *PLoS biology*, 12(7), p.e1001906.
Jamal-Hanjani, M., A'Hern, R., et al., 2015. Extreme chromosomal instability forecasts improved outcome in ER-negative breast cancer: a prospective validation cohort study from the TACT trial. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, p.mdv178.
Jamal-Hanjani, M., Quezada, S.A., et al., 2015. Translational Implications of Tumor Heterogeneity. *Clinical Cancer Research*, 21(6), pp.1258–1266.
Janssen, A. et al., 2011. Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. *Science*, 333(6051), pp.1895–1898.
Janssen, A., Kops, G.J.P.L. & Medema, R.H., 2009. Elevating the frequency of chromosome mis-segregation as a strategy to kill tumor cells. *Proceedings of the National Academy of Sciences of the United States of America*, 106(45), pp.19108–19113.
Jänne, P.A. et al., 2006. A rapid and sensitive enzymatic method for epidermal growth factor receptor mutation screening. *Clinical Cancer Research*, 12(3 Pt 1), pp.751–758.
Jänne, P.A. et al., 2013. Selumetinib plus docetaxel for KRAS-mutant advanced non-small-cell lung cancer: a randomised, multicentre, placebo-controlled, phase 2 study. *The Lancet Oncology*, 14(1), pp.38–47.
Jänne, P.A., Gray, N. & Settleman, J., 2009. Factors underlying sensitivity of cancers to small-molecule kinase inhibitors. *Nat Rev Drug Discov*, 8(9), pp.709–723.
Johnson, B.E. et al., 2014. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science*, 343(6167), pp.189–193.
Johnson, D.B., Smalley, K.S.M. & Sosman, J.A., 2014. Molecular pathways: targeting NRAS in melanoma and acute myelogenous leukemia. *Clinical Cancer Research*, 20(16), pp.4186–4192.
Jonkers, Y.M.H. et al., 2005. Chromosomal instability predicts metastatic disease in patients with insulinomas. *Endocrine-related cancer*, 12(2), pp.435–447.
Juric, D. et al., 2015. Convergent loss of PTEN leads to clinical resistance to a PI(3)Kα inhibitor. *Nature*, 518(7538), pp.240–244.
Juul, N. et al., 2010. Assessment of an RNA interference screen-derived mitotic and ceramide pathway metagene as a predictor of response to neoadjuvant paclitaxel for primary triple-negative breast cancer: a retrospective analysis of five clinical trials. *The Lancet Oncology*, 11(4), pp.358–365.
Kakimoto, Y., Yamamoto, N. & Shibahara, T., 2008. Microsatellite analysis of serum DNA in patients with oral squamous cell carcinoma. *Oncol Rep*, 20(5), pp.1195–1200.
Kalikaki, A. et al., 2008. Comparison of EGFR and K-RAS gene status between primary tumours and corresponding metastases in NSCLC. *British Journal of Cancer*, 99(6), pp.923–929.
Kan, Z. et al., 2010. Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature*, 466(7308), pp.869–873.
Kandoth, C. et al., 2014. Mutational landscape and significance across 12 major cancer types.

Nature, 502(7471), pp.333–339.
Kang, S., Bader, A.G. & Vogt, P.K., 2005. Phosphatidylinositol 3-kinase mutations identified in human cancer are oncogenic. *Proceedings of the National Academy of Sciences*, 102(3), pp.802–807.
Katayama, R. et al., 2012. Mechanisms of Acquired Crizotinib Resistance in ALK-Rearranged Lung Cancers. *Science Translational Medicine*, 4(120), pp.120ra17–120ra17.
Kawano, O. et al., 2006. PIK3CA mutation status in Japanese lung cancer patients. *Lung Cancer*, 54(2), pp.209–215.
Khan, S.H. & Wahl, G.M., 1998. p53 and pRb prevent rereplication in response to microtubule inhibitors by mediating a reversible G1 arrest. *Cancer Research*, 58(3), pp.396–401.
Kim, C. & Paik, S., 2010. Gene-expression-based prognostic assays for breast cancer. *Nature Publishing Group*, 7(6), pp.340–347.
Kinde, I. et al., 2011. Detection and quantification of rare mutations with massively parallel sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 108(23), pp.9530–9535.
Kleppe, M. & Levine, R.L., 2014. Tumor heterogeneity confounds and illuminates. *Nature Medicine*, 20(4), pp.342–344.
Koboldt, D.C. et al., 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, 22(3), pp.568–576.
Kogita, A. et al., 2015. Inter- and intra-tumor profiling of multi-regional colon cancer and metastasis. *Biochem Biophys Res Commun*, 458(1), pp.52–56.
Komarova, N.L. & Wodarz, D., 2005. Drug resistance in cancer: principles of emergence and prevention. *Proceedings of the National Academy of Sciences*, 102(27), pp.9714–9719.
Komarova, N.L. & Wodarz, D., 2004. The optimal rate of chromosome loss for the inactivation of tumor suppressor genes in cancer. *Proceedings of the National Academy of Sciences*, 101(18), pp.7017–7021.
Konecny, G.E. et al., 2004. Her-2/neu gene amplification and response to paclitaxel in patients with metastatic breast cancer. *JNCI Journal of the National Cancer Institute*, 96(15), pp.1141–1151.
Kops, G.J.P.L., Foltz, D.R. & Cleveland, D.W., 2004. Lethality to human cancer cells through massive chromosome loss by inhibition of the mitotic checkpoint. *Proceedings of the National Academy of Sciences*, 101(23), pp.8699–8704.
Koressaar, T. & Remm, M., 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics*, 23(10), pp.1289–1291.
Kosaka, T. et al., 2006. Analysis of epidermal growth factor receptor gene mutation in patients with non-small cell lung cancer and acquired resistance to gefitinib. *Clinical Cancer Research*, 12(19), pp.5764–5769.
Kronenwett, U. et al., 2006. Genomic instability and prognosis in breast carcinomas. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 15(9), pp.1630–1635.
Kronenwett, U. et al., 2004. Improved grading of breast adenocarcinomas based on genomic instability. *Cancer Research*, 64(3), pp.904–909.
Kuhlmann, J.D. et al., 2012. LOH at 6q and 10q in fractionated circulating DNA of ovarian cancer patients is predictive for tumor cell spread and overall survival. *BMC cancer*, 12(1), p.325.
Kumar, A. et al., 2014. Deep sequencing of multiple regions of glial tumors reveals spatial heterogeneity for mutations in clinically relevant genes. *Genome Biology*, 15(12), p.530.
Kumar, P., Henikoff, S. & Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, 4(7), pp.1073–1081.
Ladanyi, M. & Pao, W., 2008. Lung adenocarcinoma: guiding EGFR-targeted therapy and beyond. *Modern Pathology*, 21 Suppl 2, pp.S16–22.
Landau, D.A. et al., 2013. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*, 152(4), pp.714–726.
Lawrence, M.S. et al., 2014. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*, 505(7484), pp.495–501.
Lawrence, M.S. et al., 2013. Mutational heterogeneity in cancer and the search for new cancer-

associated genes. *Nature*, pp.1–5.
Le Pennec, S. et al., 2015. Intratumor heterogeneity and clonal evolution in an aggressive papillary thyroid cancer and matched metastases. *Endocrine-related cancer*, 22(2), pp.205–216.
Le, K. et al., 2013. Selective RAF inhibitor impairs ERK1/2 phosphorylation and growth in mutant NRAS, vemurafenib-resistant melanoma cells. *Pigment cell & melanoma research*, 26(4), pp.509–517.
Leary, R.J. et al., 2012. Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Science Translational Medicine*, 4(162), pp.162ra154–162ra154.
Leary, R.J. et al., 2010. Development of personalized tumor biomarkers using massively parallel sequencing. *Science Translational Medicine*, 2(20), pp.20ra14–20ra14.
Lee, A.J.X. et al., 2011. Chromosomal instability confers intrinsic multidrug resistance. *Cancer Research*, 71(5), pp.1858–1870.
Lee, W. et al., 2010. nature09004. *Nature*, 465(7297), pp.473–477.
Li, A. et al., 2012. Activated mutant NRas(Q61K) drives aberrant melanocyte signaling, survival, and invasiveness via a Rac1-dependent mechanism. *J Invest Dermatol*, 132(11), pp.2610–2621.
Li, A.R. et al., 2008. EGFR mutations in lung adenocarcinomas: clinical testing experience and relationship to EGFR gene copy number and immunohistochemical expression. *The Journal of Molecular Diagnostics*, 10(3), pp.242–248.
Li, H. & Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), pp.1754–1760.
Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), pp.2078–2079.
Li, M. et al., 2006. BEAMing up for detection and quantification of rare sequence variants. *Nature Methods*, 3(2), pp.95–97.
Li, R. et al., 2005. Chromosomal alterations cause the high rates and wide ranges of drug resistance in cancer cells. *Cancer genetics and cytogenetics*, 163(1), pp.44–56.
Liegl, B. et al., 2008. Heterogeneity of kinase inhibitor resistance mechanisms in GIST. *The Journal of Pathology*, 216(1), pp.64–74.
Lindell, K.O., Erlen, J.A. & Kaminski, N., 2006. Lessons from our patients: development of a warm autopsy program. *PLoS medicine*, 3(7), p.e234.
Lindström, L.S. et al., 2012. Clinically used breast cancer markers such as estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 are unstable throughout tumor progression. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 30(21), pp.2601–2608.
Lingle, W.L. et al., 2002. Centrosome amplification drives chromosomal instability in breast tumor development. *Proceedings of the National Academy of Sciences*, 99(4), pp.1978–1983.
Lissoni, P. et al., 2000. Chemotherapy and angiogenesis in advanced cancer: vascular endothelial growth factor (VEGF) decline as predictor of disease control during taxol therapy in metastatic breast cancer. *The International journal of biological markers*, 15(4), pp.308–311.
Liu, P. et al., 2012. Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing. *Carcinogenesis*, 33(7), pp.1270–1276.
Liu, Q. & Sommer, S.S., 2000. Pyrophosphorolysis-activated polymerization (PAP): application to allele-specific amplification. *BioTechniques*, 29(5), pp.1072–6– 1078– 1080 passim.
Liu, W. et al., 2009. Copy number analysis indicates monoclonal origin of lethal metastatic prostate cancer. *Nature Medicine*, 15(5), pp.559–565.
Loeb, L.A., 2001. A mutator phenotype in cancer. *Cancer Research*, 61(8), pp.3230–3239.
Lohr, J.G. et al., 2014. Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nature Biotechnology*, pp.1–8.
Long, G.V. et al., 2015. Dabrafenib and trametinib versus dabrafenib and placebo for Val600 BRAF-mutant melanoma: a multicentre, double-blind, phase 3 randomised controlled trial. *Lancet*.
Lovejoy, C.A. et al., 2012. Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway. H. S. Scott,

ed. *PLoS Genetics*, 8(7), p.e1002772.
Lynch, T.J. et al., 2004. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *New England Journal of Medicine*, 350(21), pp.2129–2139.
M'kacher, R. et al., 2010. JC human polyomavirus is associated to chromosomal instability in peripheral blood lymphocytes of Hodgkin's lymphoma patients and poor clinical outcome. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, 21(4), pp.826–832.
Maley, C.C. et al., 2006. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nature Genetics*, 38(4), pp.468–473.
Martincorena, I. et al., 2015. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*, 348(6237), pp.880–886.
Marusyk, A. et al., 2014. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature*, 514(7520), pp.54–58.
Marusyk, A., Almendro, V. & Polyak, K., 2012. Intra-tumour heterogeneity: a looking glass for cancer? *Nature Reviews Cancer*, 12(5), pp.323–334.
Mascaux, C. et al., 2005. The role of RAS oncogene in survival of patients with lung cancer: a systematic review of the literature with meta-analysis. *British Journal of Cancer*, 92(1), pp.131–139.
Massion, P.P. et al., 2009. Recurrent genomic gains in preinvasive lesions as a biomarker of risk for lung cancer. E. J. Bernhard, ed. *PLoS ONE*, 4(6), p.e5611.
May, T. et al., 2014. The limits of traditional approaches to informed consent for genomic medicine. *HEC forum : an interdisciplinary journal on hospitals' ethical and legal issues*, 26(3), pp.185–202.
McBride, D.J. et al., 2010. Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes, Chromosomes and Cancer*, 49(11), pp.1062–1069.
McClelland, S.E., Burrell, R.A. & Swanton, C., 2009. Chromosomal instability: a composite phenotype that influences sensitivity to chemotherapy. *Cell Cycle*, 8(20), pp.3262–3266.
McGranahan, N. & Swanton, C., 2015. Biological and Therapeutic Impact of Intratumor Heterogeneity in Cancer Evolution. *Cancer Cell*, 27(1), pp.15–26.
McGranahan, N. et al., 2012. Cancer chromosomal instability: therapeutic and diagnostic challenges. 13(6), pp.528–538.
McKenna, A. et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), pp.1297–1303.
Mengelbier, L.H. et al., 2015. Intratumoral genome diversity parallels progression and predicts outcome in pediatric cancer. *Nature Communications*, 6, p.6125.
Merlo, L.M.F. et al., 2010. A Comprehensive Survey of Clonal Diversity Measures in Barrett's Esophagus as Biomarkers of Progression to Esophageal Adenocarcinoma. *Cancer Prevention Research*, 3(11), pp.1388–1397.
Mettu, R.K.R. et al., 2010. A 12-gene genomic instability signature predicts clinical outcomes in multiple cancer types. *The International journal of biological markers*, 25(4), pp.219–228.
Meyer, R. et al., 2009. Overexpression and mislocalization of the chromosomal segregation protein separase in multiple human cancers. *Clinical Cancer Research*, 15(8), pp.2703–2710.
Misale, S. et al., 2012. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature*, 486(7404), pp.532–536.
Mitsudomi, T. & Yatabe, Y., 2010. Epidermal growth factor receptor in relation to tumor development: EGFR gene and cancer. *The FEBS journal*, 277(2), pp.301–308.
Mitsudomi, T. & Yatabe, Y., 2007. Mutations of the epidermal growth factor receptor gene and related genes as determinants of epidermal growth factor receptor tyrosine kinase inhibitors sensitivity in lung cancer. *Cancer science*, 98(12), pp.1817–1824.
Morris, E.J. et al., 2013. Discovery of a novel ERK inhibitor with activity in models of acquired resistance to BRAF and MEK inhibitors. *Cancer Discovery*, 3(7), pp.742–750.
Morris, L.G.T. et al., 2013. Recurrent somatic mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation. *Nature Genetics*, 45(3), pp.253–261.
Murayama-Hosokawa, S. et al., 2010. Genome-wide single-nucleotide polymorphism arrays in endometrial carcinomas associate extensive chromosomal instability with poor prognosis

and unveil frequent chromosomal imbalances involved in the PI3-kinase pathway. *Oncogene*, 29(13), pp.1897–1908.
Murtaza, M. et al., 2014. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature*, 497(7447), pp.108–112.
Nadauld, L.D. et al., 2014. Metastatic tumor evolution and organoid modeling implicate TGFBR2 as a cancer driver in diffuse gastric cancer. *Genome Biology*, 15(8), p.428.
Nakamura, H. et al., 2003. Chromosomal instability detected by fluorescence in situ hybridization in surgical specimens of non-small cell lung cancer is associated with poor survival. *Clinical Cancer Research*, 9(6), pp.2294–2299.
Nakamura, K. et al., 2011. Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Research*, 39(13), pp.e90–e90.
Nakayama, S. et al., 2009. Prediction of paclitaxel sensitivity by CDK1 and CDK2 activity in human breast cancer cells. *Breast Cancer Research*, 11(1), p.R12.
Navin, N. et al., 2010. Inferring tumor progression from genomic heterogeneity. *Genome Research*, 20(1), pp.68–80.
Navin, N. et al., 2012. Tumour evolution inferred by single-cell sequencing. *Nature*, 472(7341), pp.90–94.
Navin, N. et al., 2011. Tumour evolution inferred by single-cell sequencing. *Nature*, 472(7341), pp.90–94.
Cancer Genome Atlas Research Network, 2015. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*, 511(7511), pp.543–550.
Ni, X. et al., 2013. Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. *Proceedings of the National Academy of Sciences of the United States of America*, 110(52), pp.21083–21088.
Niikura, N. et al., 2012. Loss of Human Epidermal Growth Factor Receptor 2 (HER2) Expression in Metastatic Sites of HER2-Overexpressing Primary Breast Tumors. *Journal of Clinical Oncology*, 30(6), pp.593–599.
Nik-Zainal, S. et al., 2012. The Life History of 21 Breast Cancers. *Cell*, 149(5), pp.994–1007.
Nixon, K.C., 1999. The Parsimony Ratchet, a New Method for Rapid Parsimony Analysis. *Cladistics*, 15(4), pp.1–8.
Nowak, A.K. et al., 2004. Systematic review of taxane-containing versus non-taxane-containing regimens for adjuvant and neoadjuvant treatment of early breast cancer. *The Lancet Oncology*, 5(6), pp.372–380.
Nowell, P.C., 1976. The clonal evolution of tumor cell populations. *Science*, 194(4260), pp.23–28.
Nygaard, A.D. et al., 2013. The prognostic value of KRAS mutated plasma DNA in advanced non-small cell lung cancer. *Lung Cancer*, 79(3), pp.312–317.
Ogino, S. et al., 2005. Sensitive sequencing method for KRAS mutation detection by Pyrosequencing. *The Journal of Molecular Diagnostics*, 7(3), pp.413–421.
Ohashi, K. et al., 2013. Characteristics of lung cancers harboring NRAS mutations. *Clinical Cancer Research*, 19(9), pp.2584–2591.
Ojha, J. et al., 2015. Deep sequencing identifies genetic heterogeneity and recurrent convergent evolution in chronic lymphocytic leukemia. *Blood*, 125(3), pp.492–498.
Paez, J.G. et al., 2004. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*, 304(5676), pp.1497–1500.
Paik, P.K. et al., 2015. Next-Generation Sequencing of Stage IV Squamous Cell Lung Cancers Reveals an Association of PI3K Aberrations and Evidence of Clonal Heterogeneity in Patients with Brain Metastases. *Cancer Discovery*, 5(6), pp.610–621.
Pailler, E. et al., 2015. High level of chromosomal instability in circulating tumor cells of ROS1-rearranged non-small-cell lung cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, p.mdv165.
Pantel, K., Diaz, L.A. & Polyak, K., 2013. Tracking tumor resistance using 'liquid biopsies'. *Nature Medicine*, 19(6), pp.676–677.
Pao, W. et al., 2004. EGF receptor gene mutations are common in lung cancers from “never smokers” and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proceedings of the National Academy of Sciences*, 101(36), pp.13306–13311.
Pao, W. et al., 2005. KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. *PLoS medicine*, 2(1), p.e17.

Paradiso, A. et al., 2005. Biomarkers predictive for clinical efficacy of taxol-based chemotherapy in advanced breast cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, 16 Suppl 4(suppl_4), pp.iv14–19.

Paulsson, J. et al., 2011. Prognostic but not predictive role of platelet-derived growth factor receptors in patients with recurrent glioblastoma. *International Journal of Cancer*, 128(8), pp.1981–1988.

Pavelka, N., Rancati, G. & Li, R., 2010. Dr Jekyll and Mr Hyde: role of aneuploidy in cellular adaptation and cancer. *Current opinion in cell biology*, 22(6), pp.809–815.

Penault-Llorca, F. et al., 2009. Ki67 expression and docetaxel efficacy in patients with estrogen receptor-positive breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 27(17), pp.2809–2815.

Perera, S.A. et al., 2008. Telomere dysfunction promotes genome instability and metastatic potential in a K-ras p53 mouse model of lung cancer. *Carcinogenesis*, 29(4), pp.747–753.

Peto, R. et al., 1977. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. analysis and examples. *British Journal of Cancer*, 35(1), pp.1–39.

Pfeifer, G.P. & Hainaut, P., 2003. On the origin of G --> T transversions in lung cancer. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, 526(1-2), pp.39–43.

Pleasance, E.D. et al., 2010. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature*, 463(7278), pp.184–190.

Pratilas, C.A. et al., 2008. Genetic predictors of MEK dependence in non-small cell lung cancer. *Cancer Research*, 68(22), pp.9375–9383.

Pusztai, L., 2009. Gene expression profiling of breast cancer. *Breast Cancer Research*, 11 Suppl 3(Suppl 3), p.S11.

Rack, B. et al., 2014. Circulating Tumor Cells Predict Survival in Early Average-to-High Risk Breast Cancer Patients. *JNCI Journal of the National Cancer Institute*, 106(5), pp.dju066–dju066.

Riely, G.J. et al., 2008. Frequency and distinctive spectrum of KRAS mutations in never smokers with lung adenocarcinoma. *Clinical Cancer Research*, 14(18), pp.5731–5734.

Roberts, S.A. et al., 2013. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nature Genetics*, 45(9), pp.970–976.

Robinson, J.T. et al., 2011. Integrative genomics viewer. *Nature Biotechnology*, 29(1), pp.24–26.

Romano, E. et al., 2013. Identification of Multiple Mechanisms of Resistance to Vemurafenib in a Patient with BRAFV600E-Mutated Cutaneous Melanoma Successfully Rechallenged after Progression. *Clinical Cancer Research*, 19(20), pp.5749–5757.

Roschke, A.V. & Kirsch, I.R., 2005. Targeting cancer cells by exploiting karyotypic complexity and chromosomal instability. *Cell Cycle*, 4(5), pp.679–682.

Roschke, A.V. & Kirsch, I.R., 2010. Targeting karyotypic complexity and chromosomal instability of cancer cells. *Curr Drug Targets*, 11(10), pp.1341–1350.

Roschke, A.V. et al., 2005. Karyotypic “state” as a potential determinant for anticancer drug discovery. *Proceedings of the National Academy of Sciences*, 102(8), pp.2964–2969.

Roschke, A.V. et al., 2003. Karyotypic complexity of the NCI-60 drug-screening panel. *Cancer Research*, 63(24), pp.8634–8647.

Rosell, R. et al., 2005. Mutations in the tyrosine kinase domain of the EGFR gene associated with gefitinib response in non-small-cell lung cancer. *Lung Cancer*, 50(1), pp.25–33.

Rothe, F. et al., 2014. Plasma circulating tumor DNA as an alternative to metastatic biopsies for mutational analysis in breast cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, 25(10), pp.1959–1965.

Roylance, R. et al., 2014. Expression of regulators of mitotic fidelity are associated with intercellular heterogeneity and chromosomal instability in primary breast cancer. *Breast Cancer Research and Treatment*, 148(1), pp.221–229.

Roylance, R. et al., 2011. Relationship of Extreme Chromosomal Instability with Long-term Survival in a Retrospective Analysis of Primary Breast Cancer. *Cancer Epidemiology Biomarkers & Prevention*, 20(10), pp.2183–2194.

Rubin, M.A. et al., 2000. Rapid (“warm”) autopsy study for procurement of metastatic prostate cancer. *Clinical Cancer Research*, 6(3), pp.1038–1045.

Rudin, C.M., Hong, K. & Streit, M., 2013. Molecular characterization of acquired resistance to

the BRAF inhibitor dabrafenib in a patient with BRAF-mutant non-small-cell lung cancer. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*, 8(5), pp.e41–2.

Russnes, H.G. et al., 2011. Insight into the heterogeneity of breast cancer through next-generation sequencing. *Journal of Clinical Investigation*, 121(10), pp.3810–3818.

Samuels, Y. et al., 2004. High frequency of mutations of the PIK3CA gene in human cancers. *Science*, 304(5670), pp.554–554.

Schiller, J.H. et al., 2001. Lack of prognostic significance of p53 and K-ras mutations in primary resected non-small-cell lung cancer on E4592: a Laboratory Ancillary Study on an Eastern Cooperative Oncology Group Prospective Randomized Trial of Postoperative Adjuvant Therapy. *Journal of Clinical Oncology*, 19(2), pp.448–457.

Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. *Bioinformatics*, 27(4), pp.592–593.

Schmidt, M. et al., 2005. Ablation of the spindle assembly checkpoint by a compound targeting Mps1. *EMBO reports*, 6(9), pp.866–872.

Schütz, E. et al., 2015. Chromosomal instability in cell-free DNA is a serum biomarker for prostate cancer. *Clinical Chemistry*, 61(1), pp.239–248.

Schwartzentruber, J. et al., 2012. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature*, 482(7384), pp.226–231.

Schwarz, J.M. et al., 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods*, 11(4), pp.361–362.

Schwarz, R.F. et al., 2015. Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. C. Kemp, ed. *PLoS medicine*, 12(2), p.e1001789.

Seghers, A.C. et al., 2012. Successful rechallenge in two patients with BRAF-V600-mutant melanoma who experienced previous progression during treatment with a selective BRAF inhibitor. *Melanoma research*, 22(6), pp.466–472.

Sen, B. et al., 2012. Kinase-impaired BRAF mutations in lung cancer confer sensitivity to dasatinib. *Science Translational Medicine*, 4(136), pp.136ra70–136ra70.

Sen, S., Zhou, H. & White, R.A., 1997. A putative serine/threonine kinase encoding gene BTAK on chromosome 20q13 is amplified and overexpressed in human breast cancer cell lines. *Oncogene*, 14(18), pp.2195–2200.

Seol, H. et al., 2012. Intratumoral heterogeneity of HER2 gene amplification in breast cancer: its clinicopathological significance. *Modern Pathology*, 25(7), pp.938–948.

Sequist, L.V. et al., 2011. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Science Translational Medicine*, 3(75), pp.75ra26–75ra26.

Shah, N.P. et al., 2002. Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer Cell*, 2(2), pp.117–125.

Shah, S.P. et al., 2009. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*, 461(7265), pp.809–813.

Shah, S.P. et al., 2012. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*, 486(7403), pp.395–399.

Sheffer, M. et al., 2009. Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 106(17), pp.7131–7136.

Shipitsin, M. et al., 2007. Molecular definition of breast tumor heterogeneity. *Cancer Cell*, 11(3), pp.259–273.

Siravegna, G. et al., 2015. Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nature Medicine*, pp.1–9.

Smid, M. et al., 2010. Patterns and incidence of chromosomal instability and their prognostic relevance in breast cancer subtypes. *Breast Cancer Research and Treatment*, 128(1), pp.23–30.

Snuderl, M. et al., 2011. Mosaic Amplification of Multiple Receptor Tyrosine Kinase Genes in Glioblastoma. *Cancer Cell*, 20(6), pp.810–817.

Soh, J. et al., 2009. Oncogene mutations, copy number gains and mutant allele specific imbalance (MASI) frequently occur together in tumor cells. I. O.-L. Ng, ed. *PLoS ONE*, 4(10), p.e7464.

Sotillo, R. et al., 2010. Mad2-induced chromosome instability leads to lung tumour relapse after oncogene withdrawal. *Nature*, 464(7287), pp.436–440.

Sottoriva, A. et al., 2015. A Big Bang model of human colorectal tumor growth. *Nature Genetics*, 47(3), pp.209–216.

Sottoriva, A. et al., 2013. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 110(10), pp.4009–4014.

Stewénius, Y. et al., 2005. Structural and numerical chromosome changes in colon cancer develop through telomere-mediated anaphase bridges, not through mitotic multipolarity. *Proceedings of the National Academy of Sciences*, 102(15), pp.5541–5546.

Stoecklein, N.H. & Klein, C.A., 2010. Genetic disparity between primary tumours, disseminated tumour cells, and manifest metastasis. *International Journal of Cancer*, 126(3), pp.589–598.

Stommel, J.M. et al., 2007. Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. *Science*, 318(5848), pp.287–290.

Stratton, M.R., Campbell, P.J. & Futreal, P.A., 2009. The cancer genome. *Nature*, 458(7239), pp.719–724.

Stroun, M. et al., 2001. About the possible origin and mechanism of circulating DNA apoptosis and active DNA release. *Clinica chimica acta; international journal of clinical chemistry*, 313(1-2), pp.139–142.

Su, K.Y. et al., 2012. Pretreatment epidermal growth factor receptor (EGFR) T790M mutation predicts shorter EGFR tyrosine kinase inhibitor response duration in patients with non-small-cell lung cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 30(4), pp.433–440.

Sudo, T. et al., 2004. Dependence of paclitaxel sensitivity on a functional spindle assembly checkpoint. *Cancer Research*, 64(7), pp.2502–2508.

Suzuki, H. et al., 2015. Mutational landscape and clonal architecture in grade II and III gliomas. *Nature Genetics*, 47(5), pp.458–468.

Swanton, C., 2014. Cancer evolution: the final frontier of precision medicine? *Annals of Oncology*, 25(3), pp.549–551.

Swanton, C., 2012. Intratumor heterogeneity: evolution through space and time. *Cancer Research*, 72(19), pp.4875–4882.

Swanton, C. et al., 2009. Chromosomal instability determines taxane response. *Proceedings of the National Academy of Sciences of the United States of America*, 106(21), pp.8671–8676.

Swanton, C., Andre, F. & Mardis, E., 2014. Deciphering root causes of intrinsic BRAF inhibitor resistance in melanoma: ushering in a new genomics case reports feature for Annals of Oncology. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, 25(5), pp.917–918.

Swanton, C., Marani, M., Pardo, O., Warne, P.H., Kelly, G., Sahai, E., Elustondo, F., Chang, J., Temple, J., Ahmed, A.A., Brenton, J.D., Downward, J. & Nicke, B., 2007a. Regulators of Mitotic Arrest and Ceramide Metabolism Are Determinants of Sensitivity to Paclitaxel and Other Chemotherapeutic Drugs. *Cancer Cell*, 11(6), pp.498–512.

Swanton, C., Tomlinson, I. & Downward, J., 2006. Chromosomal instability, colorectal cancer and taxane resistance. *Cell Cycle*, 5(8), pp.818–823.

Szerlip, N.J. et al., 2012. Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRA amplification in glioblastoma defines subpopulations with distinct growth factor response. *Proceedings of the National Academy of Sciences of the United States of America*, 109(8), pp.3041–3046.

Takami, S. et al., 2001. Chromosomal instability detected by fluorescence in situ hybridization in Japanese breast cancer patients. *Clinica chimica acta; international journal of clinical chemistry*, 308(1-2), pp.127–131.

Taly, V. et al., 2013. Multiplex picodroplet digital PCR to detect KRAS mutations in circulating DNA from the plasma of colorectal cancer patients. *Clinical Chemistry*, 59(12), pp.1722–1731.

Tan, I.B. et al., 2015. High-depth sequencing of over 750 genes supports linear progression of primary tumors and metastases in most patients with liver-limited metastatic colorectal cancer. *Genome Biology*, 16(1), p.32.

Tanaka, H. et al., 2007. Lineage-specific dependency of lung adenocarcinomas on the lung development regulator TTF-1. *Cancer Research*, 67(13), pp.6007–6011.

Taron, M. et al., 2005. Activating mutations in the tyrosine kinase domain of the epidermal growth factor receptor are associated with improved survival in gefitinib-treated chemorefractory lung adenocarcinomas. *Clinical Cancer Research*, 11(16), pp.5878–5885.

Thakur, Das, M. et al., 2013. Modelling vemurafenib resistance in melanoma reveals a strategy to forestall drug resistance. *Nature*, 494(7436), pp.251–255.

Thirlwell, C. et al., 2010. Clonality assessment and clonal ordering of individual neoplastic crypts shows polyclonality of colorectal adenomas. *Gastroenterology*, 138(4), pp.1441–54–1454.e1–7.

Thorvaldsdottir, H., Robinson, J.T. & Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2), pp.178–192.

Tone, A.A. et al., 2014. Intratumoral heterogeneity in a minority of ovarian low-grade serous carcinomas. *BMC cancer*, 14(1), p.982.

Torres, E.M. et al., 2007. Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science*, 317(5840), pp.916–924.

Tsao, M.S. et al., 2007. Prognostic and predictive importance of p53 and RAS for adjuvant chemotherapy in non small-cell lung cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 25(33), pp.5240–5247.

Turajlic, S. et al., 2014. Whole-genome sequencing reveals complex mechanisms of intrinsic resistance to BRAF inhibition. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, 25(5), pp.959–967.

Turajlic, S., McGranahan, N. & Swanton, C., 2015. Inferring mutational timing and reconstructing tumour evolutionary histories. *Biochim Biophys Acta*, 1855(2), pp.264–275.

Turke, A.B. et al., 2010. Preexistence and clonal selection of MET amplification in EGFR mutant NSCLC. *Cancer Cell*, 17(1), pp.77–88.

Untergasser, A. et al., 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Research*, 40(15), pp.e115–e115.

Vakiani, E. et al., 2012. Comparative genomic analysis of primary versus metastatic colorectal carcinomas. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 30(24), pp.2956–2962.

Van Poznak, C. et al., 2002. Assessment of molecular markers of clinical sensitivity to single-agent taxane therapy for metastatic breast cancer. *Journal of Clinical Oncology*, 20(9), pp.2319–2326.

Vignot, S. et al., 2015. Comparative analysis of primary tumour and matched metastases in colorectal cancer patients: Evaluation of concordance between genomic and transcriptional profiles. *Eur J Cancer*, 51(7), pp.791–799.

Vogelstein, B. & Kinzler, K.W., 1999. Digital PCR. *Proceedings of the National Academy of Sciences*, 96(16), pp.9236–9241.

Vogelstein, B. et al., 2013. Cancer genome landscapes. *Science*, 339(6127), pp.1546–1558.

Walter, M.J. et al., 2012. Clonal architecture of secondary acute myeloid leukemia. *New England Journal of Medicine*, 366(12), pp.1090–1098.

Walther, A., Houlston, R. & Tomlinson, I., 2008. Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. *Gut*, 57(7), pp.941–950.

Wang, J.Y. et al., 2004. Molecular detection of APC, K- ras, and p53 mutations in the serum of colorectal cancer patients as circulating biomarkers. *World journal of surgery*, 28(7), pp.721–726.

Wang, K., Li, M. & Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, 38(16), pp.e164–e164.

Wang, Q. et al., 2004. BUBR1 deficiency results in abnormal megakaryopoiesis. *Blood*, 103(4), pp.1278–1285.

Wang, X. et al., 2002. Significance of MAD2 expression to mitotic checkpoint control in ovarian cancer cells. *Cancer Research*, 62(6), pp.1662–1668.

Weaver, B.A.A. & Cleveland, D.W., 2005. Decoding the links between mitosis, cancer, and chemotherapy: The mitotic checkpoint, adaptation, and cell death. *Cancer Cell*, 8(1), pp.7–12.

Weir, B.A. et al., 2007. Characterizing the cancer genome in lung adenocarcinoma. *Nature*, 450(7171), pp.893–898.

Williams, B.R. et al., 2008. Aneuploidy affects proliferation and spontaneous immortalization in

mammalian cells. *Science*, 322(5902), pp.703–709.
Winton, T. et al., 2005. Vinorelbine plus cisplatin vs. observation in resected non-small-cell lung cancer. *New England Journal of Medicine*, 352(25), pp.2589–2597.
Yelena, Y et al., 2014. Dual Inhibition of EGFR with Afatinib and Cetuximab in Kinase Inhibitor-Resistant *EGFR*-Mutant Lung Cancer With and Without *T790M* Mutations. *Cancer Discovery*, 4(9), pp 1036–1045.
Wu, X. et al., 2012. Clonal selection drives genetic divergence of metastatic medulloblastoma. *Nature*, 482(7386), pp.529–533.
Yachida, S. et al., 2010. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature*, 467(7319), pp.1114–1117.
Yamada, T. et al., 1998. Detection of K-ras gene mutations in plasma DNA of patients with pancreatic adenocarcinoma: correlation with clinicopathological features. *Clinical Cancer Research*, 4(6), pp.1527–1532.
Yang, J. et al., 2012. Intratumoral heterogeneity determines discordant results of diagnostic tests for human epidermal growth factor receptor (HER) 2 in gastric cancer specimens. *Cell biochemistry and biophysics*, 62(1), pp.221–228.
Yap, T.A. et al., 2012. Intratumor heterogeneity: seeing the wood for the trees. *Science Translational Medicine*, 4(127), pp.127ps10–127ps10.
Yates, L.R. & Campbell, P.J., 2012. Evolution of the cancer genome. *Nature Reviews Genetics*, 13(11), pp.795–806.
Yates, L.R. et al., 2015. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nature Medicine*.
Yoon, H.H. et al., 2012. Adverse prognostic impact of intratumor heterogeneous HER2 gene amplification in patients with esophageal adenocarcinoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 30(32), pp.3932–3938.
Yuan, B. et al., 2006. Increased expression of mitotic checkpoint genes in breast cancer cells with chromosomal instability. *Clinical Cancer Research*, 12(2), pp.405–410.
Yung, T.K.F. et al., 2009. Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non-small cell lung cancer patients. *Clinical Cancer Research*, 15(6), pp.2076–2084.
Zhang, J. et al., 2014. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science*, 346(6206), pp.256–259.
Zhang, N. et al., 2008. Overexpression of Separase induces aneuploidy and mammary tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 105(35), pp.13033–13038.
Zhang, X.C. et al., 2013. Tumor evolution and intratumor heterogeneity of an oropharyngeal squamous cell carcinoma revealed by whole-genome sequencing. *Neoplasia*, 15(12), pp.1371–1378.
Zhao, B., Hemann, M.T. & Lauffenburger, D.A., 2014. Intratumor heterogeneity alters most effective drugs in designed combinations. *Proceedings of the National Academy of Sciences*, 111(29), pp.10773–10778.
Zou, Z.-Q. et al., 2009. A novel dual PI3Kalpha/mTOR inhibitor PI-103 with high antitumor activity in non-small cell lung cancer cells. *International journal of molecular medicine*, 24(1), pp.97–101.