Genomic Profiling and Characteristics of Circulating Tumour Cells in Patients with Small Cell Lung Cancer

A Thesis Submitted to The University of Manchester for the Degree of PhD in the Faculty of Medical and Human Sciences.

2015

Louise Rosalyn Carter

School of Medicine

Contents

Word Count 72,886

List of Tables

List of Figures

Abstract

Genomic Profiling and Characteristics of Circulating Tumour Cells in Patients with Small Cell Lung Cancer.

A Thesis submitted to the University of Manchester for the Degree of PhD by Louise Rosalyn Carter, May 2015.

Background Small cell lung cancer (SCLC) is an aggressive, highly metastatic disease with dismal prognosis. Chemoresistance, both intrinsic and acquired, represents one of the major challenges in the management of SCLC contributing to the poor outcomes seen. Response rates to first-line chemotherapy are high, but the responses are not durable with short progression-free survival and significantly reduced response rates to further treatment seen. Patients with chemorefractory disease have particularly poor survival, even when compared to SCLC as a whole. Biopsies, particularly serial biopsies, are challenging to obtain in SCLC for research. Circulating tumour cells (CTCs) are prevalent in SCLC and represent a potential minimally invasive alternative source of tumour material for molecular analysis. The ambitious aim of this thesis was to develop methods for the molecular analysis of CTCs in SCLC, to interrogate the genomic landscape of this disease and to explore mechanisms of resistance to chemotherapy.

Methods To monitor tumour genetics in SCLC, a CTC workflow was developed using the CellSearch system for CTC enrichment followed by DEPArray CTC isolation. Using this workflow, CTCs were isolated from chemorefractory patients and chemoresponsive patients' blood samples, prior to chemotherapy and again at progression with relapsed disease. Following single-cell whole genome amplification, Sanger Sequencing, TAm-Seq, low coverage whole genome sequencing and whole exome sequencing (WES) of CTCs and CTC-derived explants (CDX) were used to investigate mutations and to generate genome-wide patterns of copy number alterations (CNA).

Results Hallmark SCLC molecular abnormalities such as TP53 mutations and copy number loss in tumour suppressor genes such as RB1 (previously identified in bulk tumour profiling), were noted in the isolated SCLC CTCs and in the patient matched CDX models in mice developed by colleagues in our group. Distinct CNA profiles were found in the CTCs isolated from patients with chemorefractory disease compared to those isolated from patients with chemoresponsive disease. A potential signature of 760 genes with statistically significant change in copy number between the two groups of patients' CTCs was identified. This signature of CNA changes were not seen in CTCs isolated when initially chemoresponsive patients relapsed when compared to the baseline samples, however new mutations were identified by WES between presentation and relapse.

Conclusion SCLC CTCs, the invasive subset of tumour cells, reflect and share the common mutational changes identified in bulk SCLC tumour sequencing. The identification of a signature of CNAs potentially associated with intrinsic resistance, yet none associated with the development of acquired resistance, suggests that there may be different mechanisms underlying these two processes. The potential research utility of CTCs in SCLC has been confirmed with the results of this thesis, which has opened up new avenues to study acquired and intrinsic chemoresistance, as well as a route to identify much needed new drug targets.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

List of Abbreviations

aCGH	Array comparative genomic hybridisation
AKT	v-akt murine thymoma viral oncogene homolog 1
ALK	Anaplastic lymphoma receptor tyrosine kinase
APC	Allophycocyanin
ATM	ATM serine/threonine kinase
ATP1A1	ATPase, Na+/K+ transporting, alpha 1 polypeptide
ATP7B	ATPase, Cu++ transporting, beta polypeptide
BCL2	B-cell CLL/lymphoma 2
BG	Apolipoprotein H (beta-2-glycoprotein I)
BIM	BCL2-like 11 (apoptosis facilitator)
BP	Base pair
BRAF	B-Raf proto-oncogene, serine/threonine kinase
BRCA2	Breast cancer 2, early onset
CAM	Cell adhesion matrix assay
CCNE1	Cyclin E1
CDH1	Cadherin 1, type 1, E-cadherin (epithelial)
CDK4	Cyclin-dependent kinase 4
CDX	CTC-derived explant
CEP	Clinical and Experimental Pharmacology
CES1	Carboxylesterase 1
CES2	Carboxylesterase 2
CES3	Carboxylesterase 3
cfDNA	Circulating free DNA
CGH	Comparative genomic hybridisation
CK	Cytokeratin
CMC	Circulating melanoma cell
CNA	Copy number alteration
COBL	Cordon-bleu WH2 repeat protein
COPD	Chronic obstructive pulmonary disease
CRC	Colorectal cancer
CREBBP	CREB binding protein
CR UK	Cancer Research United Kingdom
CT	Computerised tomography
CTC	Circulating tumour cell
CTM	Circulating tumour microemboli
CTNNB1	Catenin (cadherin-associated protein), beta 1, 88kDa
CXCL	Chemokine ligand
CXR	Chest X-ray
DAPI	4',6'-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide Triphosphate
E2F2	E2F transcription factor 2
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor

EML4	Echinoderm microtubule associated protein like 4
EMT	Epithelial to Mesenchymal transition
EP300	E1A binding protein p300
EpCAM	Epithelial cell adhesion molecule
EPHA7	EPH receptor A7
ES SCLC	Extensive stage SCLC
FACS	Fluorescence-activated cell sorting
FDA	Food and Drug Administration
FDR	False discovery rate
FFPE	Formalin-fixed Paraffin-embedded
FGFR1	Fibroblast growth factor receptor 1
FHIT	Fragile histidine triad
FISH	Fluorescence in situ hybridization
FLT3	Fms-related tyrosine kinase 3
FOLFOX	Folinic acid, Fluorouracil and Oxaliplatin chemotherapy regimen
gDNA	Genomic DNA
GEMM	Genetically engineered mouse model
GII	Genomic integrity index
GPR113	Adhesion G protein-coupled receptor F3
GSTA	Glutathione S-transferase alpha
HER2	Erb-b2 receptor tyrosine kinase 2
HG19	Human genome issue 19
HSPG2	Heparan sulfate proteoglycan 2
IHC	Immunohistochemistry
INPP5D	Inositol polyphosphate-5-phosphatase, 145kDa
IPA	Ingenuity pathway analysis
IPASS	Iressa Pan-Asia study
ISEL	Iressa Survival Evaluation in Lung Cancer Study
ISET	Isolation by size of epithelial tumour cells
KDM4E	Lysine (K)-specific demethylase 4E
KIAA1432	RAB6A GEF complex partner 1
KIT	V-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
KL	Klotho
KRAS	Kirsten rat sarcoma viral oncogene homolog
LIMMA	Linear models for microarray analysis
LRP1B	Low density lipoprotein receptor-related protein 1B
LRP5	Low density lipoprotein receptor-related protein 5
LRRK2	Leucine-rich repeat kinase 2
LS SCLC	Limited stage SCLC
MCL	Mantle cell lymphoma
MCSP	Melanoma-associated chondroitin sulphate proteoglycan
MDA	Multiple displacement amplification
MED12L	Mediator complex subunit 12-like
MET	MET proto-oncogene, receptor tyrosine kinase
miRNA	Micro ribonucleic acid

MLH1	MutL homolog 1
MLL	Lysine (K)-specific methyltransferase 2A
MLL2	Lysine (K)-specific methyltransferase 2D
mRNA	Messenger ribonucleic acid
MRP	Multidrug resistance Protein
MSE1	Glutamyl-tRNA synthetase 2, mitochondrial
MTOR	Mechanistic target of rapamycin (serine/threonine kinase)
MUC1	Mucin 1, cell surface associated
MYC	V-myc avian myelocytomatosis viral oncogene homolog
MYCL1	V-myc avian myelocytomatosis viral oncogene lung carcinoma derived homolog
MYCN	V-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog
NCI	National Cancer Institute (USA)
NEB	Nebulin
NGS	Next Generation Sequencing
NOTCH2	Neurogenic locus notch homolog protein 2 (Drosophila
NOTCH3	Neurogenic locus notch homolog protein 3
NRAS	Neuroblastoma RAS viral (v-ras) oncogene homolog
NSG	Nod scid gamma
NSCLC	Non-small cell lung cancer
OS	Overall survival
PARP	Poly (ADP-ribose) polymerase 1
PBS	Phosphate buffered saline
PC	Principal component
PCDH10	Protocadherin-10
PCA	Principal component analysis
PCI	Prophylactic cranial irradiation
PCR	Polymerase chain reaction
PDX	Patient-derived xenograft
PE	Phycoerythrin
PFS	Progression-free survival
P-gp	ATP-binding cassette, sub-family B (MDR/TAP), member 1
PHF1	PHD finger protein 1
PHLPP2	PH domain and leucine rich repeat protein phosphatase 2
PI3K	Phosphoinositide 3-kinase
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha
PPM1J	protein phosphatase, Mg2+/Mn2+ dependent, 1J
PRRC2C	Proline-rich coiled-coil 2C
PS	Performance status
PTCH1	Patched 1
PTCH2	Patched 2
PTEN	Phosphatase and tensin homolog
PTPRN2	Protein tyrosine phosphatase, receptor type, N polypeptide 2
QC	Quality control
qPCR	Quantitative PCR

qRT-PCR	Quantitative reverse transcription PCR
RAD51	RAD51 recombinase
RALBP1	RalA binding protein 1
RASSF1	Ras association (RalGDS/AF-6) domain family member 1
RB1	Retinoblastoma 1
RBC	Red blood cell
RECIST	Response evaluation criteria in solid tumours
RET	Ret proto-oncogene
RICTOR	RPTOR independent companion of MTOR, complex 2
RLF	Rearranged L-myc fusion
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
ROS1	ROS proto-oncogene 1 , receptor tyrosine kinase
RT-PCR	Reverse transcription PCR
SASH1	SAM and SH3 domain containing 1
SCLC	Small cell lung cancer
SER	Serine
SLIT2	Slit homolog 2 (Drosophila)
SMARCA2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2
SNV	Single nucleotide variation
SOX2	SRY (sex determining region Y)-box 2
SPRI	solid phase reversible immobilization
STARD9	StAR-related lipid transfer (START) domain containing 9
TAm-Seq	Tagged-amplicon deep sequencing
TECTA	Tectorin alpha
TFI	Tissue factor pathway inhibitor
TGFBR2	Transforming growth factor, beta receptor II (70/80kDa)
THR	Threonine
TJP2	Tight junction protein 2
TKI	Tyrosine kinase inhibitor
TNM	Tumour node metastasis (staging system)
TP53	Tumour protein p53
TRRAP	Transformation/transcription domain-associated protein
VALG	Veterans Administration Lung Study Group
WBC	White blood cell
WES	Whole Exome Sequencing
WGA	Whole Genome Amplification
WGS	Whole Genome Sequencing
WTA	Whole transcriptome analysis

Dedication

I would like to dedicate this thesis to Dr Geoffrey Carter and Dr Elizabeth Carter, who first inspired me to become a doctor and have been role models to me every day since.

Acknowledgement

I would like to express my gratitude to my supervisors Professor Caroline Dive and Dr Ged Brady at the Cancer Research Manchester Institute, Dr Fiona Blackhall at the Christie NHS Foundation Trust and Dr Gillian Ellison at Astra Zeneca. I am profoundly appreciative for their time, guidance and enthusiasm over the past 3 years, in addition to the invaluable opportunities that they have afforded me. I would like to acknowledge the support of Ian Waddell, my advisor. I am thankful to both Cancer Research UK and Astra Zeneca for jointly funding this Clinical Fellowship.

There are many people at both the Cancer Research Manchester Institute to whom I owe my thanks. Notably I would like to thank Dominic Rothwell for tutoring me in the laboratory and mentoring me during this whole project. His humour, patience and teaching had been invaluable to me. I would also like to thank all the Nucleic Acids Biomarkers team, Debbie Burt, Suzi Dalby, Mahmood Ayub, James Chudziak, Jenny Antonello, Dan Morris and Nigel Smith. They have all been so generous to me with their time and assistance, and made this a very memorable experience.

I would like to thank the whole of the Clinical and Experimental Pharmacology group who have made this not only a scientifically interesting and valuable time, but also an enjoyable one. I would like to thank the CTC team led by Karen Morris for all their help. I would also like to thank Dr Rob Metcalf, Dr Laura Cove Smith and Dr Matt Krebs for their support. I would like to acknowledge the invaluable assistance with the bioinformatics in this project of the Computational Biology group led by Crispin Miller, with particular thanks to Yaoyong Li and Hui Sun Leong for their ingenuity and patience. I am grateful to the core facilities, particularly Yvonne Hey and Gillian Newton, for all their advice and help with Next Generation Sequencing.

At the Christie NHS Foundation Trust I would like to thank the Lung team for supporting the Chemores trial and recruiting patients for this research. I would like to thank Lynsey Priest and Vickki Codlin for all their assistance with the Chemores trial. I would also like to acknowledge the nurses and phlebotomists for their assistance with patient recruitment and sample collection.

Lastly I am indebted to the patients from the Christie NHS Trust who were altruistically involved in this research, donating their blood samples and time at what must be the most difficult point of their lives. I am always humbled by their grace and fortitude, and their determination that future cancer patients will benefit from their experiences.

The Author

Louise Carter is a Speciality Registrar in Medical Oncology in the North Western Deanery, appointed in 2008. She attended Selwyn College, Cambridge University for her preclinical training, graduating in 2001 with a first-class honours degree. She then transferred to Imperial College of Medicine in London, to complete her clinical training. She graduated in 2004 with a distinction in Clinical Medicine. She undertook General Medical training in London and obtained Membership of the Royal College of Physicians in 2007. She then commenced her Medical Oncology Speciality Registrar training in 2008 in the North Western Deanery, predominantly based at the Christie Hospital in Manchester. She was competitively appointed to a Clinical Research Fellowship in 2012, within a scheme jointly funded by Cancer Research UK and Astra Zeneca to train Medical Oncologist in translational medicine. This involved 3 years of pre-doctoral research at the Cancer Research Manchester Institute, University of Manchester. The research conducted during this fellowship forms the basis of this thesis.

Publications Arising from this Thesis

- Hodgkinson C, Morrow C, Li Y, Metcalf R, Rothwell D, Trapani F, Polanski R, Burt D, Simpson K, Morris K, Pepper S, Nonaka D, Greystoke A, Kelly P, Bola B, Krebs M, Antonello J, Faulkner S, Priest L, **Carter L**, Tate C, Miller C, Blackhall F, Brady G, Dive C. Tumorigenicity and Genetic Profiling of Circulating Tumor Cells in Small Cell Lung Cancer. Nature Medicine Aug 2014.

National and International Presentations/Posters Arising from this Thesis

- **Carter L** , Rothwell D, Li Y, Burt D, Antonello J, Krebs M, Chudziak J, Priest L, Miller C, Ellison G, Blackhall F, Dive C and Brady G. Investigating small cell lung cancer molecular status, heterogeneity and chemoresistance utilising circulating tumour cells. Poster presentation at the IASLC SCLC Meeting, Memorial Sloan Kettering Cancer Centre, New York, USA, April 2015.
- **Carter L** , Rothwell D, Li Y, Burt D, Antonello J, Krebs M, Chudziak J, Priest L, Miller C, Ellison G, Blackhall F, Dive C and Brady G. Investigating small cell lung cancer molecular status and heterogeneity utilising circulating tumour cells. Oral presentation at the National Cancer Research Institute Cancer Conference, Liverpool, UK, November 2014. Selected as a proffered paper.

- **Carter L** , Rothwell D, Li Y, Burt D, Antonello J, Krebs M, Chudziak J, Priest L, Miller C, Ellison G, Blackhall F, Dive C and Brady G. Utilising circulating tumour cells to investigate molecular status and heterogeneity in small cell lung cancer. Oral presentation at the International Symposium on Advances in Circulating Tumour Cells, Crete, Greece, October 2014. Awarded ACTC Travel Award.

Review Articles

- Krebs MG, Metcalf RL, **Carter L**, Brady G, Blackhall FH, Dive C. Molecular analysis of circulating tumour cells – biology and biomarkers. Nat Rev Clin Oncol. 2014 Mar; 11(3):129-44.
- **Carter L**, Metcalf R, Blackhall FH, Dive C, Krebs MG. Biology and Clinical Relevance of Circulating Tumour Cells. J Thorac Dis. 2012 Oct: 4(5): 453-5.

Other Publications

- Rothwell D, Li Y, Ayub M, Tate C, Newton G, Hey Y, **Carter L**, Faulkner S, Moro M, Pepper S, Miller C, Blackhall F, Bertolini G, Roz L, Dive C, Brady G. Evaluation and Validation of a Robust Single Cell RNA-Amplification Protocol Through Transcriptional Profiling of Enriched Lung Cancer Initiating Cells. BMC Genomics, 2014 Dec, 15: 1129.
- **Carter L**, Hubner RA, Valle JW. Clinical Management of Targeted Therapies in Neuroendocrine Tumours. In Raymond E, Faivre S, Ruszniewski P (Eds,)Management of Neuroendocrine Tumors of the Pancreas and Digestive Tract. (pp. 141 – 154): Springer.
- **Carter LR**, Califano R. The Torch Trial. Transl. Lung Cancer Res 2012: 1(4):283-286.
- Califano R, Board R, Abidin A, **Carter L**, Cove-Smith L, Kulraratne B and Lorigan P. Update on clinical trials for malignant melanoma. In Morton R (Ed.), Treatment of metastatic melanoma. Croatia: InTech.

Chapter 1: Introduction

1.1 Small Cell Lung Cancer

1.1.1 Introduction to Small Cell Lung Cancer

Lung cancer continues to be the commonest cause of cancer death worldwide [1]. There are 1.6 million cases and 1.4 million deaths per year [1]. Lung cancer is broadly divided into non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC), due to the different histological features of the subtypes and the significantly different treatment paradigms. SCLC is the most aggressive form of lung cancer, with a rapid doubling time [2], and accounts for approximately 15% of all lung cancer cases in Europe and the USA [3]. Survival rates in SCLC are low with just 5% of all patients surviving 5 years (figure 1.1) [4, 5] . There has unfortunately been little improvement in survival over the last 15 years despite large numbers of clinical trials [3]. SCLC is often staged using the Veterans Administration Lung Study Group (VALG) classification in two categories, limited (LS-SCLC) and extensive disease (ES-SCLC) as opposed to using Tumour Node Metastasis (TNM) staging system. Limited disease is defined as disease confined to one hemithorax with regional lymph node metastases [6]. Patients with limited stage disease are treated with curative intent. The median survival ranges from 15 to 20 months, with 20 - 25 % of patients with LS-SCLC being alive at 5 years [7]. The poor prognosis in SCLC is in part accounted for by the fact that only one third of patients are diagnosed with limited disease [3]. Patients with extensive disease, with metastases beyond the regional lymph nodes, are treated with palliative intent and overall survival (OS) is low with a median survival of 7 to 10 months [8]. SCLC is closely associated with tobacco smoking [9]. In developed countries the peak incidence of SCLC was in the 1980s, occurring 20 years after the peak rates of smoking [7]. The incidence of SCLC has declined slowly since in developed countries; though this is not a global trend as the smoking rates remain high in some regions. Rates of smoking are increasing for example in South East Asia leading to increasing incidence of lung cancer being noted in this region [1]. Within the UK the rates of smoking also vary between regions with both the north west and north east of England continuing to have high rates of smoking and associated high incidences of SCLC [10].

SCLC generally presents with a bulky mass developing from the central airways with spread to the regional lymph nodes [7]. It is characterised by the development of metastases at an early stage which contributes to the poor survival seen. SCLC has a predilection for metastasising to the contralateral lung, liver, brain, bone and adrenal

glands [11]. Brain metastases are present in approximately 10% of patients at presentation, but up to 50% of patients develop brain metastasis over the course of their treatment, representing a significant cause of mortality in this disease [12]. SCLC is the most common cause of paraneoplastic syndromes, such as the syndrome of inappropriate antidiuresis and Lambert-Eaton syndrome [7].

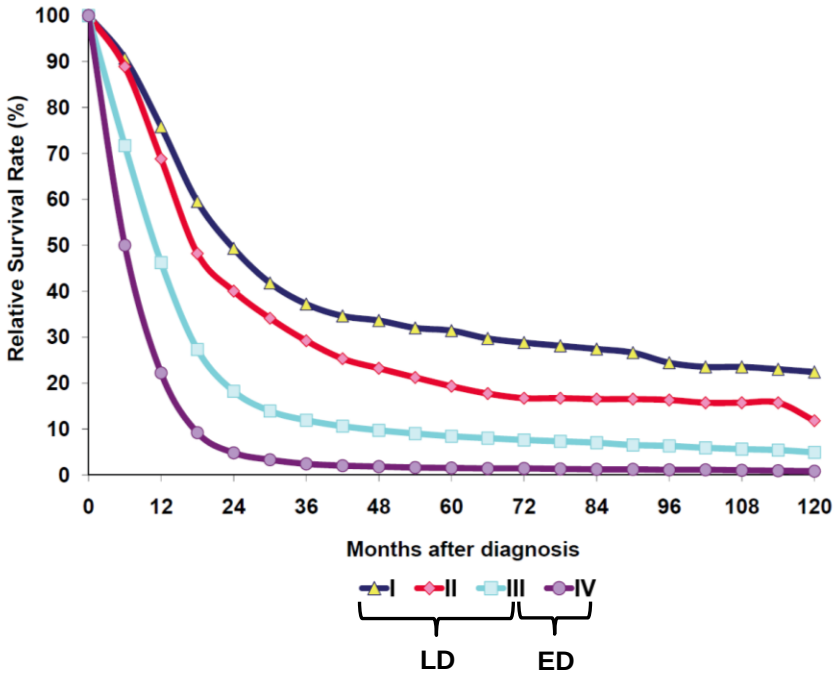


Figure 1.1 Relative survival rates according to disease stage in SCLC. (Adapted from Ries et al 2014 [13]) The relative percentage survival rates for SCLC patients divided according to stage. The data comes from the SEER program with survival data for patients from 1988 to 2001 included. The stage of patients according to the TNM and VALG classifications is indicated.

SCLC originates from neuroendocrine cells or neuroendocrine-cell precursors [14]. The diagnosis is made from histological or cytological examination of biopsies which are often small and may be compromised by crush artefacts from biopsy forceps [15]. The classical morphology of SCLC seen on biopsies is the presence of small cells which are round, oval or spindle shaped with sparse cytoplasm, high mitotic index and finely granular nuclei with inconspicuous nucleoli [16]. Immunohistochemistry confirms the neuroendocrine origin of SCLC with positive staining for neuroendocrine markers such as synaptophysin, chromogranin A and CD56 seen (figure 1.2) [17-19].

In SCLC, as in lung cancer in general, diagnoses are often made from bronchoscopic or radiologically-guided biopsies which result in small amount of tissue or cells with which to establish the diagnosis [2, 19, 20]. This means there is often limited amounts

of tissue available for research after confirming the diagnosis [21]. This issue is particularly acute in SCLC due to the rarity of surgically resected specimens in contrast to NSCLC [3, 22]. Bronchoscopic and radiologically-guided biopsies, although necessary to establish the diagnosis of cancer, have the potential to result in both morbidity, and in rare cases mortality for patients, and so their use for research alone needs to be carefully considered [23]. Recurrent samples are also needed to monitor the molecular changes associated with treatment [24]. Patients may be reluctant to undergo repeated biopsies due to fears of discomfort or pain. The limited availability of tissue and the difficulty of obtaining repeated biopsies therefore represents one of the challenges of researching SCLC.

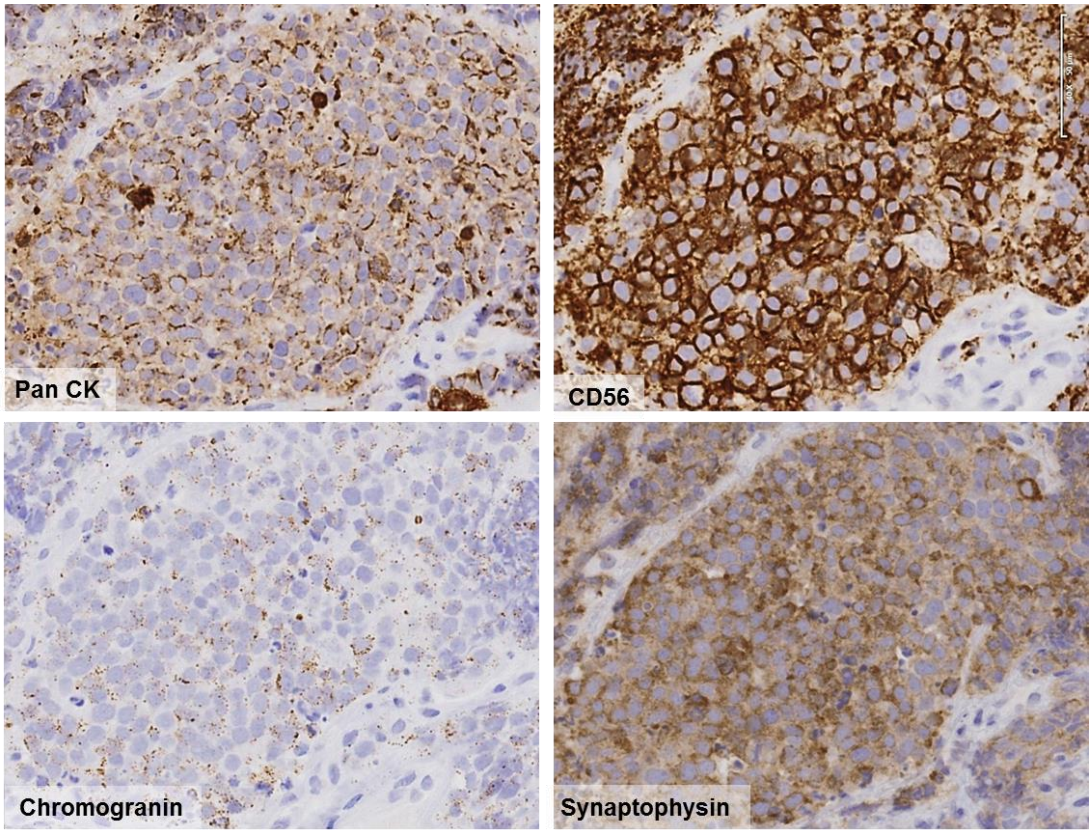


Figure 1.2 Neuroendocrine immunohistochemistry markers in SCLC (Adapted from Hodgkinson et al 2014 [24]). A SCLC biopsy stained for pan cytokeratin (CK), and the neuroendocrine markers CD56, chromogranin A and synaptophysin. Scale bar 50 μM.

1.1.2 Treatment of Small Cell Lung Cancer

Chemotherapy with platinum-based regimens remains the mainstay of treatment for SCLC, despite more than 20 years of clinical trials [25-27]. The key developments in the treatment of SCLC since the 1960s are summarised in figure 1.3. In patients with limited stage disease chemotherapy is delivered concomitantly with radiotherapy [28].

Patients are treated with 4 cycles to 6 cycles of Cisplatin and Etoposide doublet chemotherapy in addition to once or twice daily radiotherapy outside of clinical trials [3]. Prophylactic cranial irradiation (PCI) is offered on completion of the chemoradiotherapy due to the high incidence of brain metastasis in SCLC [29]. Surgical resection is rarely an appropriate treatment option in SCLC, but if carried out should be followed by adjuvant chemotherapy and PCI due to the high risk of micrometastatic disease [3].

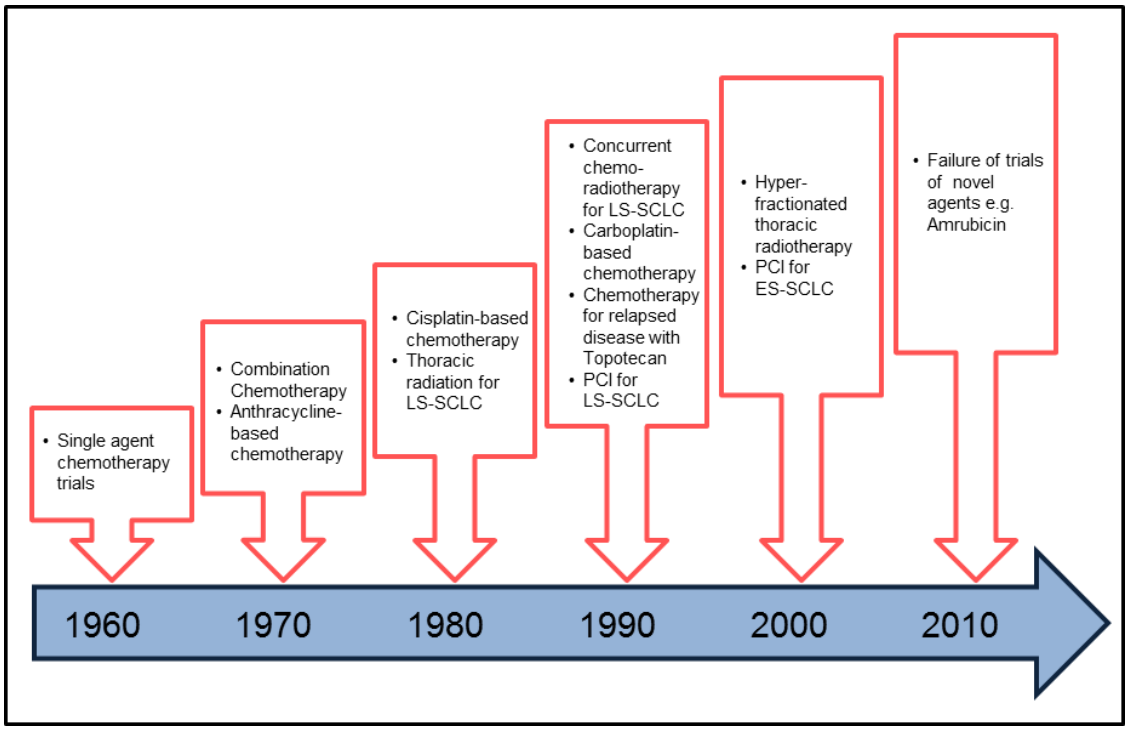


Figure 1.3 Summary of key developments in the treatment of SCLC (Adapted from Rodriguez et al 2010 [30]).

In ES-SCLC chemotherapy is delivered alone either as a single agent or as doublet chemotherapy depending on the performance status (PS) of the patient [27]. Platinum chemotherapy with or without Etoposide or Irinotecan are the commonly used regimens in the first-line setting [31]. Patients are offered PCI post chemotherapy if they respond to first-line treatment [32]. SCLC is a very chemosensitive cancer with up to 80% of patients responding to first-line therapy including 15 to 20% of patients having complete radiological responses [31, 33-35]. This contrasts to NSCLC where the response rate to platinum therapy in the first-line is between 19 and 36% [36, 37]. Unfortunately responses in SCLC are not durable and the median progression-free survival (PFS) is just 4.6 months [38]. Second-line therapies such as rechallenging with Platinum-containing regimens, Topotecan, Irinotecan or Anthracycline containing treatment are offered to patients who remain fit enough for further therapy [3, 39-41].

However, in the second-line setting the response rates to chemotherapy are much lower, at between 7 and 24% in clinical trials [3, 39, 41-43]. This contributes to the very poor OS seen in this group of patients.

Extensive numbers of clinical trials have looked at the management of SCLC and outside of the advances in radiotherapy there have been few improvements in management in recent years [44]. In the past two decades there has been increasing focus on the development of targeted agents to improve outcomes in oncology. In NSCLC there have been a number of advances leading to improvements in survival through the development of targeted agents, and through the identification of driver mutations present in subsets of NSCLC tumours to act as biomarkers for their use [45, 46]. Gefitinib (trade name Iressa), is a tyrosine kinase inhibitor (TKI) designed to target the epidermal growth factor receptor (EGFR). The initial phase III trial of Gefitinib, Iressa survival evaluation in lung cancer trial (ISEL), in an unselected population of NSCLC patients in the second-line was negative, having failed to show an improvement in OS [47]. In contrast the Iressa Pan-Asia Study (IPASS) was positive with improvements in PFS time for patients treated with Gefitinib in comparison to those receiving chemotherapy [48]. In the IPASS trial patients were selected for inclusion on the basis of clinical characteristics which increased the chance of response to EGFR TKIs such as being female or never smokers. The response rates of patients with EGFR mutations within the IPASS trial was 71.2 %, in contrast to the overall response rates to Gefitinib in the ISEL trial of 8% [47, 48]. The trials of Gefitinib in NSCLC highlight the need not only for effective therapies, but also biomarkers to guide their use. Between 2 and 7% of NSCLC tumours have a fusion gene of the echinoderm microtubule associated protein-like 4 (EML4) and the anaplastic lymphoma kinase (ALK) gene [49-52]. Crizotinib is a small molecule inhibitor of the ALK tyrosine kinase receptor. A single arm phase II trial of the use of Crizotinib in patients whose tumours have the ALK translocation had a 57% response rate, with a further 33% of patients having stable disease [53]. The US Food and Drug Administration (FDA) granted conditional approval for Crizotinib just 4 years after the ALK fusion gene was identified. This again highlights the critical role of biomarkers in the development of targeted agents.

There has been a large number of phase I, II and III trials of targeted agents in SCLC. Drugs that target the control of the cell cycle, inhibit angiogenesis, target multidrug resistance mechanisms, promote apoptosis, vaccines and immune conjugates have all been tested in SCLC without demonstrating significant improvement in survival [27, 54-56]. However, patient selection in the majority of these trials was not based on the

presence of predictive biomarker in tumour samples but instead was in unselected populations. This reflects both the limited data on the genomic profiles of SCLC that was available when the trials were designed and the paucity of druggable targets that have been identified in SCLC. It may, therefore, be that the lack of available predictive biomarkers in SCLC accounts for the failure to demonstrate benefit of targeted agents to date, particularly when one considers the example of Gefitinib in NSCLC. The challenges of SCLC for clinical trials however, extend beyond the lack of predictive markers and druggable targets with the very aggressive nature of the disease presenting its own specific issues. Patients deteriorate very quickly with SCLC which means that treatment needs to be started quickly before a patient deteriorates and may no longer be fit enough to receive it. This can make it challenging to perform all the screening investigations prior to randomising patients for a trial in a short timescale, including molecular analyses of tumour biopsies. Consequently it is possible that maintenance trials may represent an interesting time point to consider the use of targeted agents in SCLC as the time pressure for treatment would no longer be present. Given the limited advances that have been made with chemotherapy in SCLC in the past 20 years, targeted agents with appropriate biomarkers may still be one of the best strategies for improving outcomes in SCLC. However, as discussed above the limited availability of tumour tissue in SCLC is one the hurdles that needs to be overcome to be able to use molecular subtyping of SCLC as a method of stratification for clinical trials [2, 57].

1.1.3 Models of SCLC

There are three major approaches used for the study of SCLC; tumour specimens, cell lines and mouse models, including both patient-derived xenografts (PDX) and genetically engineered mouse models (GEMM). Tumour samples from patients, as already discussed, have limited availability in SCLC due to the frequent diagnosis from small biopsies or cytology specimens obtained during bronchoscopies or under radiological guidance, as surgical resections are rare. Over recent years the use of other sources of patient tumour material, such as circulating tumour cells (CTCs), has been increasingly studied as an attractive alternative to biopsies. The use of CTCs in SCLC will be discussed in more detail later in this chapter. Cell lines and mouse models have been invaluable in the study of SCLC due to the paucity of available tumour samples, but each has their own limitations as will be briefly summarised in this section.

Cell lines have been widely used in research in SCLC providing critical evidence about the molecular features of this disease [58, 59]. They have limitless replicative ability and can be used in both in vitro, and in vivo assays to investigate invasiveness and tumorigenicity [60]. They are a pure population of tumour cells which represents both an advantage and disadvantage of cell lines. It allows the study of specific genetic and epigenetic changes without any contamination from non-tumour cells [60]. However, it also means that the important interactions of stromal, immune, vascular and inflammatory cells with tumour cells are not investigated. Cell lines are easily distributed and so can be used by investigators worldwide allowing the comparison of results obtained from assays [58]. They have been used extensively in the testing of therapeutic approaches, but the results have not always been translatable to the clinic [21, 60]. There have been major concerns about the contamination of cell lines both with mycoplasma and other cell lines which may have distorted results obtained from their analyses [58]. One of the other concerns that has been raised by researchers is the possibility of genetic drift during multiple passages of cell lines which will impede the ability of cell lines to represent the original patient material from which they were derived [21].

Mouse models have been another important method used to investigate SCLC. Tumour xenografts have been created through the implantation of established tumour cells lines into immune compromised mice allowing in vivo work to be compared to the in vitro cell line research. Many of the limitations of cell lines such as the inability to study the effects of stroma remain with xenografts [61]. In creating PDX, in contrast, a small fragment of tumour is transplanted into an immune compromised mouse which ensures the tumour architecture is maintained enabling the study of the interactions of stromal tissue with tumour cells [62]. PDX may also capture the heterogeneity of patients' tumours better than GEMMs and cell lines which may be critical when investigating chemoresistance in SCLC [61]. A comparison of the gene expression in xenografts and a xenograft cell line to primary SCLC samples demonstrated that the cell line had lost the expression of genes identified in the other specimens [21]. It is therefore argued that PDX may better represent the genetic landscape of SCLC tumours than cell lines. However, there is evidence that xenografts may acquire new genomic changes and faster growth rates than the primary tumours from which they were created, which represents a limitation of this approach [21]. PDX have been extensively used to study novel therapeutics though the results do not always mirror those from the clinic [61, 63]. The lack of surgical specimens in SCLC also represents a limiting factor for the creation of SCLC PDX, particularly as they must be processed

very rapidly after surgery [21, 61]. The methods used for orthotopic implants are often technically challenging [61]. The success of establishing PDX tumours is also variable limiting the number of models that have been create [24]. PDX are generated by transplanting tumour specimens into immune-deficient mice, so by definition the role of the immune system in interacting with tumours cannot be studied.

CTCs represent a novel and exciting source of tumour tissue from which to generate PDX in SCLC [24]. CTC derived explants (CDX), unlike traditional PDX, do not have the same issues of limited availability of tumour tissue for their generation. CDX models also critically recapitulate the responses seen in patients' tumours to standard chemotherapy in SCLC, suggesting this novel approach may be very useful in investigating treatments in SCLC in the future [24]. CDX will be discussed further in chapter 4 and the results of the molecular profiling of the tumours and how they reflect SCLC presented.

GEMMs have been used to study SCLC and provided key insights into the molecular drivers of this disease. Initial models studying SCLC using knockout models of RB1 demonstrated the development of extrapulmonary neuroendocrine tumours but no lung tumours whilst knockout models of RB1 and TP53 also developed a range of neuroendocrine tumours but again not lung neuroendocrine tumours [64-66]. A subsequent mouse model was engineered with conditional alleles of RB1 and TP53 which were inactivated in lung tissue alone. The tumours which developed in this model morphologically and immunophenotypically replicated SCLC [67]. Importantly the models showed both pulmonary and extrapulmonary metastases consistent with the biology of SCLC, which the majority of PDX and previous GEMMs fail to demonstrate [67]. These models are also developed in mice with intact immune systems allowing the investigation of the impact of immune cells on the tumour, unlike in PDX and cell lines. The tumour and its environment including the immune system are however murine rather than human. GEMMs fail to capture the heterogeneity of tumours which represents a limitation, but some have argued that this allows a focus on the critical drivers of a cancer without passenger mutations clouding the genetic landscape [62]. SCLC is critically associated with smoking and these models fail to capture the range of mutations created by smoking. Although RB1 and TP53 mutations are clearly critical for the development of SCLC, smoking produces effects beyond this which these GEMMs fail to capture [68]. GEMMs are costly and time consuming to develop which will limit the number of mutations that can be modelled in this manner [61].

The various models that have been used in place of SCLC patient biopsy specimens all have their limitations as has been discussed, but nonetheless all have provided critical information which has advanced the knowledge of this disease. The different advantages and disadvantages of the models used to reflect SCLC tumours in patients mean a combination of approaches may be needed to explore this disease. Ultimately however, this also reinforces the need to improve the collection of primary patient-derived tumour material which can be analysed rather than using models as a substitute.

1.1.4 Molecular Analysis of Small Cell Lung Cancer

Advances in studying the molecular landscape of SCLC have been made in recent years, despite the limited availability of tumour tissue for research. Analysis of the genomic landscape of SCLC has revealed the presence of multiple copy number alterations and mutations. In comparison to many other tumour types, SCLC, along with other lung tumours and melanomas, have very high rates of mutations, with 7.4 million protein changing mutations seen per million base pairs in one large study [69, 70]. The high rate of mutations are felt likely to be attributed to the mutagenic effects of tobacco smoking with which SCLC is so closely associated, with estimations that one mutation is produced by every 15 cigarettes smoked [59, 70]. Sequencing of SCLC reveals an excess of G>T/C>A transversions than would be expected from the rate of mutations in evolution, which is felt to be due to the carcinogens in tobacco smoke [59, 71-73]. One of the challenges of investigating the genomic changes in SCLC, given its high mutation rate, is identifying driver mutations from amongst the background passenger mutations [70].

1.1.4.1 TP53 and RB1

Genomic changes in TP53 and RB1 have been identified in multiple studies examining SCLC and changes in these two genes are hypothesised to be critical in the evolution of SCLC [74]. TP53 is a critical tumour suppressor gene which has a crucial role in carcinogenesis through the control of the cell cycle and as such has been described as “the guardian of the genome” [75]. TP53 has a complex role in cells and acting as a transcription factor influences many cellular functions such as the cell cycle, DNA repair, apoptosis and changes in metabolism. The assessment of signals from intracellular stress and abnormality sensors can result in TP53 inducing senescence or

apoptosis in cells hence its loss being tumorigenic [76]. It is the most commonly mutated gene in human cancer with increased rates of mutations in TP53 noted in more advanced stages of cancer and more aggressive cancer types [77, 78]. Studies in SCLC have estimated that loss of heterozygosity or mutations in TP53 occur in up to 90% of SCLC cases [74, 79, 80]. The incidence of TP53 mutations in neuroendocrine tumours of the lung increases in more aggressive histopathological subtypes [81]. A diverse range of mutations, not only in their type and location but also structural impact, have been demonstrated in TP53 (figure 1.4) [78]. There is, however, a predominance of G to T transversions seen amongst the mutations present in SCLC suggesting the mutations are linked to the carcinogens in tobacco smoke [71, 82] . RB1 is also a tumour suppressor gene which regulates the G1-S cell cycle checkpoint, integrating both extracellular and intracellular signals to decide if a cell should proceed through the cell cycle [76]. Alterations in RB1 are also seen in up to 90% of SCLC tumours with changes including point mutations, small deletions, rearrangements and chromosomal loss seen [83-85].

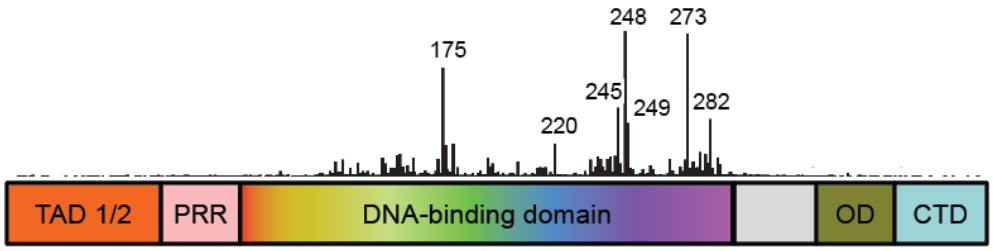


Figure 1.4 Frequency of position of mutations in TP53 according to the site of the mutation (Adapted from Joerger et al 2014 [86]). The frequency of mutations at different positions in TP53 as reported in the TP53 Mutation Database of the International Agency for Research on Cancer is demonstrated [87]. The vertical bars indicated the frequency of missense mutations in human cancers demonstrated in each residue. TAD 1/2, transactivation domain; PRR, proline-rich region; OD oligomerization domain; CTD, C terminal domain.

Genomic changes in both TP53 and RB1 were identified within an early SCLC case which was resected and then subjected to whole genome sequencing (WGS) [88]. In a comparison of resected SCLC (early stage) and autopsy samples (late stage) in a separate study, no difference was noted in the somatic copy number alterations including no difference in the loss of 17p (TP53) and 13q (RB1)[70]. This suggests that mutations in these two genes occur early in the development of SCLC. As discussed earlier a GEMM engineered to have conditional inactivation of TP53 and RB1 in lung tissue alone resulted in the development of tumours which demonstrated many of the features of human SCLC [67], supporting the idea that mutations in these two genes

occur early, and are crucial for, the development of SCLC [89]. The loss of the tumour suppressor genes TP53 and RB1 are therefore felt to be critical to the evolution of SCLC and its aggressiveness [82].

1.1.4.2 Copy Number Alterations in Small Cell Lung Cancer

Copy number alterations (CNA) changes in SCLC have been frequently noted in the analysis of tumour samples (figure 1.5). The first noted karyotypic change in SCLC was the recurrent deletion of chromosome 3p [90]. Since this initial finding multiple consistent copy number changes have been noted in SCLC. The loss of 17p, containing TP53, and 13q, containing RB1, have been frequently noted changes [70]. Other regions of frequently altered copy number include the loss of 3p and the amplification of 3q and 5p [70]. When assessing CNA changes at a gene level the loss of TP53 and RB1 are commonly noted, but in addition the loss of the tumour suppressor genes RASSF1 and FHIT, located on 3p, and PTEN, located on 10q, were also frequently seen [70, 91, 92]. A variety of amplifications which could drive the growth of SCLC tumours have been noted. Amplifications of members of the MYC oncogene family such as MYC, MYCN and MYCL1 have been seen in up to 20% of SCLC tumours [74, 92]. The amplification of members of the SOX family have also been noted, with 27% of tumours having an amplification identified in SOX2 [70, 91]. Suppression of SOX2, through the use of short hairpin RNA, was noted to block the proliferation of SCLC cell lines in which SOX2 was amplified, supporting the relevance of this amplification in the pathogenesis of SCLC [91]. EGFR and FGFR1 have been noted to be amplified in SCLC in addition to adenocarcinoma and squamous carcinoma of the lung, with FGFR1 amplification representing a potentially druggable target [2, 70]. Amplifications in BCL2, EPHA7, CCNE1 and KIT have also been noted in subsets of SCLC tumours [2, 80, 91, 93]. The analysis of changes in copy number in SCLC has therefore revealed a complex pattern of the loss of tumour suppressor genes and the amplification of oncogenes driving the growth of this cancer. Unfortunately to date the commonest copy number changes identified in SCLC, such as the loss of TP53 and RB1 and the amplification of MYC family members and SOX2, are not currently druggable targets though FGFR1 amplification represents a potential therapeutic target [2, 94].

1.1.4.3 Mutations in Small Cell Lung Cancer

As previously stated SCLC has frequent mutations identified in sequencing data, many of which will be passenger mutations. However, there have been mutations noted with high prevalence in genes likely to have a significant role in the pathogenesis of SCLC. TP53 and RB1 have had inactivating mutations frequently identified within them as previously discussed [74, 80, 85, 95]. Inactivating mutations within the tumour suppressor gene PTEN have also been noted [70, 91, 96]. There have been rare activating mutations noted with genes such as PIK3CA, EGFR, KIT and KRAS [81, 91, 94, 97, 98]. Genes encoding histone modifying enzymes such as MLL, MLL2, CREBBP and EP300 have also been noted to be frequently mutated in SCLC, suggesting a key role for histone modification in the development of SCLC [59, 70, 88, 91]. Pathways that are frequently mutated in SCLC include the Phosphatidylinositol 3-kinase (PI3K) pathway, SOX family members, hedgehog family members and DNA repair and checkpoint pathway genes [70, 91, 99].

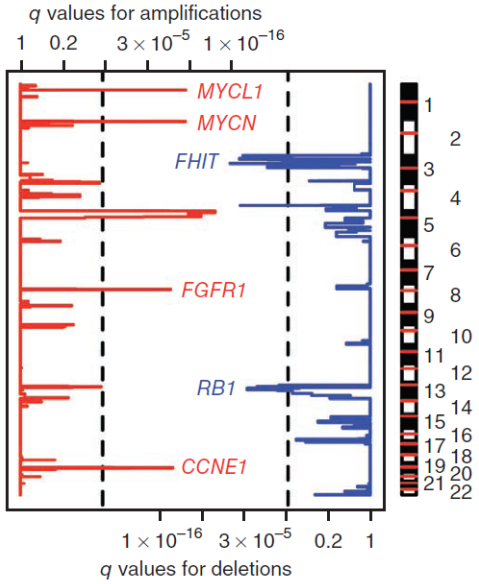


Figure 1.5 Genes with significantly altered copy number in SCLC (Reproduced from Peifer et al 2012 [70]). Copy number alterations were identified in 63 SCLC tumours. Genes with significantly amplified (red) or deleted (blue) copy number are indicated in the figure. Statistical significance for the copy number change for each gene was calculated, as indicated by the q score, with genes crossing the black dotted lines being significant.

Given the progress that has been made within NSCLC and other tumour types through the use of targeted agents the presence of potentially actionable mutations in SCLC has been investigated. EGFR mutations have been noted in a small number of SCLC cases, with 4% of tumours in one study noted to be mutated, with particular association

between cases of SCLC combined with adenocarcinomas and EGFR mutations noted [55, 98, 100]. In one study investigating 122 SCLC tumours, 4% had an EGFR mutation identified within them. An unselected trial of Gefitinib treatment in SCLC was negative, but there are a small number of cases reports of responses to EGFR TKIs in patients with EGFR mutated SCLC [101-103]. Trials of targeted therapies focussed on commonly mutated pathways in SCLC such as the PI3K, DNA repair and Hedgehog pathways have been carried out [56, 104]. Trials of mTOR inhibitors in unselected populations of SCLC, both as maintenance therapy and for the treatment of relapsed disease, were negative [105, 106]. Amuvatinib, a multi-kinase inhibitor and inhibitor of RAD 51 was trialled in combination with Cisplatin and Etoposide in patients with relapsed or refractory disease, with an overall disease control rate of 30% seen which was insufficient to initiate the next stage of the trial [107]. Vismodegib, an antagonist of the SMO receptor, was trialled in combination with Cisplatin and Etopside in the first-line without demonstrating improvement in survival [56]. Trials of other inhibitors of the hedgehog pathways are yet to report [56]. Again the failure of some of these trials may be accounted for by the lack of identified predictive biomarkers to select patients for entry to trials. Ross et al sequenced 98 SCLC tumours to investigate the frequency of actionable mutations in SCLC tumours which could potentially be used to direct patients into clinical trials of targeted therapies [94]. They identified at least one actionable mutation in 53% of the tumours sequenced. Genes in which alterations were identified included RICTOR, KIT and PIK3CA with alterations in 10%, 7% and 6% of the tumours respectively. There were mutations in each of EGFR, PTEN and KRAS in 5% of tumours analysed. This highlights the potential of sequencing of SCLC tumours to reveal drug targets but also confirms the heterogeneous range of mutations present within this tumour type.

1.1.4.4 Changes in Gene Expression in Small Cell Lung Cancer

Analysis of gene expression in SCLC has provided further insights into the biology of this tumour type. Rudin et al demonstrated frequent amplification of SOX2 in SCLC tumours [91]. Analysis of the transcriptome of 53 SCLC samples confirmed that SOX2 was overexpressed. Studies of SOX2 expression have confirmed its roles in the promotion of the pathogenesis of tumour development from lung epithelial cells [108]. Analysis of the transcriptome of SCLC has also identified multiple fusion transcripts [22, 91]. In one study a recurrent RLF-MYCL1 fusion transcript was noted in one tumour sample and four cell line samples [91]. The RLF-MYCL1 fusion resulted in the

expression of a fusion protein with 79 amino acids from RLF fused to MYCL1 which lacked its first 27 amino acids. Within the clinical specimen that the fusion was identified in, MYCL1 was also noted to be overexpressed. In a second study 60 fusion transcripts were noted [22]. However, no in-frame fusions were detected in more than one of the 42 specimens analysed suggesting there was not a single key gene affected. Gene expression and proteomic profiling of SCLC revealed higher expression of a number of E2F1-regulated proteins such as EZH2 [109]. E2F1 targets such as EZH2 and Chk1 are currently being evaluated as therapeutic targets in other cancers, suggesting a further potential avenue of research for SCLC. In the same study PARP1 was noted to have higher expression suggesting a potential role for PARP inhibition in SCLC. PARP inhibitors were demonstrated to have single agent activity in preclinical testing in SCLC [109]. Initial reports have suggested that PARP inhibitors such as BMN-673 also have single agent activity in clinical trials in SCLC [2]. There are several other clinical trials on-going with PARP inhibitors representing an area of significant interest in SCLC.

1.1.4.5 Conclusions of Molecular Changes in Small Cell Lung Cancer

The genomic profiling of SCLC has revealed the complex landscape of a highly mutated disease. Almost universal changes, both in copy number and mutations, have been noted in TP53 and RB1 suggesting they play a critical role in SCLC tumorigenesis. Common CNA changes such as the loss of the tumour suppressor genes PTEN, FHIT and RASSF1 and the amplification of members of the MYC and SOX families have been identified. Frequent mutations in genes encoding histone modifying enzymes have also been seen, suggesting a role in the development of SCLC. Potential therapeutic options such as the inhibition of PARP and KIT have also been revealed through analysis of SCLC tumours. It should, however, be noted that the two landmark studies of next generation sequencing (NGS) of SCLC tumours were only published in 2012 [70, 91]. To date less than 140 SCLC tumours have been subjected to whole exome sequencing (WES) in contrast to more than 1000 NSCLC tumours [2]. In the cancer genome atlas alone data is available from over 1000 invasive breast carcinomas which have been WES. This suggests that there remain significant amounts of research into the genomic landscape of SCLC that are required to investigate this complex disease, particularly when the high rates of mutation seen in this disease is considered. Analysis of SCLC tumours may also be required at the individual level to personalise therapy given the high mutation rate in this disease. The

results of selected key studies of SCLC patient samples using NGS are summarised in table 1.1.

Analysis	Samples	Observations	Reference
WES, WGS, Transcriptome sequencing	29 exomes, 2 genomes, 15 transcriptomes of SCLC tumours	Inactivation TP53 and RB1. Recurrent mutations in genes encoding histone modifiers. Mutations in PTEN, SLIT2 and EPPHA7. FGFR1 amplification.	Peifer et al. Nat Genet 2012 [70]
WES, Transcriptome sequencing, CNA analysis	36 SCLC and matched normal tissue pairs, 17 matched SCLC and lymphoblastoid pairs and 4 primary tumours and 23 cell lines.	Inactivation and loss of TP53 and RB1. SOX2 amplification in 27% of samples. Recurrent RLF-MYCL1 fusion.	Rudin et al. Nat Genet 2012 [91]
CNA analysis, Transcriptome Sequencing	58 SCLC tumour samples with matched normal tissue pairs for CNA analysis and 42 pairs for transcriptome sequencing	MYC family amplification. KIAA1432 amplification. 60 fusion transcripts but not recurrent.	Iwakawa et al. Gene Chromosome Canc 2013 [22]
WES and CNA analysis	51 WES and 47 CNA analysis of SCLC tumours and matched normal tissue pairs	High frequency of TP53 and RB1 amplifications. Frequent MYC family amplifications. Genetic alterations in PI3K/AKT/mTOR pathway in 36% of tumours.	Umemura et al. J Thorac Oncol 2014 [99]
Methylated CpG Island Recovery Assay	18 SCLC tumours (with 5 matched normal tissues) and 5 cell lines	73 gene targets methylated in >77% SCLC samples. Methylated genes enriched for homeobox, transcription factors, neuronal fate commitment and neuronal differentiation	Kalari et al. Oncogene 2013 [110]

Table 1.1 Selected NGS studies of SCLC patient samples.

1.2 Chemoresistance

Chemoresistance is one of the key challenges limiting the effectiveness of oncological therapies. Drug resistance is thought to lead to the failure of treatment in more than 90% of patients with metastatic cancers [111]. It may also lead to the failure of adjuvant chemotherapy due to the presence of resistant micrometastases [112]. In researching chemoresistance multiple mechanisms have been revealed. Common mechanisms of drug resistance across tumour types include accelerated efflux of drugs, decreased activation of drugs or increased inactivation of drugs [113-115]. Tumour cells may evade drug-induced apoptosis as a form of pleiotropic drug resistance [116]. Many chemotherapy agents act by inducing DNA damage and tumours may develop resistance by decreasing the ability of chemotherapy to damage DNA, increasing DNA repair mechanisms or an increased ability to tolerate damage

[117]. Tumours may also become resistant to drugs through alteration or bypass of a drug target [118, 119].

Resistance can be either intrinsic or acquired [117]. Intrinsic resistance refers to tumours which have de novo resistance to chemotherapy without prior exposure to that treatment. Acquired resistance in contrast refers to the development of resistance to a chemotherapy in a tumour which was initially sensitive to this same treatment. Unfortunately in the process of acquiring resistance to one drug the tumour may become cross-resistant to a range of other chemotherapies [112]. Whilst acquired resistance evolves under the direct selection pressure of treatment, intrinsic resistance develops in tumours without the influence of treatment. The molecular basis of intrinsic resistance is therefore unlikely to represent specific changes producing resistance to an individual drug, but it likely to represent changes which provide advantages for the cells such as improved growth but also act to confer resistance to chemotherapy such as impaired control of apoptosis.

In SCLC chemotherapy resistance represents a major challenge. Although response to first-line therapy is high as discussed earlier, for the 20% of patients who are resistant to first-line therapy the response rate to further therapy is very low and their prognosis is very poor [41]. However, unfortunately, even for the patients who respond to first-line therapy the durability of responses is low and response rates to second-line therapy are dramatically reduced, indicating the development of resistance [3]. This represents a significant change in the behaviour of SCLC to chemotherapy, changing from a very chemosensitive tumour type in which patients have rapid and marked tumour reduction from chemotherapy to one in which the tumours are very chemoresistant. Tackling the causes of chemoresistance may therefore yield significant therapeutic improvements for patients.

When considering chemotherapy response in SCLC patients are often separated in to two groups, chemorefractory and chemoresponsive. In this thesis chemorefractory patients are defined using clinical criteria as those whose tumour progresses within 3 months of completion of therapy whilst chemoresponsive patients are defined as those whose tumour progresses after longer than 3 months (figure 1.6) [3, 120, 121]. Chemorefractory disease is therefore equated with intrinsic resistance whilst the relapse of chemoresponsive disease is equated with the development of acquired resistance in this thesis. It is acknowledged that there are caveats associated with the use of these definitions. Whilst the definitions used for chemorefractory and chemosensitive disease are clinical definitions, the categories of intrinsic and acquired

resistance refer to biological differences in the behaviour of tumours. It is also important to remember that SCLC is not binary in nature as this categorisation implies, but that the response to chemotherapy is on a spectrum. For example the definition of refractory disease used in this thesis encompasses both those patients whose tumours progressed whilst receiving treatment and those who progressed over 2 months after completing treatment. Equating the relapse of chemoresponsive disease with acquired resistance is also further complicated by the fact that at relapse after first-line chemotherapy up to 20% of patients may respond to a rechallenge of platinum-containing chemotherapy [3].

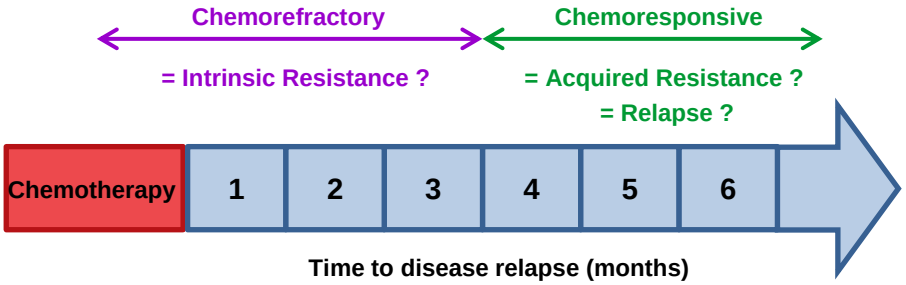


Figure 1.6 Chemorefractory and chemoresponsive disease in SCLC. Chemorefractory disease is defined as tumours which progress within 3 months of completion of chemotherapy whilst chemoresponsive disease is defined as tumours that progress more than 3 months after the completion of chemotherapy. Chemorefractory disease is equated with intrinsic resistance whilst the relapse of chemoresponsive disease is associated with acquired resistance. However, chemorefractory and chemoresponsive disease are defined using clinical criteria whilst intrinsic and acquired resistance refer to biological differences within tumours.

Studies of chemoresistance in SCLC have revealed the development of multiple different mechanisms, as in other tumour types [116, 117]. The mechanisms of resistance to platinum agents have been extensively studied but elucidating the most clinically relevant mechanisms remains difficult [116, 122-124]. Multidrug resistance proteins (MRP), ATP-dependent efflux pumps with broad substrate specificity, are expressed in a high proportion of SCLC cell lines and are thought to be a mechanism of chemoresistance in some tumours [125]. Patients with SCLC whose tumour samples had positive immunohistochemistry (IHC) staining for MRP1 or MRP2 had lower response rates to platinum-based combination chemotherapy than those whose tumours had negative staining in two studies [126, 127]. The expression of MRP1 was noted to increase in relapsed tumours after platinum and etoposide chemotherapy compared to the values seen in chemonaïve tumours suggesting a link with acquired resistance to chemotherapy [126]. Another potential mechanism of resistance in SCLC is increased expression of p glycoprotein (P-gp), an efflux pump. Overexpression of P-

gp is seen in SCLC cell lines which have been selected for resistance through exposure to Paclitaxel, Etoposide and Anthracyclines but not in cells lines that were not selected for resistance [128-132]. When analysing P-gp expression in tumour samples, higher expression was seen in patients with poorer response and survival [117]. The levels of P-gp were also noted to be increased in tumours which had been exposed to prior chemotherapy compared to pre-treatment levels [117]. However, trials of the P-gp inhibitors Verapamil and Megestrol acetate with chemotherapy did not improve patients' outcomes [133, 134].

Evasion of drug-induced apoptosis is another potential mechanism for chemoresistance in SCLC. B-cell lymphoma-2 (BCL-2), a family of proteins with a role in the regulation of apoptosis, are felt to have roles in both malignant transformation and drug resistance [135]. BCL-2 is known to be frequently amplified and overexpressed in SCLC tumours [136] and also in drug resistant SCLC cell lines [137]. Despite preclinical evidence of the response of SCLC xenografts to BCL-2 inhibitors, trials of anti-BCL2 agents have yet to deliver clear clinical benefits [54, 136]. Etoposide, a topoisomerase inhibitor, is frequently used in the treatment of SCLC. Both decreased expression by IHC and mutations within topo II-α have been seen in SCLC tumours exposed to Etoposide and resistant SCLC cell lines [117]. The presence of cancer stem cells, loosely defined as a population of cells with the ability to self-renew and repopulate tumours, within cancers has been another proposed mechanism of drug resistance. It is proposed that cancer stem cells may have a number of properties that promote their chemoresistance such as increased DNA damage repair and drug efflux [138]. Cancer stem cells may also be more quiescent than other cells so are not affected to the same degree by chemotherapies as the rapidly proliferating cells within tumours [138].

There have been multiple potential mechanisms identified in SCLC, as in other malignancies for chemoresistance. Given the number of mechanisms of resistance that occur, investigation of individual patients' tumours are likely to be necessary to personalise their therapy to overcome drug resistance. Given the presence of heterogeneity within tumours it is also possible that multiple mechanisms of resistance may be present within a single patient's tumour and methods which can explore this heterogeneity may therefore be required [139]. As in SCLC the major problem is acquired resistance, as opposed to intrinsic resistance, serial tumour samples will be required to reveal the evolution of resistance mechanisms in response to therapy as these will not be present in the diagnostic biopsy. The difficulty of obtaining repeated biopsies is again a challenge for the investigation of chemoresistance in SCLC.

1.3 Circulating Tumour Cells

Within oncology there has been an increased focus on the delivery of personalised medicine to improve patient outcomes [140, 141]. This requires the development, not only of novel targeted therapies, but also the identification of molecular biomarkers within tumours to optimise therapy selection [142]. The evolution of molecular changes within tumours associated with the development of resistance need to be monitored to enable this optimisation [143]. The success of personalised medicine therefore depends on the ability to repeatedly sample tumours over the course of patients' treatment. The challenges of obtaining biopsies in lung cancer patients have been discussed above and are particularly acute when considering the need for serial biopsies. Other limitations of tumour biopsies include the inability of a single biopsy to capture the heterogeneity within a patient's primary tumour and between primary and metastatic sites [144]. With the drive towards personalised medicine and the challenges of obtaining repeated tissue biopsies there has been increased interest in the potential of "liquid biopsies" such as CTCs and circulating free DNA (cfDNA) [145]. CTCs have the potential to act as a minimally invasive method of monitoring tumours [146]. In 1869 Thomas Ashworth identified microscopically CTCs in the blood of a patient with metastatic cancer, and hypothesised that they may have a role in the development of metastases [147]. As 90% of patients die of their metastatic disease, CTCs are clearly a very significant population of tumour cells to study [148]. CTCs can be shed from primary tumours and metastatic sites, with traffic of CTCs between the primary tumour, metastases and the bone marrow such that they may capture the heterogeneity of tumours [149, 150]. CTCs are arguably therefore a very relevant source of tumour material to enable its serial characterisation.

1.3.1 Enrichment and Isolation of Circulating Tumour Cells

One of the challenges of studying CTCs is their identification, enrichment and isolation, as within a millilitre of blood there may be as little as one CTC amongst ~10^7 white blood cells [151, 152]. Over 40 technologies have been developed to detect CTCs with new technologies being regularly published [153]. A summary of selected key CTC detection technologies is given in table 1.2. CTC enrichment techniques are based on exploiting the physical or biological differences between CTCs and WBCs [151]. The physical properties exploited by CTC enrichment technologies include size, density and membrane capacitance and conductance [154-156] whilst the biological

Technology	Method of CTC enrichment	Method of CTC detection	Notes
Antibody-based capture assays			
CellSearch®[157, 158]	EpCAM-coated ferromagnetic beads	Immunocytochemistry for cytokeratin, CD45 and DAPI	FDA approved in advanced breast, prostate and colorectal cancer
MagSweeper[159]	EpCAM-coated magnetic beads enriched using magnetic rod	Microscopic visualization	Live cells can be isolated
GILUPI cell collector[160]	Functionalized EpCAM-coated medical wire	Immunocytochemistry for EpCAM, cytokeratin and DAPI	*In vivo* collection
CTC chip[161] and Herringbone chip[162]	EpCAM-coated microposts and chip surface	Immunocytochemistry for cytokeratin, CD45 and DAPI +/- tumour-specific markers	Microfluidic microchip technology
CTC iChip®[163]	Magnetic bead capture combined with microfluidic inertial focusing	Immunocytochemistry or RT-PCR	Positive selection by EpCAM or negative selection by CD45.
Ephesia CTC-chip[164]	Functionalized magnetic beads combined with microfluidics	Immunocytochemistry for cell surface and nuclear markers	Flexibility with capture antibody
AdnaTest[165]	Immunomagnetic beads with MUC1-coupled and EpCAM-coupled antibodies	Multiplex RT-PCR for panel of genes (*MUC1*, *HER2* or *EpCAM*)	Cell lysis means quantification of tumour cell number is not possible
Physical characteristic-based assays			
ISET®[154]	Filtration based on cell size	Immunocytochemistry or FISH	Nonepithelial cells can be isolated
Dean Flow Fractionation[166]	Size-based selection using centrifugal force	Immunocytochemistry for cytokeratin, EpCAM, CD45 and Hoechst	Nonepithelial cells can be isolated
Dielectrophoretic field-flow fractionation[156]	Membrane capacitance	Immunocytochemistry	CTCs selected are viable
JETTA Microfluidic Chip [167]	Size and deformability based microchannel device	Immunocytochemistry	CTCs segregated in chambers
Functional assays			
EPISPOT assay[24]	CD45 depletion and short-term culture in plates coated in antibody against MUC-1, PSA or cytokeratin-19	Immunofluorescence secondary antibodies to MUC-1, PSA or cytokeratin-19	Detection of only viable CTCs
CAM[168]	Density gradient centrifugation and cells applied to CAM for short-term culture	Immunocytochemistry for cell-surface markers	Detection of only viable CTCs
Other assays			
ImageStream® [169]	Pre-enrichment by any method of choice	Flow cytometry-based imaging using multi-marker immunofluorescence	Cells can be imaged for up to 10 cell-surface or intracellular markers
High-throughput fluorescent scanning[170]	Red cell lysis and density gradient centrifugation	Immunocytochemistry of cell surface and nuclear markers	Nonepithelial cells can be isolated on a slide
DEPArray™[156]	Requires pre-enrichment step	Fluorescence imaging. Movement of cells within chip by electric field changes	Isolation of purified single cells for downstream analysis

Table 1.2 Selected technologies for CTC detection. (Adapted from Krebs et al 2014 [152])

differences include tumour cell surface marker expression and invasiveness [157, 158, 168]. Further steps to detect CTCs after their enrichment are required such as immunostaining and microscopy or polymerase chain reaction (PCR) based detection [171]. CTC technologies can also be considered in terms of whether the enrichment of CTCs is marker dependent or independent. Marker-dependent technologies enrich CTCs according to the expression of specific antigens whilst marker independent technologies use other criteria such as size to distinguish CTCs and WBCs [152].

The most widely used method of CTC enrichment is the CellSearch platform (Janssen Diagnostics) which has been used in large-scale clinical trials. It is the only CTC technology to have undergone the rigorous assay validation to lead to FDA approval [157, 158]. The CellSearch platform is a semi-automated technology which enriches CTCs from 7.5 ml of blood based on epithelial cell adhesion molecule (EpCAM) expression using ferromagnetic beads coated in anti-EpCAM antibodies. CTCs are then detected with immunostaining with a CTC being defined as being a cell of greater than 4 μm in size with the morphological characteristics of a tumour cell, and with positive expression of phycoerthrin-conjugated anti-cytokeratin (CK-PE) and 4'6'-Diamidino-2phenylindole (DAPI) but negative expression of the leukocyte marker allophycocyanin-conjugated anti-CD45 antibody (CD45-APC) (figure 1.7). Landmark trials proved the prognostic significance of the enumeration of CTCs by CellSearch in breast, prostate and colorectal cancer (CRC) [172-174]. One key feature resulting in the success of CellSearch was the development of the CellSave preservative tube (Janssen Diagnostics) which maintains cell surface markers for 96 hours allowing the centralised analysis of blood samples in CTC clinical trials. Many other techniques have been developed to enrich CTCs utilising EpCAM expression such as the CTC-chip [161], Herringbone chip [162], IsoFlux [175] (Fluxion), MagSweeper [159] (Stanford University) and the GILUPI cell collector [160] (GILUPI Nanomedizin). None of these other technologies have, however, been through the detailed process of assay validation and clinical qualification of a CTC-based biomarker [176]. The GILUPI cell collector is a functionalized medical wire coated in anti-EPCAM antibody magnetic beads. It is inserted into the antecubital vein of a patient for 30 minutes which allows the sampling of a larger blood volume than many other CTC technologies, potentially increasing the number of CTCs captured.

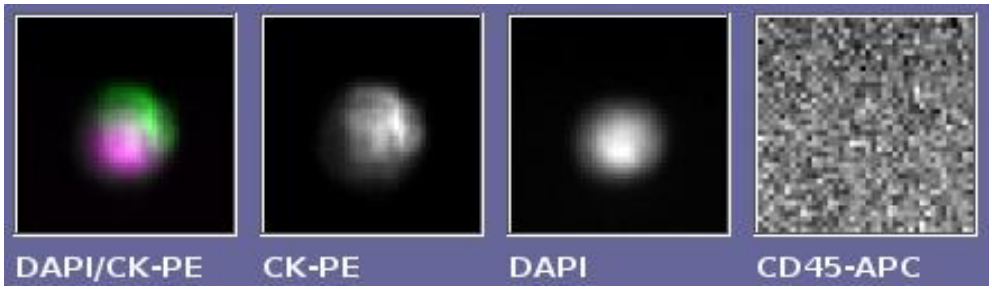


Figure 1.7 A CTC presented in the image gallery of CellSearch. CTCs are enriched using the CellSearch technology using ferromagnetic beads coated in anti-EpCAM antibody. The captured cells are immunofluorescently stained and then imaged with the images presented to the user in a gallery. CTCs are defined as cells of greater than 4 μM in size, with the morphological features of a tumour cell, and positive staining for CK-PE and DAPI, but negative staining for CD45-APC.

Epithelial to mesenchymal transition (EMT) is a reversible dynamic process involved in the development of metastasis in cancer [177]. Within normal tissue cells are tightly bound to each other and the basement membrane, but during EMT cells take on a more mesenchymal phenotype which enables increased motility and invasion. Cells that have undergone EMT are argued to lead to the generation of metastases either alone, or through cooperation with epithelial cells [152]. The process of EMT has been demonstrated in CTCs such that a subset of CTCs will have downregulation of epithelial markers including EpCAM [178, 179]. Technologies that do not rely on EpCAM expression are therefore required to enrich more mesenchymal CTCs. Depletion of CD45-positive leukocytes either on its own or combined with other methods such as red blood cell lysis and density centrifugation can preserve CTCs irrelevant of marker expression [180-182]. Marker independent technologies such as ISET which isolates CTCs through size based filtration [154], the spiral biochip which uses centrifugal forces for size based CTC enrichment [166] and the CAM (cell adhesion matrix) methodology (Vita Assay) which enriches invasive viable CTCs [168], all have the potential to enrich EpCAM positive and negative CTCs.

The enrichment steps described all result in persistent leukocyte contamination, unsurprising given the millions of WBCs initially present in a blood sample and the potentially small number of CTCs [152]. It however represents a significant challenge for the genomic and transcriptomic analysis of CTCs due to the presence of the wild-type DNA from leukocytes. PCR-based methods have been used to detect the CTC signal in enriched CTC samples, such as cytokeratin-19 mRNA detection in breast cancer [183]. However, to be able to detect the genetic profile of pure CTCs further isolation steps are required. Examples of isolation steps used include physical

micromanipulation with visualisation by microscopy followed by the isolation of cells of interest with pipettes or micromanipulators [184]. The DEPArray is another technique suitable for the isolation of a pure population of CTCs [185, 186]. It is a semi-automated technology which images the cells and then cells of interest are physically moved and isolated for further analysis via a series of electrostatic cages. CTC technology is continually evolving and this represents a very dynamic field. However, for technologies to become clinically established they must undergo assay validation to confirm they are reliable, robust and reproducible. The costs of CTC isolation and the lack of consensus in the field for which technologies are most appropriate to use represent potential barriers to the wider clinical use of CTCs. Despite these hurdles the potential of CTCs to enable the study of cancers, and inform clinical management, suggests rich rewards if these challenges are overcome.

1.3.2 Circulating Tumour Cell Research in Lung Cancer

The potential clinical utility of CTCs has been explored in studies in both SCLC and NSCLC. **In NSCLC** CTC counts by CellSearch have been shown to have prognostic significance [187, 188] as in breast, prostate and colorectal cancer [172-174]. In the laboratory of Professor Dive a study examined CTC numbers in 101 patients with NSCLC using CellSearch. Patients with more than 5 CTCs in 7.5 ml of blood were demonstrated to have a worse PFS and OS than those with less than 5 CTCs [187]. A separate study of CTC numbers in 43 NSCLC patients' blood enumerated using CellSearch also demonstrated poorer PFS and OS in patients who had more than 5 CTCs compared to those who had less than 5 [188].

Higher numbers of CTCs are seen with more advanced disease in both NSCLC and SCLC, and therefore their potential use in the diagnosis of lung cancer has been explored [187, 189, 190]. Early diagnosis of lung cancer could lead to more patients being eligible for curative therapy so this represents an important area of research. A study investigated the presence of CTCs by ISET in 168 patients under follow up with a diagnosis of chronic obstructive pulmonary disease (COPD), a smoking related illness [191]. In the study 3% of the patients with COPD had CTCs identified in their blood samples and so had annual CT surveillance, leading to the diagnosis of lung cancer from 1 to 4 years later. The patients were all eligible for surgical resection of their lung cancer. No CTCs were identified in control smoker (n = 42) and non-smoker groups (n = 35). The investigation of patients with abnormal imaging to confirm whether the diagnosis is lung cancer can be challenging, particularly if the lesions are difficult to

biopsy due to their anatomical location. For some patients, to confirm the cause of an isolated lung nodule surgical resection is necessary. The potential use of CTCs in the diagnostic pathway of these patients is therefore of interest. A study examining CTCs in 150 patients referred for investigation of a possible lung cancer diagnosis utilised CellSearch for enumeration. The presence of CTCs was not able to distinguish patients with lung cancer (n = 125) and those with a non-malignant diagnosis (n = 25). There were CTCs identified in 4 of the 25 patients with a non-malignant diagnosis in this study which may have influenced the result given the small number of patients in this group [190]. The identification of circulating tumour microemboli (CTM), tumour cell clusters, however improved the diagnostic accuracy of clinical and imaging data when investigating possible early stage NSCLC in a separate study [192]. The identification of CTMs could therefore aid the investigating clinician to decide whether the risks of biopsies or surgical resection are warranted in the investigation of a lung nodule identified on imaging. CTCs and CTM may therefore have a role in improving the diagnosis of lung cancer in high-risk populations.

In NSCLC the use of targeted therapies has led to the improvement in patient outcomes [193, 194]. The challenge of identifying the relevant molecular biomarkers in patients' biopsy samples to personalise therapy remains an issue. The potential of CTCs to act as liquid biopsies for the identification of markers of response to targeted therapies has been explored. EGFR mutations were identified in the CTCs of 19 out of 20 patients with EGFR mutated NSCLC using a PCR-based assay following CTC enrichment [118]. The development of the T790M resistance mutation was also identified in four patients who had serial monitoring of blood samples, exemplifying the utility of a blood-borne biomarker. Two studies have also explored the use of CTCs for the identification of the ALK fusion gene [195, 196]. In both studies CTCs were enriched using ISET microfiltration and the ALK fusion gene identified by fluorescent in situ hybridisation. In a separate study EML4-ALK fusion transcripts were identified from enriched CTC samples from NSCLC patients by reverse transcription real time PCR (qRT-PCR) [163]. Interestingly breast cancer studies addressing HER2 expression in CTCs have demonstrated discordant HER2 status with the primary tumour being positive but CTCs being negative for HER2 expression [197]. Patients with HER2 positive tumours but negative CTCs did not see significant improvement in PFS with anti-HER2 therapy. These examples highlight the potential utility of CTCs for directing the use of targeted therapies, and to explore the biology and evolution of cancer.

CTCs have also been used to investigate response to chemotherapy. Hirose at al investigated the relationship between CTC number and response to gemcitabine and carboplatin combination chemotherapy in NSCLC patients [198]. There was no difference in response between patients with CTCs identified and those without CTCs. A study in NSCLC demonstrated a correlation with the change in CTC number after treatment with erlotinib and pertuzumab and radiographic response, suggesting CTCs may have a role as in addition to imaging to assess response to therapies [199]. Studies in other tumour types have also suggested that CTCs could be used to investigate response to chemotherapy. In pancreatic adenocarcinoma gene expression profiles associated with response to different chemotherapy agents were assessed in CTCs, and retrospectively compared to patients' responses to chemotherapy [200]. The results demonstrated that the model could be used to predict chemotherapy responses, highlighting a potential use for CTCs. In breast cancer Yu et al demonstrated that cell lines could be created from CTCs isolated from six patients with metastatic breast cancer and used for drug sensitivity screens, demonstrating how CTCs could be used in real time to personalise patients' therapy [201].

CTCs have also been used to investigate resistance to chemotherapy and targeted therapies. A study assessed resistance to chemotherapy by correlating the expression of MRP genes in CTCs with response to chemotherapy [202]. The CTCs were enriched using Dynabeads coated in an antibody to EpCAM and RNA for analysis was then isolated from the enriched CTC sample. Of the 105 patients recruited, including 10 with NSCLC, 54 had CTCs present to evaluate. Overexpression of individual MRP genes was hypothesised to lead to resistance to specific chemotherapies; for example, overexpression of MRP7 is linked with resistance to Taxanes and Vinca alkaloids. The response of patients to the chemotherapies they received was predicted based on the pattern of expression of MRP genes, and in 98% of the patients correlated with the observed response on imaging in the patients. A further study by the same group assessing the expression of MRP genes in CTCs from metastatic breast cancer patients confirmed the association of a shorter PFS in patients with expression of the drug resistance signature identified [203]. However, the association of ALDH1 expression, a marker for stemness, was also investigated. Of the patients with ALDH1 expression over 76% had expression of a drug resistant signature, and a correlation between the number of MRPs expressed and ALDH1 expression was also noted. These studies suggest that CTCs can be used to explore a number of mechanisms associated with resistance. The potential utility of CTCs to explore drug selection and resistance, and to personalise therapies for patients is apparent from these studies.

In SCLC the number of CTCs identified by CellSearch is greater than in any other disease type assessed to date with a range from 0 to 44,896 CTCs in 7.5mls of blood and CTCs having been identified in 85% of patients analysed [204]. SCLC therefore potentially represents an ideal cancer with which to investigate the utility of CTCs. As in NSCLC the baseline pre-treatment CTC number has been shown to be prognostic in SCLC [189, 204, 205]. A study by Hou et al of 97 patients with SCLC investigated the significance of CTC number as enumerated by CellSearch [204]. Patients with more than 50 CTCs in 7.5 ml blood had a statistically significant reduction in OS at 5.4 months, in contrast to 11.5 months in those with less than 50 CTCs (figure 1.8). In a separate study by Naito et al, CTC number as enumerated by CellSearch was investigated in 51 patients, again confirming poorer OS in patients with greater numbers of CTCs, but in this study the number of CTCs distinguishing favourable and unfavourable prognosis was 8 [205]. Of note in the study by Hou et al two thirds of the patients had extensive stage disease whilst just half the patients in the study by Naito et al had extensive stage disease which may have influenced the number of CTCs that distinguished unfavourable and favourable patients within each study. The change in CTC number with chemotherapy as measured after one cycle has also been demonstrated to be prognostic [187, 189, 204, 206, 207] with two studies concluding that the change in CTC number rather than the initial number of CTCs was of the greatest prognostic significance [189, 206]. Additionally the presence of CTCs after chemotherapy, and at relapse, has been shown to be prognostic [205]. The presence of CTM has also been noted to be of prognostic significance in SCLC, with patients with one or more CTM identified by CellSearch having a worse prognosis than those with no CTM identified. Multiple studies have confirmed the prognostic significance of CTCs in SCLC suggesting that CTCs are a critical group of cells to research in this disease.

The potential of changes in CTC numbers to act as a surrogate marker for response to chemotherapy has also been investigated. In a study by Hiltermann et al in SCLC although the change in CTC number with one cycle of chemotherapy was associated with survival, it was not associated with radiological response [189]. In total the CTCs numbers in 59 patients, consisting of 21 with LS SCLC and 38 with ES SCLC, were analysed. The patients with LS SCLC were all treated with concurrent chemoradiotherapy whilst those with ES SCLC were treated with chemotherapy alone. In a separate study Fu et al investigated 112 patients with stage III SCLC who were radically treated with concurrent chemoradiotherapy, noting the number of CTCs after one cycle of chemotherapy did correlate with response [208]. The discrepancy

between these two studies may relate to the different stages of the patients included with all the patients in the study by Fu et al having stage III disease whilst in the study by Hiltermann et al the patients ranged from stage 2 to stage 4. This difference in the patient populations in the studies influenced the treatments received, with all the patients with limited stage disease having concurrent chemoradiotherapy whilst patients with extensive disease receive chemotherapy alone. It therefore may follow that the correlation of change in CTC numbers to the two different treatment approaches may vary. The potential of CTCs to be used in SCLC to guide therapy and investigate resistance is as yet largely unexplored but the evidence from other disease types suggest this may be a rewarding area for further research.

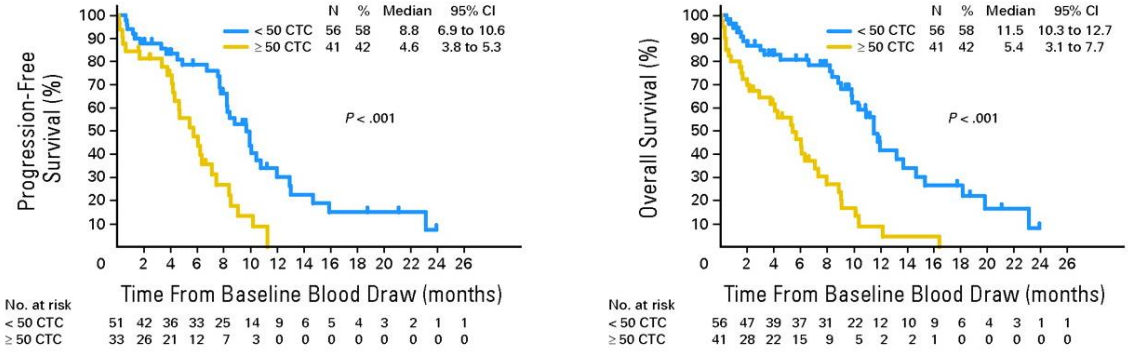


Figure 1.8 Kaplan-Meier curves for PFS and OS of SCLC patients with fewer than 50 or greater than or equal to 50 CTCs as enumerated by CellSearch in baseline blood samples (Adapted from Hou et al 2012 [204]).

1.3.3 Molecular Profiling of Single Circulating Tumour Cells

Tumours are complex structures consisting not only of cancer cells but also non-cancerous cells such as fibroblasts, endothelial cells and macrophages. The sequencing of tumours therefore results in a signal which is an admixture from cancerous and non-cancerous cells [139]. The wild-type signal from the non-cancerous cells may hide or distort the relevant tumour specific signals [152]. Beyond this tumours are also very heterogenous with variation between the primary tumour and metastases noted [144]. Even within a single tumour mass there are multiple subpopulations of cells present [209]. The dominant clone present within a tumour will change over time in response to different pressures such as therapies [210]. Bulk profiling of tumours is unable to explore this heterogeneity unlike single cell profiling [139]. The heterogeneity in tumours is particularly relevant when assessing response

and resistance to therapies as the rare subpopulations of cells containing for example resistance mechanisms, may be critical to a patient's overall response to therapy.

Although the potential benefits of single-cell analysis are evident there are significant challenges present. Single-cell analysis requires the accurate isolation of individual cells of interest [211]. For CTCs, as discussed earlier, this requires not only the enrichment of this rare cell population from the haemopoeitic cells but also a process for the accurate identification and isolation of CTCs [152]. Single cells contain very limited nucleic acid so the methods for amplification of the genome or transcriptome are necessary for their analysis [211]. A number of methods for whole genome amplification (WGA) of single cells have been developed. WGA methods can broadly be divided into methods using multiple displacement amplification (MDA) [212, 213], PCR-based amplification [214, 215] or methods that combine displacement amplification and PCR [216]. Methods for whole transcriptome analysis (WTA) have also been developed and include PCR-based methods [217, 218], MDA [219] and in vitro transcription-based amplification of reverse transcribed messenger RNA (mRNA) [220]. All the methods for WGA and WTA have limitations and produce errors which needs to be carefully considered when analysing the results of these studies, and methods for their reduction need to be developed [221]. Errors in WGA include amplification bias, allelic drop outs and preferential allelic amplification, with each method having its own particular challenges [211]. WTA can also cause bias in transcript coverage, base copying errors and inaccurate quantification of low abundancy transcripts [211]. Beyond amplification techniques methods for genotyping, NGS, array analysis and the bioinformatics approaches to analyse the data from single cells are required, and need to be reliable and reproducible if the data is to be of value [152]. Despite these challenges single-cell analyses have been developed and have revealed interesting biological changes in tumours, and they have the potential to be a significant tool in oncology research [139].

Initial approaches to the molecular analysis of CTCs focussed on analysing enriched CTCs which contained variable numbers of contaminating leukocytes. Examples of this approach include establishing a CTC "specific" gene signature when profiling CTC enriched and CTC depleted samples from patients with metastatic cancer, which in a test group was able to distinguish samples from patients with cancer from normal controls [222]. Analysis of gene expression in enriched CTC samples from patients with metastatic breast cancer revealed 55 mRNAs and 10 micro RNAs (miRNAs) which were more abundantly expressed in patients with 5 or more CTCs in 7.5 ml of blood. Interestingly clustering analysis of the CTCs revealed 5 distinct clusters with some

clusters associated with a greater risk of developing visceral and non-visceral metastases. Some studies have focussed on profiling groups of isolated CTCs rather than single CTCs in an effort to reduce bias from errors in single-cell amplification. Magbanua et al have used this approach to analyse CTCs from both patients with breast and prostate cancer [223, 224] demonstrating change in copy number in a breast cancer patient over time. This approach, however, fails to capture the heterogeneity of tumours by amalgamating the signal from individual cells.

Studies on single CTCs remain in their infancy and have focused on the technical feasibility of the analyses. Single CRC CTCs isolated from an enriched population by micromanipulation were sequenced for KRAS, BRAF and PIK3CA mutations after WGA. PIK3CA mutations were identified in four patients' CTCs, KRAS mutations in one patient's CTCs and no BRAF mutations were identified in the five patients' CTCs analysed [225]. Heterogeneity was noted in one patient's CTCs with two different PIK3CA mutations identified. In a separate study of CRC, massively parallel sequencing of CTCs and tumour samples for a panel of 68 CRC-related genes identified driver mutations in the CTCs that were present in the corresponding tumour samples [226]. Interestingly additional mutations identified within the CTCs were subsequently identified in the tumour when further deep sequencing was performed, suggesting they may be present in a small subclone of tumour cells.

Gene expression profiling studies of single cells have been performed but face the added challenge of the ease of degradation of RNA, and the concerns that processing may alter expression profiles. Ramsköld et al developed the Smart-Seq protocol for sequencing single-cell mRNA [218]. They confirmed the technique was highly reproducible, with low technical variation and demonstrated improved ability to detect rare transcripts compared to other single-cell methods. They utilised the Smart-Seq protocol to analyse individual melanoma CTCs, demonstrating gene expression profiles that were similar to melanoma cell lines and melanocytes, but different from leukocytes and non-melanoma cell lines. They also demonstrated that CTCs had low levels of expression of MHC class I genes which may provide a mechanism by which CTCs avoid immune detection. Analysis of gene expression of CRC CTCs also identified another method of immune avoidance with increased expression CD47, an antiphagocytic mechanism, noted [227].

In lung cancer a study assessed the amplification of CTCs following CellSearch enrichment and FACS sorting, noting 21% of the CTCs identified by CellSearch could be isolated and then amplified [228]. In 2013 a study by Ni et al examined CNAs and

mutations in CTCs isolated from patients with NSCLC and SCLC [229]. Analysis of samples from one patient with NSCLC including the primary tumour, liver metastases and CTCs identified 44 single nucleotide variants (SNVs) and indels amongst the CTCs of which all but 6 were present in the tumour biopsy samples. The study demonstrated that although there was heterogeneity of CNA profiles between patients, and between disease types, there was not heterogeneity amongst an individual patient's CTCs. Sequential profiling of one patient with SCLC revealed changes in mutations but not CNA patterns over time. In response to treatment with chemotherapy 23 mutations were significantly increased including 6 genes associated with ATP binding. Summaries of key single cell CTC studies of clinical samples are given in tables 1.3 and 1.4, with mutational profiling studies in table 1.3 and gene expression profiling studies in table 1.4. The highlighted single cell studies of CTCs confirm the potential utility of molecular profiling of CTCs to investigate cancers including SCLC, and hopefully to provide a method to monitor and personalise therapy for individual patients.

Cancer type	Methods	Observations
Breast, Prostate and GI [230]	Micromanipulation of CK or EpCAM-positive cells from blood, bone marrow and lymph nodes; CGH and single-stranded conformational polymorphism	Heterogeneity of chromosomal aberrations and *TP53* mutations in minimal residual disease
Melanoma [231]	Immunomagnetic enrichment using anti-MCSP antibody and micromanipulation of CMCs; Single-cell CGH	CGH of CMCs revealed hierarchical clustering by patient suggesting clonal origin
Breast [232]	Density gradient centrifugation and micromanipulation Array CGH and qRT-PCR	Heterogeneity of EGFR copy number
Colorectal [185]	CTCs isolated by density gradient centrifugation and with DEPArray; DNA Sanger sequencing and pyrosequencing	Heterogeneity of *KRAS* mutations between primary tumours, metastases and CTCs
Colorectal [225]	CellSearch enrichment and micromanipulation of CTCs; Sanger sequencing after amplification	Inter-CTC heterogeneity of driver mutations observed
Colorectal [226]	CellSearch enrichment and micromanipulation of CTCs; Array CGH and NGS	Multiple CNAs in CTCs; Mutations demonstrated in CTCs present in tumours at subclonal level
NSCLC, SCLC [229]	CellSearch enrichment and micropipetting of CTCs; NGS for CNA and WES	Heterogeneity of CNA patterns between patients and disease types but not within CTCs from an individual patient. Change in mutations over time in a patient's CTCs
Prostate [233]	Enrichment of CTCs on NanoVelcro chip and laser capture microdissection WES	Proof of principle experiment on use of the technology. Higher number of shared mutations between CTCs than between CTCs and WBCs
Breast [234]	Isolated CTCs with MagSweeper and micromanipulation; Targeted DNA preamplification and PIK3CA sequencing	Heterogeneity of PIK3CA mutation between CTCs and tumour cells. Change in proportion of CTCs with PIK3CA mutations over time in one patient
Prostate [235]	High definition CTC assay and manipulation for isolation NGS for CNA analysis	Change in predominant CNA CTC clone present with different treatments.
Breast [236]	CellSearch enrichment and FACS isolation Array CGH	CNAs typical of metastatic breast cancer identified in CTCs
Breast [237]	CellSearch enrichment and DEPArray isolation Sanger sequencing and Array CGH	Molecular heterogeneity of metastatic breast CTCs
Lung [24]	CellSearch enrichment and DEPArray isolation Low pass WGS and Sanger sequencing	CNA changes typical of SCLC identified in CTCs. Common mutations identified in CTCs and CDX from same patients samples. (Results chapter 4)

Table 1.3 Selected clinical CTC single cell mutational profiling studies (Adapted from Krebs et al 2014 [152])

Cancer type	Methods	Observations
Breast [238]	Isolated CTCs with MagSweeper and micromanipulation; High-throughput microfluidic-based qRT-PCR array	Demonstrated two major subgroups of CTCs, one with increased metastasis-associated genes and one with high expression of EMT associated genes
Melanoma [218]	Isolated CTCs with MagSweeper; Use of Smart-Seq for mRNA sequencing	Reproducible gene expression assay; Putative CMCs gene-expression profile similar to melanocytes and melanoma cell lines but distinct from WBCs
Prostate [239]	Isolated CTCs with MagSweeper and micromanipulation; RNA-Seq for gene-expression analysis	Despite RNA degradation, prostate-specific and cancer-specific transcripts have been identified
Prostate [184]	CTCs isolated with filtration and micromanipulation; High-throughput microfluidic-based qRT-PCR array	Heterogeneous upregulation of EMT-associated gene expression, particularly in castration-resistant disease
Lung, Prostate, Breast, Melanoma [163]	iChip enrichment and micromanipulation qRT-PCR	CTCs positive for EpCAM alone and EpCAM and Vimentin from single patient consistent with partial EMT
Colorectal [227]	CTC enrichment and micromanipulation Single-cell Array CGH and qPCR	Tumour and CTC matched CNA profiles Increased CD47 expression in CTCs as possible immune avoidance mechanism
Prostate [240]	Isolated CTCs with MagSweeper and micromanipulation; Low pass WGS, WES and RNA-Seq	Early trunk and metastatic trunk mutations from primary tumour and lymph node metastasis found in CTCs
Pancreatic, Breast, Prostate [241]	CD45 positive cell depletion and micromanipulation RNA-Seq	Highly expressed extra cellular matrix protein genes in CTCs

Table 1.4 Selected clinical CTC single cell gene expression profiling studies (Adapted from Krebs et al 2014 [152])

1.4 Perspectives

SCLC continues to have a very poor prognosis with the majority of patients surviving less than one year from diagnosis. Trials have failed to lead to sustained improvements for patients over the last 20 years outside of radiotherapy. This is in part associated with the rapid development of profound acquired chemoresistance that occurs in the majority of patients with SCLC. The use of targeted agents has failed to produce the same improvements in survival as seen in other tumour types, though this may be due to the lack of biomarkers for patient selection. Investigation of the molecular drivers of SCLC which could be targeted by new therapeutics has been hampered by the lack of tumour tissue available for research. The two landmark studies investigating the molecular landscape of SCLC from tumour tissue samples were only published in 2012 [70, 91]. In SCLC there remains, therefore, the significant potential to improve patient outcomes through new treatment approaches.

With the increasing focus on the personalisation of therapy in oncology there has been increased interest in liquid biopsies such as CTCs. In SCLC the role of CTCs as prognostic biomarkers has been established. There has been progress in other tumour types to use CTCs to investigate response and resistance to therapy and to inform on the biology of the diseases. However, there has been only one study published in 2013 which has attempted to use CTCs to molecularly profile SCLC [229]. The focus of this PhD thesis was therefore to establish the molecular profiling of SCLC CTCs, and to address whether CTCs reflect the molecular landscape seen when profiling tumour samples. The key issue of whether CTCs could be used to explore intrinsic and acquired resistance to chemotherapy which has the potential to lead to improvement in outcomes was also explored.

1.5 Overall Goal and Objectives of the Thesis

The overall goals of this thesis were to establish a robust protocol for the genomic profiling of single CTCs in patients with SCLC, and to examine the changes in genomic profiles in intrinsic and acquired resistance in SCLC.

1.5.1 Specific Objectives of the Thesis

1) To optimise an approach for the WES of single and pooled SCLC CTCs.

2) To genomically profile CDX and compare the profiles to SCLC CTCs isolated from the same patients' samples.
3) To investigate the use of TAm-Seq, a focussed NGS methodology, on SCLC CTCs.
4) To assess how SCLC CTC CNA and mutation profiles reflect published data from SCLC tumour samples.
5) To investigate if differences in baseline CTC CNA and mutation profiles can be identified between SCLC patients who are responsive and those who are resistant to first-line platinum-based chemotherapy.
6) To investigate if changes in CTC CNA and mutation profiles can be identified in patient samples taken at baseline and at the development of relapsed disease after treatment with platinum-based chemotherapy to inform on the development of acquired resistance.

Chapter 2: Materials and Methods

2.1 Patients and Clinical Samples

Patients with SCLC were consented to the CHEMORES (Chemotherapy resistance) study, a prospective biomarker programme for the investigation of blood-borne biomarkers in lung cancer (http://www.cancerresearchuk.org/about-cancer/trials/a-study-understand-more-why-chemotherapy-treat-lung-cancer-can-stop-working-chemores). The study is run at the Christie Hospital NHS Trust with Dr Fiona Blackhall as the principal investigator. The study had been prospectively approved by the NHS North West 9 Research Ethical Committee (REC reference 07/H1014/96). The study allowed the acquisition and analysis of CTCs, plasma and serum samples and tissue from diagnostic biopsies from patients with lung cancer, although no analysis of biomarkers apart from CTCs is presented in this thesis.

Patients who had treatment naïve histologically, or cytologically, confirmed SCLC were eligible for recruitment to the study. Patients needed be referred to the tertiary cancer centre for a treatment assessment and be able to provide full written consent. Exclusion criteria included patients in clinical trials of new investigational medicinal products unless there was an agreement that the patient could also consent to this study, and that their clinical data could also be used for this study. Clinical data such as age, smoking status and histological subtype was collected. Further information about clinical outcomes was also collected.

2.1.1 Clinical Data

Baseline data

1. Demographic data – date of birth, gender, ethnic origin, smoking status, height and weight.

2. Date of diagnosis.

3. Diagnostic pathology – specimen type and method of procurement e.g. endobronchial biopsy, fine needle aspirate of lymph node, histological subtype and immunoprofile.

4. Disease stage at study entry with sites of known metastases.

5. Haemoglobin, white blood cell count, platelet count, sodium, creatinine, lactate dehydrogenase, aspartate aminotransferase and alkaline phosphatase.

7. Performance status (WHO scale).

Follow-up data

Follow-up data was obtained from case notes until the date of death, including:

1. Type of treatment received; number of cycles and doses.

2. Response to treatment: progressive disease, stable disease, partial or complete response using RECIST criteria where possible.

3. Grade III and IV toxicities.

4. Febrile neutropaenia, neurological toxicity, auditory toxicity.

5. Date of progression/relapse.

6. Treatment for relapse.

7. Date and cause of death.

2.1.2 Sample Collection

Thirty millilitres of blood were taken at baseline and again when the patient relapsed post first-line therapy or subsequent lines of therapy. Two 10 mls vacutainers with CellSave preservative, one 10 mls vacutainer with EDTA added are taken at each time point. At baseline, prior to receiving chemotherapy, an additional 3 mls of blood was taken into an EDTA containing vacutainer. The blood samples were taken following standard operating procedures and good clinical practice guidelines. The results relating to the processing of the second CellSave sample and the 10ml EDTA sample will not be presented in this thesis. Surplus FFPE tissue from the initial diagnostic biopsy was also obtained for further genomic analysis if available.

2.2 CTC Enrichment and Isolation

2.2.1 CTC Enrichment and Enumeration

Blood samples were collected in 10ml CellSave preservative vacutainers and stored at room temperature for up to 96 hours. The blood samples were then processed using the CellTracks® AutoPrep® system, the CellSearch™ Circulating Tumour Cell kit and the CellTracks® Analyser II (Janssen Diagnostics) according to the manufacturer's instructions. Briefly this involved transferring 7.5 mls of blood into a conical tube and

then adding 6.5 mls Circulating Tumour Cell Kit Buffer. The conical tube was then centrifuged at 800g for 10 min. to fractionate the sample into the cell layer and buffer/plasma layer. The tube was then placed into the CellTracks® AutoPrep® system and processed. This involved an automated system in which the plasma/buffer layer was aspirated before the ferrofluid nanoparticles, which are coated with anti-EpCAM antibodies, were added to the cell layer and incubated. A magnetic field was applied and all the ferrofluid nanoparticle bound cells were attracted to the sides of the tube and the remaining unbound cells were aspirated. The magnetic field was then removed and the remaining cells were re-suspended in a buffer which contained a permeabilisation reagent and the fluorescently-labelled antibodies. The antibodies used were phycoerthrin-conjugated anti-cytokeratin (CK-PE) antibody which identifies epithelial cells, allophycocyanin-conjugated anti-CD45 antibody (CD45-APC) which identifies WBCs and 4'6'-Diamidino-2phenylindole (DAPI) which identifies the nuclei. The magnetic field was reapplied to allow aspiration of the excess staining reagents before the remaining cells were re-suspended in the MagNest™ cell presentation device. The MagNest™ device was then transferred to the CellTracks® Analyser II after a minimum of 20 min. incubation in the dark. The surface of the cartridge was scanned for each of the fluorescent filters and a gallery of images of the DAPI and CK-PE positive events were presented to the user. The number of CTCs per 7.5 ml blood was then assessed using the criteria that a cell is a CTC if it has the morphological features of a tumour cell, is greater than 4 µm in size, and is positive for CK and DAPI staining but negative for CD45.

2.2.2 CTC Isolation

After scanning on the CellTracks® Analyser II system the cartridges used during the CellSearch CTC enumeration were removed from the MagNest™ device and stored in the dark at 4 °C for further processing. Both single and groups of cells were isolated using the DEPArray system (Silicon Biosystems) according to the manufacturer's instructions. The CTCs and WBCs were initially aspirated from the CellSearch cartridge using gel loading tips pre-rinsed in 2% BSA in PBS solution. The CellSearch cartridge was then rinsed twice with 325 µl SB115 buffer. The sample was then centrifuged at 1000 g for 5 min in a swinging-bucket rotor centrifuge and the cells were washed in 1ml SB115 buffer and centrifuged again. The final sample was re-suspended in 14 µl of SB115 buffer before being loaded into the DEPArray cartridge in addition to 800 µl SB115 buffer. The sample was automatically injected into the

cartridge (9.26 µl) and the cells were organised into the electrical cages generated by the preprogramed electric field. The cartridge was then imaged using three fluorescent filter cubes (PE, APC and DAPI) in addition to bright field images. Images of cells were then presented to the user who selected cells of interest which were then routed automatically to the parking area. Individual cells or groups of cells were then moved into the recovery area and flushed into a 200 µl PCR tube with drops of buffer. The sample was then centrifuged for 10 min. at 14,100 g in a fixed rotor centrifuge before the addition of 100 µl of phosphate buffered saline (PBS). The sample was then centrifuged again at 14,100 g for 25 min. and the supernatant removed leaving a sample with a final volume of 1 to 2 µl. The samples were frozen at -80 °C prior to further processing.

2.3 Whole Genome Amplification of CTC DNA

Whole genome amplification (WGA) was performed using the Ampli1™ WGA kit version 01 (Silicon Biosystems) according to the manufacturer's instructions to generate a 50 µl WGA product. For the initial lysis step the sample was incubated at 42 °C with the proteinase K containing lysis reaction mix for a minimum of 15 hours before the proteinase K was inactivated by heating to 80 °C for 10 min. The DNA was then digested into smaller fragments at 37 °C for 3 hours using an MSEI restriction endonuclease with a final step of heating the sample to 65 °C for 5 min. The preannealing reaction mix was prepared and then for 51 min. dropping 1 °C per min. from 65 °C to 15 °C. The preannealing reaction mix was added to the sample and then incubated at 15 °C for a minimum of 12 hours to ligate the adaptors to the DNA fragments. The sample was then amplified in the primary PCR reaction as per table 2.1. The PCRs were performed using the Veriti thermocycler (Life technologies). For amplification of CDX DNA or genomic DNA from leukocytes 1 µl of DNA at 1 ng/µl concentration was used for WGA.

The WGA products were assessed by loading and resolving 1 µl on a 1 % agarose gel with GelStar™ nucleic acid gel stain (Lonza), a fluorescent DNA stain added. The WGA products were purified with the High Pure PCR Product Purification kit (Roche) which uses spin columns for the purification of DNA. The samples were processed according to the "Purification of PCR products in solution after amplification" protocol, eluting into a 50 µl final volume. The DNA concentrations were measured with the

Qubit fluorometer (Invitrogen) using the Quant-iT™ dsDNA HS assay kit as per the manufacturer's instructions, adding 1 µl of DNA.

Cycle Number	Temperature (°C)	Hold	Additional time and temp
1	68	3 min.	
14	94	40 sec.	*= +1 sec./cycle
	57	30 sec.	
	68	90 sec.*	
8	94	40 sec.	** = +1 °C/cycle *= + 1 sec./cycle
	57**	30sec.	
	68	1 min. 45 sec.*	
22	94	40 sec.	*= +1 sec./cycle
	65	30 sec.	
	68	1 min. 53 sec.*	
1	68	3 min. 40 sec.	

Table 2.1 Thermal cycling conditions for WGA primary PCR.

2.3.1 Whole Genome Amplification Quality Control Assay

The quality of the WGA product was investigated using the Ampli1™ QC kit (Silicon Biosystems) as per the manufacturer's instructions. A multiplex PCR to assess 4 amplicons of different lengths in chromosomes 5q, 6q, 12p and 17p was performed to assess the quality of the amplification product. For the PCR reaction 1 µl of the unpurified WGA product was added to a 10µl reaction and processed using the Veriti thermocycler. The initial PCR step was heating the sample to 95 °C for 4 min. followed by 32 cycles of 95 °C for 30 sec., 58 °C for 30 sec. and 72 °C for 90 sec. with a final elongation step of 72 °C for 7 min. The PCR products were then visualised on a 1.5% agarose gel containing a fluorescent DNA stain to enable comparison of the size and number of the obtained PCR products to the expected products. The total number of amplicons identified on the gel per sample were used to calculate the genome integrity index (GII) ranging from 0 to 4 which was used to indicate the success of the WGA [237].

2.4 DNA Extraction

CTC tumours (3 – 5 mm^3) were disaggregated using a sterile scalpel. DNA was then isolated using the QiaAmp DNA mini kit (Qiagen) as per the manufacturer's instructions using the DNA purification from tissues protocol. The sample was eluted into a final volume of 50 µl. DNA from 1 ml of whole blood was extracted using the QIAamp DNA blood mini kit (Qiagen) as per the manufacturer's instructions using the DNA purification from blood or body fluids protocol (spin protocol). The sample was eluted into a final volume of 50 µl. The DNA concentrations were measured with the Qubit fluorometer (Invitrogen) using the Quant-iT™ dsDNA BR assay kit as per the manufacturer's instructions, adding 1 µl of DNA.

2.5 DNA Library Preparation

The WGA products (250ng) were digested to remove the Ampli1 amplification primers with an MSE1 endonuclease (New England Biolabs) which digests DNA at the restriction-site motif TTAA. The samples were processed as per the manufacturer's instruction except the reaction was only incubated at 37 °C for 30 min. The sample was then immediately purified with the High Pure PCR Product Purification kit (Roche). The samples were processed according to the "Purification of PCR products in solution after amplification" protocol, eluting into a final volume of 100 µl. The samples were then sonicated with the Bioruptor UCD-200 (Diagenode) for 11 cycles (30 sec. on, 30 sec. off) to produce fragments of 300 – 350 bp in length. The DNA fragment length was checked by a Bioanalyzer (2100 Bioanalyzer, Agilent Life Sciences and Chemical Analysis) using the High Sensitivity DNA kit (Agilent Technologies). The DNA concentrations were measured with the Qubit fluorometer (Invitrogen) using the Quant-iT™ dsDNA HS assay kit as per the manufacturer's instructions, adding 1 µl of DNA.

DNA libraries were prepared using NEBNext® Ultra™ DNA Library Prep Kit For Illumina® with 50ng of DNA added per library preparation. The samples were processed according to the manufacturer's instructions. Briefly the initial NEBNext End prep reaction was carried out on the Veriti thermocycler with the samples incubated at 20 °C for 30 min. then incubated at 65 °C for 30 mins. In the adaptor ligation reaction mix the NEBNext adaptor for Illumina was diluted 1 in 10 as the input of DNA into the library preparation was less than 100ng. The reaction mix was otherwise prepared as

per the standard instructions and the samples incubated at 20 °C for 15 min. User enzyme was added to each sample and the samples incubated for 15 min. at 37 °C. The samples were then cleaned up using the Cleanup of Adaptor-ligated DNA without Size Selection protocol, using Agencourt® AMPure® XP magnetic particle solution (Beckman Coulter Company), solid phase reversible immobilization (SPRI) beads. For the final PCR amplification reaction each sample was barcoded with a unique index primer using NEBNext® Multiplex Oligos for Illumina® Index Primer sets 1 and 2. In the initial step of the PCR amplification the sample is heated to 98 °C for 30 sec. followed by 10 cycles in which the sample was heated to 98 °C for 10 sec. then 65 °C for 30 sec. and 72 °C for 30 sec. For the final step of the reaction the sample is heated to 72 °C for 5 min. The samples were then cleaned up using Agencourt® AMPure® XP magnetic particle solution as per the cleanup of PCR Amplification instructions.

The library concentrations were measured with the Qubit fluorometer (Invitrogen) using the Quant-iT™ dsDNA HS assay kit as per the manufacturer's instructions, adding 1 µl of DNA. The size distribution of the libraries was then checked using the 2100 Bioanalyzer with the High Sensitivity DNA kit. The libraries were also quantified using the KAPA Library Quantification Kit (KAPA Biosystems). The samples were processed as per the alternative protocol from the manufacturer in which a 10 µl reaction volume for the PCR rather than a 20 µl reaction volume was used. The cycling conditions used were an initial step of 95 °C for 5 min., followed by 35 cycles of 95 °C for 30 sec. and 60 °C for 45 sec. The PCR reactions were performed using a LightCycler® 96 Real-Time PCR system (Roche).

2.6 Next Generation Sequencing

Whole genome sequencing (WGS) of tumour tissue was carried out on the Illumina HiSeq2500® instrument (Illumina) using the TruSeq® PE cluster kit V2 and the TruSeq® SBv3 chemistry. The samples were initially clustered using the cBOT™ with the TruSeq® Rapid Duo cBOT Sample loading kit prior to sequencing. WGS of the CTCs, unamplified tumour DNA and amplified tumour DNA was carried out on the Illumina® MiSeq® desktop sequencer (Illumina) and the Illumina® NextSeq ™500 desktop sequencer (Illumina). The MiSeq reagent kit V2 (300 cycles) was used with the MiSeq® desktop sequencer with 11 to 15 samples multiplexed to be sequenced together. The NextSeq™ 500 Mid Output kits (300 cycles) and the NextSeq™ 500

High Output kits (300 cycles) were used with the NextSeq 500™ desktop sequencer with 24 samples multiplexed to be sequenced together. The sequencing reads were based called, filtered by quality metrics and aligned to the human reference genome (HG19) as per the manufacturer's recommendations. Whole exome sequencing (WES) was carried out on the Illumina HiSeq2500® instrument (Illumina) using the the TruSeq® Rapid PE cluster kit (HS kit) and the TruSeq® Rapid SBS kit (HS kit, 200 cycles).

2.7 PCR and Sanger Sequencing

Point mutations identified by NGS were confirmed by amplifying samples with gene-specific primers designed with OligoArchitect™ Online (http://www.sigmaaldrich.com/technical-documents/articles/biology/oligoarchitect-online.html). The list of the genes and the primer sequences are detailed in table 2.2 (used in chapter 4) and 2.3 (used in chapter 5). All the primers were validated using genomic DNA extracted from normal healthy volunteer blood. The PCR reactions were carried out in a total volume of 20 µL containing 4 – 10ng of template DNA, 1 unit of Taq DNA polymerase (Roche), 2.0µl of PCR reaction buffer with $MgCL^2$ 10X concentration (Roche), and a final concentration of 250 µM of each dNTP(Promega) and 200 nM of the both the specific forward and reverse primers. The PCRs were performed using the Veriti thermocycler. The initial step involved heating the sample to 95 °C for 2 min. before processing the samples for 35 cycles of 95 °C for 30 sec., 59 °C for 30 sec., 72 °C for 60 sec. and a final extension at 72 °C for 7 min. The PCR products were loaded and resolved on a 2.0% agarose gel with a fluorescent DNA stain added. The PCR products were purified using the QIAquick PCR purification kit (Qiagen) as per the manufacturer's instructions with a final elution volume of 30 µl.

Sanger sequencing of the samples was performed utilising the BigDye Terminator v3.1 cycle sequencing kit, a fluorescent dideoxy terminator sequencing method. It was performed on a 3130 xl Genetic Analyzer (Applied Biosystems) using the same primers used in the PCR reactions. ABI sequencing files were analysed using CHROMAS software (http://technelysium.com.au/?page_id=13) and publically available web based BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and alignment tools MEGA (http://www.megasoftware.net/).

Gene	Locus	Chr	Position	dbSNP_id	Reference	Report	Forward Primer Sequence	Reverse Primer Sequence
ALK	ALK_2	2	29416481	rs1881420	T	C	TAAACCAGGAGCCGTACGTT	AAAGAAACCCACAGCTGCAG
ATM	ATM_1	11	108175462	rs1801516	G	A	TCAAACTATTGGGTGGATTTGTT	GTGTCGAAGACAGCTGGTGA
EGFR	EGFR_1	7	55229255	rs2227983	G	A	TGCTGTGACCCACTCTGTCT	AACCTCCTACCCCTCCAGAA
ERBB2	ERBB2_4	17	37884037	rs61552325	C	G	TGGGTCACCTTCTCTTGACC	CAAGTACTCGGGGTTCTCCA
ERBB4	ERBB4_1	2	212251695		C	A	TTCCTCATCCAGCTCTCCTC	TACCGAGATGGAGGTTTTGC
FGFR2	FGFR2_1	10	123278286		C	T	TGTTTTGGCAGGACAGTGAG	TGTCTGTTCTAGCACTCGGG
FLT3	FLT3_1	13	28624294	rs1933437	G	A	TTGAATGATCACCTACGCAGT	AAAGGAGGAAAAAGTGCTTCA
HNF1A	HNF1A_3	12	121416622	rs1169289	C	G	TGCAAGGAGTTTGGTTTGTG	GGATCAGTGCCTCTTTGCTC
HNF1A	HNF1A_5	12	121437221	rs1169304	T	C	TATGCTCATCACCGACACCA	CTGGAGGCCTCAGTGTCTG
IDH1	IDH1_1	2	209108317	rs34218846	C	T	CCAACCCTTAGACAGAGCCA	GGTCATTTGGTTGTGGTGGG
JAK3	JAK3_1	19	17953950	rs55778349	G	C	GGCTCTCACCTGACAGTCTT	AACAGGGCTTGAAGTTGGGT
KIT	KIT_1	4	55593464	rs3822214	A	C	AGAGCAAATCCATCCCCACA	TGTCTCAGTCATTAGAGCACTC
MLH1	MLH1_1	3	37053568	rs1799977	A	G	TTTGCTGGTGGAGATAAGGTT	TCGACATACCGACTAACAGCA
RB1	RB1_1	13	49033826		T	TT	GCGATTTCATGATTTGAAAAA	ACTTGGTCCAAATGCCTGTC
RET	RET_1	10	43610119	rs1799939	G	A	TACCACAAGTTTGCCCACAA	CCTCACCAGGATCTTGAAGG
TP53	TP53_1	17	7577548	rs28934575	C	A	GGGTCAGAGGCAAGCAGA	CTTGGGCCTGTGTTATCTCC
TP53	TP53_2	17	7578490		A	C	CTCACAACCTCCGTCATGTG	CACTTGTGCCCTGACTTTCA
TP53	TP53_3	17	7579472	rs1042522	G	C	TTTTCTGGGAAGGGACAGAA	GTCCCAAGCAATGGATGATT
TP53	TP53_4	17	7578190		T	C	GAGAAAGCCCCCCTACTGC	AGCATCTTATCCGAGTGGAAGG
TP53	TP53_5	17	7577046		C	A	TGTCCTGCTTGCTTACCTCG	GCCTCTTGCTTCTCTTTTCCT
TP53	TP53_TS1	17	7578461		C	A	AGCTGCTCACCATCGCTA	CCAACTGGCCAAGACCT
TP53	TP53_TS2	17	7577548		C	A	GGGGTCAGAGGCAAGCAG	CTTGGGCCTGTGTTATCTCC
TP53	TP53_TS3	17	7578442		T	C	AGCTGCTCACCATCGCTA	CCAACTGGCCAAGACCT
TP53	TP53_TS4	17	7578490	.	A	C	TGTGCTGTGACTGCTTGTAG	TGCCCTGACTTTCAACTCTGT
TP53	TP53_TS5	17	7578536		T	C	TGTGCTGTGACTGCTTGTAG	TGCCCTGACTTTCAACTCTGT
TP53	TP53_TS6	17	7577093		C	G	TGTCCTGCTTGCTTACCTCG	GCCTCTTGCTTCTCTTTTCCT

Table 2.2 Sanger sequencing primers for chapter 4.

Gene	Chromosome	Position	Reference	Report	Forward Primer Sequence	Reverse Primer Sequence
BRAF	7	140534539	G	T	AATCTCTGGGGAACGGAACT	TTTGTTGGGCAGGAAGACTC
LRP1B	2	141135833	T	A	TTTTATCCCCAGAGCCACAG	CTTGCCGCACTTATTGGACT
LRP1B	2	141777669	G	T	TGAGAGAGGGCCTATTATTCACA	TTCCTATGGCCATCATTGGT
LRP5	11	68174189	G	A	TGCTGGGCTGTTGATGTTTA	CTTTGAGGCAGGAACAGAGG
PCDH10	4	134072249	T	A	ACGGTGAGGTCGTGTACTCC	GAAAGGCACGTCTCCCAGTA
TP53	17	7578536	T	C	GTTTCTTTGCTGCCGTCTTC	ACACGCAAATTTCCTTCCAC

Table 2.3 Sanger sequencing primers for chapter 5.

2.8 Reamplification of DNA Libraries

The development of a method for the reamplification of DNA libraries is described in chapter 3. The final method used for the reamplification of DNA libraries is described in this section. The reamplification PCRs were performed using 20 ng of DNA library in a 50 µl reaction containing Herculase II Rxn buffer (5x concentration), dNTPs (250 nM final concentration), PE1 and PE2 primers (5 µM final concentration), 1 µl Herculase II Fusion DNA polymerase and nuclease free water. The samples were incubated on the Veriti thermocycler with an initial step of heating to 98 °C for 30 sec. followed by 6 cycles in which the sample was heated to 98 °C for 10 sec. then 65 °C for 30 sec. and 72 °C for 30 sec. For the final step of the PCR the sample is heated to 72 °C for 5 min. The samples were then cleaned up using SPRI beads, as per the cleanup of PCR Amplification instructions in the NEBNext® Ultra™ DNA Library Prep Kit. The reamplified library concentrations were measured with the Qubit fluorometer using the Quant-iT™ dsDNA HS assay kit as per the manufacturer's instructions, adding 1 µl of DNA.

2.9 In-Solution Target Enrichment for WES

The optimisation of the method used for in-solution target enrichment for WES is described in chapter 3. The final method used for in-solution target enrichment is detailed in this section. The in-solution capture was performed using the SureSelect^{XT} Target Enrichment System for Illumina Paired-End Sequencing Library using version 1.5 of the protocol which was adapted for use with the amplified libraries. The Sure Select Human All Exon V5 baits were used for the target enrichment. Reamplified NEBNext® Ultra™ DNA Libraries were used as the input for the target enrichment rather than creating libraries using the SureSelect kit. For the target enrichment 187.5 ng of each of four samples with different indexes were multiplexed for the hybridization reaction. The four samples were completely lyophilized using the Eppendorf concentrator 5301, a vacuum concentrator, at 30 °C and then resuspended in 3.4 µl of nuclease free water to create the prepped DNA library mix. The hybridization reaction was then carried out as per the protocol with the creation of a hybridisation buffer from SureSelect Hyb # 1, 2, 3 and 4 reagents. Separately the RNase block dilution (1 in 4 dilution) was created and 2 µl of this were combined with 5 µl of the SureSelect Human All Exon V5 library to create the capture library mix. The 3.4 µl of the prepped DNA libraries were combined with the SureSelect block master mix and heated to 95 °C for

5 min. and then held at 65 °C. The hybridisation buffer was then separately heated to 65 °C for 5 min. and then held at 65 °C. The capture library mix was then separately heated to 65 °C for 2 min. before 13 µl of the hybridisation buffer was added. The entire prepped library mix was then added to the capture library and mixed well by pipetting. The PCR tubes were then sealed well before the sample was incubated at 65 °C for 24 hours in the Veriti thermocycler with the lid heated to 105 °C.

Following the 24 hour hybridisation incubation the samples were processed to select the hybridised capture DNA using Dynabeads MyOne Streptavidin T1 magnetic beads which had been prepared as per the protocol. The samples were processed as per the Select hybrid capture with SureSelect protocol. The captured libraries were then amplified with 14 µl of the captured on bead DNA amplified in a 50 µl reaction containing Herculase II Rxn Buffer (5 x concentration), 250 nM concentration of dNTPs, 1 µl Herculase II Fusion DNA polymerase, primers and nuclease free water. Unlike the SureSelect protocol the PE1 and PE2 primers (sequences in table 3.1) were used for the amplification with a final concentration of each primer of 5 µM in the PCR reaction. The libraries were already indexed so the addition of a further index in this reaction was not required. The PCR cycling conditions on the Veriti thermocycler were 98 °C for 2 min. followed by 12 cycles of 98 °C for 30 sec., 57 °C for 30 sec. and 72 °C for 1 min. followed by a final cycle of 72 °C for 10 min. The samples were then purified as per the protocol using Agencourt AMPure XP beads. The final enriched samples were quantified post purification using the Qubit fluorometer and assessed using the 2100 Bioanalyzer with the high sensitivity DNA assay chip. The libraries were also quantified using the KAPA Library Quantification Kit as described in section 2.5 and stored at -80 °C prior to dilution for sequencing.

2.10 Bioinformatics and Statistical Considerations

The bioinformatics analyses were performed with Dr Yaoyong Li and Dr Hui Sun Leong (bioinformaticians in the RNA Biology Group/ Computational Biology Group).

2.10.1 WGS Analysis of Tumours

Reads from the Illumina HiSeq runs were aligned using SMALT (http://smalt.sourceforge.net/ with default strategies:

http://www.sanger.ac.uk/resources/software/smalt/) to the human genome (version hg19). To enable the identification of potential contaminating reads of murine origin, the same data was also aligned to the mouse reference genome (version mm9). Any reads which aligned to both the human and mouse genomes were then discarded. Single nucleotide variants (SNVs) and short indels were identified within each sample from aligned paired-end reads using GATK [242]. Duplicate read removal, realignment around known indels, base- and variant quality score recalibration [243] were performed as pre- and post-processing. Variant calling was performed using unifiedGenotyper with default settings (https://www.broadinstitute.org/gatk/guide/tooldocs/org_broadinstitute_gatk_tools_walkers_genotyper_UnifiedGenotyper.php). The putative SNVs and indels identified by GATK were then annotated using ANNOVAR [244].

2.10.2 Copy Number Alteration (CNA) Analysis of WGA Products

Illumina MiSeq® and NextSeq® whole genome data for CTCs, WBCs and tumours were aligned to the human genome using SMALT. Copy number alterations in the HiSeq®, MiSeq® and NextSeq™ data were identified using FREEC [245] with the adaptive window size. Mappability data for HG19 with an edit distance of 1 were downloaded from the FREEC web site (http://bioinfo-out.curie.fr/projects/freec/). An estimated copy number was assigned to every cytoband of the human genome (version hg19) in each sample. The weighted mean of the overlapping copy number estimates (as computed by FREEC) that map to the given cytoband was calculated and passed to Circos [246]. FREEC predicted copy number data was averaged across cytobands, as before, and both median and predicted copy number data were imported into MeV [247] to generate the Principal component analysis (PCA) data (median centering mode with recommended MeV algorithm). The copy numbers were also mapped to genome coordinates using the Bioconductor package annmap to provide ENSEMBL version 70 annotation [248] and clustered in MeV using Pearson Correlation average linkage.

2.10.3 Identification of Cancer-Related Genes

On 3rd February 2014 the geneRIF database was downloaded from the NCBI web site (ftp://ftp.ncbi.nlm.nih.gov/gene/GeneRIF/). All human genes with RIF text containing at least one of the 10 key words or word stems ("carcinogen", "cancer", "carcinoma",

"tumor", "leukemia", "tumour", "oncogen", "leukaemia", "oncolog", "malignan") were considered cancer-related. In total 6,682 cancer-related protein-coding genes were identified by mapping ENSEMBL gene ID and gene symbol in the geneRIF database to ENSEMBL (v70). Only autosomal chromosomes were used in CNA analysis so 268 genes on the X, Y or mitochondrial chromosome and were removed from the list leaving a total of 6414 cancer-related genes.

2.10.4 Quality Control Analysis of Copy Number Alteration Data

During the evaluation of the initial WGA/CNA approach, matched single and pooled WBCs and CTCs were analysed. Six WBC samples and six CTC samples were subjected to WGA and NGS with the resultant Illumina MiSeq data analyzed for CNA at the cytoband and cancer-related gene level. From this evaluation, a small number of potentially unreliable loci (0.8% for cytobands and 1.1% for cancer genes) with reported loss or gain in at least 3 WBC samples were identified. These loci were subsequently removed from the further analysis of CNA data from the CDX and CTC samples, which reduced the number of cancer-related genes analysed to 6,341 (Chapter 4). The loci identified as unreliable are documented in table 2.4. When the larger cohort of CTCs and WBCs were analysed the same analysis for potentially unreliable loci was expanded to use the 48 WBC samples from this sample set (Chapter 5). A locus was called potentially unreliable if it was lost or amplified in more than 16 of the 48 samples (greater than 33%). In total 0.98% of the cytobands, 0.84% of cancer-related genes and 0.96% of protein-coding genes were removed from further analysis. The loci identified as unreliable are documented in table 2.5.

2.10.5 LIMMA Analyses of Copy Number Data

The bioconductor package LIMMA (version 3.20.9) was used to analyse CNA data generated from the amplified CTCs [249]. When comparing CTCs isolated from the chemoresponsive and chemorefractory patients the median ratio CNA data generated from the 19336 protein-coding genes was used as the input for the analysis. Genes that had significant difference in copy number in the CTCs from the two patient groups were identified as those with an absolute fold change of greater than 1.5, a difference in median copy number between the two groups using a false discovery rate (FDR) of 5 %. The method of Benjamini and Hochberg was used to estimate the FDR and correct for multiple hypotheses testing [250]. The same analysis was used to identify the

differences in copy number between the CTCs isolated at relapse from the chemoresponsive patients and the CTCs from the chemorefractory patients. When contrasting the CTCs isolated at the baseline and relapse time point samples from the chemoresponsive patients the analyses were carried out on an individual patient basis using the same criteria.

	Cytoband	**Cancer genes**
Total	811	6,414
Number of the excluded	7	73
Excluded cytobands or cancer genes	2q12.3; 4q13.2; 6p11.2; 9q21.12; 11p15.5; 16q11.2; 18p11.32	ADCYAP1; ARID3A; ASCL2; AXIN1; BRSK2; BSG; C19orf6; CACNA1H; CD151; CD81; CDHR5; CIRBP; CTSD; ELANE; FSTL3; GNRHR; GPX4; GZMM; HBA1; HBA2; HINT1; HMHA1; HRAS; IFITM1; IFITM3; IGF2; INS; IRF7; KCNQ1; KISS1R; KLF9; LSP1; MBD3; MOB2; MPG; MSLN; MUC2; MUC5AC; MUC5B; MUC6; MUM1; ORC6; PDIA2; PIDD; PKP3; PLK5; PNPLA2; POLR2E; PRTN3; PTBP1; RAB11FIP3; RASSF7; REEP6; RHBDF1; RNF126; RNH1; RPLP2; SCT; SIRT3; SOX8; SSTR5; STK11; STUB1; TALDO1; TCF3; TH; THOC1; TNNI2; TSPAN4; TYMS; USP14; VPS35; YES1

Table 2.4 Cytobands identified as potentially unreliable and removed from CNA analysis in chapter 4.

	Cytoband	Cancer-related genes	Protein-coding genes
Total	811	6,414	19523
Number excluded	8	54	187
Excluded cytobands or cancer genes	2q12.3; 2q21.2; 4q13.2; 6p11.2; 9q21.12; 17q23.1; 18p11.32; 19q11	AGBL2; ARID3A; BIRC6; BSG; C19orf6; C1QTNF4; CACNA1H; CDC27; CEBPD; CELF1; CIRBP; CLTC; ELANE; FSTL3; GPR39; GPX4; GZMM; HINT1; HMHA1; KIAA0146; KISS1R; LTBP1; MCM4; MSLN; MUM1; NDUFS3; PLK5; POLR2E; PPM1D; PRKDC; PRR11; PRTN3; PTBP1; PTPMT1; PTPRJ; PTRH2; RAD51C; REEP6; RNF126; RPS6KB1; SOX8; SSTR5; STK11; STUB1; THOC1; TPSAB1; TPSG1; TRIM37; TYMS; UBE2V2; USP14; USP32; VMP1; YES1	ABCA7; AC002558.2; AC004528.1; AC005329.1; AC009041.2; AC027307.3; ADAMTSL5; AGBL2; ANKRD30BL; APC2; APPBP2; ARID3A; ATP5D; AZU1; BIRC6; BSG; C16orf13; C17orf64; C18orf56; C19orf21; C19orf24; C19ORF24; C19orf25; C19orf26; C19orf6; C1QTNF4; C1QTNF8; C2CD4C; C9orf135; CA4; CACNA1H; CCDC78; CDC27; CDC34; CDC42SE2; CEBPD; CELF1; CENPC1; CETN1; CFD; CHTF18; CIRBP; CLTC; CLUL1; CNN2; COLEC12; CTD-2510F5.6; DAZAP1; DHX40; DUX4; DUX4L2; DUX4L3; DUX4L4; DUX4L5; DUX4L6; DUX4L7; EDAR; EFNA2; ELANE; ENOSF1; FAM173A; FAM180B; FAM195A; FBXL16; FGF22; FNBP4; FRG1; FRG2; FRG2C; FSTL3; GAMT; GDPD1; GNG13; GPR39; GPX4; GRIN3B; GZMM; HAGHL; HCN2; HEATR6; HINT1; HMHA1; HRASLS5; HYDIN; JMJD8; KBTBD4; KBTBD4; KIAA0146; KISS1R; LMF1; LPPR3; LTBP1; LYRM7; MADCAM1; MAMDC2; MCM4; MED16; METRN; MIDN; MIER2; MSLN; MSLNL; MTCH2; MUM1; MYL4; NARFL; NDUFS3; NDUFS7; NUP160; ODF3L2; OR4A47; OR4B1; OR4C3; OR4C5; OR4F17; OR4S1; OR4X1; OR4X2; PALM; PCSK4; PLK5; POLR2E; POLRMT; PPAP2C; PPM1D; PPM1E; PRKDC; PRR11; PRR25; PRSS57; PRTN3; PTBP1; PTPMT1; PTPRJ; PTRH2; R3HDM4; RAB40C; RAD51C; REEP6; RHBDL1; RHOT2; RNF126; RNFT1; RP11-15E18.4; RP11-178C3.1; RP11-683L23.1; RPS15; RPS6KB1; RPUSD1; SBNO2; SH3RF3; SHC2; SKA2; SLC22A10; SLC22A24; SLC22A25; SLC22A9; SMC5; SMG8; SOX8; SSTR5; STAP1; STK11; STUB1; TEX14; THEG; THOC1; TPGS1; TPSAB1; TPSD1; TPSG1; TRIM37; TTC27; TUBD1; TYMS; UBA6; UBE2V2; USP14; USP32; VMP1; WDR18; WDR24; WDR90; WFIKKN1; YES1; YPEL2; ZNF717

Table 2.5 Cytobands identified as potentially unreliable and removed from CNA analysis in chapter 5.

2.10.6 Whole Exome Sequencing Analysis

WES data from the Illumina Hiseq® reads for the CTC, WBC and tumour samples were aligned to the human genome (hg 19) using SMALT with default strategies. SNVs and short indels were identified from the aligned paired-end reads using GATK for each sample covering the target regions of the enrichment protocol. Duplicate read removal, realignment around known indels, base and variant quality score recalibration were then performed as both pre and post processing. Variant calling was performed using unifiedGenotyper with the default settings. Putative SNVs and Indels were then annotated using ANNOVAR. Of the 34 samples on which WES was performed 11 were WBC samples. Any SNVs or Indels identified in the WBC samples were removed from further analyses of all the samples to reduce possibility of including data that resulted from amplification and sequencing errors.

2.10.7 Statistical Analyses

Statistical analyses were performed using GraphPad Prism version 6 for Windows, (GraphPad Software, California USA, www.graphpad.com). Comparison of CTC numbers, progression free survival times and overall survival times between the chemorefractory and chemoresponsive patients were performed using the Mann-Whitney U test. Comparison of the CTCs numbers at baseline and when patients relapsed after first-line chemotherapy was performed using the Wilcoxon matched pairs signed rank tests. Probability values of $p \leq 0.05$ were considered significant and were two sided.

2.11 Summary of Methods for the Processing of CTCs Isolated from Patients with SCLC

A summary of the workflow used for the processing of CTCs isolated from patients with SCLC is detailed in figure 2.1.

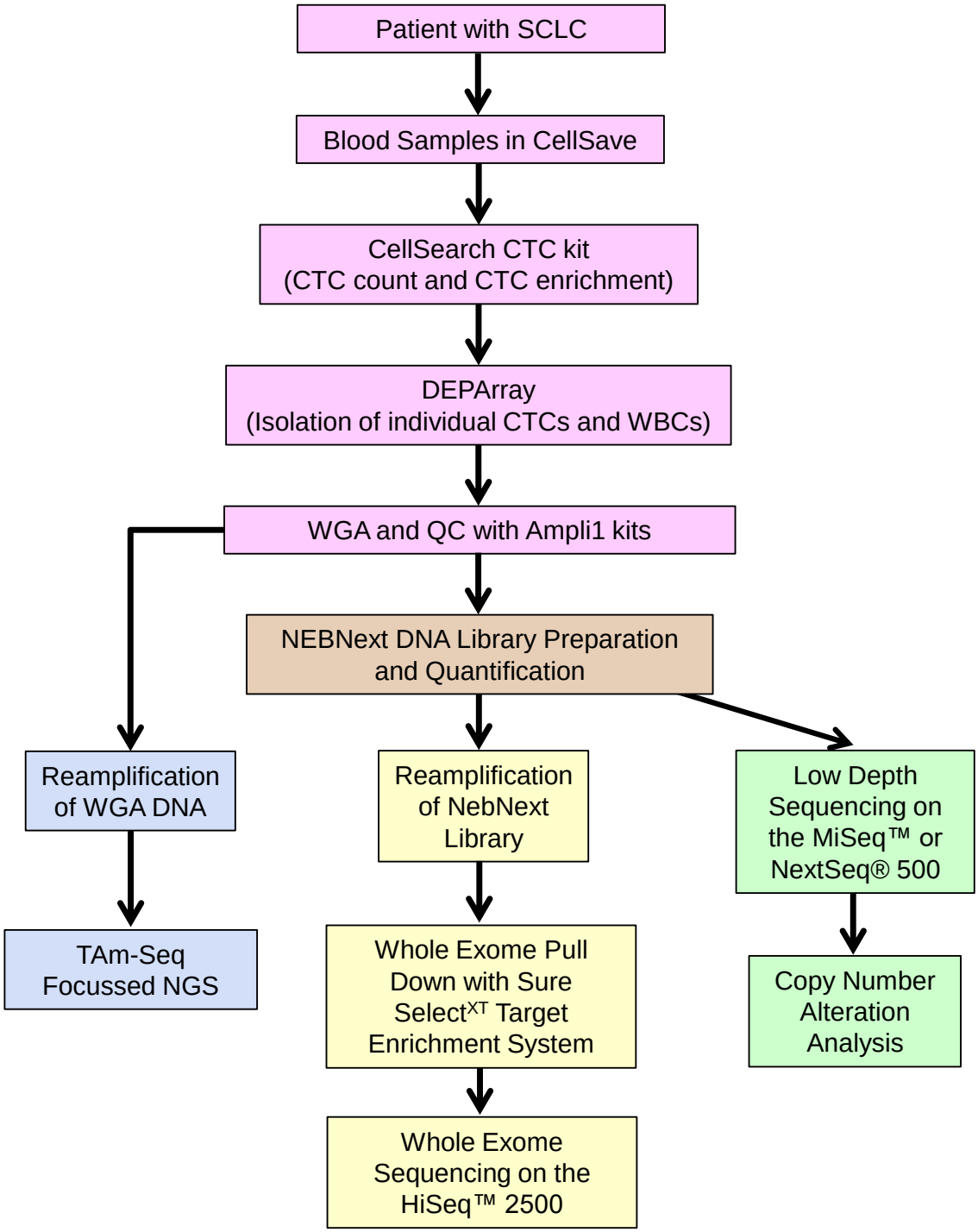


Figure 2.1 Summary for the workflow for the genomic profiling of CTCs.

Chapter 3: Optimisation of Workflow for the Reamplification of NGS Libraries and WES of CTCs

3.1 Introduction

NGS platforms, which use massively parallel sequencing techniques, have been commercially available since 2005 [251]. The initial sequencing of the human genome by Sanger sequencing required a multicentre collaborative approach through the Human Genome project [252]. It cost an estimated $10 – 25 million to perform and took several years. NGS in contrast enables the rapid generation of large quantities of sequencing data in a short space of time with sequencing of the whole genome, exome or transcriptome possible. The costs of NGS are lower per base than Sanger sequencing and have continued to fall allowing expanded roles for these techniques, though the costs still engender some limitations to its use [253]. The insights into the genomic architecture of cancers from NGS has transformed cancer research and these insights are increasingly being transferred into the clinical management of patients [254].

The first example of WGS in lung cancer was performed on two cell lines in 2008 and revealed 306 germline mutations and 103 somatic changes[255]. There have been multiple examples of WGS of lung cancer since, including the Cancer Genome Atlas in which squamous cell carcinoma of the lung was identified as having the highest frequency of mutations from the 12 cancer types studied [77]. WES enables detailed sequencing of transcribed regions of the genome whilst having reduced costs compared to achieving the same depth of WGS. WES is particularly relevant for investigating diseases as more than 85% of known disease-causing mutations are located within the protein-coding regions of the genome [256]. In the study of lung cancer WES of tumour tissue has been used to investigate frequently mutated genes in adenocarcinomas, squamous cell carcinomas and SCLC [70, 91, 257-260]. TP53 was confirmed as being frequently mutated in all these subtypes of lung cancer but other genes such as MLL and MLL2, histone methyltransferases were also noted to be mutated in all the histological subtypes. These studies highlight the potential utility of WES for the investigation of the biology of lung cancer but are limited by the need for tumour biopsies.

The genomic analysis of CTCs would allow the investigation of molecular changes in SCLC without the necessity for multiple biopsies. In contrast to the bulk profiling of biopsies, the profiling of individual CTCs provides the possibility of exploring the

heterogeneity that is increasingly known to be present in tumours at a single cell level [144]. WES of CTCs would allow the exploration of the subclones of cells in the circulation which are potentially producing metastases. The use of targeted agents is increasingly important in the treatment of cancers [261]. Targeted agents are designed to exploit the molecular abnormalities within cancers, but the degree of heterogeneity of these abnormalities may influence patients' responses. WES of CTCs would therefore enable the investigation of these molecular abnormalities at a single cell level and allow the change in mutations to be monitored serially over the course of a patient's treatment. At the time of commencing this research there were no published examples of CTC WES. Subsequently a small number of research articles using WES of CTCs have been published with examples including CTCs from both prostate and lung cancer [229, 233, 240]. In two patients with prostate cancer WES of CTCs and the previously resected primary tumours was performed by Lohr et al, with the identification in the CTCs of early trunk mutations present in the primaries, highlighting the potential utility of WES of CTCs [240]. WES of CTCs in lung cancer identified large similarity with the metastatic but not the primary tumour of a patient with lung cancer, supporting the hypothesis that CTCs are the subgroup of cells responsible for metastases [229].

The establishment of a technique for the WES of CTCs from patients with SCLC was one of the major aims of this thesis. WES requires samples to be enriched for the exonic regions in preference to the intronic regions using a target enrichment strategy, to enable increased depth of sequencing of the exome at lower costs. There are multiple methods for target enrichment including PCR amplification, molecular inversion probes, hybrid capture (in-solution capture or microarray based genomic selection) [262, 263]. Whole exome capture using the SureSelect^{XT} Target Enrichment System for Illumina Paired-End Sequencing Library, a method for in-solution target enrichment, was well established in the core facilities at the CRUK Manchester Institute. This technique was therefore selected as the method for exome enrichment to be optimised for CTC WES. Although cheaper than WGS, for WES there are costs associated with both the target enrichment and the sequencing. Monitoring changes in patients' CTCs over the course of their therapy requires multiple samples to be analysed per patient. Adapting methods to allow the simultaneous processing of multiple samples would increase the number of samples that could be analysed per patient, whilst also reducing the costs associated with target enrichment. Optimisation of the workflow associated with the in-solution target enrichment to fit in with the

established workflow for CNA of CTCs and to reduce overall costs was therefore undertaken.

WES of CTCs is complementary to the data generated from CNA assessment of CTCs. It can provide focussed data about individual SNVs and Indels to contrast to the global view of changes in copy number provided by CNA analysis. The generation of CNA and WES data from a single cell could potentially provide the greatest amount of information. This is because it would allow the coexistence of CNA changes and mutations to be analysed. Unfortunately, even when single cells are WGA, due to the limited DNA available and the large input of DNA needed for WES it can be very difficult to generate sufficient DNA for both CNA and WES analyses. Maximising the amount of information that can be generated from a single CTC reduces the number of blood samples patients must be subjected to, and minimises the costs associated with CTC enrichment, isolation and amplification. The development of a method to reamplify the WGA DNA from single CTCs for further analysis were therefore required to enable CNA and WES of single CTCs from patients with SCLC. As DNA polymerases introduce errors during PCR [264], it was important to develop methods of reamplification which limit the potential for the introduction of these errors.

3.2 Aims

1) To develop and evaluate methods for the reamplification of WGA DNA and WGA NGS libraries suitable for WES of CTCs.

2) To establish a method for the in-solution capture for target enrichment to enable WES of CTCs.

3.3 Methods

General methods are detailed in chapter 2. Figure 2.1 provides an overview of the CTC processing workflow used. Details on individual processes can be found as follows

1) Patient and clinical sample collection - see section 2.1.
2) CTC enrichment and isolation – see section 2.2.
3) WGA of CTC DNA - see section 2.3.
4) DNA library preparation - see section 2.5.

5) Final method for reamplification of DNA libraries - see section 2.8.
6) Final method for in-solution capture method for exon target enrichment - see section 2.9.
7) Next generation sequencing - see section 2.6.
8) Bioinformatics and statistical considerations – see section 2.10.

3.4 Results

3.4.1 Method Development for Reamplification of DNA for WES

One of the aims of this thesis was to enable CNA analysis and WES of the same single CTC to be performed. Methods to reamplify the WGA DNA to generate sufficient DNA as input for both these techniques were therefore required. DNA could be reamplified at two possible points during the CTC processing workflow to generate sufficient DNA for WES. The purified WGA DNA or the DNA libraries could both be sources of DNA for reamplification. A comparison of the use of DNA reamplified at both these time points in the workflow was therefore undertaken.

3.4.1.1 Reamplification of WGA DNA

To assess reamplification of WGA DNA four DNA samples (three tumour DNA samples and one pool of WBCs) were WGA as per the method in section 2.3. These samples were then reamplified using the Ampli1™ WGA kit (Silicon Biosystems) according to the re-amplification of Ampli1™ WGA product protocol. Briefly 1 µl of the purified WGA product was reamplified using the Ampli1™ enzyme and buffer reamplification mix in a total volume of 50 µl. The samples were then incubated on the Veriti thermal cycler. An initial cycle of 94 °C for 60 sec., then 60 °C for 30 sec. and then 72 °C for 120 sec. was followed by 10 cycles of 94 °C for 30 sec., 60 °C for 30 sec. and 72 °C for 120 sec. (increasing by 20 sec./cycle). The PCR amplicons were visualised on a 1% agarose gel with GelStar™ nucleic acid gel stain added (data not shown). For each sample the expected smear of DNA from 300 to 1200 bp in size was seen. The DNA concentrations were measured with the Qubit fluorometer using the Quant-iT™ dsDNA HS assay kit as per the manufacturer's instructions, adding 1 µl of DNA. The total yield of DNA ranged from 1.44 µg to 2.20 µg. For each sample 0.5 µg was then digested using an MSE1 endonuclease as detailed in section 2.5. Both MSE1 digested DNA and undigested DNA (250ng) were then sonicated and 50ng used to make DNA libraries using the NEBNext® Ultra™ DNA Library Prep Kit.

3.4.1.2 Reamplification of DNA libraries

To assess the effects of reamplification of the DNA libraries NEBNext™ Ultra DNA libraries made from the same four WGA samples (three WGA tumour DNA samples and one WGA pool of WBCs) as used in section 3.4.1.1 were reamplified. The DNA libraries were reamplified using NEBNext High fidelity PCR Master Mix (New England Biosciences). The total PCR reaction volume was 50 µl to which 1 µl of DNA library was added with 25 µl of NEBNext High fidelity PCR Master Mix, nuclease free water and forward and reverse primers. The forward and reverse primers used were the PE1 and PE2 primers used in the final PCR amplification in the NEBNext® Ultra™ DNA Library Prep Kit (sequences in table 3.1). Three final concentrations of PE1 and PE 2 primers (1 µM, 5 µM and 10 µM) were investigated. The samples were incubated on the Veriti thermal cycler with an initial step of 98 °C for 30 sec. followed by 15 cycles in which the sample was heated to 98 °C for 10 sec. then 65 °C for 30 sec. and 72 °C for 30 sec. For the final step of the reaction the sample is heated to 72 °C for 5 min. The samples were then cleaned up using Agencourt® AMPure® XP magnetic particle solution, a type of SPRI beads, as per the cleanup of PCR Amplification instructions in the NEBNext® Ultra™ DNA Library Prep Kit.

Primer Name	Primer Sequence
PE1	AATGATACGGCGACCACCGAGATCT
PE2	CAAGCAGAAGACGGCATACGAGAT

Table 3.1 PE1 and PE2 primer sequences

The DNA concentrations of the reamplified libraries were measured with the Qubit fluorometer adding 1 µl of DNA. The size distribution of the reamplified libraries was also checked using the 2100 Bioanalyzer with the High Sensitivity DNA kit. These results were compared to the Qubit fluorometer and Bioanalyzer results from the original DNA libraries for these samples (table 3.2). There was a trend to a higher concentration of DNA libraries produced from using 5 and 10 µM concentrations of the primers compared to using 1 µM concentration. There was also a trend to a slightly higher bioanalyzer fragment size for the reamplified samples than the original samples. As there was a trend to higher concentrations of DNA libraries with higher concentrations of the primers, and as the in-solution capture mechanisms require up to 1µg of DNA, the final concentration of 10µM for primers was chosen to use in subsequent reamplifications.

Samples name	Concentration of primer in reamplification reaction (µM)	Concentration of DNA library (ng/µl)	DNA yield in 28 µl library (ng)	Bioanalyzer average fragment size (bp)
Original WBC	NA	15.7	438.5	346
Original T1	NA	20.4	571.2	337
Original T2	NA	9.1	254.2	321
Original T3	NA	10.1	282.8	340
WBC	1	26.2	733.6	402
T1	1	26.6	744.8	420
T2	1	24.8	694.4	410
T3	1	25.8	722.4	403
WBC	5	26.4	739.2	389
T1	5	29.2	817.6	415
T2	5	30.0	840.0	394
T3	5	27.2	761.6	386
WBC	10	35.0	980.0	389
T1	10	27.6	772.8	435
T2	10	29.6	828.8	362
T3	10	26.8	750.4	438

Table 3.2 Comparison of the effect of primer concentration on the yield of DNA when reamplifying NGS libraries. DNA libraries for three WGA tumour samples (T1, T2 and T3) and one WGA pooled WBC sample (WBC) were reamplified using NEBNext high fidelity PCR enzyme and three different final concentrations of PE1 and PE2 primers (1, 5 and 10 µM). The concentrations of the final libraries were measured using the Qubit fluorometer and the average fragment sizes were measured using a bionanalyzer. These results were then compared to the original DNA libraries (original WBC, T1, T2 and T3, shaded in grey) which had not undergone further reamplification.

3.4.1.3 Comparison of Reamplification of WGA DNA and DNA Libraries

To compare the effect of the two methods of reamplification, reamplification of the WGA DNA and reamplification of DNA libraries, on downstream analyses a comparison of CNA data generated from the different sample types was undertaken. DNA reamplified from WGA DNA which had either undergone MSE1 digestion or had no digestion prior to DNA library preparation was compared to reamplified NEB libraries and the original samples. The preparation of these samples is described in sections 3.4.1.1 and 3.4.1.2. The samples are summarised in table 3.3. Note samples T2 and T3 originate from the same patients samples. The DNA libraries were quantified, multiplexed and sequenced on the Illumina MiSeq® as described in chapter 2 (section 2.6). The CNAs in 6,414 cancer-related genes were then calculated as per the method in section 2.10.2. Hierarchical clustering of the CNA results of the reamplified and original samples was then generated (figure 3.1).

Sample Type	Processing Method			
	Original sample	Reamplifed WGA DNA with MSE1 digestion	Reamplified DNA without MSE1 digestion	Reamplified NEB libraries
WGA pooled WBC	W	W_A	W_B	W_C
WGA Tumour 1	T1	T1_A	T1_B	T1_C
WGA Tumour 2	T2	T2_A	T2_B	T2_C
WGA Tumour 3	T3	T3_A	T3_B	T3_C

Table 3.3 Samples used in CNA analysis to compare methods of reamplification. Summary of samples used for CNA analysis describing the sample type and processing methods used for reamplification. The names given to each sample in the subsequent analysis are detailed in the table.

Each sample type, irrespective of the reamplification method used, had an overall unique CNA pattern with, for example, minimal copy number changes seen in the WBC sample in comparison to the three tumours. Between the three tumours different CNA patterns were seen with T1 having less regions of amplification than T2 or T3. There were some subtle differences seen between the samples reamplified with different techniques, for example there is a small region of amplification seen in the pair of reamplified WGA samples of T3 that is not seen in the original sample or the reamplified DNA library (highlighted with green arrow). When examining the hierarchical clustering of the samples, each sample clusters together with the other samples from the same original source irrespective of the method of reamplification used. In all cases the reamplified DNA library groups in the same branch of the dendrogram as the original libraries, unlike the reamplified WGA DNA samples. To try to assess the degree to which samples have been altered from the original samples through the reamplification process a T-test was used to compare the samples looking for cytobands that have significantly different copy numbers when compared to the original dataset. For the reamplified WGA DNA 2.2% of the cytobands differed from the original results if the samples were subjected to MSE1 digestion and 3.0% were significantly different for the samples that underwent no MSE1 digestion. In comparison just 0.2% of the reamplified cytobands differed significantly when comparing the original CNA analysis to the analysis of the reamplified DNA libraries. This suggests that the reamplified DNA libraries most closely reflect the original DNA libraries, although it is accepted that this is based on the results of only one experiment. One of the key aims of the reamplification process was to enable CNA

analysis and WES to be performed on the same CTC to allow comparison of the two results. It is therefore important the two methods are performed on the most similar input material to allow comparisons to be made. For this reason and due to the fact it would also reduce the costs associated with a second DNA library preparation it was decided that reamplification of the DNA libraries was the most appropriate workflow.

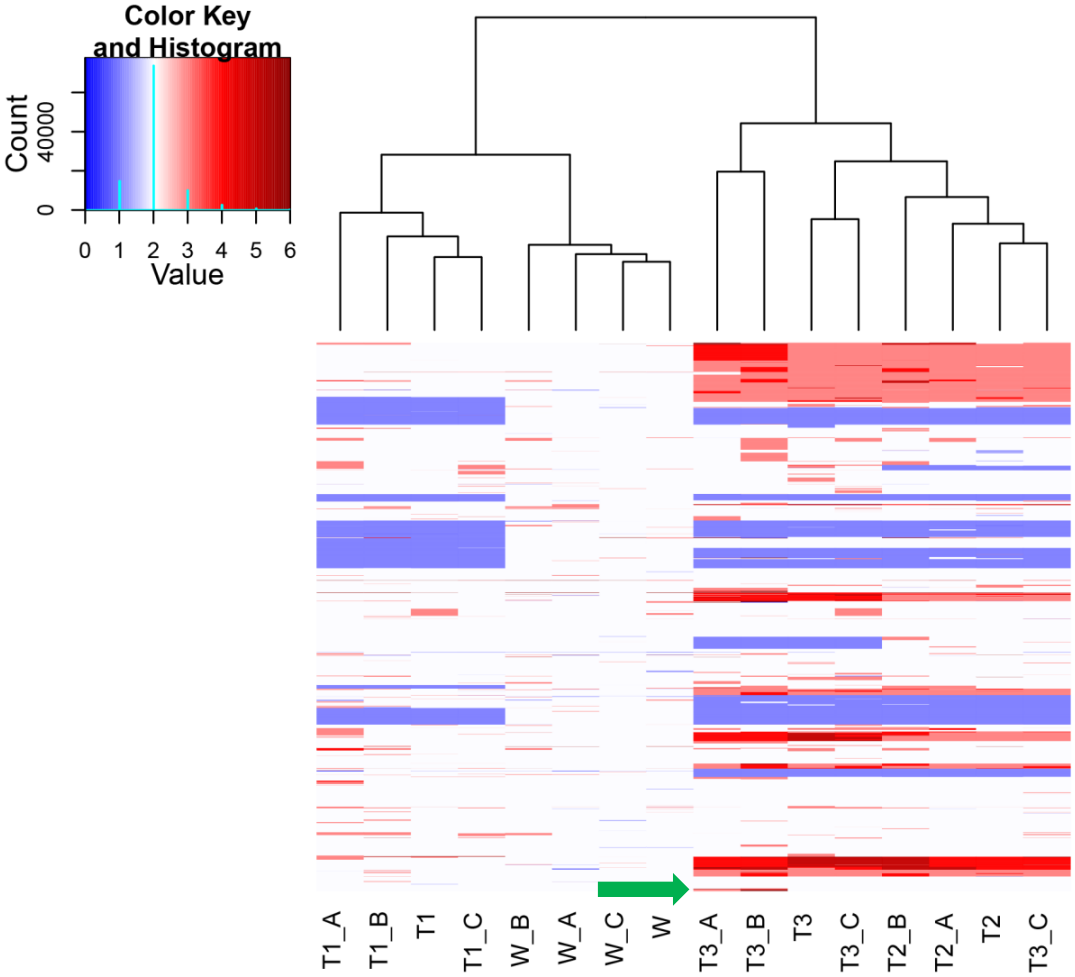


Figure 3.1 Hierarchical clustering of CNA data comparing methods of reamplification. A comparison of the effect of reamplification methods on CNA analysis of 4 samples was performed. Three WGA tumour DNA samples (T1, T2 and T3) and a WGA pooled WBC sample (W) were reamplified according to the re-amplification of Ampli1™ WGA product protocol. This was followed with (A) or without (B) MSE1 digestion and DNA libraries were prepared. The DNA libraries of the original samples were also reamplified using NEBNext high fidelity PCR enzyme. The eight reamplified WGA samples were multiplexed with four reamplified DNA libraries (C) and sequenced on the Illumina Miseq®. The CNA analysis from the 12 reamplified samples was compared to the CNA data from the original samples. Regions of amplified copy number are red whilst regions of loss are denoted in blue and regions with no change are coloured white. Hierarchical clustering of the CNA data for the 16 samples was then performed. Note samples T2 and T3 come from the same patient's samples. The green arrow highlights a region of amplification seen in the reamplified WGA DNA samples produced from T3 which was not seen in the original or reamplified DNA libraries.

3.4.1.4 Optimisation of Reamplification of DNA Libraries

Optimisation of the reamplification of the DNA libraries was performed to ensure the maximum efficiency of the reamplification of DNA was achieved. Reamplification using the High fidelity DNA polymerase enzyme (New England Biosciences), using the method as described in section 3.4.1.2, was compared to reamplification using the Herculase II fusion DNA polymerase enzyme (Agilent Technologies). DNA libraries made from 9 WGA CTC samples from a patient with SCLC were pooled and diluted to 10 ng/µl concentration, as measured using the Qubit fluorometer. A comparison of the effect of the DNA polymerase enzyme used, the concentration of the primer used and the temperature for annealing was carried out. Both High fidelity DNA polymerase enzyme and Herculase II fusion DNA polymerase enzyme were compared, with samples processed using an annealing temperature of both 65 °C and 57 °C compared for the latter enzyme. The effect of the concentration of PE1 and PE 2 primers was compared, varying from 0.3 µM to 10 µM final concentrations in a 50 µl final reaction volume. The PCR conditions for samples processed using the High Fidelity DNA Polymerase enzyme were as described in section 3.4.1.2. The PCR conditions for samples processed using the Herculase II Fusion DNA polymerase enzyme were as described in section 3.4.1.2 except the annealing temperature was varied for one set of samples from 65 °C to 57 °C. The DNA concentrations of the different reamplified libraries produced were measured with the Qubit fluorometer, adding 1 µl of DNA (Table 3.4).

		High Fidelity DNA Polymerase (65 °C)	**Herculase II Fusion DNA Polymerase (65 °C)**	**Herculase II Fusion DNA Polymerase (57 °C)**
Final Primer Concentration (µM)	**Replicate**	**Final Concentration of Reamplified DNA Library (ng/µl)**		
10	1	29.8	142	141.6
	2	0.2	134.2	144.4
5	1	30.8	122.2	115.8
	2	38.8	140.2	119.8
1	1	43.6	89	81
	2	49.4	81.4	80.2
0.3	1	49.4	29.6	29.8
	2	39.8	34.2	31.6

Table 3.4 Comparison of the effect of the DNA polymerase enzyme, primer concentration and annealing temperature on the efficiency of reamplification of DNA libraries. Replicates of DNA libraries prepared from WGA CTCs from a patient with SCLC were reamplified using either High fidelity DNA polymerase or Herculase II fusion DNA polymerase. The effect of different concentrations of the PE1 and PE2 primers (0.3, 1, 5 and 10 µM final

concentrations) on the efficiency of reamplification was assessed. When assessing Herculase II fusion DNA polymerase annealing temperatures of 65 °C and 57 °C were investigated. Following the reamplification PCR the concentration of the DNA samples were measured using a Qubit fluorometer.

Post quantification with the Qubit fluorometer, the reamplified samples were purified using SPRI beads, as per the cleanup of PCR Amplification instructions in the NEBNext® Ultra™ DNA Library Prep Kit. The PCR products were then visualised resolving 8 µl of sample on a 1.5% agarose gel with a fluorescent DNA stain added (Figure 3.2). The concentration of the DNA was higher post the reamplification PCRs when the Herculase II fusion DNA polymerase enzyme was used in comparison to the High fidelity DNA polymerase enzyme. For future reamplifications it was therefore decided to use the Herculase II fusion DNA polymerase.

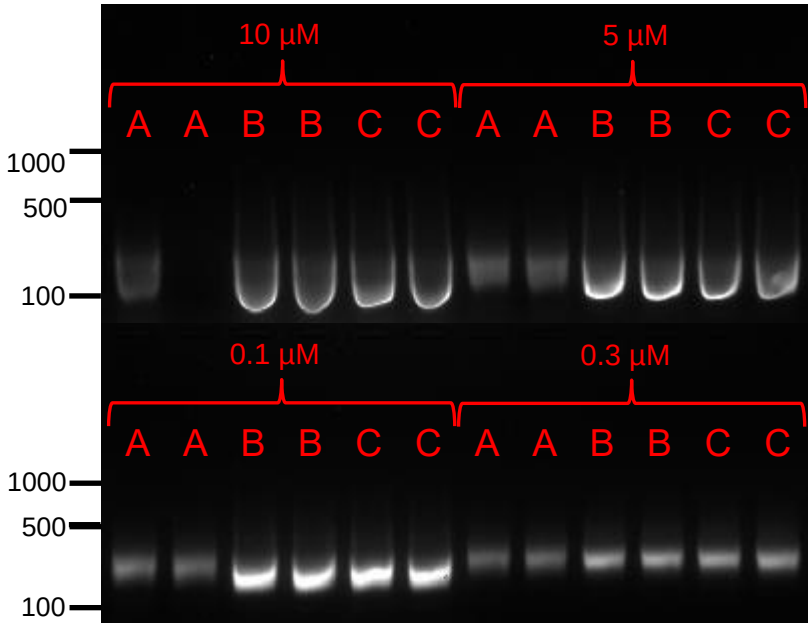


Figure 3.2 Comparison of the effect of the DNA polymerase enzyme, primer concentration and annealing temperature on the efficiency of reamplification of DNA libraries. Replicates of DNA libraries prepared from WGA CTCs from a patient with SCLC were reamplified using either High fidelity DNA polymerase (A) or the Herculase II fusion DNA polymerase. The effect of different concentrations of the PE1 and PE2 primers (0.3, 1, 5 and 10 µM) on the efficiency of reamplification was also investigated. When assessing the Herculase II Fusion DNA polymerase annealing temperatures of 65 °C (B) and 57 °C (C) were examined. Following the reamplification PCR the samples were purified using SPRI beads and 8 µl of the sample visualised on a 1.5% agarose gel with GelStar™ nucleic acid gel stain. The final primer concentration is indicated above the samples in µM. Markers for the 100, 500 and 1000 base pair ladders are indicated.

When assessing the effect of the annealing temperature there was not a significant difference between 65 °C and 57 °C. It was therefore decided that as using an

annealing temperature of 65 °C had increased specificity this temperature would be used in the reamplification PCRs for annealing. It was demonstrated that when the final concentration of the PE 1 and 2 primers was 5 or 10 µM in comparison to 0.3 or 1 µM the concentration of the final reamplified DNA was higher. The concentration of DNA was not significantly higher when 10 µM final concentrations of PE1 and PE2 primers were used in comparison to 5 µM. It was therefore felt that using a final concentration of 5 µM of the PE1 and PE2 primers provided a good balance between efficiency of the reamplification PCR and the cost of the reaction and that this would therefore be used in future experiments.

An experiment was then performed to investigate the optimal number of cycles for the PCR reaction. Reamplification PCR reactions were performed using 20 ng of pooled DNA libraries from WGA CTCs as described above. The PCR was performed using Herculase II fusion DNA polymerase, Herculase II Rxn buffer (5x concentration), dNTPs (250 nM final concentrations), PE1 and PE2 primers (5 µM final concentrations) and nuclease free water to a final volume of 50 µl. The PCR cycling conditions were as described in 3.4.1.2, except the number of cycles of amplification was varied (5, 10, 15, 20 and 25 cycles), with three DNA replicates processed for each number of cycles.

Following the reamplification PCR the DNA concentrations were measured using a Qubit fluorometer (Table 3.5). The total yield of reamplified DNA increased with the number of cycles of PCR. There was however not a logarithmic increase with smaller increases in DNA seen beyond 15 cycles. In total 750 ng to 1 µg of purified DNA is needed for the planned in-solution exome capture method. It was therefore felt that more than 10 cycles of PCR would be unnecessary and that minimising the number of cycles of reamplification, and therefore the risks of polymerase errors, would be beneficial.

Number of cycles of reamplification in PCR	Replicate	Concentration (ng/µl)	Total Yield of DNA from PCR reamplification (ng)
5	1	31.2	1560
	2	27.4	1370
	3	24	1200
10	1	93.8	4690
	2	93.8	4690
	3	85.4	4270
15	1	134.4	6720
	2	132.2	6610
	3	89.8	4490
20	1	158.4	7920
	2	120.8	6040
	3	172.8	8640
25	1	88	4400
	2	153.2	7660
	3	79.2	3960

Table 3.5 Investigation of the effect of the number of cycles of PCR during reamplification of DNA libraries on DNA yield. Reamplification PCR using WGA CTC DNA libraries and Herculase II fusion DNA polymerase were carried out, with final concentrations of the PE1 and PE2 primers of 5 µM. The number of PCR cycles of reamplification were altered ranging from 5 to 25 with three replicates processed using each number of PCR cycles. The concentration of DNA from each PCR reaction was measured with the Qubit fluorometer and the total yield of the DNA calculated.

To further assess the impact of cycle number on the reamplification of DNA libraries, DNA libraries were made from 6 tumour samples (as per sections 2.4 and 2.5). These samples were then reamplified for 6 or 10 cycles of PCR using the reamplification protocol described above. The amount of template DNA in the PCR was 20ng for all but two of the samples in which 40ng was used. The samples were purified using SPRI beads and the concentrations in the final 28 µl volume measured using the Qubit fluorometer. The concentrations and yields of the DNA library are detailed in table 3.6. The size distribution of the DNA fragments were also measured using the bioanalyzer and representative results for two of the tumour samples are shown in figure 3.3 as an example. The concentrations of DNA achieved were higher if the PCR was carried out for 10 rather than 6 cycles for all the samples where 20 ng input of DNA was used. However, the bioanalyzer results for the samples were distorted if the PCR was carried out for 10 cycles with 2 peaks noted in the reamplified product rather than the expected single peak. There were also increased numbers of larger sized fragments causing a distortion in the bioanalyzer trace when the samples were reamplified for 10 cycles as

opposed to 6 cycles. These products are likely to reflect multimer products that are generated during the reamplification process as the primer concentrations are becoming limited. The bioanalyzer traces for the samples that had been reamplified for 6 cycles of PCR had the expected pattern with a single peak. It was therefore decided that despite the lower yield of DNA achieved when using 6 cycles of PCR, it was the appropriate number to use in further experiments. Using 6 cycles of PCR limited the number of cycles of amplification in which errors could occur and did not distort the size distribution of fragments. The final method used for the reamplification of DNA libraries is described in section 2.8.

Tumour Sample	6 cycles		10 cycles	
	Concentration of reamplified DNA (ng/µL)	Total yield of reamplified DNA (ng)	Concentration of reamplified DNA (ng/µL)	Total yield of reamplified DNA (ng)
T1	52.4*	1467.2	14.6	408.8
T2	14.2	397.6	75.8	2122.4
T3	46.8	1310.4	65.2	1825.6
T4	16.4*	459.2	17.2	481.6
T5	13.4	375.2	62.6	1752.8
T6	17.4	487.2	45	1260

Table 3.6 Comparison of the concentrations of tumour DNA libraries reamplified for 6 or 10 cycles. Reamplification PCRs using WGA tumour DNA libraries and Herculase II Fusion DNA polymerase, with final concentrations of PE1 and PE2 primers of 5 µM, were carried out. The input of DNA was 20 ng for the PCR (* indicated for 40 ng input was used instead of 20ng). The number of cycles of amplification in the PCRs was either 6 or 10. The samples were purified with SPRI beads. The concentration of DNA was measured with the Qubit fluorometer and the total yield of DNA calculated.

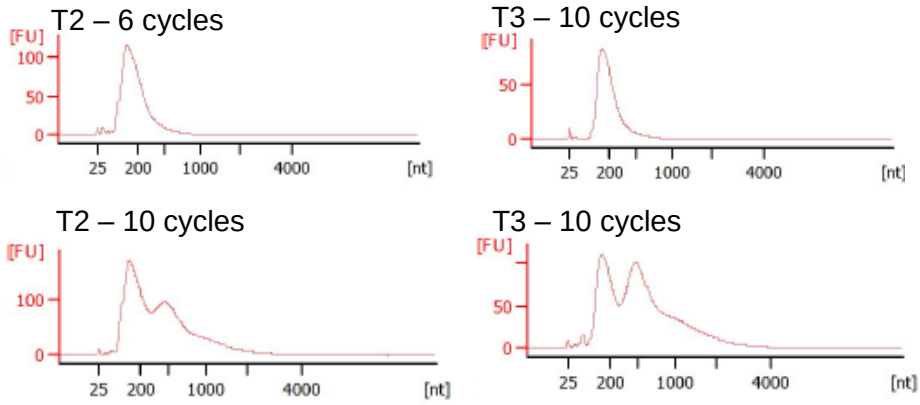


Figure 3.3 Comparison of the bioanalyzer results of tumour DNA libraries reamplified for 6 or 10 cycles. Reamplification PCRs using WGA CTC DNA libraries and Herculase II Fusion DNA polymerase, with final concentrations of PE1 and PE2 primers of 5 µM, were carried out. The number of cycles of amplification in the PCRs was either 6 or 10. The samples were

purified with SPRI beads. The size distribution of DNA fragments was measured for the samples with the bioanalyzer and examples of the results for 2 tumours (T2 and T3) are shown.

3.4.2 Optimisation of Exome Target Enrichment Using In-Solution Capture.

3.4.2.1 In-Solution Capture of Samples

Aim 1 of this thesis (chapter 1) was to develop methods to perform WES of CTCs to enable mutations in protein-coding genes in SCLC to be explored. Achieving this aim would require exome target enrichment from amplified DNA. The SureSelect^{XT} Target Enrichment System for Illumina Paired-End Sequencing Library was chosen as the method for target enrichment as its use was well established in the core facilities of the CRUK Manchester Institute, and the method could be adapted to fit with the CTC workflow. It was therefore necessary to ensure that it was possible to use the amplified DNA as the input for the chosen method of in-solution capture, and to ensure the results were not distorted when compared to unamplified material. To this end, an experiment comparing WES after in-solution capture of both unamplified and amplified material was performed. To increase the number of samples that could be processed per patient and to reduce the costs associated with in-solution capture investigation of multiplexing samples for target enrichment was completed. A comparison was carried out of WES of amplified samples that had been multiplexed prior to the in-solution capture with those in which the in-solution capture was performed on individual samples, to ensure there was no contamination of reads or deterioration in the quality of the target enrichment.

The target enrichment was performed using the SureSelect^{XT} Target Enrichment System for Illumina Paired-End Sequencing Library using version 1.5 of the protocol (section 2.9). The Sure Select Human All Exon V5 baits were used for the target enrichment. CTC-derived explant (CDX) tumours from four patients samples, leukocyte genomic DNA and CTCs and WBCs isolated from one patient with SCLC were processed for this experiment. A summary of the samples used for the experiment is given in table 3.7. To enable the assessment of the effect of amplification on the data generated from WES a CDX tumour and paired leukocyte genomic sample were amplified using the Ampli1 kit as described in section 2.3 so a comparison of the data from the amplified and unamplified samples could be made. To assess the effect of multiplexing the samples for the target enrichment DNA libraries from an amplified CDX tumour, a paired amplified leukocyte sample, amplified CTC and WBC pools from a patient with SCLC underwent in-solution capture as single

samples. The same four samples were also multiplexed prior to the in-solution capture as a comparison. The data from CDX tumours from two patients with SCLC samples which had been WGS were also used for a comparator to the WES data from unamplified samples.

Sample Name	Sample Type	Ampli1 Amplified	Singleplex or Multiplex In-solution Target Enrichment	Sequencing Method
M1-L-WGS	Tumour	Unamplified	Not Applicable	WGS
M1-R-WGS	Tumour	Unamplified	Not Applicable	WGS
M2-WGS	Tumour	Unamplified	Not Applicable	WGS
M1L-UnA	Tumour	Unamplified	Singleplex	WES
M1R-UnA	Tumour	Unamplified	Singleplex	WES
M2-UnA	Tumour	Unamplified	Singleplex	WES
M3L-UnA	Tumour	Unamplified	Singleplex	WES
M3R-UnA	Tumour	Unamplified	Singleplex	WES
M4-UnA	Tumour	Unamplified	Singleplex	WES
M3gDNA-UnA	Leukocyte genomic DNA	Unamplified	Singleplex	WES
M3gDNA-A	Leukocyte genomic DNA	Amplified	Singleplex	WES
M3L-A	Tumour	Amplified	Singleplex	WES
pCTC-A	CTC	Amplified	Singleplex	WES
pWBC-A	WBC	Amplified	Singleplex	WES
M3gDNA-M	Leukocyte genomic DNA	Amplified	Multiplex	WES
M3L-M	Tumour	Amplified	Multiplex	WES
pCTC-M	CTC	Amplified	Multiplex	WES
pWBC-M	WBC	Amplified	Multiplex	WES

Table 3.7 Summary of samples used in assessment of in-solution target enrichment for WES. Samples used for experiments comparing the effect of using amplified or unamplified DNA, and single versus multiplexed in-solution target enrichment, on the efficiency of exome capture for sequencing. A comparison to three CDX tumour samples that had previously been WGS was also made.

For the unamplified CDX tumours 750 ng of NEBNext® Ultra™ DNA Libraries were used as the input for the target enrichment, whilst for the amplified samples 750 ng of the reamplified NEBNext® Ultra™ DNA Libraries were used. For the multiplexed samples 187.5 ng of reamplified DNA libraries of each of the four samples with different indexes were multiplexed for the target enrichment. The samples were completely lyophilized using a vacuum concentrator and then each was resuspended in 3.4 µl of nuclease free water. The samples were then processed according to the SureSelect^{XT}

Target Enrichment System for Illumina Paired-End Sequencing Library protocol with a hybridisation reaction set up and the samples incubated for 24 hours (section 2.9). The samples were then processed to select the hybridised capture DNA using Dynabeads MyOne Streptavidin T1 magnetic beads. The captured libraries were then amplified as per the SureSelect protocol except the PE1 and PE2 primers (sequences in table 3.1) were used for the amplification, with a final concentration of each primer of 5 µM in the PCR reaction. The libraries were already indexed so the addition of a further index in this reaction was not required. Following purification of the libraries the samples were quantified with both the Qubit fluorometer and the KAPA Library Quantification Kit (section 2.5) and diluted for sequencing. The samples were sequenced on the Illumina HiSeq as per section 2.6 of the methods section and the bioinformatics analysis of the WES data carried out as per the methods section 2.10.5.

3.4.2.2 Results of In-Solution Capture Experiment

The data generated from sequencing the in-solution exome capture samples on the HiSeq was processed using SMALT, to align the reads to the human genome and GATK was used to identify SNVs and Indels as described in section 2.10.5. The total number of reads per sample was identified and ranged from 41,070,745 to 67,944,802 for the WES samples and 1,508,716,202 to 1,807,359,070 for the WGS samples (Table 3.8). The number and percentage of uniquely mapped reads was also calculated for each sample, with a uniquely mapped read being where both the paired end reads uniquely align to the genome and have a mapping quality score of greater than 5. The percentage of uniquely mapped reads ranged from 65.3% to 88.1% for the unamplified samples. The percentage of reads for the amplified samples was very low at just 1.8 to 4.8%. Analysis of the reads from the amplified samples revealed large changes in the GC content amongst the first 24 reads, but that the later reads were as expected with even representation of each of the 4 nucleotides (Figure 3.4). It was also noted that more than 30% of the reads had the same sequence in the first 24 bases “AGTGGGATTCCTGCTCTCAGTTAA”. This sequence corresponds to the Ampli1 adaptor sequence, which was used during the WGA process, suggesting the MSE1 digestion to remove this was incomplete. The content bias in the first 24 reads caused problems with the alignment of the data by SMALT leading to the low number of uniquely mapped reads seen. This problem was compounded as SMALT was unable to align short reads and as some of the reads were short, just 7 bases in one case, SMALT did not align those reads or any of the subsequent reads, leading to the low

alignment of reads seen. To address these issues the initial 25 bases of each read were trimmed and the option of trimming of Ns where the bases were not called was disabled so that shorter reads would not interfere with all subsequent reads. The number of and percentage of uniquely mapped reads significantly improved with these measures. The percentage of uniquely mapped reads now ranged from 70.1% to 74.2% for the amplified samples, which was consistent with the results seen for the unamplified samples, as described in Table 3.8.

Sample Name	Sample Type	Total Reads	Uniquely Mapped Reads	Percentage of Uniquely Mapped Reads	Percentage of Reads Covering Exome
M1-L-WGS	Unamplified Tumour	1,807,359,070	1,509,488,068	83.52%	NA
M1-R-WGS	Unamplified Tumour	1,736,245,528	1,498,910,524	86.33%	NA
M2-WGS	Unamplified Tumour	1,508,716,202	1,329,097,546	88.09%	NA
M1L-UnA	Unamplified Tumour	41,744,634	32,430,600	77.69%	30.73%
M1R-UnA	Unamplified Tumour	41,347,205	35,995,355	87.06%	29.58%
M2-UnA	Unamplified Tumour	43,338,795	35,546,938	82.02%	29.80%
M3L-UnA	Unamplified Tumour	42,272,030	33,817,841	80.00%	28.50%
M3R-UnA	Unamplified Tumour	42,475,836	31,634,335	74.48%	29.19%
M4-UnA	Unamplified Tumour	41,070,745	29,533,909	71.91%	28.20%
M3gDNA-UnA	Unamplified gDNA	43,919,215	28,660,262	65.26%	26.59%
M3gDNA-A	Amplified gDNA	49,717,806	36,017,510	72.44%	49.15%
M3L-A	Amplified Tumor	48,748,476	35,596,308	73.02%	60.51%
pCTC-A	Amplified CTC	54,064,722	39,575,676	73.20%	60.14%
pWBC-A	Amplified WBC	61,897,340	43,407,972	70.13%	60.98%
M3gDNA-M	Multiplexed Amplified gDNA	49,312,222	36,035,164	73.08%	50.06%
M3L-M	Multiplexed Amplified Tumour	67,944,802	50,386,722	74.16%	61.86%
pCTC-M	Multiplexed Amplified CTC	44,041,308	32,110,018	72.91%	60.11%
pWBC-M	Multiplexed Amplified WBC	58,432,864	41,597,006	71.19%	62.27%

Table 3.8 Summary of sequencing results for in-solution capture optimisation experiment. DNA, both amplified and unamplified, from CDX tumours, WBCs and CTCs, was enriched for exonic regions using in-solution capture, with samples either enriched singly or multiplexed prior to enrichment. Sequencing data was then generated using the HiSeq. Three tumour samples were also processed on the HiSeq to generate WGS results. The data was then analysed to assess the total number of reads, uniquely mapped reads, the percentage of uniquely mapped reads and the percentage of reads covering the exome.

It was hypothesised that using amplified DNA as opposed to unamplified DNA as the input for the in-solution target capture may reduce the efficiency of the exome enrichment. To assess the efficiency of the in-solution capture enrichment, the

percentage of on-target reads covering the exome out of the total number of reads per sample, which included the reads covering intronic regions, was calculated (Table 3.8). The average percentage of on target reads covering the exome varied, with 28.9% of the reads for the unamplified samples covering the exome whilst for the amplified samples the average percentage was 58.1%. There was no significant difference between the amplified samples processed singly or multiplexed for the in-solution capture, with an average of 57.7 and 58.6% of the reads covering the exome respectively. This suggests that both using amplified DNA and multiplexing samples for the in-solution capture do not deleteriously affect the efficiency of the in-solution capture.

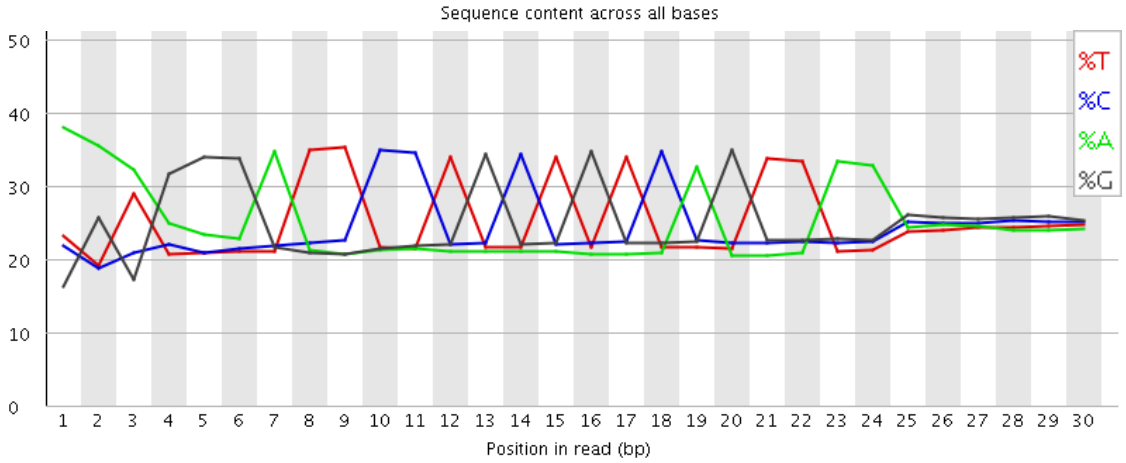


Figure 3.4 Sequence content of WES of amplified samples. Sequence content per base of the sequencing data run on the HiSeq from amplified CDX tumour, WBC and CTC DNA samples which had been enriched for exonic regions using in-solution capture. The figure shows the percentage of reads which are T, C, A or G per base for the initial 30 bases of the sequence. The percentage of reads for each of the bases is stable from read 25 for the rest of the sequence.

To further investigate the effect of using amplified DNA and multiplexing samples as input for the in-solution capture, the sequencing coverage of the target regions of the in-solution capture baits was calculated (Table 3.9). The sequencing coverage of the target regions was calculated by assessing the percentage of the bases which were sequenced out of all the bases in the target region from the Sure Select protocol. For the assessment of the coverage of the target regions by the WGS samples, 10% of the total number of uniquely mapped reads were used due to the very high number of reads and the computing time involved in assessing this size of data. For all other sequencing metrics for the WGS data the total number of reads was used. For both WES and WGS of the unamplified samples the coverage was good with an average of 96.4 % of the target bases being sequenced. However, the coverage of the target

regions when sequencing the amplified DNA was lower, with an average of 64.8% of the target bases being sequenced. There was no difference between coverage of the target region when the single and multiplex in-solution captured samples were sequenced (64.8 % and 64.7 %). The Ampli1 amplification protocol uses the MSE1 enzyme to digest DNA at the restriction-site motif TTAA. The DNA fragments are therefore created in a non-random manner, based on the distribution of TTAA sites within the genome. The consistent fragmentation of the DNA in this manner may mean that some of the in-solution capture baits are unable to anneal to the sample DNA, leaving regions of the exome for which the target DNA is not captured.

Sample Name	Sample Type	Uniquely Mapped Reads	Coverage of Target Regions	SNVs Identified	Indels Identified
M1-L-WGS	Unamplified Tumour	1509488068	93.83%	42961	1930
M1-R-WGS	Unamplified Tumour	1498910524	94.02%	36753	1858
M2-WGS	Unamplified Tumour	1329097546	93.83%	33697	1929
M1L-UnA	Unamplified Tumour	32430600	97.40%	87565	4571
M1R-UnA	Unamplified Tumour	35995355	97.79%	75331	5033
M2-UnA	Unamplified Tumour	35546938	96.68%	70339	4392
M3L-UnA	Unamplified Tumour	33817841	97.79%	69984	4659
M3R-UnA	Unamplified Tumour	31634335	96.85%	65098	4170
M4-UnA	Unamplified Tumour	29533909	97.32%	132733	4427
M3gDNA-UnA	Unamplified gDNA	28660262	98.35%	71768	4327
M3gDNA-A	Amplified gDNA	36017510	69.14%	44798	2589
M3L-A	Amplified Tumor	35596308	67.38%	50382	2458
pCTC-A	Amplified CTC	39575676	61.46%	33369	2098
pWBC-A	Amplified WBC	43407972	61.17%	35173	2218
M3gDNA-M	Multiplexed Amplified gDNA	36035164	60.63%	32775	2174
M3L-M	Multiplexed Amplified Tumour	50386722	60.41%	30430	1916
pCTC-M	Multiplexed Amplified CTC	32110018	68.45%	53663	2645
pWBC-M	Multiplexed Amplified WBC	41597006	69.15%	44455	2615

Table 3.9 Summary of coverage of target regions and the identification of mutations in the in-solution capture experiment. DNA, both amplified and unamplified, from CDX tumours, WBCs and CTCs was enriched for exonic regions using in-solution capture, with samples either enriched singly or multiplexed prior to enrichment. Sequencing data was then generated using the HiSeq. Three unamplified CDX tumour samples were also processed on the HiSeq to generate WGS results. The coverage of the target regions for the in-solution capture baits was assessed for each sample. The number of SNVs and Indels as identified by GATK within the target regions was also calculated for each sample.

It was hypothesised that the reduced sequencing coverage of the target regions was due to the effect of the MSE1 digestion during the amplification process, which influenced DNA fragment size causing some DNA fragments to be unable to anneal to the baits, so regions of the exome were not captured for sequencing. If this was the cause of the reduced coverage of the target regions in the amplified samples, high concordance between the regions not sequenced between the samples would be expected. To investigate this hypothesis the regions of the exome not sequenced for each of the amplified samples were identified, with a region being called unsequenced if there were less than 5 reads. The concordance between these regions was then assessed for each of the samples with an average of 83% (range 71 to 96%) of the unsequenced regions being common to each sample. This supports the hypothesis that the unsequenced regions in the amplified DNA are consistent and may be as a result of the effects of the digestion of the DNA by the MSE1 endonuclease causing amplicons to be unable to anneal to the target enrichment baits. There will be some variation beyond this as the coverage of individual bases between samples will vary slightly by chance, with for example a single base being sequenced five times in one sample but four times in another sample causing lack of concordance in this calculation. However, the majority of the regions that are not sequenced are consistent and predictable between all these amplified samples.

The total number of SNVs and Indels in each sample as identified by GATK was also noted for each sample sequenced (Table 3.9). There were concerns that using amplified samples could increase the number of SNVs and Indels identified due to errors occurring during the amplification process. However, the number of both SNVs and Indels identified by GATK in the unamplified samples was higher than in the amplified samples. The average number of SNVs and Indels identified in the unamplified samples was 81831 and 4511 respectively, whilst in the amplified samples the average number of SNVs and Indels was 40631 and 2339. This is likely to reflect the lower coverage of the target regions in the amplified samples in contrast to the unamplified samples. It was also of concern that multiplexing the samples for the enrichment could lead to cross contamination so SNVs or Indels from one sample could be attributed to other samples, potentially increasing the number of SNVs or Indels identified. There was, however, no consistent difference between the numbers of mutations identified in the samples which had been enriched singly as opposed to those which had been multiplexed. The average number of SNVs was 40931 and Indels was 2341 in the samples which were enriched singly, whilst in the samples that were multiplexed for enrichment the average number of SNVs was 40331 and the

average number of Indels was 2338. This again suggests that there is no deterioration of the quality of the WES data when the in-solution capture is carried out on multiplexed samples.

As discussed above one concern of multiplexing samples was that there could be cross contamination between the samples, with reads from one sample being wrongly attributed to another sample, which would compromise the data analysis. To investigate this, unique mutations in 40 genes potentially mutated in SCLC were identified in the singleplex enriched samples, M3gDNA-A, M3L-A, pCTC-A and pWBC-A. In M3gDNA-A 3 unique SNVs were noted (in GPR113, LRRK2 and RET), in M3L-A 5 unique SNVs were noted (in LRRK2, MED12L, TP53 and TRRAP), and 1 unique SNV was noted in each of pCTC-A (in TP53) and pWBC-A (in LRKK2). The presence of these mutations was then investigated in all the samples which had been multiplexed prior to enrichment (Table 3.10). It was hypothesised that if the multiplexing did not cause problems with assigning reads to the correct sample the unique mutations from each sample would only be identified in the multiplexed version of this sample. This was found to be correct, with the unique mutations only ever noted in the corresponding multiplexed samples suggesting no contamination of reads. One unique mutation in TP53 identified in M3L-A was not found in M3L-M which may reflect the read depth at this region or an error of sequencing. Overall these results demonstrate that it is feasible to use the amplified DNA in the in-solution capture and that multiplexing samples prior to the target enrichment does not affect the efficiency of the enrichment process. The SureSelect^{XT} Target Enrichment System for Illumina Paired-End Sequencing Library was therefore used for subsequent experiments with 4 samples multiplexed together prior to the in-solution capture. The final method used for the in-solution capture target enrichment of multiplexed samples is described in section 2.9.

	M3gDNA-A		M3L-A		pCTC-A		pWBC-A	
	Expected	Observed	Expected	Observed	Expected	Observed	Expected	Observed
M3gDNA-M	3	3	0	0	0	0	0	0
M3L-M	0	0	5	4	0	0	0	0
pCTC-M	0	0	0	0	1	1	0	0
pWBC-M	0	0	0	0	0	0	1	1

Table 3.10 Identification of unique SNVs from the singleplex enriched samples in the multiplexed enriched samples. DNA from CDX tumours, WBCs and CTCs was enriched for exonic regions using in-solution capture, either singly or multiplexed together. Sequencing data was then generated using the HiSeq. Unique mutations in 40 selected genes potentially mutated in SCLC were identified in the singleplex enriched samples. The presence of these SNVs was then investigated in all the multiplex samples.

3.5 Discussion

The WES of CTCs in SCLC was one of the key aims of this thesis as it would allow the analysis of different patterns of mutations between subgroups of patients to enable the investigation of response to treatments. To enable meaningful data to be generated, the experimental process for WES of the CTCs needs to be robust and accurate, with the minimal introduction of errors. The optimisation of the reamplification method to generate sufficient DNA for WES maximised the efficiency of this approach, whilst minimising the number of cycles of PCR in which errors could be introduced. The results presented in this chapter also demonstrated that amplified DNA can be used for the in-solution target enrichment without deleteriously affecting the downstream sequencing apart from the target coverage.

The development of NGS has enabled the rapid generation of large amounts of sequence data from samples, allowing the investigation of the genome, epigenome and transcriptome. NGS has allowed the exploration of the drivers of cancers and revealed new potential treatment targets. As discussed in the introduction the NGS of SCLC tumours has revealed patterns of common mutations, providing information about the biology of this disease [70, 91]. The genomic profiling of CTCs has the potential to provide exciting information about the biology of SCLC, and to inform on the responses to treatment in SCLC, through a minimally invasive technique, as opposed to tissue biopsies. However, the profiling of single cells and small numbers of cells, although providing a unique insight into the heterogeneity of tumours at a single cell level, presents additional challenges due to the minimal amount of nucleic acids present in a

single cell. Amplification processes are therefore required to generate sufficient material to act as the input for NGS techniques, which may need to be adapted accordingly. The effects of using amplified DNA as the input for NGS on the sequencing data generated also needs to be assessed. Techniques to amplify sufficient DNA to act as an input for exome enrichment, and techniques for the enrichment of exonic regions were therefore developed and evaluated.

DNA from single CTCs was WGA after isolation to increase the amount of DNA available for downstream analyses. To explore potential SCLC associated changes in copy number and point mutations the aim was to perform both CNA and WES on the same CTCs. This would allow correlations between the identified alterations in copy number and point mutations within an individual cell to be made. This, however, required sufficient amounts of DNA from single or small numbers of CTCs to be generated to use as the template for both analyses. Techniques to reamplify the WGA DNA were therefore required. A comparison of the effects of reamplifying the WGA DNA and the WGA DNA libraries was therefore made, to determine which approach introduced the least amount of technical variation. CNA data was generated from 4 samples in which both the WGA DNA and the WGA DNA libraries had been reamplified, and a comparison to the original data made. Hierarchical clustering of the CNA data grouped all the samples of one type together, irrelevant of the reamplification method used, but the reamplified WGA DNA libraries were consistently grouped most closely with the original libraries indicating less variation. Assessment of differences in the CNA cytoband data were also made between the original samples and the reamplified samples. This revealed just 0.2% of the cytobands differed significantly between the original and the reamplified libraries, compared to 2.2% and 3.0% for the reamplified WGA DNA with and without the MSE1 digestion step. The library preparation in addition to the reamplification process may influence this result. To enable valid correlations to be drawn between the CNA and WES data from a CTC it was important that the input DNA differed as little as possible for these two techniques. The data showed that the reamplified WGA DNA libraries most closely resembled the original libraries, probably in part because the effects of the MSE1 digestion, purification, sonication and library preparation are common to these two sample sets unlike when the WGA DNA is reamplified.

DNA produced from WGA will contain errors due to the inherent error rate of the DNA polymerases [228]. To decrease these errors the aim of the reamplification process was to minimise the number of cycles of PCR used during which additional errors could be produced, whilst maximising the efficiency of the process. To improve the efficiency

of the reamplification of WGA DNA libraries, investigation of the effects of primer concentration, the DNA polymerase enzyme used, the annealing temperature and the number of cycles of PCR was undertaken. The effects on both the final amount of DNA produced and the library fragment sizes were analysed to allow optimisation of the method used. The final method used for subsequent experiments therefore balanced the DNA yield produced, whilst also ensuring there was no distortion of the libraries generated.

WES has the potential to reveal the key mutations present in cancers which drive their growth, spread and response to therapy. WES requires a target enrichment step to ensure that the exonic regions of the genome are sequenced in preference to the rest of the genome. At the time of commencing this thesis there was no published research on the use of WES in CTCs. For the target enrichment step the SureSelect^{XT} Target Enrichment System for Illumina Paired-End Sequencing Library, a form of in-solution target enrichment, was chosen. The three published examples of WES of patients' CTCs that have come out during the course of this study, as opposed to spiked cell lines, all also used in-solution target enrichment. Lohr et al. and Ni et al. both used the SureSelect^{XT} Human All Exon Library as baits, whilst Zhao et al. used the SeqCap EZ Human Exome Library baits (Roche). Lohr et al. and Ni et al. both performed CNA and WES of CTCs, with Ni et al. using the same libraries for CNA and WES, whilst Lohr et al. used separate library preparations for both analyses.

As there was no published data on the use of WES of CTCs at the time of my project, an assessment on the effect of the use of amplified material with the chosen in-solution capture method for exome enrichment was undertaken. Unamplified and amplified CDX tumour and WBC samples were enriched using the in-solution capture method and sequence data generated on the HiSeq. The sequencing metrics of these samples were compared and initially the percentage of uniquely mapped reads was much lower for the amplified than the unamplified samples. This was due to both the presence of the Ampli1 adaptor, which is added to the ends of every DNA fragment during the WGA process, and short reads causing problems with the SMALT aligner. Therefore, for the amplified samples the initial 25 reads, covering the adaptor were trimmed and the SMALT aligner criteria were adapted to avoid removal of Ns if bases could not be called enabling the analysis of short reads. As a result of these adaptations the percentage of uniquely mapped reads was comparable in both the amplified and unamplified WES samples. Surprisingly the target enrichment was more efficient in the amplified than the unamplified samples with 58.1% of the reads covering the exome as opposed to 28.9% in the unamplified samples in this experiment. There was variability

seen between the efficiency of the enrichment of samples between all the subsequent experiments using this method of in-solution capture (appendix 1 table A1.4), so this data is unlikely to actually represent that the efficiency of target capture is improved with the amplified DNA in comparison to the unamplified DNA. However, there was no evidence from this experiment that the use of the amplified DNA deleteriously affected the efficiency of the target enrichment compared to the unamplified DNA.

One of the major differences noted between the unamplified and amplified samples WES results was the degree of coverage of the target regions in the exome in the sequencing data. The average coverage of the target regions for the unamplified WES samples was 96.4%, whilst for the amplified samples the average coverage was reduced with only 64.8% of the target region sequenced. Analysis of the regions that were not sequenced in all of the amplified samples revealed an 83% overlap between the samples. This supports the hypothesis that the MSE1 digestion used in the amplification protocol causes predictable problems with downstream reactions, such as annealing of DNA to the baits due to fragmentation of target regions. This issue with the Ampli1 amplification protocol has been noted by other researchers [237]. However, these regions are predictable and have significant overlap between the different amplified samples and over 64% of the target areas are sequenced. As one of the aims is to compare the mutations present in one subgroup of patients with SCLC, but absent in another, the fact that these unsequenced regions are predictable and consistent between samples, rather than entirely stochastic, enables these comparisons to be made. It should also be noted that other methods of WGA have problems with sequencing such as amplification biases, for example, due to GC content, allelic drop out and preferential allelic amplification, so problems with sequence coverage are not unique to the Ampli1 method used.

To increase the number of samples that could be processed per patient and reduce the costs associated with the target enrichment, an assessment of the effect of multiplexing samples prior to the in-solution capture was made. Comparison of WES of amplified samples which had been enriched for exonic regions singly was made to WES of the same four samples which had been multiplexed prior to the in-solution capture. The sequencing metrics revealed no significant difference in the percentage of uniquely mapped reads, percentage of reads covering the exome as opposed to the rest of the genome or the coverage of the target regions between the two sets of samples. There was also critically no evidence of cross contamination of reads between the multiplexed samples. Unique mutations were identified in the WES data of the singleplex samples were only ever seen in the corresponding multiplexed samples, supporting the

applicability of this technique. Of note Ni et al. also multiplexed four samples prior to the target enrichment for the WES of CTCs when using the SureSelect^{XT} Human All Exon Library [229].

The data presented in this chapter provide evidence supporting the methodology developed for the WES of SCLC CTCs. It was important to establish that the use of amplified DNA is compatible with the target enrichment used, and not deleterious to the sequencing results generated, to engender confidence in results generated using this technique. In SCLC WES of CTCs has the potential to provide critical information about metastasis biology and treatment response. In chapter 5 the methods presented here were applied to DNA derived from SCLC CTCs, to enable the investigation of resistance to chemotherapy, highlighting the applicability of this methodology.

Chapter 4: Genomic Profiling of SCLC CDX and CTCs

4A. Genome-Wide Profiling of SCLC CTCs and CDX

4A.1 Introduction

SCLC is a very aggressive form of cancer associated with poor survival for patients. There have been minimal advances in the treatment of SCLC over the last 20 years so improvements in treatments for patients with SCLC represent a critical area for research. This will require new techniques to enable the investigation of the genetic landscape of this disease and to enable the investigation of the molecular events underpinning both responses, and resistance to standard chemotherapies and novel targeted agents. These studies are hampered in SCLC as there is limited fresh tumour tissue available not only to establish the clinical diagnosis of SCLC, but to have further tumour tissue available for research such as molecular analysis. This is because in SCLC surgical resection is rarely considered an appropriate treatment. Diagnoses are often made from cytological samples limiting the number of cells available for further analyses. It should also be acknowledged that lung biopsies have the risk of significant morbidity and in rare cases mortality associated with them so repeated biopsies for research purposes should be considered with caution [23]. The high necrotic content of cells in SCLC biopsies as well as their small size creates additional challenges for their molecular analysis [265]. The existing models available for the investigation of SCLC including cell lines, PDX and GEMMS, and their limitations were discussed in chapter 1. There is clear need for both improved models to study SCLC and also additional methods to study SCLC at an individual patient level. The approach of creating CTC derived explants (CDX) therefore could provide a much needed novel model to investigate key areas of interest in SCLC [24] whilst the genomic analysis of CTC samples from same patients provides an additional method of interrogating their cancer.

4A.1.1 Generation of CDX Models

CDXs were created by the preclinical team in CEP utilising CTCs isolated from patients with SCLC with the full methods detailed in appendix 8 [24]. Briefly, blood was drawn from patients with SCLC prior to administration of chemotherapy with 10 mls taken into EDTA vacutainers for the creation of the CDX models. A parallel sample of 10 mls of blood was taken in CellSearch vacutainers, to allow the enumeration of the number of EpCAM/cytokeratin positive CTCs using the CellSearch platform. CTCs were enriched

using the RosetteSep Human Circulating Epithelial Tumour Cell Cocktail (Stem Cell Technology) from the EDTA blood samples as per the protocol. The cell pellets were combined with HITES medium and matrigel and were injected subcutaneously into one or both flanks of female NOD.Cg-PrKdc^{scid}112rg^{tm1Wjl}/SzJ (NSG) mice. Mice were monitored for palpable tumour growth and when the tumour burden reached 1,000 m^3, or the animal had signs of ill health, the animal was killed. The tumours were harvested for further analysis including molecular characterisation.

Using this approach CDXs were generated from 4 of the initial patients' CTCs that were sampled confirming the tumorigenicity of CTCs (table 4.1). Two of the patients whose samples generated CDXs responded to chemotherapy and two had chemorefractory disease, as defined in chapter 1, as they progressed within 3 months of completion of chemotherapy. To demonstrate the wider utility of the CDX models it was important to see if they reflected the patient's response to treatments. Three of the 4 CDXs models were tested to see if they mirrored the patients' response to chemotherapy. Thirty NSG mice were implanted with passage 4 of CDX models 2, 3 and 4. The mice were treated with intraperitoneal Cisplatin and Etoposide or vehicle once tumours reached more than 200mm^3. The tumour volumes were then monitored until tumours reached 4 times the initial tumour volume or the animal demonstrated signs of ill health. The responses of the CDX tumours reflected those seen in the patients from whose samples they were derived. CDX2 and 4 tumours derived from chemorefractory patients both had poor responses to treatment unlike CDX3 tumours, which was derived from a chemoresponsive patient's samples, in which the tumours declined in size significantly with chemotherapy treatment. This result supported the utility of the CDX models to not only reflect the tumorigenicity of CTCs but to also allow the investigation of the biology of SCLC and therapeutic responses.

Beyond the creation of the CDX further analysis was necessary to confirm that the tumours created are consistent with SCLC, and reflect the commonly noted genomic alterations found in this disease. This was particularly critical as there are concerns that cell lines and GEMMs fail to capture the complexity of SCLC tumours. As chemoresistance is a significant challenge in the management of SCLC, contributing to the poor patient survival seen, the utility of CDX to investigate this issue was assessed. The differences in responses to therapies in these two groups of patients may be driven by underlying differences in genomic alterations. Contrasting the genomic aberrations in CDX created from these two groups of patients may therefore provide new insights into chemoresistance in SCLC.

The genomic profiling of the CTCs isolated from patients with SCLC also represents a novel method of studying the biology of this cancer and of profiling individual patients' tumours. The enumeration of EpCAM positive CTCs by CellSearch, and the change in the number of CTCs with one cycle of treatment have both been shown to be prognostic in SCLC [204]. This suggests that these are a clinically very relevant population of cells in SCLC. However, there has been no evidence in SCLC to date that it is this subgroup of CTCs that are producing metastases in patients and are therefore tumorigenic. Investigation of whether the genomic profiles of the paired CTC sample, enriched using the CellSearch system, reflect the profiles of the CDX tumours created from the same patient's blood samples is therefore important to investigate this. Establishing robust methodology for the profiling of single CTCs in SCLC patients also provides an alternative technique to using biopsies to interrogate genomic alterations on an individual patient basis.

Widespread changes in copy number have been frequently noted in the investigation of SCLC [22, 70, 90, 91, 99, 266, 267]. Since the initial consistent deletion of chromosome 3p was noted, many further areas of the genome have been identified as frequently altered such as the loss of 13q and 17p and the amplification of 3q and 5p. The amplification of MYC family members is also frequently seen in SCLC and may be driving the growth of these tumours [2]. CNA analyses also allow a global view of the aberrations driving SCLC. The investigation of CNA in SCLC therefore represents a very relevant method of profiling SCLC and one of the aims of this thesis was to investigate CNA in both the CDX and CTCs. In addition to the frequent changes seen in copy number in SCLC, TP53 and RB1 are also noted to be commonly mutated [74, 80, 83, 85]. Deletion of TP53 and RB1 are used in the generation of GEMMs modelling SCLC indicating these are critical mutations for the development of SCLC and so very relevant to study [67]. The investigation of TP53 and RB1 point mutations in the CDX and CTCs was also therefore undertaken. The genomic profiling of the CDX and paired CTC samples was therefore an important aim of this thesis and the results of these experiments will be described in this chapter.

The results presented in this section of chapter 4 describe research published by Hodgkinson et al in Nature Medicine which represent research across the clinical and experimental pharmacology group (CEP) [24]. I was involved in two areas of the research in this paper. I was involved in setting up and gaining the ethical approval for the clinical trial, the subsequent recruitment of patients to the trial and the collection of their clinical data. I also performed the molecular analysis of both CDX tumours and

CTCs isolated from the same patients samples with the results of these experiments discussed in this chapter.

4A.2 Aims

1) To investigate CNAs in SCLC CDX models.
2) To compare the CNAs seen in the SCLC CDX models to published data from SCLC tumours.
3) To investigate TP53 and RB1 mutation status in the SCLC CDX models.
4) To investigate CNA and point mutations in TP53 in the CTCs isolated from the SCLC patients whose blood samples were used to create CDX.
5) To compare the genomic profiles of CDX models with CTCs isolated from the same patients with SCLC.

4A.3 Methods

General methods are detailed in chapter 2. Figure 2.1 provides an overview of the CTC processing workflow used. Details on individual methods used in this section can be found as follows

1) Patient and clinical sample collection – see section 2.1.
2) CTC enrichment and isolation – see section 2.2.
3) WGA of CTC DNA - see section 2.3.
4) DNA Extraction - see section 2.4.
5) DNA library Preparation - see section 2.5.
6) Next Generation Sequencing - see section 2.6.
7) PCR and Sanger Sequencing - see section 2.7.
8) Bioinformatics and statistical considerations – see section 2.10.

4A.4 Results

4A.4.1 Patient Characteristics

Blood samples were obtained from six patients recruited as previously described in section 2.1 between August 2012 and February 2013. The patients all had histologically or cytologically confirmed ES SCLC and were chemotherapy naïve. There were four females and two males recruited to the study (table 4.1). The median age of the patients was 69 years old (range 56 to 78). The patients all had strong histories of tobacco smoking, as would be typical of patients diagnosed with SCLC,

with a mean smoking history of 47 pack years. The patients ranged from performance status 1 to 3. The patients all received treatment with platinum-based chemotherapy. Three of the patients had chemoresponsive disease and three patients had chemorefractory disease as they had documented disease progression within 3 months of completion of chemotherapy. The four patients whose samples generated CDX had CTC counts in the paired blood sample processed using CellSearch of 458 or greater whilst the two patients in whom CDX did not develop had CTC counts of 20 and 222.

Patient	CTC count / 7.5ml[1]	Metastatic sites	Chemo-responsive/ refractory	Patient Survival (months)[2]	Time from CTC implantation to palpable tumor (months)
P1	458	Bone, lung, lymph node	Responsive	7.3	4.4
P2	1625	Bone, brain, meningeal	Refractory	3.5	2.4
P3	507	Lymph node	Responsive	9.7	4.3
P4	1376	Liver, lung, lymph node	Refractory	0.9	3.2
P5	222	Liver, lymph node, pancreas	Refractory	1.7	No tumor at 13.7 (mouse culled)
P6	20	Lymph node, pleura	Responsive	13.4[3]	No tumor at 12.3 (mouse culled)

Table 4.1 Clinical characteristics of SCLC donors for CDX project (Reproduced from Hodgkinson et al 2014 [24]).
[1] CTC count performed on CellSearch platform.
[2] From date of CTC sample blood draw.
[3] Patient alive at last follow up.

4A.4.2 Genomic Analysis of CDX Tumours

Genomic analyses of tumour tissue isolated from the four CDX models and related patient samples were performed (figure 4.1). The CDX models and associated

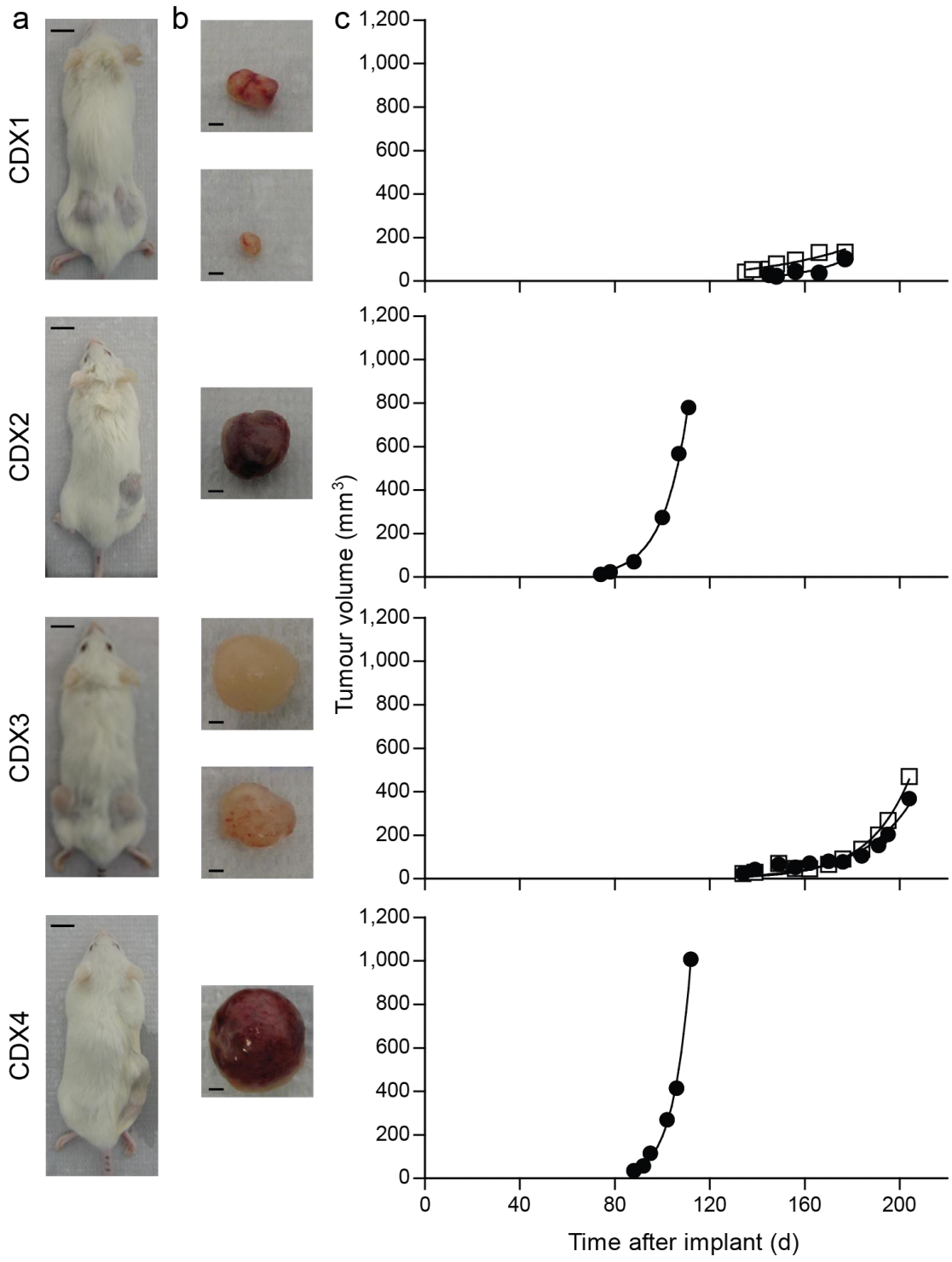


Figure 4.1 CDX models created from four patients with SCLC (Reproduced from Hodgkinson et al 2014 [24]). CTCs enriched from the blood of patients with SCLC were injected in to one or both flanks of immune compromised mice. The tumour-bearing mice from passage 1 are shown in panel a, with a scale bar indicating 1 cm. In panel b the CDX tumours at death are shown, with scale bar indicating 2 mm. In panel c the change in tumour volume after implant is shown in days. The black circles indicated the right flank tumour whilst the white squares indicate the left tumour and the solid line indicates the exponential growth line of best fit.

samples which underwent genomic analysis are detailed in table 4.2 with the nomenclature used to describe them in this chapter stated. CTCs were implanted in to both flanks of CDX1 and 3 resulting in two tumours, one from each flank whilst for CDX2 and 4 CTCs were only implanted into one flank resulting in just one tumour to analyse. WGS of the six tumour specimens and a genomic leukocyte DNA sample from patient 3 was carried out and used to generate CNA data as described in sections 2.6 and 2.10.2. All the CDX tumours analysed had multiple regions of change in copy number seen in contrast to the leukocyte genomic DNA sample M3G which had very few alterations in copy number (figure 4.2). Each CDX tumour had a unique pattern of losses and gains seen in the CNA analysis. There were however common regions of change across the six tumour samples reflecting the frequent changes seen in SCLC e.g. deletions in 10q, which contains PTEN, 13q, which contains RB1, and 17p, which contains TP53 [70, 91]. Interestingly the loss of 3p, the first commonly noted CNA seen in SCLC was seen in all the samples apart from CDX4 and M3G[90]. The regions of amplification were less consistent between the tumours.

Patient	CDX generated	CDX Model	Leukocyte Genomic DNA Sample	Paired CTCs profiled	Single CTCs	Pool of CTCs	Pool of WBCs
P1	2 tumours (left and right flanks)	CDX1 L and CDX1R	NA	No	NA	NA	NA
P2	1 tumour	CDX2	NA	Yes	P2CTC1-6	P2CTCP1-2	P2WBC
P3	2 tumours (left and right flanks)	CDX3L and CDX3R	Yes (M3G)	No	NA	NA	NA
P4	1 tumour	CDX4	NA	Yes	P4CTC1-2	P4CTCP1	P4WBC
P5	NA	NA	NA	No	NA	NA	NA
P6	NA	NA	NA	No	NA	NA	NA

Table 4.2 Summary of the samples generated and processed from the SCLC patients.
The CDX generated from and leukocyte genomic DNA, CTCs and WBC samples isolated from the six patients' samples used for the investigation of the CDX. The nomenclature used to describe these samples is given in the table.

CNA changes were analysed in the seven samples, six CDX tumours and one genomic control, across 13 genes identified as being commonly amplified or deleted in SCLC in previous publications [70, 91] (table 4.3). The presence of deletions in PTEN, RB1, TP53, FHIT and RASSF1 was assessed whilst amplifications in MYCL1, BCL2, CCNE1, MYCN, SOX2, EGFR, MYC and FGFR1 were also examined. In M3G all 13 genes had a normal copy number of 2, as expected. However, consistent loss of an

allele for each of RB1, TP53 and PTEN were seen in all the CDX tumours, whilst the loss of RASSF1 and FHIT was seen in all but CDX4. Regions of amplification were not as consistently seen across the six tumour samples with for example SOX2 amplification only identified in the four tumours from CDX1 and 3, whilst MYC amplification was only seen in the tumours from CDX1. These findings would be typical of the research of copy number change in SCLC in which regions of loss are common amongst all tumours with for example the loss of TP53 and RB1 being felt to be critical changes in the development of SCLC[67]. However, the amplifications potentially driving SCLC are more variable, with for example SOX2 amplification seen in only 27 % of the samples analysed in one large study [91].

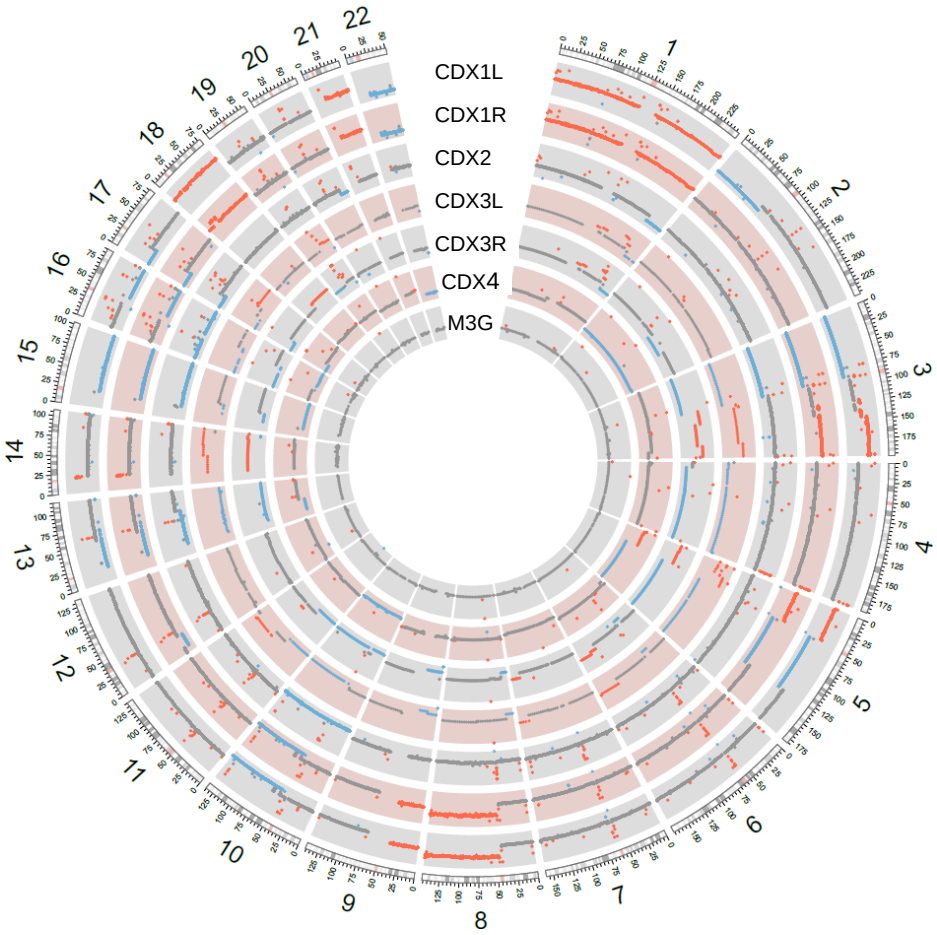


Figure 4.2 Circos plot of CNA data from CDXs and genomic sample from Patient 3 (Reproduced from Hodgkinson et al 2014 [24]). WGS of the left (L) and right (R) flank tumours of CDX1 and CDX3, the single flank tumours for CDX2 and CDX4 and the genomic sample from patient M3 (M3G) was performed. CNA analysis of the WGS data was then performed. Gains are seen in red and losses seen in blue whilst regions of no change are grey.

		Tumor DNA no WGA							WGA input 1 ng Genomic DNA						
		CDX1L	CDX1R	CDX2	CDX3L	CDX3R	CDX4	M3G	CDX1LA	CDX1RA	CDX2A	CDX3LA	CDX3RA	CDX4A	M3GA
Increased	*MYCL1*	3	3	2	2	2	2	2	3	3	2	2	2	2	3
	BCL2	4	3	2	2	3	2	2	5	3	2	2	3	2	2
	CCNE1	2	2	2	6	6	2	2	2	2	2	6	6	2	2
	MYCN	1	2	2	2	2	1	2	1	2	2	2	2	1	2
	SOX2	4	3	2	4	4	2	2	4	3	2	4	4	2	2
	EGFR	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	MYC	4	4	2	2	2	2	2	3	4	2	2	2	2	2
	FGFR1	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Decreased	*PTEN*	1	1	1	1	1	1	2	1	1	1	1	1	1	2
	RB1	1	1	1	1	1	1	2	1	1	1	1	1	1	2
	TP53	1	1	1	1	1	1	2	1	1	1	1	1	1	2
	FHIT	1	1	1	1	1	2	2	1	1	1	1	1	2	2
	RASSF1	1	1	1	1	1	2	2	1	1	1	1	1	4	2

Table 4.3 CNA analyses of 13 genes frequently amplified or lost in SCLC (Adapted from Hodgkinson et al 2014 [24]). Comparison of CNA of 13 frequently altered genes in the unamplified CDX and genomic samples to the amplified CDX and genomic samples (suffix A added to sample name). Gains are coloured red, losses are coloured blue and samples with two copies of a gene are coloured white.

Comparison of the left and right flank tumours of CDX 1 and 3 demonstrated they were largely similar. In CDX1L and CDX1R large regions of amplification present in chromosomes 9 and 21 are seen in both tumours but no other samples. In CDX3L and CDX3R regions of amplification in chromosome 14 and deletion in chromosome 9 are also seen in both the tumours, but no other samples. There were, however, some differences between the tumours isolated from the two flanks of the same mouse, for example CDX3R had a loss in 19p but normal copy number in 11q whilst CDX3L had normal copy number in 19p but an amplification seen in 11q. On a gene level a deletion is seen in CDX1L but not CDX1R in MYCN whilst an amplification is seen in BCL2 in CDX3R but not CDX3L. This may provide evidence of heterogeneity of the CTC populations which generated the tumours in the two flanks. It could, however, also represent that evolution of the tumours has occurred post implantation of the CTCs in the two flanks.

As two of the patients responded to platinum-based chemotherapy and two were refractory, and these responses were mirrored in the three CDX models tested, comparison of any differences between the chemoresponsive and chemorefractory tumours was made. In the CNA analysis of the two chemoresponsive CDX (1 and 3) to the two chemorefractory CDX (2 and 4) different patterns of losses and gains were seen. In CDX1 and CDX3 significant regions of both amplifications and losses were seen whilst in CDX2 and CDX4 a loss predominant pattern was seen. When examining the 13 genes frequently lost or amplified in SCLC in the chemoresponsive CDX there are multiple examples of both losses and amplifications. In the chemorefractory CDX tumours, in contrast, only deletions are seen. This analysis is only based on the results of analysing four CDX but provides an interesting difference to be validated in future chemoresponsive and chemorefractory samples.

Amplification of CTCs is necessary to generate enough DNA to act as the template for further analyses. To assess the impact of WGA on CNA data, WGA of 1ng input of DNA from the 6 CDX tumours and the M3G was performed using the Ampli1 system as described in section 2.3. DNA libraries of the samples were then made and low depth WGS performed on the MiSeq as described in sections 2.5 and 2.6. CNA analysis of the sequencing data from the seven samples was generated and compared to the data generated from the unamplified samples (table 4.3). In contrast to the unamplified M3 genomic sample in which the 13 frequently altered genes in SCLC all had just two copies, the amplified M3 genomic sample had an amplification identified in MYCL1 but otherwise all the genes had two copies. It is possible this represents an artefact from the amplification process. The patterns of copy number in the CDX tumours were

similar in both the amplified and unamplified samples. The regions of loss seen in PTEN, RB1, TP53, FHIT and RASSF1 were consistent between the amplified and unamplified CDX samples. In the unamplified CDX4 sample RASSF1 had two copies whilst in the amplified sample this gene was amplified. All the other genes that were amplified or deleted across the CDX tumours were seen in both the unamplified and amplified samples.

4A.4.3 TP53 Mutation Status in CDX Models

The tumour suppressor genes TP53 and RB1 are highly mutated in SCLC, with aberrations seen in up to 90% of cases [74, 80, 83, 85]. It is felt that changes in these two genes are critical in the evolution of SCLC with copy number changes in both these genes and mutations of TP53 being found in an early stage SCLC tumour that was WGS[88]. An investigation of whether mutations were present in these two genes in the CDX tumours was therefore undertaken. Patient-specific mutations were identified in both TP53 and RB1 from the NGS data and confirmed by Sanger sequencing (table 4.4). Each of the CDX models had a TP53 mutation sequenced which was unique to that model, in keeping with the diversity of mutations previously demonstrated in TP53 [78]. The mutations in TP53 were common to both the left and right flank tumours of CDX1 and CDX3, supporting the idea that these are early changes in the tumour evolution. RB1 mutations were only seen in CDX1 and CDX2 but not CDX3 and CDX4. However, all four models had loss of copy number in RB1 seen. The genomic sample M3G was wild-type for both TP53 and RB1 mutations, as would be predicted.

			Tumour Sequence[#]					
Gene	Nucleotide Change	Amino Acid Change	CDX1L	CDX1R	CDX2	CDX3L	CDX3R	CDX4
RB1	c.1597G>T	p.E533*	Mut	Mut	WT	WT	WT	WT
RB1	c.1963_1964insT	p.Y655Lfs*13	WT	WT	Mut	WT	WT	WT
TP53	c.733G>T	p.G245C	Mut	Mut	WT	WT	WT	WT
TP53	c.440T>G	p.V147G	WT	WT	Mut	WT	WT	WT
TP53	c.263A>G	p.Y220C	WT	WT	WT	Mut	Mut	WT
TP53	c.892G>T	p.E298*	WT	WT	WT	WT	WT	Mut

Table 4.4 TP53 and RB1 mutations in CDX models (Reproduced from Hodgkinson et al 2014 [24]). TP53 and RB1 mutations were identified in the CDX models by NGS and confirmed by Sanger sequencing. A mutation is coloured in purple (MUT) whilst the wild-type sequence is coloured turquoise (WT) in the table.

4.A.4.4 Verification of Single WBC CTC Approach

Both amplification and NGS can introduce errors into sequencing results [212, 215, 268-270]. As was discussed in section 3.4.2.2, many of the errors generated from the amplification process are likely to be consistent and predictable in nature as a result of the effects of MSE1 digestion of the DNA [237]. Errors introduced during NGS may be stochastic in nature or may represent the varied ability to sequence regions of the genome, such as GC rich regions [270]. An initial comparison of WBC and CTC samples was undertaken to assess the single cell CNA workflow used in an effort to reduce reproducible errors from amplification and sequencing being incorporated into subsequent analyses. Matched single and pooled WBCs and CTCs were isolated from two patients with SCLC who were not included in the CDX experiments. The samples were WGA and DNA libraries were then prepared which were subjected to NGS as described in chapter 2. The Illumina Miseq® data generated was then used to analyse CNAs within the samples. The CNA data generated from both cytoband and cancer-related genes were examined within principal component analyses (PCA) plots (figure 4.3). Clear separation of both pools and single CTCs and WBCs was seen supporting the approach used to identify these two types of cells. The WBCs and CTCs clustered with the same cell type, irrespective of if they were single or pooled samples. Potentially unreliable cytobands or genes were identified by examining the CNA data from the six WBC samples. Any cytobands or genes in which three or more of the samples deviated from a copy number of two were identified as potentially errors. Using this approach 0.8% of the cytobands and 1.1% of the cancer genes were removed from further analysis.

4A.4.5 Comparison of Genomic Profiles of CDX and CTCs Isolated from Patients with SCLC

A comparison of the genomic properties of the CDX to CTCs isolated from blood samples from the same patients was undertaken. The aims of the genomic profiling of the CTCs were to assess how they reflected the published profiles of SCLC tumours and to confirm that the profiles differed from WBCs processed in the same manner. An analysis to see if the CTCs profiles reflected the CDX tumours would also allow the investigation of the hypothesis that the CTCs enumerated by CellSearch potentially are tumorigenic in nature. This may provide further supporting evidence for why this group of cells impact on the prognosis of patients with SCLC. CTCs were enriched using the CellSearch system, and then single and pools of both CTCs and WBCs (to provide a germline sample) were isolated using the DEPArray from patients 2 and 4 for WGA, as

described in section 2.2. The cells were then amplified and a quality control analysis performed to assess the success of amplification (section 2.3). From patient 2's sample 8 single CTCs, 2 pools of 10 CTCs and 2 pools of 10 WBCs were isolated of which all but 1 of the single CTCs were successfully amplified (figure 4.4A). For further processing for NGS 6 single CTCs, 2 pools of 10 CTCs and 1 pool of 10 WBCs were used. From patient 4's sample 8 single CTCs, 1 group of 10 CTCs, 3 single WBCs and 1 pool of WBCs were isolated. Of the isolated cells only 4 of the single CTCs, the pool of CTCs, 1 single WBC and the pool of WBCs amplified successfully when the QC was performed (figure 4.4B). For NGS analysis 2 single CTCs, the pool of CTCs and the pool of WBCs were processed further.

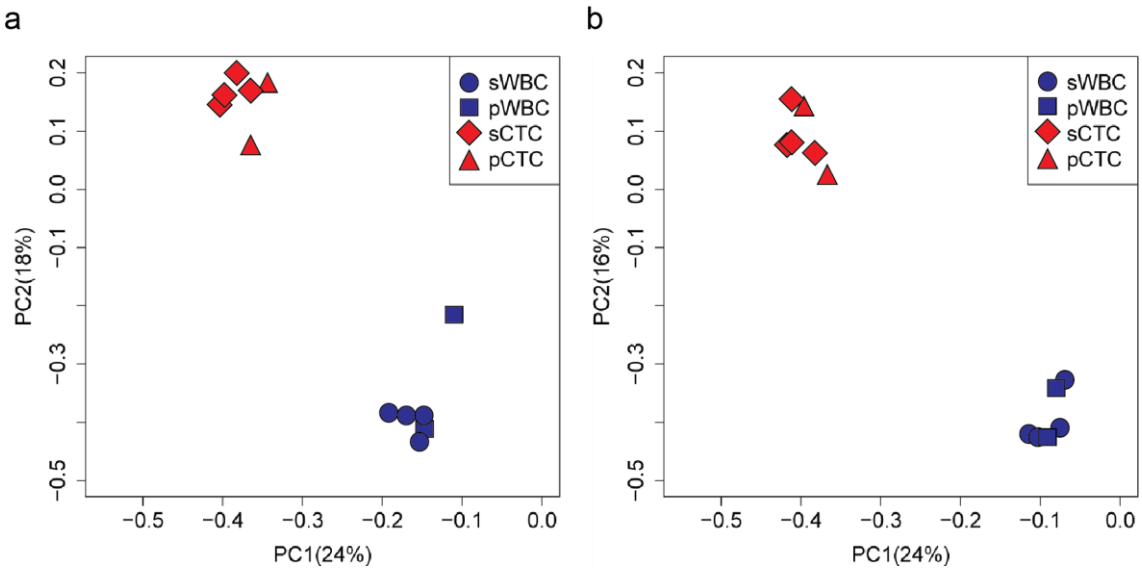


Figure 4.3 PCA of CNA data from matched WBC and SCLC CTC samples. (Reproduced from Hodgkinson et al. 2014 [24]) From two patients with SCLC, two pooled WBC samples (pWBC blue squares), four single WBCs (sWBC blue circles), two pooled CTC samples (pCTC red triangles) and four single CTCs (sCTC red diamonds) were analysed for CNAs. PCA of the CNA data from either cytobands or cancer-related genes was then generated.

a) PCA of cytoband CNA data with the percentage variance for the components given in parenthesis.
b) PCA of cancer gene CNA data with percentage variance for the components given in parenthesis.

CNA data was generated for the CTC samples as it was for both the CDX tumours and the amplified CDX tumours from the 6,341 cancer-related genes, to allow comparison of the patterns seen. Hierarchical clustering of the CNA data was carried out with these samples (figure 4.5). The CNA profiles of the WBCs were markedly different to all the CDX and CTCs with very minimal CNAs seen. These copy number changes may represent stochastic artefacts from the amplification process which did not occur

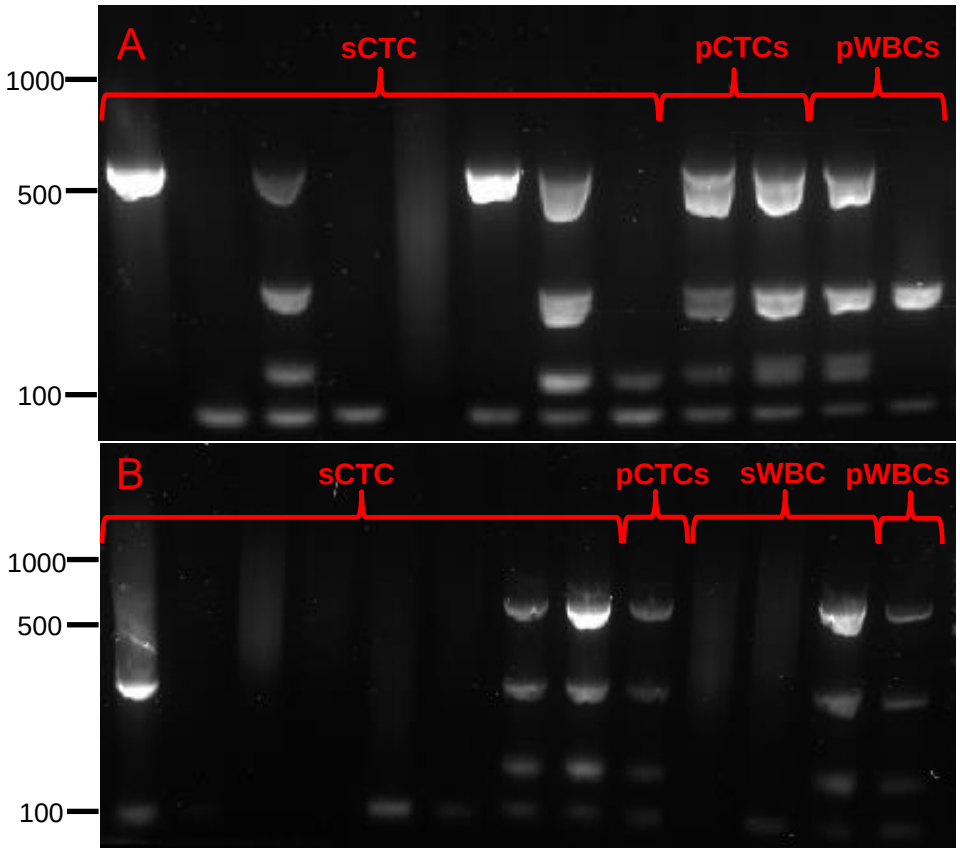


Figure 4.4 Ampli1 QC of CTCs and WBCs from Patients 2 and 4.
Single CTCs (sCTC), pools of CTCs (pCTCs), single WBCs (sWBC) and pools of WBCs (pWBCs) were isolated from patients 2 and 4's blood samples using the DEPArray, after enrichment by CellSearch. The cells were amplified using the Ampli1 WGA kit. The quality of the amplification was then assessed with the Ampli1 QC kit. The Ampli1 QC kit is a multiplex PCR for 4 amplicons. The PCR products are assessed on 1.5% agarose gel containing a fluorescent DNA stain. The expected products sizes are 91, 108-166, 299 and 614 bp in size. The results of the QC for patient 2 are in panel A and the results for patient 4 are in panel B.

in the WBCs used during the quality control analysis in section 4A.4.4. The two WBC samples cluster together in the hierarchical clustering and cluster separately to all the other samples. This highlights the ability of the approach used to distinguish CTCs and WBCs for isolation and amplification. The CTC samples from each patient and the relevant CDX model group together in the hierarchal clustering with the amplified tumours clustering next to the unamplified tumours. This demonstrates the similarity of the CTCs to the CDX tumours generated from the paired samples. The CNA profile of the CTCs isolated from a patient reflected the CNA profile of the CDX tumours with similar patterns of gains and losses seen. This suggests that there are similarities between the CTCs profiled and those which created the tumours and therefore may support the idea that these CTCs are potentially tumorigenic. There was however some heterogeneity of the CNA profile seen between the CTCs and CDX tumours from

a patient, for example there are regions of amplification seen in the CTCs from patient 2 but not in the CDX2. This may be due to evolution of the tumour post implantation. It may also suggest that only a subpopulation of the CTCs present lead to the formation of tumours. There is heterogeneity not only between the CTCs and CDXs from a patient but also within the CTCs isolated from an individual patient. CTC 5 in patient 2 had a markedly different profile to the other samples from patient 2, and although it clusters with the other samples from patient 2, it is in a distinct branch. Of note the DEPArray images of CTC 5 also suggested it was a larger cell and had a larger nucleus than the other five single CTCs processed from this patient, but in other respects had the expected staining and appearance of a CTC. The variation in size of CTC however may have contributed to some of the differences seen.

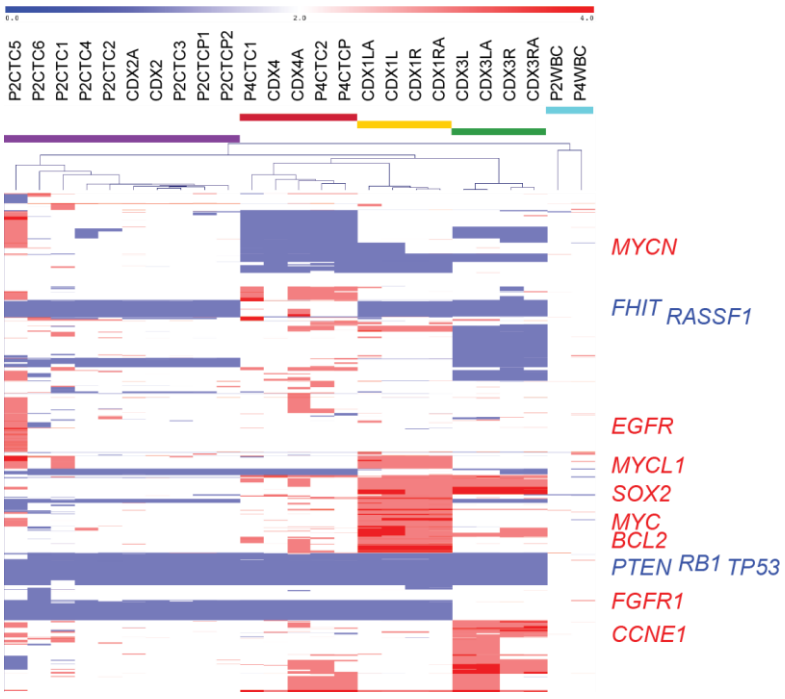


Figure 4.5 Hierarchical clustering of CNA data from the CDX models and the corresponding patients' CTCs. (Reproduced from Hodgkinson et al 2014 [24]). Single CTCs, and pools of CTCs and WBCs and 1ng of DNA isolated from the CDX tumours were WGA and NGS of the DNA libraries performed. CNA analysis was performed on the amplified CTCs and WBCs and both the unamplified and amplified CDX tumour DNA, using sequencing data from 6,341 cancer-related genes. Gains are coloured red, losses are coloured blue and regions with no change in copy number are coloured white. The positions of 13 genes commonly amplified or lost in SCLC are indicated to the right of the heatmap. Hierarchical clustering of the samples copy number analysis was carried out.

PCA of the CNA data highlighted the similarities between the CTCs isolated from a patient and the CDX model generated from the patient's CTCs. The CTCs from patient 4 and CDX 4 all clustered tightly together (figure 4.6). The majority of the CTCs from

patient 2 cluster with CDX 2, apart from CTC 5 whose different CNA profile was discussed above. This again suggests there is some heterogeneity seen amongst the CTCs. The two WBCs samples from patients 2 and 4 cluster together and cluster separately to the tumours and CTCs, confirming they have a distinct profile.

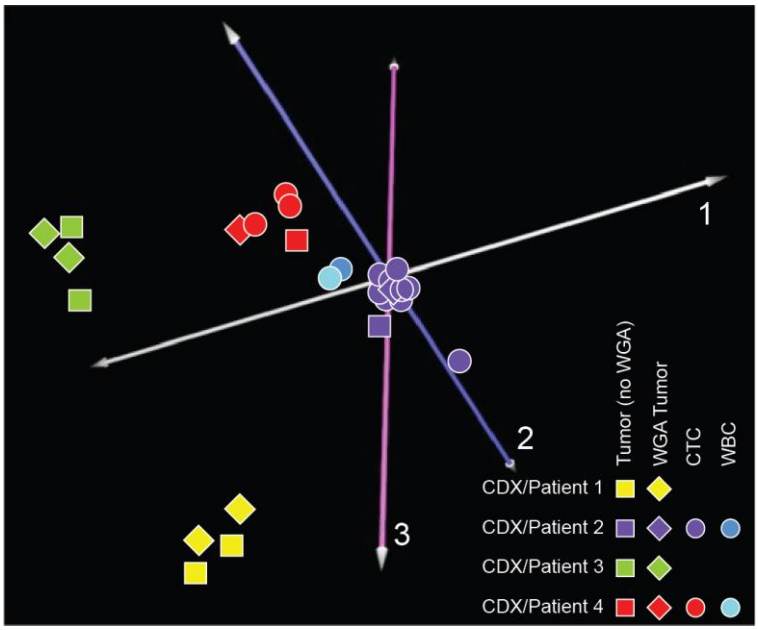


Figure 4.6 PCA of CNA data from CDX models and patient CTCs. (Reproduced from Hodgkinson et al 2014 [24]). Single CTCs and pools of CTCs and WBCs and 1ng of DNA isolated from the CDX tumours were whole genome amplified. DNA libraries were prepared from the samples and NGS performed. CNA analysis was performed on the amplified CTCs and WBCs and both the unamplified and amplified CDX tumour DNA using sequencing data from 6,341 cancer-related genes. PCA analysis of components 1, 2 and 3 was then performed.

The pattern of changes in copy number in the 13 commonly altered genes in SCLC previously examined in the CDX models was also investigated in the CTCs isolated from patients 2 and 4 (table 4.5). The WBCs isolated from the two patients had no copy number changes in the 13 genes examined. The CNA patterns of the CTCs were similar to the relevant CDX models, particularly with respect to loss of copy number. There was some heterogeneity of amplifications seen, for example MYCL1 was not amplified in CDX2 but was amplified in P2 CTC1 and P2 CTC5. An amplification was seen in P2CTC4 in BCL2 but not in the other CTCs profiled from this patient.

Mutations in TP53 and RB1 had previously been identified in CDX2 and CDX4 (section 4A.4.3). The mutations present in TP53 across the two patients' CTCs and mutations in RB1 in patient 2's CTCs were analysed by Sanger sequencing, to see if the same

		CDX2	CDX2A	P2CTC1	P2CTC2	P2CTC3	P2CTC4	P2CTC5	P2CTC6	P2CTCP1	P2CTCP2	P2WBC	CDX4	CDX4A	P4CTC1	P4CTC2	P4CTCP	P4WBC
Increased	*MYCL1*	2	2	3	2	2	2	3	2	2	2	2	2	2	2	2	2	2
	BCL2	2	2	2	2	2	3	2	1	2	2	2	2	2	2	2	2	2
	CCNE1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	MYCN	2	2	2	2	2	2	3	2	2	2	2	1	1	1	1	1	2
	SOX2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	EGFR	2	2	2	2	2	2	3	1	2	2	2	2	2	2	2	2	2
	MYC	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	FGFR1	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2
Decreased	*PTEN*	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2
	RB1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	2
	TP53	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2
	FHIT	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2
	RASSF1	1	1	1	1	1	1	1	1	1	1	2	2	4	2	2	2	2

Table 4.5 CNA analyses of the 13 genes frequently amplified or lost in SCLC in CDXs and CTCs (Adapted from Hodgkinson et al 2014 [24]). Comparison of copy number in 13 genes in the unamplified CDX and genomic samples to the amplified CDX and genomic samples (suffix A added to sample name) and the WGA CTCs isolated from the corresponding patients samples. DNA libraries were prepared and the CNA data generated using sequencing data from 6,341 cancer-related genes. Gains are coloured red, losses are coloured blue and genes with two copies are coloured white.

mutations could be identified as in the CDX (figure 4.7 and 4.8). The TP53 and RB1 mutations identified in CDX 2 were identified in all the CTCs from patient 2 in which the Sanger sequencing was successful whilst the WBC samples had the wild-type sequence for both these alleles. The TP53 mutation identified in CDX4 was also confirmed as present in patient 4's CTCs, but absent from the WBCs. This again supports the idea that TP53 and RB1 mutations are early changes in SCLC as they were identified in all the samples analysed from a patient.

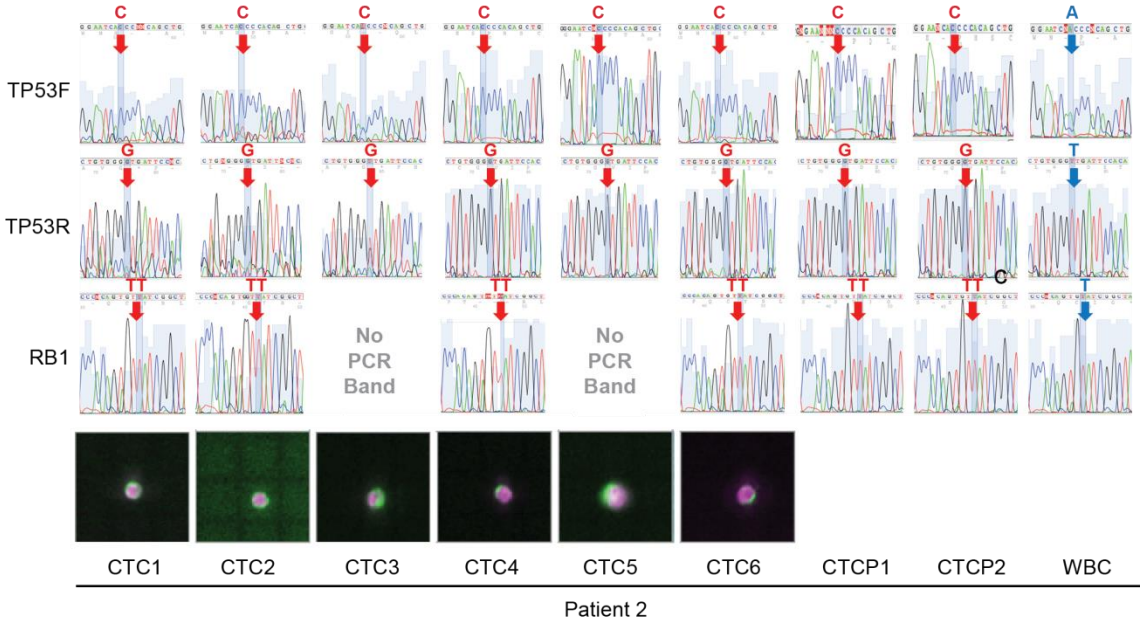


Figure 4.7 TP53 and RB1 mutations in Patient 2's CTCs. (Reproduced from Hodgkinson et al 2014 [24]). Single CTCs and pools of CTCs and WBCs isolated from patient 2's blood samples were WGA. Sanger sequencing was performed to investigate whether TP53 and RB1 mutations identified in CDX2 were present in the CTCs and WBCs isolated from patient 2's paired blood sample. The composite images of the cytokeratin and DAPI staining from the DEPArray for the single CTCs are presented. Sanger sequencing was carried out in duplicate with representative traces shown. Red arrows indicate somatic mutations whilst blue arrows indicate the wild-type allele.

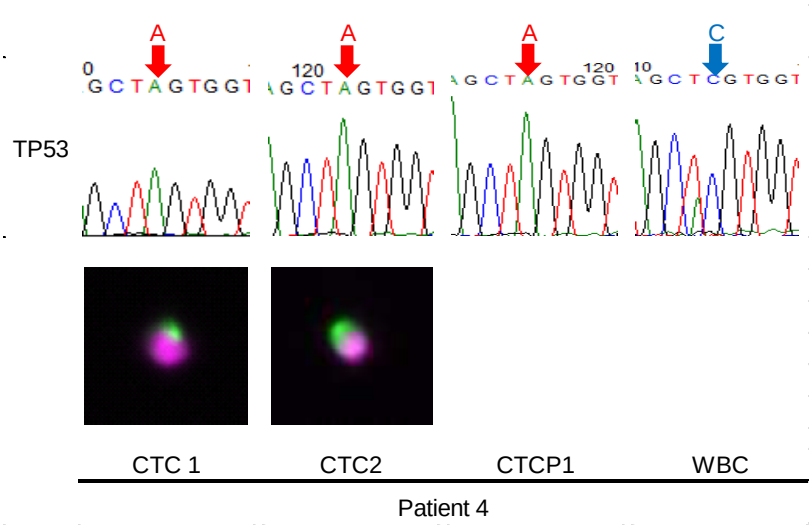


Figure 4.8 TP53 mutations in Patient 4's CTCs. Single CTCs and pools of CTCs and WBCs isolated from patient 4's blood samples were WGA. Sanger sequencing was performed to investigate whether TP53 mutations identified in CDX4 were present in the CTCs and WBCs isolated from patient 4's paired blood sample. The composite images of the cytokeratin and DAPI staining from the DEPArray for the single CTCs are presented. Sanger sequencing was carried out in duplicate with representative traces shown. Red arrows indicate somatic mutations whilst blue arrows indicate the wild-type allele.

4B Focussed Next Generation Sequencing of SCLC CTCs

4B.1 Introduction

Analyses of both CDX and CTCs provide methods for studying the biology of SCLC, potentially revealing important functional or druggable aberrations. To expand these analyses out to larger groups of patients and monitor changes over time 'clinic friendly' methods for genomic analyses are required. WGS captures mutational changes across the entire coding and non-coding genome allowing the identification of new mutations. However, WGS has significant sequencing costs and the analysis of the data requires significant bioinformatics input. Traditional Sanger sequencing provides a low costs means of sequencing, but in each reaction only a single amplicon of interest can be examined. Focussed NGS allows the analysis of a panel of genes of interest providing a compromise in terms of coverage and costs of sequencing. As only a small region of the genome is covered with focussed NGS, increased depth of sequencing can be achieved without dramatically increasing the cost of the sequencing. However, unlike with WGS in which a wide range of potential genetic abnormalities are assessed, in focussed NGS a panel of genes of interest must already have been identified for further assessment.

One type of focussed NGS, tagged-amplicon deep sequencing (TAm-Seq), is a robust method initially developed to analyse mutations within a targeted panel of genes from cfDNA [271], thereby enabling amplification and sequencing of targeted regions of the genome with small inputs of DNA. Primers were designed to sequence short overlapping amplicons in the coding regions of TP53 and PTEN and selected regions of EGFR, BRAF, KRAS and PIK3CA. The DNA is initially amplified in a 15 cycle preamplification step with a pool of target-specific primer pairs which allow preservation of the alleles in the template DNA. The regions of interest are then amplified in singleplex PCRs. A further PCR is then performed to attach sequencing adaptors and sample-specific barcodes before 100-base single-end sequencing is performed. This approach was used with cfDNA isolated from the blood of patients with advanced ovarian cancer to identify TP53 mutations, with all the mutations identified by digital PCR also identified by TAm-Seq, supporting the utility of this approach in the analysis of clinical samples [271]. An EGFR mutation was identified in the cfDNA from one of the ovarian cancer patients which had not been identified within the initial ovarian biopsy. The EGFR mutation was subsequently identified in an omental metastasis demonstrating that this approach could be used to identify potential treatment options for patients. This finding further highlights the utility of liquid biopsies to reveal genomic alterations from multiple sites rather than just the single biopsy site. TAm-Seq has not however been used to analyse mutations present in CTCs to date having thus far only being applied to FFPE tissue and cfDNA.

One of the broader aims of the thesis was to identify from CNA analysis and WES genes that were differentially altered in patients with SCLC according to their response to therapy. If a panel of mutations were identified through WES in SCLC CTCs, focussed NGS would provide a method of examining these changes in a wider cohort of patients. Focussed NGS would be more ‘clinic friendly’ in terms of the costs and time for sample processing and bioinformatic analysis than either WGS or WES of CTCs. Focussed NGS would therefore have the potential to be applied routinely to CTCs isolated from SCLC patients, for example, to assess whether a patient would be likely to be chemoresponsive or chemorefractory. The results of a feasibility study of whether TAm-Seq could be used to investigate mutations in a panel of genes in SCLC CTCs are presented in this chapter as a demonstration of the potential applicability of focussed NGS.

4B.2 Aims

1) To establish the use of TAm-Seq in CTCs isolated from patients with SCLC.
2) To compare CNA and TAm-Seq data from CTCs isolated from patients with SCLC.

4B.3 Methods

General methods are detailed in chapter 2. Figure 2.1 provides an overview of the CTC processing workflow used. Details on individual methods used in this section which are described in the Materials and Method chapter can be found as follows

1) Patient and clinical sample collection – see section 2.1.
2) CTC enrichment and isolation – see section 2.2.
3) WGA of CTC DNA – see section 2.3.
4) DNA Extraction – see section 2.4.
5) DNA library Preparation – see section 2.5.
6) Next Generation Sequencing – see section 2.6.
7) PCR and Sanger Sequencing – see section 2.7.
8) Bioinformatics and statistical considerations – see section 2.10.

4B.3.1 Reamplification of WGA DNA

CTCs and WBCs were isolated and WGA from six patients with SCLC blood samples as per sections 2.2 and 2.3. In total 66 samples of cells, including single and pools of both CTCs and WBCs were processed (see table 4.6 for a summary of the CTC and WBC samples analysed). Of note single WBCs were collected from patient 6 alone as the protocol of the cells isolated per patient was altered at this point to enable the analysis of single WBCs as controls for the single CTCs in addition to the pools of WBCs previously isolated. The WGA DNA from the CTCs and WBCs was then reamplified using the Ampli1™ WGA kit (Silicon Biosystems) according to the re-amplification of Ampli1™ WGA product protocol. Briefly, 1 µl of the purified WGA product was reamplified using the Ampli1™ enzyme and buffer reamplification mix in a total volume of 50 µl. The samples were then incubated on the Veriti thermal cycler. An initial cycle of 94 °C for 60 sec., then 60 °C for 30 sec. and then 72 °C for 120 sec. was followed by 10 cycles of 94 °C for 30 sec., 60 °C for 30 sec. and 72 °C for 120 sec. (increasing by 20 sec/cycle). The reamplified WGA products were purified with the High Pure PCR Product Purification kit (Roche). The samples were processed according to the “Purification of PCR products in solution after amplification” protocol,

eluting into a 50 µl final volume. The DNA concentrations were measured with the Nanodrop ND-1000 spectrophotometer as per the manufacturer's instructions with the yield of DNA ranging from 1.3 µg to 3.5 µg.

			Number of cells recovered for analysis			
Patient	**Number of CTCs identified by CellSearch**	**Number of CTCS identified by DEP-Array**	**Single CTC**	**Pool of 10 CTCs**	**Single WBC**	**Pool of 10 WBC**
1	170	56	8	1	0	2
2	1356	295	8	3	0	2
3	20815	343	8	2	0	2
4	1408	111	8	2	0	2
5	522	63	8	2	0	2
6	1200	312	8	1	3	1

Table 4.6 CTC counts by CellSearch and DEPArray for six SCLC patients whose CTCs were analysed by TAm-Seq. CTCs were enumerated in 7.5mls of blood for each of six patients with SCLC using the CellSearch platform. Individual CTCs and WBCs were then identified on the DEPArray and isolated as single cells or pools of cells.

4B3.2 TAm-Seq Analysis

The TAm-Seq method and its analysis has been described previously [271] and was performed by collaborators T Forshew and N Rosenfeld and colleagues at the Cancer Research (CRUK) Cambridge Institute. TAm-Seq analysis consisted of the following steps - primer design, preamplification for TAm-Seq, target-specific amplification on the access array microfluidic system, sequencing adaptor and barcode primer addition, quantification and clean-up of DNA library, sequencing and analysis of sequencing data. Briefly, target specific primers were designed using Primer3 with 5' universal primer sequences (termed CS1 and CS2). Amplicons were designed to sequence regions of the TP53, PTEN, EGFR, PIK3CA, NRAS, BRAF, CTNNB1 and KRAS genes. These genes were selected as they are frequently mutated in lung cancer (both NSCLC and SCLC). Amplicons were designed to be situated within TTAA sites due to the MSE1 digestion of the DNA during the WGA method used [215]. The primer

sequences are detailed in appendix 2. Barcode primers were designed to include the PE1 or PE2 sequences used in Illumina cluster generation, a 10 bp barcode followed by either CS1 or CS2. For the preamplification reactions 1 – 5 µl of the target DNA was amplified with 50 nM of each forward and reverse target-specific primer. The samples were incubated at 95 °C for 10 min., and then underwent 15 cycles of heating to 95 °C for 15 sec. and 60 °C for 4 min. The samples were then treated with Exo-SAP-it (Affymetrix) as a clean-up step.

The target-specific amplification was performed on the Access Array IFC (Fluidigm) with individual primer pairs loaded in to the primer inlets. For the amplification 1 µl of each sample was loaded after being diluted 5 fold, and processed with the following thermal cycling conditions - 35 cycles of amplification (50 °C 2 min., 70 °C 20 min., 95 °C 10 min., 10 cycles of 95 °C 15 sec., 60 °C 30 sec., 72 °C 60 sec., 2 cycles of 95 °C 15 sec., 80 °C 30 sec., 60 °C 30 sec., 72 °C 60 sec., 8 cycles of 95 °C 15 sec., 60 °C 30 sec., 72 °C 60 sec., 2 cycles of 95 °C 15 sec., 80 °C 30 sec., 60 °C 30 sec., 72 °C 60 sec., 8 cycles of 95 °C 15 sec., 60 °C 30 sec., 72 °C 60 sec., 5 cycles of 95 °C 15 sec., 80 °C 30 sec., 60 °C 30 sec., 72 °C 60 sec., 1 cycle of 72 °C for 3 min. The harvested product (1 µl) was then transferred to 2 PCR plates and diluted 100 fold. The sequence adaptors and barcodes were then added to the samples with the samples amplified heating to 95 °C for 10 min. then 15 cycles of heating to 95 °C for 15 sec., 60 °C for 30 sec. and 72 °C for 4 min. followed by one cycle of heating to 72 °C for 3 min. The samples were then analysed using the 2100 BioAnalyzer (Agilent). The samples were pooled together and purified with SPRI beads. The libraries were then checked again with the BioAnalyzer and subjected to cluster generation and sequenced using the Illumin GAIIx sequencer. The reads were then demultiplexed and aligned to the human genome (hg 19) using bwa-short [272]. The aligned reads were then assigned to each amplicon and non-reference alleles identified, assessing the duplicates for each sample.

4B.4 Results

4B.4.1 Patient Characteristics

Blood samples were obtained from six patients with ES SCLC recruited as described in section 2.1 from October to November 2012. The patients all had histologically or cytologically confirmed SCLC and were chemotherapy naïve. There were three men and three women recruited for the analysis of SCLC CTCs utilising TAm-Seq. The patients had a median age of 63 years and 6 months (range 57 – 73) (table 4.7). The

range of performance statuses of the patients was from 1 to 3. The patients were all heavy smokers with the number of pack years ranging from 30 to 130. The patients all subsequently received chemotherapy being treated with either carboplatin alone or in combination with etoposide. Response to treatment was determined with a CT scan at the end of treatment with chemotherapy, with two patients responding to therapy and three being refractory to chemotherapy. One patient developed neutropenic sepsis and died before a response assessment had been carried out. OS ranged from 1 month to 14.3 months for the six patients recruited.

Patient	Age	Gender	PS	Metastatic Sites	Smoking status (pack years)	First Line therapy	Chemo-responsive /refractory	Overall survival (months)[1]
1	70	M	2	pleural effusion, lymph nodes	130	Carboplatin/ Etoposide	Refractory	3.7
2	64	F	2	lung	80	Carboplatin/ Etoposide	NA *	1
3	61	M	1	liver, bone, skin	90	Carboplatin/ Etoposide	Refractory	4.2
4	63	F	2	brain, liver bone, pleural effusion	84	Carboplatin/ Etoposide	Refractory	3.5
5	73	M	3	adrenal, lymph nodes	30	Carboplatin	Responsive	6.5
6	57	F	2	liver, bone, lymph node	64	Carboplatin/ Etoposide	Responsive	14. 3

Table 4.7 Characteristics of six patients with SCLC whose CTCs were analysed using TAm-Seq

[1] From date of CTC sample blood draw.

* Died after two cycles of chemotherapy from neutropenic sepsis before response assessment carried out.

4B.4.2 Isolation of CTCs from SCLC Patients for Analysis

CTCs were enumerated on the CellSearch system, which is used to identify EpCAM and CK positive CTCs, with the number of CTCs ranging from 170 to 20815 per 7.5mls of blood for the six patients analysed (table 4.6). The enriched CTC samples were

then processed on the DEPArray to isolate pure samples of CTCs or WBCs as per section 2.2. Lower numbers of CTCs are identified on the DEPArray than CellSearch, due to loss in processing and the dead volume of the cartridge, with the counts ranging from 56 to 343. This represents from 2 to 33% of the number of CTCs identified by CellSearch. The cells isolated from each enriched CTC sample for amplification are detailed in table 4.6. In total 66 of the single and pools of cells amplified were taken forward for TAm-Seq analysis based on the amplification quality control analysis and to ensure representation of each patients samples.

4B.4.3 Amplicon Detection using TAM-Seq

Re-amplified DNA from 66 samples of single and pools of cells isolated from six patients was analysed utilising TAm-Seq to investigate mutations in a test panel of genes. In total 45 single CTCs, 11 pools of CTCs, 2 single WBCs and 8 pools of WBCs were analysed from the 6 patients utilising amplicons covering the 8 genes of interest (table 4.8). The average read depth was approximately 1000-3000 times depth for each amplicon. Successful sequencing of an amplicon was defined as an amplicon being sequenced more than 50 times. The average percentage of amplicons successfully sequenced across the six patients was 78% for the single cells and 92% for the pools of cells. The efficiency of sequencing the amplicons varied amongst the 8 genes with just 57% of the CTNNB1 amplicons from single CTCs being successfully sequenced whilst in PIK3CA 90% of the amplicons were successfully sequenced (table 4.8). There was also a sample specific effect with 95% of the amplicons from patient 6 being sequenced whilst from patient 2 only 78% were sequenced (table 4.8). These results highlight the fact that TAm-Seq can be used to sequence SCLC CTCs for a panel of amplicons of interest, but that there will be both variability in the success of individual amplicons and different patients' samples.

4B.4.4 Detection of TP53 mutations Using TAM-Seq in SCLC CTCs

Unique TP53 mutations were identified in each of the patients' CTCs but no mutations were identified in the other seven genes sequenced with TAM-Seq. The TP53 mutation were identified in the pools of CTCs from each patient by TAm-Seq but the WBCs were all wild-type for TP53 mutations (see table 4.9A). The point mutations were all missense mutations located between exons 5 and 8 in the DNA binding domain.

Gene	Number of Amplicons	Patient 1		Patient 2		Patient 3		Patient 4		Patient 5		Patient 6	
		single cell (n=8)	pool of cells (n=3)	single cell (n=8)	pool of cells (n=3)	single cell (n=8)	pool of cells (n=3)	single cell (n=6)	pool of cells (n=3)	single cell (n= 7)	pool of cells (n=4)	single cell (n=8)	pool of cells (n=4)
BRAF	1	62.5	66.7	75.0	100.0	87.5	66.7	83.3	33.3	100	100	100	75
CTNNB1	1	37.5	100	50.0	66.7	75.0	100	50.0	100	42.9	100	87.5	100.0
EGFR	12	69.8	100.0	65.6	77.8	81.5	88.9	80.6	100.0	92.9	100	100	87.5
KRAS	4	75.0	83.3	71.9	100	80.6	100	87.5	100	92.9	100	100	100
NRAS	2	81.3	83.3	75.0	100	83.3	100	83.3	66.7	92.9	100	100	100
PIK3CA	2	93.8	100	81.3	100	88.9	100	75.0	100	100	100	100	87.5
PTEN	5	85.0	86.7	72.5	93.3	82.2	93.3	63.3	93.3	97.1	95.0	97.5	95.0
TP53	15	45.0	100	51.7	71.1	60.7	88.9	40.0	100	66.7	100	87.5	100

Table 4.8 Percentage of TAm-Seq amplicons successfully sequenced. TAm-Seq analyses was performed on single and pools of reamplified WGA CTCs and WBCs isolated from patients with SCLC. The success of sequencing amplicons from the 8 genes of interest for both single and pools of cells was calculated, with an amplicons being classified as successfully sequenced if there were more than 50 reads generated for that amplicon. The number of single and pools of cells processed for each patient is indicated in the table. The number of amplicons sequenced per gene is also indicated in the table. The success of the sequencing is indicated with samples with low percentages of samples successfully sequenced indicated in blue and high percentages of samples successfully sequenced in red.

A

Patient numbers	1			2			3				4			5				6			
Sample	pCTC1	pWBC1	pWBC2	pCTC1	pCTC2	pWBC1	pCTC1	pCTC2	pWBC1	pWBC2	pCTC1	pCTC2	pWBC1	pCTC1	pCTC2	pWBC1	pWBC2	pCTC1	sWBC1	sWBC2	pWBC2
TP53 p.V157F																					
TP53 p.G245C																					
TP53 p.Y163C																					
TP53 p.V147G																					
TP53 p.K132E																					
TP53 p.R282P																					

B

Patient numbers	1								2								3								4						5							6							
CTC number	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	1	2	3	4	5	6	7	1	2	3	4	5	6	7	8
TP53 p.V157F																																													
TP53 p.G245C																																													
TP53 p.Y163C																																													
TP53 p.V147G																																													
TP53 p.K132E																																													
TP53 p.R282P																																													

Table 4.9 TP53 mutations identified by TAm-Seq in CTCs isolated from six patients with SCLC.

A TP53 mutations identified in pools of CTCs and both pools and single WBCs from six patients with SCLC.
TP53 mutations were identified by TAm-Seq in pools of reamplified WGA CTCs and WBCs from six patients with SCLC.

B TP53 mutations identified in single CTCs isolated from six patients with SCLC.
TP53 mutations identified by TAm-Seq in reamplified WGA single CTCs from six patients with SCLC.

Mutant homozygous results are brown, heterozygous results are blue whilst wild-type sequences are green. Failed sequencing is shown in yellow (less than 50 reads for an amplicon). The results were confirmed by Sanger sequencing.

Key

Wild type
Mutant Homozygous
Heterozygous
Failed

The specific TP53 mutation identified in the pools of CTCs in each patient was also identified in some of the single CTCs analysed from that patient (table 4.9B). There were no examples of more than one TP53 mutation being identified in an individual patient's samples. However, amongst the single CTCs heterogeneity of TP53 mutations were noted, with for example patient 6's CTCs being mutant homozygous, heterozygous and wild-type for the TP53 p.R282P mutation. This suggests that the heterogeneity noted in mutations in primary tumours is also present in CTCs. The TP53 mutations identified by TAm-Seq were all confirmed by Sanger sequencing to be present in the CTCs but absent from the WBCs supporting that this approach is both sensitive and specific. These data confirm that TAm-Seq can successfully be used to analyse mutations in a panel of genes of interest in CTCs isolated from SCLC patients.

4B.4.5 Comparison of Results of CNA and TAm-Seq from CTCs Isolated from Patients with SCLC

DNA from the CTCs and WBCs isolated from the six patients with SCLC was also analysed for CNA in addition to the TAm-Seq analysis (sections 2.5, 2.6 and 2.10.2). The CNA data were generated from sequencing the samples using the MiSeq, analysing 6,341 selected cancer-related genes. Hierarchical clustering of the CNA data generated from the CTCs was then performed (figure 4.9). In the hierarchical clustering the CTCs isolated from each patient clustered together, confirming the similarity of all the CTCs isolated from a patient. There was clear heterogeneity of the CTC CNA profiles seen between the six patients examined. There were few consistent regions of genes amplified across the CTCs analysed but there was one large region of losses present in all the samples from the six patients. Genes that had consistent loss of copy number amongst the CTCs included TP53, FHIT and RASSF1. This suggests that for these six patients the CTCs analysed had all lost one copy of TP53 as indicated by the CNA profiles and the other copy has a mutation in the DNA binding domain as demonstrated in the TAm-Seq analysis. These data demonstrate the additive value of both the CNA and TAm-Seq data in the investigation of the genomic profiles and potentially the biology of SCLC CTCs. The heterozygous TP53 mutations seen in the TAm-Seq results for P2 sCTC5, P6 sCTC 3 and P6 sCTC 5 represent discrepancies between the TAm-Seq and CNA data. This may reflect the errors from the reamplification process. The sequencing to generate the CNA results was performed using the original WGA DNA whilst the TAm-Seq was performed using reamplified WGA DNA. For the WES the DNA libraries were felt to yield the most accurate results when reamplified as opposed to reamplification of the WGA DNA.

However, for TAm-Seq the DNA libraries would not be appropriate input material as they already had barcodes and sequencing adaptors added. The assessment of TAm-Seq was a feasibility study to investigate its use for sequencing CTCs and therefore the compromise of using the reamplified WGA DNA was felt to be acceptable.

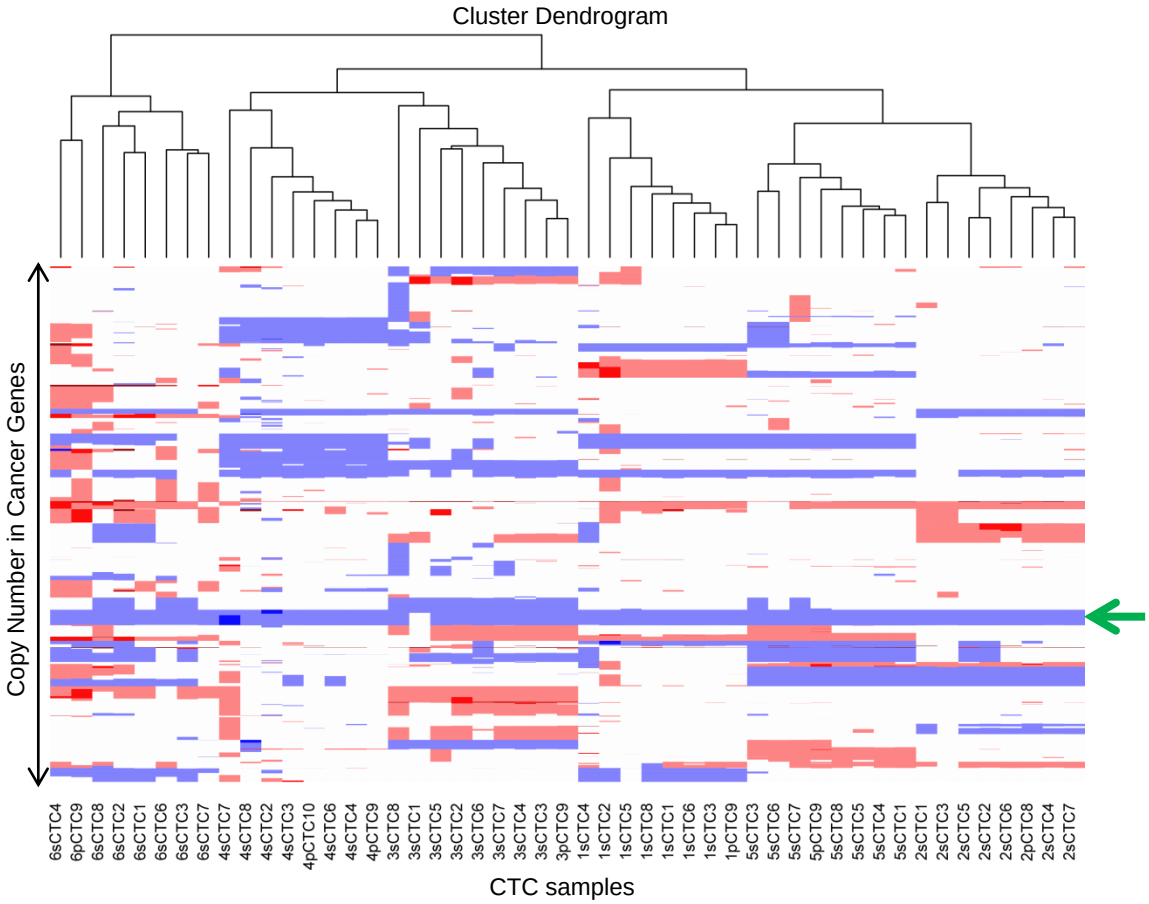


Figure 4.9 Hierarchical clustering of copy number values from CTCs isolated from six patients with SCLC. Copy number data was generated from 6,341 selected cancer-related genes from the sequencing of single and pools of CTCs isolated from six patients. The CNA data were then subjected to hierarchical clustering. Regions of amplified copy number are red whilst regions of loss are denoted in blue and regions with no change are coloured white. The green arrow indicates a common region of deletion seen across the entire sample set.

There was heterogeneity between the CNA profiles from individual CTCs isolated from a patient, for example on visual inspection of patient 6's results marked heterogeneity between the individual CTC CNA profiles was noted. This heterogeneity seen amongst some of the individual CTCs isolated from a patient is highlighted in a PCA generated from this CNA data (figure 4.10). Principal components 1 and 2 were plotted against each other and account for 37 % of the variation seen amongst the samples. Patients 1 to 5s' CTCs cluster in groups with small amounts of scatter demonstrated between the individual CTCs. However, patient 6's CTCs are widely scattered on the PCA plot

reflecting the marked heterogeneity of this patient's CNA profiles. It is also interesting to note that in the TAm-Seq analysis this patient's CTCs had the most variation in the TP53 sequencing results. Patient 6 also had considerably longer survival than the other patients at 14.3 months compared to less than 7 months for the other five patients. Although this represents an interesting observation no conclusion can be drawn as this marked heterogeneity was noted only in a single patient. Overall this data demonstrates not only that TAm-Seq analysis of SCLC CTCs is technically possible but also that the data produced can generate hypotheses for further analyses, such as the noted association between increased heterogeneity and increased survival.

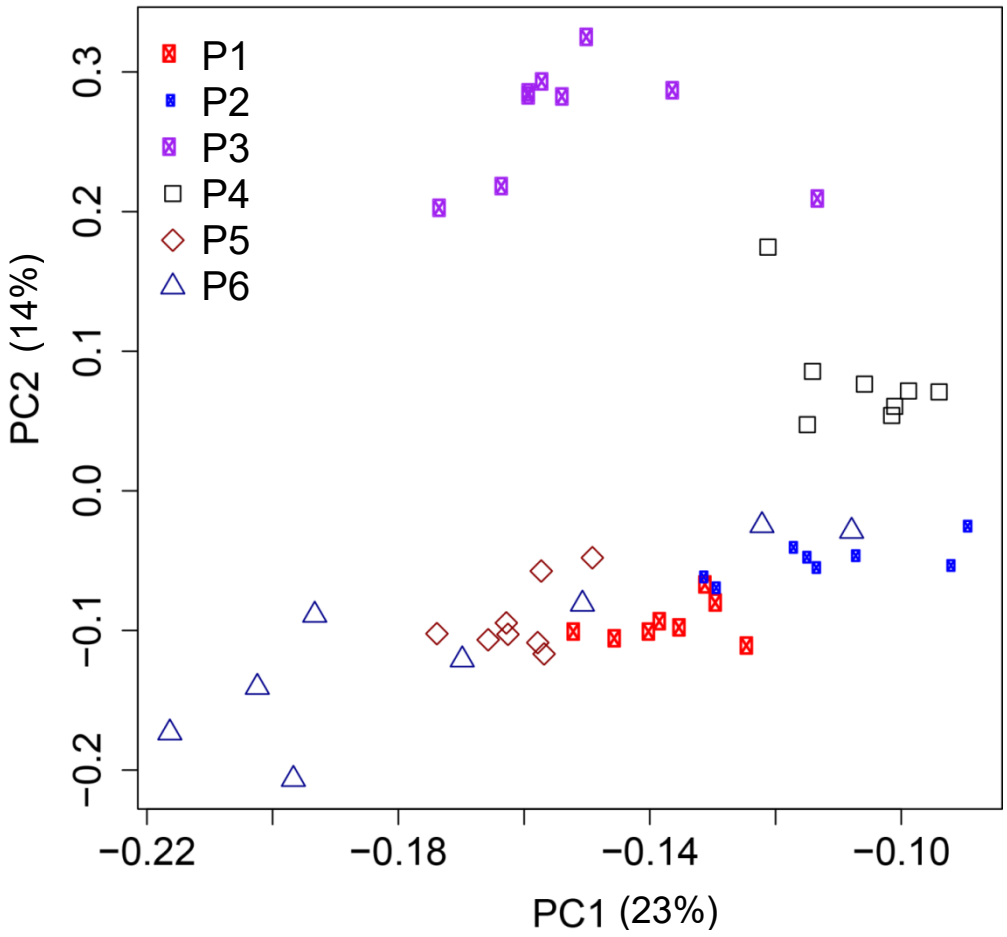


Figure 4.10 PCA of CNA data from CTCs isolated from six patients with SCLC. CNA data generated from 6,414 selected cancer-related genes were analysed from sequencing CTCs isolated from six patients with SCLC. PCA was then performed with principal components 1 and 2 visualized. The percentage of the variance accounted for by the principal components is given in parenthesis.

4.C Discussion

Research in SCLC is hampered by the lack of biopsy tissue available for research such as molecular analysis [2]. In this chapter results supporting the potential utility of both CDX tumours and CTCs for the molecular analysis of SCLC have been presented. CTCs have been hypothesised to be the subset of tumour cells with metastatic potential [153]. The tumorigenic potential of CTCs was previously demonstrated by the creation of CTC derived tumours in immune compromised mice using CTCs isolated from patients with luminal breast cancer [273]. The generation of four CDX models utilising CTCs isolated from patients with SCLC confirm the applicability of this result to SCLC [24]. From the six initial patients recruited to donate samples for the generation of CDX, four CDX mouse models were generated. The paired CellSearch samples for these four patients all had a CTC count of greater than 450 CTCs in 7.5mls of blood. In contrast in breast cancer tumours grew from only three of the 110 patients' samples implanted. When creating CDX from breast cancer CTCs only samples with more than 1109 CTCs developed tumours whereas in SCLC CDX developed when the paired blood samples had 458 or more CTCs. Although it would be interesting to hypothesise that these results reflect the aggressive nature of SCLC in contrast to luminal breast cancer, the methods and endpoints used in the two studies differed precluding this conclusion.

Molecular analysis of the CDX tumours was important to ensure that the tumours generated were consistent with SCLC, and that the method of processing and in vivo growth had not selected for tumours that did not reflect the material from which they originated. As copy number changes have been frequently seen in the molecular analysis of SCLC tumours CNA was chosen for the molecular analysis of the CDX tumours [70, 90, 91]. Comparison to common changes noted in the large studies of SCLC tumour biopsies by Peifer et al and Rudin et al, which were discussed in detail in chapter 1, were carried out [70, 91]. Genomic profiling of the CDX tumours demonstrated common changes in copy number that had been seen in SCLC tumours in previously published work. The frequently noted loss of copy number in PTEN, TP53 and RB1 was seen in all the CDX tumours [70, 91]. The commonly seen amplifications such as SOX2 and MYCL1 were seen in some, but not all, the CDX tumours, reflecting the previously published data in which there are very commonly noted deletions, but amplifications of specific genes are present in only some of the SCLC tumours analysed [70, 91]. The CNA data generated from the CDX confirmed they reflected the genetic architecture of clinical SCLC samples. To further investigate

how the CDX tumours represented the molecular profiles seen in SCLC tumours in published work, the presence of TP53 and RB1 mutations was investigated. TP53 and RB1 were chosen due to the high rates of mutations in these genes noted in previous research [82]. The generation of GEMMS with neuroendocrine tumours which were morphologically and immunophenotypically similar to SCLC through inactivation of TP53 and RB1 also provides further weight to the importance of these two genes in SCLC [67]. Unique TP53 mutations were identified in all the models by NGS and confirmed using Sanger sequencing. In contrast RB1 mutations were only demonstrated in two of the four models analysed. The mutations in TP53 and RB1 again confirm that the CDX models reflect the molecular abnormalities commonly noted in SCLC. These data support the fact the CDX can be used to investigate SCLC as they capture the major molecular changes potentially driving this cancer.

Two CDX models, CDX 1 and CDX 3 had paired tumours from the two flanks of the NSG mice in which they developed. This allowed a comparison of the molecular changes between the two tumours. The losses of copy number in 10q, 13q and 17p leading to loss of PTEN, RB1 and TP53 were common to the tumours from both flanks. The point mutations in TP53 and RB1 were also found in the tumours from both flanks. There were, however, differences seen between the tumours in both flanks such as the identification of an amplification in chromosome 11q in CDX3L, but not in CDX3R. These differences may reflect heterogeneity in the CTCs injected into both flanks of the mice resulting in tumours with some differences in their profiles. This would suggest that CDX tumours are able to reflect heterogeneity present in primary tumours [144]. The other explanation for these changes is that homogenous CTCs were injected into each flank, but there was evolution of the tumours post implantation resulting in the divergence of some CNAs.

One of the key investigations of analysing the CDX models was that the responses of the CDX tumours mirrored the responses of the patients' tumours to platinum and etoposide combination chemotherapy. Molecular analysis of the CDX tumours derived from the chemoresponsive and chemorefractory patients' samples were critically also able to identify differences in the CNA profiles. The CDX models from patients 1 and 3 who responded to chemotherapy had regions of both losses and amplifications in copy number, whilst CDX models 2 and 4 from patients who were refractory to chemotherapy, had a more loss predominant pattern of copy number changes, potentially reflecting the differing biology. Although there are only samples from four patients this result acts as a proof of principal of the potential clinical utility of the CDX models and their molecular analysis. It highlights a potential use for the CDX models in

which both standard chemotherapies and novel drugs could be tested and the molecular changes associated with different responses analysed. As the CDX models mirrored patient responses to standard therapy hopefully results from these analyses may be translatable to the clinic, unlike analyses in cell lines in SCLC [21, 274, 275].

To assess the validity of the approach used for the enrichment and isolation of CTCs, a comparison of the CNA profiles from WBCs and CTCs isolated from two patients with SCLC was made. Critically the CTCs have distinct profiles to the WBCs, indicating the approach used can accurately discriminate between the two cell types. One of the challenges of utilising single cells such as CTCs for analysis is the need for an efficient and effective amplification method. The DNA polymerase used during amplification produces errors as discussed in chapter 3 [264]. NGS also introduces errors, with for example, GC rich regions of the genome being poorly sequenced [276]. The WBCs were therefore used as controls to enable the identification and removal of potentially erroneous results from further analysis. WBCs would be expected to have normal copy number throughout the genome, so regions that persistently deviated from this amongst the WBC samples were removed from further analysis as they may be sites prone to errors. Although there is potentially the loss of information using this method it does provide increased stringency for the analyses of the remaining data. It should also be noted that the common CNA patterns seen in the CDX, both unamplified and WGA, and the single CTCs isolated from the same patient's blood sample provides a separate method of validating the methodology used for the single CTCs analysis.

A detailed molecular analysis of individual CTCs isolated from two of the patients' blood samples, whose parallel samples were used to generate CDX models 2 and 4, was carried out. Copy number data was generated from single CTCs, and pools of CTCs and WBCs to contrast to the CNA profiles of the CDX tumours. The WBCs from the two patients clustered separately to all the tumour derived samples on both hierarchical clustering and a PCA of the CNA data, confirming a distinct profile for the tumour material as opposed to the genomic samples (figures 4.5 and 4.6). The copy number profiles of the CTCs again reflected commonly seen changes in SCLC such as the loss of TP53, RB1 and PTEN. The TP53 and RB1 mutations identified in the CDX models were also identified in all the CTCs from patients 2 and 4 in which Sanger sequencing was successful. This similarity in key molecular changes supports the potential utility of CTCs for the investigation of SCLC in place of tumour tissue. It also suggests that changes in TP53 and RB1 are relatively early changes in the evolution of SCLC as they are present in all the samples analysed from an individual patient [88].

This is supported by the generation of GEMMs with tumours similar to SCLC by inactivating TP53 and RB1 as discussed previously [67].

The CTCs and CDX tumours from a patient clustered together when hierarchical clustering of the CNA data from over 6000 cancer genes was performed (figure 4.5). This supports the common origin of these two samples, as does the identification of the identical TP53 and RB1 mutations in the CDX and CTCs generated from a patient's blood samples. It also provides support for the hypothesis that the CTCs enumerated by CellSearch originated from the same initial population of CTCs that produced the CDX. However, there were some differences seen in the CNA profiles of the individual CTCs and the CDX tumours from the same patients. This may reflect the fact that there are EpCAM negative CTCs involved in the generation of CDX as well as EpCAM positive CTCs. An alternative hypothesis is that only a subpopulation of the EpCAM positive CTCs are tumorigenic and that the profiles of only these CTCs are captured in the CDX tumours. Patient-specific CNA profiles in the CTCs and CDX were seen though there was some heterogeneity seen between individual CTCs from a patient. This may reflect the fact that the individual CTCs analysed could have originated from different areas of the primary tumour or from different metastases [149, 150].

The need for new models to investigate SCLC is marked given the poor survival of this group of patients. SCLC CDX represent an exciting method of studying SCLC with potential for use in the investigation of the biology of SCLC and the analysis of new therapies. To complement these in vivo experiments it was necessary to develop robust methodologies for the genomic analysis of the CDX, which have been demonstrated in this chapter. The paired genomic analysis of CDX and CTCs are novel data which have provided insights in to SCLC and potentially chemoresistance. The different patterns of CNA in the chemorefractory and chemoresponsive patients' CDX is of particular interest given the significant issue chemoresistance represents in the treatment of SCLC. The analysis of CDX generated from SCLC CTCs is on-going within Professor Dive's laboratory, including a collaboration with Dr C Rudin at Memorial Sloan Kettering Cancer Center in which both CDX and PDX generated from chemorefractory and chemoresponsive patients' samples are being analysed to see if the same differences in CNA patterns are noted.

Identification of genes differentially mutated in subgroups of patients with SCLC by WES was one of the aims of this thesis. However, for the translation of these findings into the clinic the use of focussed NGS would potentially make the analysis quicker in terms of processing time and bioinformatics and more 'clinic friendly'. The adoption of

the use of targeted gene panels such as the Sequenom OncoCarta and LungCarta panels in the clinic has already highlighted the clinical applicability of the analysis of panels of genes of interest [277]. TAm-Seq is a method of focussed NGS which allows the amplification and deep sequencing of amplicons using a very small amount of fragmented DNA as the input [271]. The potential clinical utility of TAm-Seq had been demonstrated by Forshew et al. using cfDNA [271]. TAm-Seq can be used to investigate unknown mutations in the amplicons of interest unlike, for example, digital PCR. It also is suitable for use with fragmented DNA such as that created during the amplification of CTCs unlike the Oncocarta panel which requires specific sizes of DNA fragments for mass spectrometry. However, there was no published work using TAm-Seq with CTCs and so an assessment of its potential utility in SCLC CTCs was undertaken. Although both cfDNA and CTCs can be analysed from blood samples, CTCs were chosen for this analysis for a number of reasons. When analysing individual CTCs the coexistence of mutations within an individual cell can be analysed and correlated with the CNA changes from that cell. Analysis of the cfDNA in contrast does not allow the correlation of specific mutations as the DNA analysed is from multiple tumour cells. When analysing cfDNA there is contamination from non-tumour DNA released from normal cells into the blood stream, whilst analysis of an individual CTC allows analysis of purely tumour DNA.

DNA from CTCs isolated from blood samples from six patients with SCLC were analysed by TAm-Seq. Amplicons from the eight genes of interest, selected as they are mutated in lung cancer, were designed to occur within TTAA sites due to the WGA method used. TAm-Seq analyses of these amplicons were successful with an average of 78% success for single cells and 92% for pools of cells, though there was variation seen both between amplicons and between patients. The variability of the success of sequencing between different amplicons is likely to reflect a number of factors such as the efficiency of the primers, the GC content of the DNA, the secondary structure of the DNA and the success of WGA amplification of the region of DNA. The variation in the success of sequencing between patients is also likely to reflect the success of the WGA of the DNA from the different samples.

Unique TP53 mutations were identified in each of the six patients' CTCs, each of which occurred in the DNA binding domain, the commonest location for TP53 mutations in previous studies [78]. This confirms that TAm-Seq of CTCs can identify relevant mutations. These results also highlight one of the positive attributes of TAm-Seq in that amplicons can be designed to cover the entire coding region of a gene and therefore are able to sequence the wide range of mutations seen in TP53.

Heterogeneity of the TP53 mutations identified by TAm-Seq was noted which is surprising given the ubiquity of TP53 mutations in SCLC. This may reflect genuine variability amongst regions of the tumours which is then reflected in the CTCs or it is possible that it represents an artefact from the reamplification process or sequencing such as allele dropout.

CNA analysis of the CTCs from the same six patients' CTCs was also carried out, with the CTCs from each patient clustering together on hierarchical clustering of the CNA data from over 6000 cancer genes. A large common region of loss was seen across the six patients with the loss of copy number seen in genes including TP53, RASSF1 and FHIT. There was, however, much more variability of amplifications in the CNA data seen, as was also noted in the CDX models and associated CTCs. There were again differences seen in the CNA profiles both between patients and within the CTCs from an individual patient. This again suggests that single CTC analysis captures the complexity and heterogeneity of SCLC tumours.

The results presented demonstrate that TAm-Seq can be used to analyse mutations in a panel of genes of interest in CTCs in SCLC. There is no published data to date showing that TAm-Seq can be used to sequence CTCs as opposed to cfDNA so this represents a novel finding. Currently, in clinical practice tumour biopsy samples are analysed to look for genes of interest to direct therapy. As the amplicons can be designed to sequence any regions of interest for TAm-Seq, the potential applicability of this approach for the analysis of CTCs from clinical samples is evident as a research approach, and perhaps eventually as a clinical tool. The identification of TP53 mutations, the heterogeneity of mutations amongst individual CTCs and the correlations between TAm-Seq and CNA data demonstrate that this approach can generate meaningful data to investigate SCLC.

Successful analysis of SCLC CTCs was demonstrated using Sanger sequencing, focussed NGS using TAm-Seq and WGS in this chapter. The data demonstrate genetic changes in the SCLC CTCs and CDX in common with previously published results from SCLC tumour biopsies and cell lines. They also capture the heterogeneity both between the CTCs from different patients and within the CTCs isolated from an individual patient. These results as a whole support the use as a research tool of the molecular analysis of CTCs for the investigation of SCLC.

Chapter 5: Genomic Profiling of SCLC CTCs to Investigate Chemoresistance

5.1 Introduction

Circulating tumour cells (CTCs) have the potential to act as a liquid biopsy to allow monitoring of tumours [145]. As blood tests are minimally invasive they are more acceptable to patients, and in comparison to tissue biopsies, far easier to repeat longitudinally along the course of a patient's treatment. This is particularly relevant when considering biopsies in lung cancer which can be challenging and can cause morbidity, and in rare cases mortality for patients [23]. Crucially, blood sampling reflects the current genetic status of the tumour and will include evolving differences that will not be present in the historic tumour biopsy. As CTCs are found in higher numbers in SCLC than any other tumour type as yet investigated, it represents an ideal disease to explore their clinical utility [152]. The overarching aim of this thesis was to explore how genomic analysis of CTCs could be exploited to study inherent (in chemorefractory patients) and acquired (in relapsed chemoresponsive patients) resistance mechanisms.

Common changes in SCLC have been identified from studies which carried out genomic profiling of tumours as introduced previously in section 1.1.4. TP53 and RB1 are the most frequently altered genes with both the loss of copy number and mutations detected in both genes [74]. Genomic alterations in these two genes are felt to be critical for the development of SCLC [67]. Commonly occurring CNAs are also found in SCLC across several other genes; for example, the tumour suppressor genes FHIT, PTEN and RASSF1 are also frequently lost [70, 91]. There are also amplifications seen in a number of genes in subsets of SCLC patients such as SOX2, BCL2 and members of the MYC family [70, 91]. At the chromosome level there are frequent losses seen in 3p, 13q and 17p as well as amplifications of 3q and 5p noted in SCLC tumour biopsies [70, 90]. In addition to CNAs, there are frequent mutations identified in SCLC. It is one of the mostly highly mutated tumour types with approximately 7.4 protein changing mutations per million base pairs, occurring most likely due to the mutagenic effects of smoking [70]. Although many mutations occurring in SCLC may be passenger mutations due to this high mutation rate, recurrent mutations in a number of genes have been noted such as genes encoding histone modifying enzymes, members of the PI3K pathway and SOX family members [91]. Despite the fact that

there will be many patient-specific mutations due to the high mutation rate, there are clearly key alterations that are frequently seen across SCLC patients' tumours and which are candidates to be of biological significance. It is therefore very relevant to see if these same genomic changes that have been noted frequently in analyses of SCLC biopsies can also be identified in SCLC CTCs. This would provide formal evidence that CTCs can be used to explore common changes in SCLC and see how tumours evolve during and after therapy.

There has been a focus on the development of targeted therapies in oncology in recent years. Evidence from other disease types has confirmed this can lead to improvement in patient outcomes, though as yet no targeted agent has been successfully trialled successfully in SCLC [54]. By their very nature, targeted agents are designed to tackle an aberrant pathway that is driving a cancer's growth and progression [140, 141]. However, if this pathway is not altered in a specific cancer the targeted agent is unlikely to be an effective treatment. Consequently, there is increasing evidence that targeted agents should be used in subsets of patients identified with a specific predictive biomarker. Given the failure of targeted agents in clinical trials in SCLC this is particularly relevant [27, 54]. However, the identification of predictive biomarkers in the initial diagnostic biopsy may no longer be relevant if a targeted therapy is being considered beyond the first line of therapy, as this biomarker may no longer be present, or resistance mechanisms to this drug may have evolved [278]. The potential evolution of tumours would necessitate repeat biopsies with associated risk of discomfort and harm to patients [23]. CTCs potentially represent an ideal medium to analyse predictive biomarkers over the course of patients' therapy [118]. It is important to establish therefore that the methods of analysing CNA and WES used can identify mutations in clinically relevant, actionable genes in CTCs. Assessing whether the development of new mutations over the course of patients treatment can be detected in their CTCs will also help assess the utility of CTCs as potential predictive biomarkers.

CTCs have the potential to enable investigation of the biology of tumours such as SCLC as well as acting as prognostic and predictive biomarkers [152]. One of the key areas in SCLC in which CTCs may facilitate progress is to further understanding of chemoresistance, a critical contributor to the very poor survival seen in this disease [117]. In SCLC approximately 20% of patients do not respond to first-line chemotherapy and therefore demonstrate intrinsic resistance to chemotherapy [3, 31]. Unfortunately trials show that these patients, who will be referred to as chemorefractory from this point, also have a low chance of responding to any further therapy [3]. This contrasts to the remaining 80% of SCLC patients, who will be referred to as

chemoresponsive, who actually respond quickly and well to first-line therapy with up to 20% of patients having complete radiological responses to treatment [31]. There is therefore a marked difference in behaviour to chemotherapy in the chemorefractory and chemoresponsive patients. This raises the question of whether CTCs could be used to identify genomic differences between these two groups of patients. This may enable the key changes associated with intrinsic resistance to be revealed and possibly provide new therapeutic avenues to explore. It may also enable the development of a predictive biomarker for intrinsic resistance which could be used to identify the patients with chemorefractory disease. This would enable patients to make informed decisions about the use of chemotherapy, given the lower than expected response rate that they would have compared to the average patient with SCLC. It would also enable oncologists to consider first-line clinical trials for this patient group, as they far less likely to respond to standard therapy.

Unfortunately the 80% of patients with chemoresponsive disease do not have durable responses to first-line chemotherapy, with the median PFS being just 4.6 months [38]. At this point response rates in second-line trials are significantly lower at just 7 to 24% [3, 39, 41-43]. This indicates the development of acquired resistance limiting further treatment options, highlighting a marked change in behaviour of SCLC, going from being a highly chemoresponsive disease to a highly chemoresistant disease in these patients. Given the high initial response rate to first-line chemotherapy the development of acquired resistance represents a significant contributor to the poor outcome of SCLC patients. It represents a key area for research as increased knowledge of this issue could result in improved treatment options. Contrasting the genomic aberrations in SCLC CTCs at baseline from chemoresponsive patients, to those present in CTCs isolated when they develop progressive disease, has the potential to reveal changes associated with the development of acquired resistance. Exploring the mechanisms of acquired resistance may also reveal different options for therapies in the second-line which could be more effective. Investigation of chemoresistance using tumour biopsies and cell lines has revealed numerous potential causes that develop including decreased activation of drugs, increased inactivation of drugs and the evasion of drug-induced apoptosis [111]. It is likely that there is not a single overarching mechanism for the development of acquired resistance in SCLC, but that each patient has unique genetic changes that drive disparate resistance mechanisms. Establishing a method by which CTCs could be used to explore patient-specific mechanisms for the development of chemoresistance would provide a very relevant way to personalise therapy decisions.

Copy number changes have been frequently seen in SCLC as discussed above. It is clear that CNA develop frequently in SCLC and so their assessment is potentially very relevant to its research. It also provides a global level of assessing changes across the genome. In contrast WES would allow the investigation of smaller scale alterations such as point mutations and Indels. WES enables sequencing of just the protein-coding genes, in which more than 85% of known disease-causing mutations occur [256]. Applying these two approaches to the genomic profiling of CTCs will enable a detailed analysis of the genomic aberrations which may be driving SCLC chemoresistance. As discussed in chapter one, the significant challenge of applying these techniques to single cells, given the small quantity of nucleic acids present, remains. The analysis of single cells however may provide the possibility of assessing tumour heterogeneity which again could be very relevant for the investigation of resistance, particularly acquired resistance. Subclones of tumour cells in the initial tumour may be resistant to chemotherapy and become more dominant in later time points.

5.2 Aims

1) To assess the CNA and mutations profiles in SCLC CTCs and compare them to published data from SCLC tumours.
2) To investigate if CNA and WES of CTCs can be used to identify changes in potentially clinically actionable genes and to monitor changes in these mutations over time.
3) To investigate differences in the CNA profiles of CTCs isolated from chemo-naïve patients with chemoresponsive and chemorefractory disease.
4) To investigate differences in the mutations identified by WES in CTCs isolated from chemo-naïve patients with chemoresponsive and chemorefractory disease.
5) To investigate changes in the CNA profiles of CTCs isolated at baseline and at the development of relapsed disease after chemotherapy, from patients with initially chemoresponsive disease.
6) To investigate changes in the mutations identified by WES in CTCs isolated at baseline and again at the development of relapsed disease after chemotherapy, from patients with initially chemoresponsive disease.

5.3 Methods

General methods are detailed in chapter 2. Figure 2.1 provides an overview of the CTC processing workflow used. Details on individual methods used in this section can be found as follows;

1) Patient and clinical sample collection – see section 2.1.
2) CTC enrichment and isolation – see section 2.2.
3) WGA of CTC DNA - see section 2.3.
4) WGA quality control assay – see section 2.3.1.
5) DNA extraction - see section 2.4.
6) DNA library preparation – see section 2.5.
7) Next generation sequencing – see section 2.6.
8) PCR and Sanger sequencing - see section 2.7.
9) Reamplification of DNA libraries - see section 2.8.
10) In-solution target enrichment for WES – see section 2.9.
11) Bioinformatics and statistical considerations – see section 2.10.

5.4A Results and Discussion: CNA Analysis and WES of SCLC CTCs

5.4A.1 Patient Characteristics

Patients were recruited at a tertiary cancer centre from September 2012 to March 2014 as discussed in section 2.1. The 35 patients recruited all had histologically or cytologically confirmed SCLC and had not yet received any chemotherapy treatment. Of the 35 patients, 20 had chemoresponsive disease and 15 had chemorefractory disease, having progressed within 3 months of completion of their first-line chemotherapy. The clinical characteristics of the chemoresponsive and chemorefractory patients are summarised in table 5.1. All but three of the patients recruited had extensive stage SCLC. There were 20 women and 15 men recruited to the study. The patients were all former or current smokers, which is typical of SCLC patients. Twenty one of the patients recruited received only one line of chemotherapy, with 12 receiving second-line chemotherapy, and two patients receiving third-line chemotherapy. As would be expected, PFS was shorter for the chemorefractory patients at 3.4 months compared to 9.9 months for the chemoresponsive patients. The median OS was also shorter for the chemorefractory patients at 4.9 months in contrast to 14.2 months for the chemoresponsive patients. Analysis of both the PFS and OS

with a Mann Whitney U test confirmed that the longer PFS and OS times seen in the chemoresponsive patients was statistically significant ($p < 0.0001$). The chemorefractory patients had poorer performance status and all had ES SCLC.

Patient Characteristic	Chemoresponsive Patients	Chemorefractory Patients
Number of Patients	20	15
Age		
Median	64	69
Range	35 - 82	56 – 82
Sex		
Female	12	8
Male	8	7
Stage at Diagnosis		
Limited	3	0
Extensive	17	15
Smoking History		
Mean Pack Years	45	47
Range Pack Years	15 - 80	10 – 130
Baseline WHO PS		
0	2	0
1	10	6
2	6	6
3	2	3
Number of Sites of Metastases		
0	1	0
1	6	4
2	4	6
3+	9	5
Lines of Chemotherapy		
1	7	14
2	11	1
3	2	0
Progression Free Survival (months)		
Median	9.9	3.4
Range	5.1 - 22	0.7 - 5.0
Overall Survival (months)		
Median	14.2	4.9
Range	6.6 - 22	0.9 - 10.2

Table 5.1 Characteristics of chemoresponsive and chemorefractory SCLC patients.

Detailed molecular analysis was performed on CTCs isolated from a subset of 10 of these patients, five of whom had chemorefractory disease and five of whom had chemoresponsive disease. The first five patients recruited with chemorefractory disease, and the first five with chemoresponsive disease with successfully amplified CTC samples as described in section 5.A4.2 were molecularly profiled. The clinical characteristics of this subgroup of 10 patients were comparable to the whole cohort of patients (n = 35) and representative of SCLC patients (table 5.2). The 10 patients all had ES SCLC. The patients' median age at diagnosis was 66 years and 6 months, ranging from 57 to 78 years old. The patients all also had strong smoking histories

having smoked for 20 to 130 pack years. All but four of the patients received only first-line chemotherapy. The four patients who received more than one line of chemotherapy were all chemoresponsive, with three receiving just second-line chemotherapy and one receiving both second- and third-line chemotherapy. Patients 5 and 8 both had documented radiological response to second-line chemotherapy unlike patients 6 and 10. Patient 5 relapsed within 3 months of completing second-line treatment whilst patient 8 responded for more than 3 months. Patient 8 received but did not respond to third-line chemotherapy. Summaries of the treatment histories of the five chemoresponsive patients are given in figure 5.13. The difference in PFS and OS between the chemoresponsive and chemorefractory patients remains statistically significant for the 10 patient subset ($p = 0.0079$ for both analyses).

Patient Number	Chemoresponsive or Chemorefractory	Age	Sex	Smoking History (pack years)	PS	Metastatic sites	Lines of Chemotherapy	Progression-free survival (months)	Overall survival (months)
P1	Chemorefractory	70	Male	130	2	nodes, pleural effusion	1	3.7	4
P2	Chemorefractory	61	Male	90	1	nodes, bone, liver, soft tissue	1	4.7	4.4
P3	Chemorefractory	65	Female	84	2	bone, brain, liver, pleural effusion	1	3.1	3.5
P4	Chemoresponsive	73	Male	30	3	nodes, bone, adrenal	1	5.1	6.6
P5	Chemoresponsive	57	Female	64	2	nodes, bone, liver	2	8.1	14.3
P6	Chemoresponsive	59	Male	40	1	liver, pleural effusion	2	6.2	10.5
P7	Chemorefractory	78	Female	27	2	nodes, liver, pleura	1	0.7	0.9
P8	Chemoresponsive	68	Female	67.5	2	nodes, liver	3	6.2	16.2
P9	Chemorefractory	69	Female	20	3	nodes, bone, liver, pleura	1	2.8	5.3
P10	Chemoresponsive	63	Female	23	2	bone, liver	2	6.1	7.6

Table 5.2 Clinical characteristics of the selected 10 SCLC patients whose CTCs were isolated and subjected to molecular analysis.

5.4A.1.1 CTC Enumeration

Blood samples were taken from all 35 patients prior to them receiving chemotherapy with the CTCs enumerated by CellSearch (in 7.5 ml of blood) as described in section 2.2. Blood samples were taken when relapse was diagnosed after first-line chemotherapy in 24 of the patients, and when the patients relapsed after second-line chemotherapy in four patients. Patients' relapses were diagnosed based on computerised tomography scans (CT) or infrequently, based on chest xrays (CXR) which documented radiological progression of their cancers. The CTC counts at baseline and the two relapse time points are detailed in table 5.3. The median number of CTCs at baseline in the chemoresponsive patient group was 412 (range 0 – 4061) whilst the median number of CTCs in the chemorefractory patients was 2182 (range 6 – 20815). Although the median number of CTCs is higher in the chemorefractory group, the difference seen is not statistically significant (p = 0.4332).

CTCs by CellSearch	Chemoresponsive Patients	Chemorefractory Patients
CTCs at Baseline (T1)	n= 20	n= 15
Median	412	2182
Range	0 - 4061	6 - 20815
CTCs at relapse post first-line chemotherapy (T2)	n = 18	n = 6
Median	212	319
Range	0 - 1126	0 - 1796
CTCs at relapse post second-line chemotherapy (T3)	n = 4	
Median	794	NA
Range	0 - 1522	NA

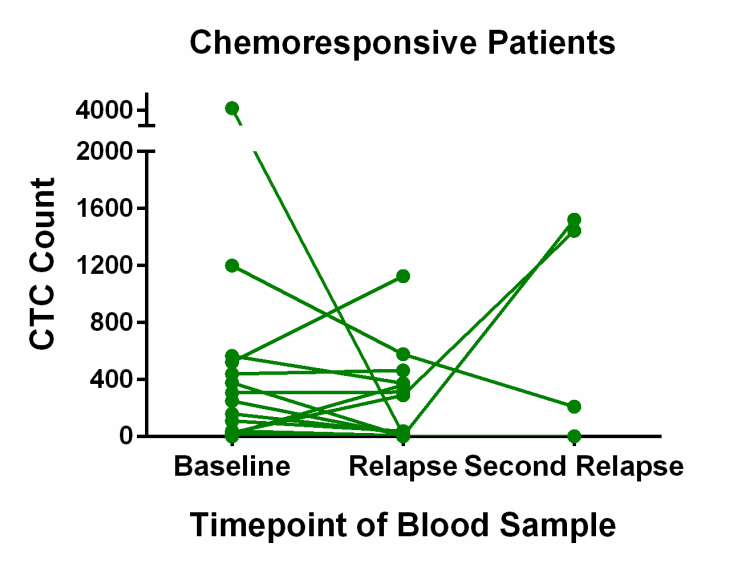
Table 5.3 CellSearch CTC counts for chemoresponsive and chemorefractory SCLC patients. CTC counts were enumerated by CellSearch in 7.5 ml of blood in patients with SCLC. CTC measurements were performed at baseline prior to the patient receiving chemotherapy, and in selected patients when radiological progression was diagnosed after first or second-line chemotherapy.

The number of CTCs at relapse after first-line chemotherapy was lower than at baseline in both the chemoresponsive and chemorefractory groups, at 212 and 319 respectively, but this again was not statistically significant (p = 0.3329). The number of CTCs may be lower at relapse because the patients are being diagnosed with relapse when they have less disease burden than when they initially presented as they are under oncological follow up. The number of CTCs at relapse will also be influenced by

the manner of a patient's relapse; for example one patient who relapsed with brain metastases had 7687 CTCs at baseline but just 38 at relapse. It is possible that the CTCs that remain when patients relapse, despite being fewer in number, are a more aggressive subset of CTCs. Further exploration of the differences in these CTCs on a molecular level is therefore very relevant to investigating potential mechanisms of resistance. Only four chemoresponsive patients had enumeration of their CTCs after receiving second-line chemotherapy with a median number of CTCs of 794 in 7.5ml of blood. When observing the change of CTC counts between baseline and relapse on an individual patient basis, it is apparent that the change in the number of CTCs at relapse is very variable between patients (figure 5.1). In some patients the number of CTCs goes up, in others it goes down and in some patients it remains static showing there is an individual pattern of change in CTCs per patient. It therefore may be that there are molecular changes in the CTCs rather than the number of CTCs per se at relapse that are critical for the aggressiveness of the disease at this stage. The CTC numbers for the 10 patients whose CTCs were molecularly analysed are given in table 5.4.

5.4A.2 Isolation and Amplification of CTCs from SCLC Patients for Molecular Analysis

CTCs were isolated from blood samples taken from 10 patients with SCLC for molecular analysis from the total of 35 patients whose CTCs were enumerated (tables 5.2 and 5.4). The patients selected were the first five patients with chemorefractory disease with greater than five CTCs enriched in their CellSearch cartridge, with CTCs isolated by DEPArray of which a minimum of three CTC samples had successful WGA amplification (GII of ≥2). The same criteria were used to select the chemoresponsive patients for analysis, except samples meeting the criteria were required both at baseline and at the diagnosis of relapsed disease after first-line chemotherapy. After enumeration by CellSearch single CTCs, pools of CTCs, single WBCs and pools of WBCs were isolated as described in section 2.2 using the DEPArray. The samples were then WGA and a quality control assay performed as in section 2.3 to establish the genome integrity index (GII). When an appropriate sample had been collected from a patient, a leukocyte gDNA sample was also amplified with 1 ng of DNA being added to the WGA reaction.



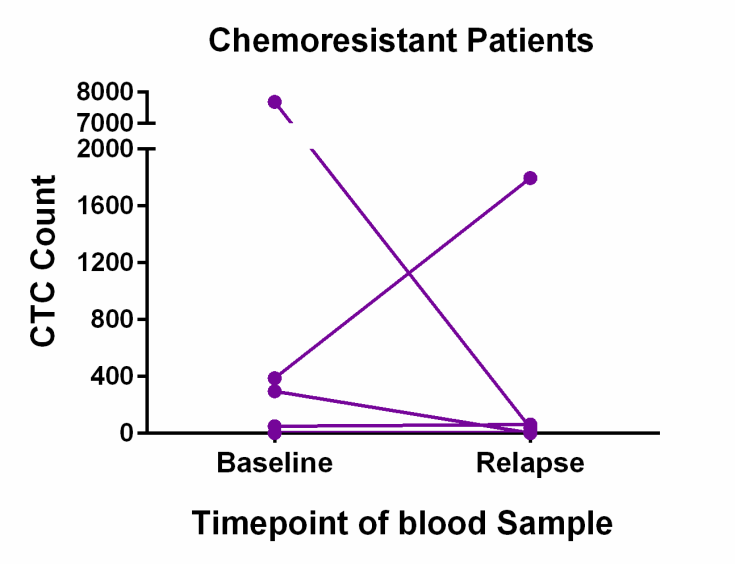


Figure 5.1 CTC counts in chemoresponsive patients and chemorefractory patients at baseline and relapse. CTC numbers were enumerated by CellSearch in 7.5 ml of blood. CTCs were measured at baseline prior to receipt of chemotherapy, at relapse after first-line chemotherapy, and if applicable at relapse after second-line chemotherapy.

5.4A.2.1 Quality Control of WGA CTCs

A quality control assay was performed to assess all the WGA samples as described in section 2.3.1. The assay is a multiplexed PCR with the success of the amplification of sample being scored as between 0 and 4 loci which are translated in to the genome integrity index (GII) [237]. Samples with higher GII scores are considered to have greater chance of successfully being used in molecular analyses. In total, across the 10 patients, 271 single and pools of cells were amplified of which 140 cells (52%) had a GII of 4 and 49 (18%) had a GII of 3. A further 29 cells (11%) had a GII of 2 and 25

(9%) had a GII of 1. There were 28 cells (11%) with failed amplification with a GII of 0. Up to 15 samples consisting of single and pools of CTCs, WBCs and gDNA were selected per time point per patient for CNA analysis. In total 149 CTC, 44 WBC and 4 gDNA samples were processed for CNA. The cells were ranked according to GII from each patient's samples and the cells with the highest GII scores were selected first for molecular analysis. However, only a small number of WBCs were selected per patient despite good amplification as they were used as controls for the analyses. A summary of the GII scores for the samples used for the molecular analyses is given in table 5.5. The GII scores from the quality control assay results of all the individual CTC and WBC samples for which CNA analysis was performed are given in appendix 1 table A1.3.

Patient Number	CTC number at baseline	CTC number at relapse post first-line chemotherapy	CTC number at relapse post second-line chemotherapy	OS (months)
1	170			4
2	20815			4.4
3	1625			3.5
4	522	1126		6.6
5	1200	578	209	14.3
6	250	12		10.5
7	1376			0.9
8	27	289	1445	16.2
9	204			5.3
10	307	312		7.6

Table 5.4 CellSearch CTC counts for 10 SCLC patients whose CTCs underwent molecular analysis. CTC counts were enumerated by CellSearch in 7.5 ml of blood in patients with SCLC. CTC measurements were performed at baseline prior to the patient receiving chemotherapy, and in selected patients when radiological progression was diagnosed after first- or second-line chemotherapy.

Following the selection of the WGA samples for detailed molecular analysis 61% of the samples had a GII of 4, 19% had a GII of 3, 12% had a GII of 2 and 7% had a GII of 1. The pools of cells had better GII scores than the single cells with 84% having a GII of 4 in contrast to 54% of the single cells. The WBCs also had improved amplification in contrast to the CTCs, with 80% of the single WBCs having GII scores of 4 in contrast to 49% of single CTCs. However, this figure may be affected by the fact that only smaller numbers of WBCs were selected for molecular analysis from the total pool of available samples, in contrast to the CTCs, enabling the selection of those with higher quality. This was found to be true, as when the same analysis was applied to the unselected cohort of every amplified cell from the 10 patients, the WBCs had only slightly improved

GII scores in contrast to the CTCs, as 48% of the amplified single WBCs had a GII score of 4 in contrast to 44% of the single CTCs.

Sample Type	Number of samples	Genome Integrity Index			
		4	3	2	1
All Samples	197	121 (61%)	38 (19%)	24 (12%)	14 (7%)
Single Cells	150	81 (54%)	35 (23%)	22 (15%)	12 (8%)
Pools of Cells	43	36 (84%)	3 (7%)	2 (5%)	2 (5%)
All CTCs	149	79 (53%)	34 (23%)	22 (15%)	14 (9%)
Single CTCs	125	61 (49%)	31 (25%)	21 (17%)	12 (10%)
Pools of CTCs	24	18 (75%)	3 (13%)	1 (4%)	2 (8%)
All WBCs	44	38 (86%)	4 (9%)	2 (5%)	0 (0%)
Single WBCs	25	20 (80%)	4 (16%)	1 (4%)	0 (0%)
Pools of WBCs	19	18 (95%)	0 (0%)	1 (5%)	0 (0%)
gDNA	4	4 (100%)	0 (0%)	0 (0%)	0 (0%)

Table 5.5 Summary of GII quality control scores for amplified samples used for molecular analysis. Post WGA of samples the Ampli1 QC assay was performed generating the GII. A summary of the range of GIIs for the different types samples used for further molecular analysis is given in the table. The percentage of each sample type with a specific GII is indicated in parentheses.

The CNA data from patient 1 highlights that there is variability visually in the CNA profiles amongst the CTC samples from patients, with three single CTCs having a markedly different appearance to the other six samples and to each other (figure 5.2). This is highlighted by the node height of the hierarchical clustering, with the six similar samples having node heights from 0.929 to 0.669 whilst the other three samples had node heights of 0.552, 0.454 and 0.358. As these CTCs have GII scores of 2, 3 and 4 respectively this variation would not appear to be linked to this QC. Analysis of the CNA data from the entire data set found that the CTCs that had the most variable CNA amongst samples had no correlation with the GII, number of reads or percentage of uniquely mapped reads. The CNA images from CTCs also did not alter with increased number of reads per sample. There was, therefore, no consistent technical cause identified to account for the variability evident. This may, therefore, represent biological variation amongst CTCs. As there was no consistent criteria to exclude these samples nor was it apparent they should be excluded as it could represent biological variation all the CNA samples were included for analysis.

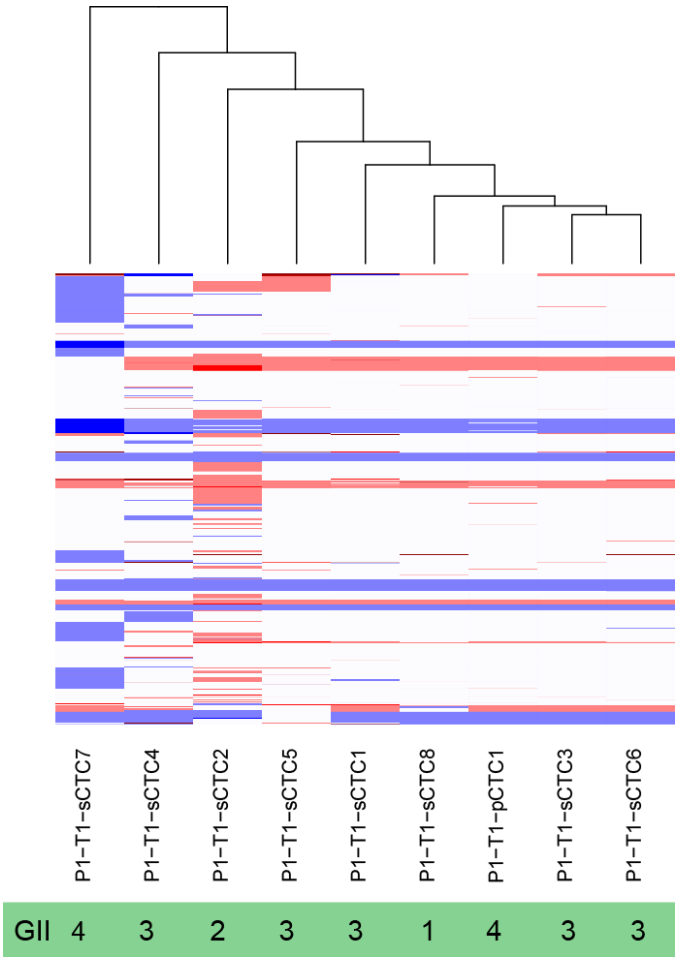


Figure 5.2 CNA analysis of CTCs isolated from Patient 1. CTCs isolated from blood samples from patient 1 were subject to WGA and then a QC assay performed to generate the GII scores. DNA libraries generated from the CTCs were subject to WGS to generate CNA data. Hierarchical clustering of the CNA data from the CTCs was carried out. Regions of gain of copy number are red, regions of loss are blue and regions with no change are white.

A summary of the CTC and WBC samples isolated per patient and those on which CNA analysis was performed is given in table 5.6. The samples were sequenced on the Illumina MiSeq or Next Seq 500 sequencers and CNA analysis performed as described in section 2.10.2. The number of uniquely mapped reads from the WGS used per sample to generate the CNA data is given in appendix 1 table A1.3. Heatmaps of the CNA profiles of all the CTCs isolated from each patient are found in appendix 3.

Number of each sample type

Patient	Timepoint of Blood Sample	Single CTC isolated	Single CTC analysed	Pool CTCs isolated	Pool CTCs analysed	Single WBC isolated	Single WBC analysed	Pool WBCs isolated	Pool WBCs analysed	gDNA sample
P 1	T1	8	8	1	1	0	0	2	2	No
P 2	T1	8	8	2	2	0	0	2	2	No
P 3	T1	8	6	2	2	0	0	2	1	No
P 4	T1	8	7	2	1	0	0	2	2	No
	T2	8	8	1	1	4	4	1	1	
P 5	T1	8	8	1	1	3	2	1	1	Yes
	T2	8	7	2	2	4	1	1	1	
	T3	10	8	2	2	9	1	3	1	
P 6	T1	8	8	1	1	5	3	0	0	Yes
	T2	9	8	0	0	3	2	1	1	
P 7	T1	8	2	1	1	3	0	1	1	No
P 8	T1	7	4	0	0	5	4	1	1	Yes
	T2	10	7	3	3	9	1	1	1	
	T3	10	9	3	2	4	2	1	1	
P 9	T1	8	7	1	1	4	2	1	1	No
P 10	T1	10	10	2	2	10	1	3	1	Yes
	T2	10	10	2	2	10	2	3	1	

Table 5.6 CTCs isolated and analysed per patient. CTCs and WBCs were isolated by DEPArray post CellSearch enrichment from 10 patients (from five patients at more than one time point). The cells were then WGA and selected CTCs taken forwards for molecular analysis. The number of cells that were isolated and amplified and those that were taken forwards for molecular analysis are indicated in the table, in addition to if a patient had a leukocyte genomic DNA sample amplified. T1 is the baseline time point prior to first-line chemotherapy. T2 is the relapse time point after first-line chemotherapy whilst T3 is the relapse after second-line chemotherapy time point.

5.4A.2.2 WES of SCLC CTCs

WES was performed on 34 samples from the 197 samples from which DNA libraries were made from the 10 SCLC patients. Due to cost restraints WES could not be performed on the entire set and so a representative set of samples from the 10 patients was chosen, including a WBC sample from each patient to act as a control. The samples used for WES are detailed in appendix 1 table A1.4. Post exon enrichment, the samples were sequenced on the HiSeq and then mapped to the human genome and analysed as described in section 2.10.6. The number of uniquely mapped reads for each of the WES samples generated on the HiSeq is also given in appendix 1 table A1.4, ranging from 2,823,702 to 69,226,810 uniquely mapped reads per sample. The efficiency of the exonic enrichment is also indicated with the percentage of the reads covering the target area ranging from 23.62% to 64.93% for the WES samples. The

WBC samples were used as controls for the sequencing. Any SNV or Indel identified in the WBCs was excluded from further analysis in all the WES samples as it was felt to potentially represent a sequencing artefact as described in section 2.10.6. In total across the 23 CTC samples 16,545 non synonymous SNVs and 961 Indels were therefore analysed from the WES sequencing data. This equates to 8.5 protein changing mutations per million base pairs, based on the size of target region of the exon enrichment protocol. This is similar to the rates quoted by Peifer et al reporting 7.4 protein changing mutations per million base pairs in SCLC biopsies [70]. Confirmatory Sanger sequencing was carried out on nine wild-type and nine mutated SNV loci identified from the WES data (appendix 4). The overall confirmatory rate was 94.4%.

5.4A.3 Comparison of the Genomic Profiles of SCLC CTCs and Published SCLC Tumour Biopsies

Knowledge of the molecular landscape of SCLC has recently been extended significantly with two landmark papers published in 2012 [70, 91]. Common patterns of copy number change and mutated genes were identified across 88 SCLC tumour biopsies and surgically resected specimens. A comparison of the CNA and WES data from the SCLC CTCs to the published SCLC tumour data was therefore undertaken. The aim of the comparison was to see if common changes identified in SCLC tumour biopsies were also demonstrated in SCLC CTCs. This would support the use of CTCs as an appropriate source of tumour DNA for the research of SCLC.

5.4A.3.1 Comparison of CNA Changes in SCLC CTCs and Published SCLC Tumour Biopsies

CNA changes have been commonly documented in SCLC research. An analysis of the CNA changes frequently seen in the CTCs and comparison to those frequently seen in tumours was carried out. In SCLC tumour biopsies loss of 3p, 17p and 13q and the amplification of 3q and 5p have been frequently noted [70]. From the 149 CTCs I isolated from the 10 SCLC patients, copy number changes in these five key chromosomes were assessed. In the 149 CTCs, cytobands in 3p were lost in 79%, cytobands in 17p were lost in 72% and cytobands in 13q were lost in 62% of the CTCs. Amplification of cytobands in 3q were seen in 77% of the CTCs analysed and 5p was amplified in 51% of the CTCs analysed. These results reflect the previously published data though the loss of 13q had been reported more frequently [70]. The

chromosomes which had most frequent alteration in copy number amongst the CTCs analysed was also investigated, as opposed to assessing the frequency of those commonly altered in tumour biopsies. Although 3p was the chromosome which had the commonest loss of cytobands within it, consistent with published data, cytobands in 15q were also frequently lost. The most frequently amplified cytobands were in 3q as seen in the published data but also in 18q and 18p.

The change in copy number at a gene level was also investigated to examine potential drivers of the development of SCLC. In SCLC tumours 13 genes have been noted to commonly have changes in copy number in the two landmark papers [70, 91]. RB1, TP53, PTEN, FHIT and RASSF1 were frequently noted to be lost whilst SOX2, MYC, EGFR, FGFR1, BCL2, MYCL1, MYCN and CCNE1 were frequently amplified in published research. The copy number changes in these 13 genes were assessed amongst the 149 CTCs analysed (figure 5.3). Although there was some variability seen, the copy number changes noted in the CTCs within these 13 genes broadly reflected that seen in published data. In the majority of the CTCs, RB1, TP53, PTEN, FHIT and RASSF1 had loss of copy number. The amplification of SOX2, MYC, EGFR, FGFR1, BCL2, MYCL1, MYCN and CCNE1 was more variable. This reflects the published data in which the amplification of individual oncogenes were seen in subsets of the SCLC tumours analysed suggesting different pathways may be driving growth in the different tumours. The most commonly lost genes amongst the CTCs are those on 3p whilst the most commonly amplified are on 18p and q reflecting the common cytoband amplifications seen amongst the CTCs. This data suggests that CTCs do reflect the common CNA changes observed amongst SCLC tumours.

5.4A.3.2 Comparison of Mutations Identified in WES of SCLC CTCs and Published SCLC Tumour Biopsies

Mutations in both the tumour suppressor genes TP53 and RB1 are hallmarks of SCLC [74]. Other genes such as PTEN, PIK3CA, SOX family members and genes encoding histone modifying enzymes have also been noted to be recurrently mutated in SCLC [70, 91]. I compared the frequently seen mutations in the CTCs to those identified in published tumour biopsy derived data. For this analysis, the commonly mutated genes in the WES of the 10 patients' baseline samples were investigated with genes with either point mutations or Indels in five or more patients CTCs identified. In total 19 genes were mutated in five to seven patients' CTCs (table 5.7). Ten of the 19 frequently mutated genes were identified as being cancer-related (using the criteria in 2.10.3).

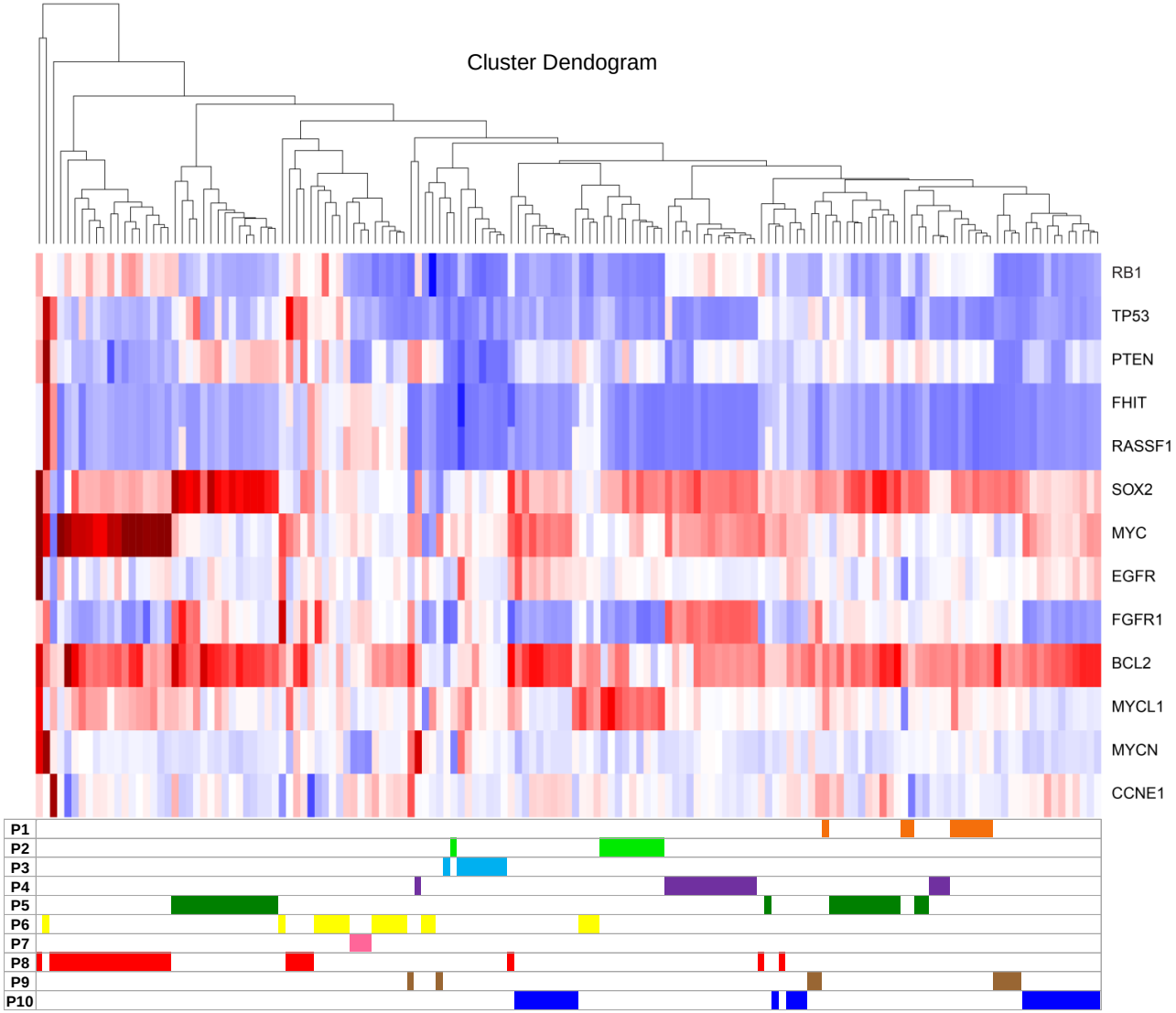


Figure 5.3 Copy number alteration analysis of 13 frequently altered genes in 149 CTCs.

Single CTCs and pools of CTCs were WGA and NGS of the DNA libraries performed. CNA analysis was then carried out to investigate changes in 13 genes frequently altered in SCLC. Hierarchical clustering of the CTCs was performed. On the plot regions of loss are coloured blue and regions of amplification are coloured red. Regions with no change in copy number are coloured white. The patient from whom each CTC was isolated is indicated in the panel below the copy number data.

Gene	No of patients gene mutated in CTCs	Gene mutated in Peifer et al [70]	Gene mutated in Rudin et al [91]	Cancer-related Gene
AHNAK2	6		Yes	
ALPK2	5	Yes	Yes	
CELSR1	5	Yes		
FCGBP	6	Yes	Yes	Yes
FRG1	5	Yes		
HYDIN	6	Yes	Yes	
INPP5D	5		Yes	Yes
LRP1B	5	Yes	Yes	Yes
NAV2	5		Yes	Yes
NEB	5		Yes	
PDE4DIP	7	Yes	Yes	Yes
PRRC2C	5			Yes
SRCAP	5		Yes	Yes
STARD9	6		Yes	Yes
TJP2	5	Yes	Yes	Yes
TP53	6	Yes	Yes	Yes
VPS16	5	Yes	Yes	
XIRP2	5	Yes	Yes	
ZFHX4	6		Yes	

Table 5.7 Commonly mutated genes in the CTCs of SCLC patients. Genes mutated due to either a SNV or Indel in five or more of the 10 patients whose SCLC CTCs were analysed by WES. The presence of mutations in these genes in the tumour samples sequenced in the Peifer and Rudin SCLC tumour genome profiling publications was also assessed. Whether genes are identified as cancer-related according to a search of the geneRIF database (section 2.10.3) is also noted.

The 19 previously described frequently mutated genes were also analysed using QIAGEN's Ingenuity® Pathway Analysis (IPA®) (www.qiagen.com/ingenuity) (core analysis with standard settings) (appendix 6 table A6.1). In these analyses all 19 genes were reported as cancer associated genes. Cellular functions affected by these genes include the cell cycle (TP53, INPP5D and STARD9) and cellular assembly and organisation (TP53, INPP5D, NEB and TJP2). All of the genes were reported to be mutated in tumour samples in the previously described studies except for PRRC2C (table 5.7) [70, 91]. LRP1B was mutated in five of the patients CTCs and was also mutated in 15 samples in both the Peifer and Rudin publications [70, 91]. LRP1B has been identified as a putative tumour suppressor gene with roles in cell adhesion and migration, lipoprotein catabolism, neuronal process outgrowth and early embryonic development [279, 280]. The identification of the mutation of LRP1B in multiple SCLC cases suggests the function of LRP1B in SCLC warrants further investigation.

In addition to identifying those genes that were frequently mutated amongst the SCLC CTCs (table 5.7), an assessment of whether genes or pathways identified in studies of SCLC tumour biopsies as being commonly mutated were also mutated in the CTCs was undertaken (table 5.8). TP53 and RB1 as stated previously are frequently identified as being mutated in SCLC biopsies [70, 91]. TP53 was mutated in 6 of the patients CTCs analysed. No RB1 mutations were seen which suggests this region may be poorly covered in the WES as a high rate of RB1 mutations that has been noted in other studies of SCLC [70, 91].

A

Gene	Number of patients CTCs mutated in
TP53	6
RB1	0
COBL	1
CREBBP	3
EP300	1
EPHA7	1
MLL	0
MLL2	0
SLIT2	2

B

Pathway/Gene Family	Number of Patients CTCs mutated in
PI3K pathway	3
Mediator Complex	3
Notch and Hedgehog family	4
SOX family	2
Glutamate receptor family	4

Table 5.8 Commonly mutated genes and pathways identified in publications analysing SCLC tumour biopsies. The number of patients' CTCs containing mutations in genes (A) and pathways/ gene families (B) identified as frequently mutated in SCLC tumour biopsy publications is detailed in the tables [70, 91].

Mutations in MLL, MLL2, CREBBP, EP300, SLIT2, COBL and EPHA7 were identified in SCLC tumours by Peifer et al [70]. Three patients had mutations in CREBBP, two patients had mutations in SLIT2 and mutations were identified in one patient's CTCs for each of EP300 and COBL. Rudin et al identified frequently mutated pathways or gene families in their analysis of SCLC tumours [91]. In the 10 patient's CTCs assessed in my study, genes in the PI3K pathway had mutations in three patients, the mediator complex genes (multiprotein complex that functions as transcriptional coactivator) were mutated in three patients, Notch and Hedgehog family members were mutated in four

patients, SOX family members were mutated in two patients and the glutamate receptor family members were mutated in four patients. It is therefore possible to use CTCs to identify commonly mutated genes and pathways found in SCLC tumour tissue.

5.4A.4 Identification of Genomic Alterations in Potentially Clinically Actionable Genes in CTCs

5.4A.4.1 Identification of Genomic Alterations in Potentially Clinically Actionable Genes in CTCs Isolated at Baseline from Patients

Another area of interest in sequencing SCLC tumours is the detection of clinically actionable mutations which could be used to direct targeted therapy. An analysis of whether the baseline CTCs from SCLC patients could be used to identify genomic aberrations in genes of clinical interest is therefore of relevance. As previously discussed Ross et al identified actionable mutations in 53% of 98 SCLC tumour biopsies they sequenced [94]. An initial analysis of the baseline CTCs of the 10 SCLC patients was undertaken to see if there were any mutations (SNVs or Indels) or changes in copy number (amplifications or deletions) in potentially actionable genes. Potentially actionable genes were identified from FDA approved targeted therapies and a search of relevant literature. Pragmatic criteria for the identification of SNVs, Indels, amplifications and deletions were adopted as this was a proof of principle analysis. If a SNV or Indel was present in a single CTC in a gene of interest it was noted to be mutated in a patient. As more CTCs were analysed for CNA than WES more stringent criteria were used and so if an amplification or deletion was present in 2 or more CTCs the gene was noted to be altered in a patient. This was a proof of principle analysis and therefore the pragmatic criteria of using just one gene for SNV and Indel identification was used, though it is acknowledged that this would be insufficient to accurately identify mutations from single cells, particularly in a clinical setting. The genes were assessed for the presence of any mutation, rather than the presence of the classical mutations associated with the use of a targeted therapy, for example the presence of any BRAF mutation rather than just the presence of the V600E mutation was noted. This was to demonstrate the proof of principle that these clinically relevant areas of CTCs genomes could be sequenced using the techniques developed and to assess the rate of mutations in these clinically relevant genes.

Genomic aberrations were identified in potentially actionable genes in CTCs from all 10 patients (table 5.9). The most common aberrations noted were copy number changes which were seen in BCL2, PIK3CA, RICTOR and PTEN. This may reflect the fact that

there were more CTC samples available for the analysis of alteration in copy number than had been whole exome sequenced. The gene with the greatest number of patients with SNVs or Indels identified in their CTCs was ROS1 which was mutated in three patients. SNVs or Indels were identified in the CTCs of two patients in RET, NOTCH2, NOTCH3 and PTCH2. These results clearly demonstrate the fact that CTCs could be used to sequence clinically actionable targets for inclusion of patients with SCLC into targeted trials. The copy number changes in BCL2, ERBB2, FGFR1, MET, and RICTOR are all potentially actionable with targeted agents in trials [54, 94, 281-283]. The EGFR mutation identified in patient 4 was P848L in exon 21 which has been identified in NSCLC cases but may not respond to EGFR TKIs [284]. Interestingly patient 2 had a stop gain mutation in PTCH1 (p.Q1366X) and was also noted to have amplification of SMO in their CTCs. This suggests they may respond to Vismodegib, a hedgehog pathway inhibitor which has been used in basal cell carcinoma further highlighting how CTCs could be used to direct targeted therapy [285, 286]. The identification of the MET amplifications in patients 2 and 10 also demonstrates how CTCs could be used to identify resistance mechanisms for targeted therapies such as EGFR TKIs [119].

5.4A.4.2 Changes in Mutations in Potentially Clinically Actionable Genes in Longitudinal Patients CTC Samples

One of the potential clinical uses that has been discussed for CTCs is monitoring clinically actionable mutations over the course of a patient's treatment, using CTCs to assess genomic changes that could be used to direct targeted therapies or to understand acquired mechanisms of resistance. To investigate this possibility a comparison of the mutations identified in potentially actionable genes in the baseline and relapse CTC samples from patients was undertaken. Summaries of the treatment histories of these five patients with SCLC are given in figure 5.13. Mutations either SNVs, Indels or changes in copy number were identified in the CTCs from the baseline samples and the relapse samples from the five chemoresponsive patients in clinically actionable genes as in section 5.4A.4.1. The mutations identified were again not the classical mutations for use of the specific targeted agents but act as a proof of principle about the methods used for sequencing and the potential utility of longitudinal monitoring using CTCs. The genes which contain mutations identified in the five patients' CTCs are detailed in table 5.10.

Gene Name	Total number of patients with mutation in gene	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
ALK	1	0	0	0	1	0	0	0	0	0	0
BCL2	8	1	1	0	1	1	1	0	1	1	1
BRAF	1	0	0	0	0	0	0	0	1	0	0
BRCA2	1	1	0	0	0	0	0	0	0	0	0
CCND1	4	1	0	0	0	1	1	0	1	0	0
EGFR	1	0	0	0	1	0	0	0	0	0	0
EPHA3	3	0	0	0	0	1	0	1	0	1	0
ERBB2	5	0	1	0	1	1	0	0	1	1	0
FGFR1	2	0	0	0	1	0	0	0	0	1	0
FLT1	1	0	0	0	0	0	0	0	0	1	0
KIT	1	1	0	0	0	0	0	0	0	0	0
MAP2K3	1	0	0	0	1	0	0	0	0	0	0
MET	3	0	1	0	1	0	0	0	0	0	1
MTOR	1	0	0	0	0	1	0	0	0	0	0
MYCL1	5	1	1	1	0	0	1	0	1	0	0
NF1	2	0	0	0	1	0	1	0	0	0	0
NOTCH1	1	0	0	0	1	0	0	0	0	0	0
NOTCH2	2	0	1	0	1	0	0	0	0	0	0
NOTCH3	2	1	0	0	0	0	1	0	0	0	0
PIK3CA	7	1	1	0	1	1	0	0	1	1	1
PTCH1	1	0	1	0	0	0	0	0	0	0	0
PTCH2	2	0	0	0	1	0	0	0	0	0	1
PTEN	6	0	1	1	1	0	0	1	1	1	0
RET	2	0	0	0	1	1	0	0	0	0	0
RICTOR	6	0	0	0	0	1	1	1	1	1	1
ROS1	3	1	0	0	0	1	0	0	0	1	0
SMO	3	0	1	0	1	0	0	0	0	0	1
TSC1	1	0	0	0	1	0	0	0	0	0	0

Key

Loss
Gain
SNV
Indel
SNV and Gain

Table 5.9 Mutations in potentially clinically actionable genes identified in the baseline CTCs of SCLC patients. WGS to assess changes in copy number and WES of DNA libraries created from amplified single and pools of CTCs was carried out. SNVs (green), Indels (purple), gains of copy number (red), loss of copy number (blue) and both SNV and gains of copy number (brown) were assessed in the CTCs of 10 patients with SCLC, in potentially clinically actionable genes.

	P4		P5			P6		P8			P10	
Gene Name	**T1**	**T2**	**T1**	**T2**	**T3**	**T1**	**T2**	**T1**	**T2**	**T3**	**T1**	**T2**
ALK	1	1	0	0	0	0	0	0	0	0	0	1
ATM	0	0	0	0	0	0	0	0	1	0	0	0
BCL2	1	1	1	1	1	1	1	1	1	1	1	1
BRAF	0	0	0	0	0	0	0	1	0	0	0	0
BRCA2	0	1	0	1	0	0	0	0	1	0	0	0
CCND1	0	0	1	1	1	1	1	1	1	1	0	0
EGFR	1	0	0	0	0	0	1	0	0	0	0	0
EPHA3	0	0	1	1	1	0	1	0	1	0	0	0
ERBB2	1	1	1	1	1	0	1	1	1	1	0	0
FGFR1	1	1	0	0	1	0	1	0	0	0	0	0
FLT1	0	0	0	0	0	0	1	0	0	0	0	0
FLT3	0	0	0	0	0	0	0	0	0	1	0	0
KDR	0	0	0	0	1	0	0	0	0	0	0	0
KIT	0	0	0	1	0	0	0	0	0	0	0	1
MAP2K3	1	0	0	0	0	0	1	0	0	0	0	0
MET	1	0	0	1	1	0	0	0	1	0	1	1
MTOR	0	0	1	1	1	0	0	0	0	0	0	0
MYCL1	0	0	0	0	1	1	0	1	1	1	0	1
NF1	1	0	0	0	1	1	1	0	0	0	0	0
NOTCH1	1	1	0	0	0	0	0	0	0	0	0	0
NOTCH2	1	1	0	0	1	0	0	0	0	0	0	0
PDGFRA	0	0	0	1	0	0	0	0	0	0	0	0
PIK3CA	1	1	1	1	1	0	1	1	1	1	1	1
PTCH1	0	1	0	0	0	0	0	0	0	0	0	0
PTCH2	1	0	0	0	0	0	0	0	0	0	1	0
PTEN	1	1	0	0	0	0	0	1	1	1	0	1
RAF1	0	0	0	0	0	0	0	0	1	0	0	0
RICTOR	1	1	1	1	1	1	1	1	1	1	1	0
ROS1	0	0	1	1	1	0	0	0	1	1	0	0
SMO	1	0	0	0	1	0	0	0	0	0	1	1
TNFSF11	0	0	0	0	0	0	0	0	0	0	0	1
TSC1	1	1	0	0	0	0	0	0	1	0	0	0
VHL	0	0	0	1	0	0	0	0	0	0	0	0

Key

Loss
Gain
SNV
Indel
SNV and Gain

Table 5.10 Mutations in potentially clinically actionable genes identified in the CTCs of SCLC patients in baseline and relapse samples. WGS to assess changes in copy number and WES of DNA libraries created from amplified single and pools of CTCs was carried out.

SNVs (green), Indels (purple), gains of copy number (red), loss of copy number (blue) and both SNV and gains of copy number (brown) were assessed in the CTCs of five patients with SCLC. CTCs were isolated at baseline (T1), at relapse after first-line chemotherapy (T2), and if patients responded to second-line chemotherapy at the diagnosis of relapse after second-line therapy (T3).

Many of the mutations identified remain constant between the different time points, however changes in some of the mutations were noted. Some mutations were not identified in the relapse CTC samples that were identified in the baseline CTC samples from patients, for example a EGFR mutation sequenced in patient 4s CTCs at baseline was no longer present in the relapse time point CTCs. This suggests the mutation may only have been present in some of the patients' CTCs or it may have only been present in a subclone of the CTCs which are not present at relapse. Mutations were also identified in the relapse time points' samples, T2 and T3, but not the baseline time point samples, T1, such as the development of a PTCH1 mutation in the relapse CTC samples from patient 4, but not the baseline CTC sample. In patient 5 a stopgain mutation in MET (p.S1338X) had developed at relapse in the CTCs which was not found in the baseline CTC sample. This highlights that CTCs can be used as a method of monitoring tumour evolution in addition to acting as a method of identifying mutations in clinically actionable genes longitudinally.

5.4A.5 Discussion

CTCs have already been demonstrated to be a prognostic biomarker in SCLC [204]. The investigation of CTCs' utility to research the biology of this disease has been limited. The broad aim of this chapter was to explore some of the potential clinical and research uses for SCLC CTCs beyond prognosis. Biopsies, as previously discussed, can be challenging to obtain in lung cancers that are of sufficient size not only to establish histological diagnosis, but also to have tissue remaining for research [2]. As surgical resections are rarely performed in SCLC in contrast to NSCLC this issue is particularly acute [3]. Serial assessments of disease are necessary to explore the development of mechanisms of resistance to chemotherapy that evolve in tumours during treatment. Patients may be reluctant to consider further biopsies due to fear of discomfort, particularly if their disease is progressing and there are limited additional choices of therapy. It is also relevant when considering biopsies for purely research purposes to note that there is a risk of morbidity, and in rare cases mortality, from lung biopsies [23]. Liquid biopsies such as CTCs are therefore a potentially very relevant and attractive research, and clinical tool in SCLC. Although the attractions of using

CTCs as a liquid biopsy are evident there remain significant challenges to overcome to reveal their potential.

5.4A.5.1 Quality Control of CTC Analyses

One of the key issues for CTCs is the small numbers of cells to profile in a standard blood draw. This requires the techniques to profile single or small numbers of cells containing very limited amounts of nucleic acids. This necessitates amplification steps to generate sufficient input DNA for downstream molecular analysis. Due to the inherent errors which can occur during amplification using DNA polymerases, inaccurate representation of some regions of the genome will be observed [264]. The method of amplification used will also influence the errors beyond those produced by the DNA polymerase alone [211]. In my study, the Ampli1 WGA kit was used, which is a PCR-based method of amplification. During processing the DNA is digested using an MSE1 restriction endonuclease resulting in digestion of the DNA at TTAA sites in the genome. This results in some fragments being less efficiently amplified due to their larger size, influencing the coverage of some regions of the genome. The quality of the product of the amplification also critically depends on the quality of the DNA input to the PCR reaction. CTCs isolated from the blood stream may be undergoing apoptosis resulting in DNA degradation which could result in poor amplification [179, 287]. The quality of the amplification of a single cell is therefore dependent on both the input DNA and the amplification process.

The Ampli1 QC assay was developed as a method of assessing the DNA produced using the Ampli1 WGA kit. It is a multiplex PCR assessing the amplification of 4 loci. Polzer at al. investigated the use of the Ampli1 QC assay as a quality control of the WGA product using the results to assign a GII ranging from 0 to 4, with higher numbers indicating better quality amplified DNA [237]. An analysis of the success of three molecular biology assays, Sanger sequencing, qPCR and aCGH, was compared to the GII of the input material. They noted improved success of the assays if the samples used as input had a higher GII score, though they were able to successfully perform the assays with samples with all GII scores. The results presented in this chapter also support the findings that successful CNA can be performed with samples with lower GII. When assessing the CTCs isolated from patient 1 as an example, CTCs with GII scores from 1 to 4 were included in the analysis. When comparing the CNA data generated from the CTC samples, it is apparent that the CTC with a GII score of 1 groups together with the majority of the other CTCs, including the pool of CTCs that

had a GII of 4, as it had a very similar CNA appearance (figure 5.2). This would support the fact that for SCLC CTCs the GII score is not an absolute when it comes to the success of downstream molecular analyses, as noted in the discussed paper. As there are only 4 amplicons tested to represent the success of amplification of DNA across the entire genome there will be limitations to the ability of the QC assay to capture the success of the amplification. When considering SCLC in particular the use of an amplicon covering TP53 may also influence the success of the QC PCRs, given the high rates of loss and mutation of this gene in SCLC. Sample selection in this research was therefore based on choosing the cells of each type, CTCs and WBCs, with the highest GII for analysis available, but samples with lower GIIs were processed to allow the assessment of more CTCs per patient.

In addition to using the GII for sample selection, there were further attempts to control the quality of the data used in the final analyses. WBCs, both single and pools, were isolated from the same blood samples as the CTCs and processed together downstream. They were used as controls for the CTCs from each patient. On average one would expect WBCs to have normal copy number with two copies of each locus on the genome. The WBC data was therefore analysed to reveal any areas which deviated from a copy number of two across the samples analysed. The aim was to identify regions which were prone to problems with amplification and sequencing so may be more likely to contain erroneous results in the CTCs from technical, rather than biological variation. When considering the WES data any SNV or Indel identified within a WBC was deemed to be unreliable so excluded from the entire sample set. This may result in the loss of biologically relevant data but was felt to allow increased confidence in the remaining data used for the analyses. There will, from the nature of amplification and sequencing, still be stochastic errors within the individual CTCs analysed. However, excluding the analysis of changes in potentially clinically actionable genes, the analyses were based on comparing similarities or differences across groups of CTCs rather than in single samples. The same technical errors would need to occur across groups of CTCs but not be present in WBCs to be included. It was therefore hypothesised that the methods employed would reduce the incorporation of technical errors in to the final analyses.

Analysing the results from individual patients did show some samples with markedly different CNA profiles from the other CTCs from the same patient e.g. single CTC 2, 4 and 7 from patient 1 (figure 5.2). It is possible this is due to technical variation or else represents biological variation such as a cell in G2 or a cell which has a different genomic profile as it comes from a separate area of the tumour from the others

sampled. To investigate if CTCs which appeared visually different to the other CTCs from a patient were likely to represent a consistent technical problem, an examination of their GII, uniquely mapped reads, percentage of uniquely mapped reads out of the total number of reads and any alteration with increasing the depth of the reads was undertaken. As there was no consistent pattern that could be noted to exclude these cells it was felt they were not all linked to any single technical variability. It was therefore felt that consistent criteria could not be created to exclude them from analyses, nor if they represent biological variation was it clear it would be advisable to exclude them, so all the cells were included in the final analyses. It should again, however, be noted that the majority of the analyses were based on comparing data from multiple samples rather than drawing conclusions about data from individual samples.

5.4A.5.2 Comparison of Genomic Profiles of SCLC CTCs and Published Tumour Biopsy Samples.

The research from tumour biopsies to date suggests that there are key genomic processes driving the development and growth of SCLC such as aberrations in TP53 and RB1 [2]. To utilise CTCs in place of tumour biopsies in studies it is important that they reflect the general genomic landscape of SCLC. The presence of commonly noted changes in copy number and mutations from SCLC tumours were investigated in the CTCs from the 10 patients whose samples were analysed in this chapter. The frequently noted chromosomal changes in tumour biopsies such as the loss of 3p and 17p and amplification of 3q were noted in the majority of the CTCs analysed. The losses of specific tumour suppressor genes such as TP53, RB1 and PTEN were also noted across the majority of the CTCs processed (figure 5.3). Analyses of tumour biopsies have revealed a variety of genes that are commonly amplified in some, though not all SCLC tumours. The results of the analyses of the CTCs again reflected this, with for example amplification in BCL2, SOX2, and MYC family members seen in some of the CTCs analysed (figure 5.3). The most frequently amplified genes in this sample set were on 18q, 18p and 3q. Although the changes in 3q have been frequently noted 18q and 18p have not been noted to be frequently amplified in the two large studies of SCLC tumour biopsies [70, 91]. The analysis of CNA changes in CTCs has therefore captured the commonly noted changes in tumours. This supports the idea that CTCs could be used to investigate the major processes driving SCLC in research.

WES of CTC and WBC samples from the same 10 SCLC patients was also carried out. Analysis of genes commonly noted to be mutated across the set of 10 patients' CTCs was carried out. TP53 was one of the most frequently mutated genes, as would be expected in SCLC. Interestingly each of the TP53 mutations was identified in only one patient's CTCs as was noted in the TAmSeq analysis (Section 4B.4.4.) supporting previous research in which a diverse range of mutations in TP53 have been described [78]. Surprisingly no RB1 mutations were noted despite their reported high frequency in SCLC, though RB1 loss of copy number was frequently noted. Given the high rate of RB1 mutations in SCLC this may reflect a technical issue in sequencing this region, as discussed in chapter 3, with poor exon enrichment in this region of the genome. Mutations in the introns of RB1 which result in loss of function have been noted in research, and would not have been sequenced in this analysis but this is unlikely to account for the overall lack of RB1 mutations noted [288]. This suggests that the approaches used do have a limitation in being able to sequence RB1 adequately. Mutations in other frequently mutated genes and pathways identified from other studies, such as genes encoding histone modifying enzymes and the PI3K pathway, were also identified amongst the patients CTCs sequenced in this research (table 5.8). This again reflects the fact that many frequently noted genetic aberrations in SCLC can be identified in CTCs in addition to tumour tissue.

Genes mutated in more than five of the 10 patients' CTCs at baseline were identified from the WES data, to identify the potentially frequently mutated genes in SCLC CTCs. Of the 19 genes which were identified all but PRRC2C had been noted to be mutated in SCLC samples in either the tumour biopsies profiled by Peifer et al or those profiled by Rudin et al (table 5.7) [70, 91]. LRP1B was mutated in five of the SCLC patients in this CTC study and was noted to be mutated in 15 of the samples sequenced in each of the two large studies of SCLC tumour biopsies. This suggests that LRP1B is frequently mutated in SCLC and so may have an important role in the progression of some tumours. It is a member of the low density lipoprotein receptor family and known to have roles in a wide range of processes including cell adhesion and migration, lipoprotein catabolism, neuronal process outgrowth and early embryonic development [279, 280, 289]. It has been noted to be mutated in several tumour types such as NSCLC, triple negative breast cancer, ovarian and renal cancer, and is hypothesised to act as a tumour suppressor gene in these cancers [289-292]. LRP1B had been identified as being inactivated by genetic and transcript alterations in almost 50% of NSCLC cell lines analysed in a study suggesting a link with lung cancer pathogenesis [279]. In a separate study, mutations were noted in four of 23 NSCLC cell lines and

alterations in transcripts seen in 11 of 36 NSCLC cell lines, but in only one of the 19 SCLC cell lines studied [293]. However, its frequent mutation in this CTC analysis and in the two studies of SCLC tumour samples suggests it may have a key role as a tumour suppressor gene in SCLC pathogenesis as well as NSCLC. The deletion or downregulation of LRP1B has been linked with acquired Doxorubicin chemoresistance in ovarian cancer [292]. However, it was mutated in two of the chemoresponsive as well as three of the chemorefractory patients' CTCs in this study suggesting it is not playing a role in chemoresistance in response to platinum in these patients. The identification of the frequent mutations in LRP1B confirms that CTCs could be used to identify new genetic aberrations that warrant further investigation to determine the role they are playing in SCLC.

5.4A.5.3 Identification of Genomic Alteration in Potentially Clinically Actionable Genes in SCLC CTCs

The ability of CTCs to act as liquid biopsies in clinical trials to identify selected biomarkers makes them an attractive option to study in place of tumour tissue. Entry to clinical trials is increasingly contingent on the presence of specific predictive biomarkers, such as EGFR mutations. This necessitates the genetic analysis of tumour samples from each patient. As already discussed, the biopsies in SCLC are often small, with limited material available after establishing the histological diagnosis for genetic analysis. If trial entry is being considered after first-line therapy the diagnostic biopsy may also no longer be relevant to analyse given the change in the tumour genomes that can occur under the pressure of treatment. CTCs could therefore be used in place of a repeat biopsy of the primary tumour or the metastases. It would also be useful to monitor tumours longitudinally to allow the investigation of the development of resistance mechanisms that can develop in response to targeted agents as this would provide valuable knowledge for drug discovery. Analysis of the baseline CTC samples from the 10 patients was carried out to see if they could be used to investigate CNA, SNVs and Indels occurring in genes of clinical interest. The results presented confirm that CTCs can be used to identify a range of genetic alterations in genes of clinical interest (table 5.9). Within each of the 10 patients changes in potentially targetable genes were identified. Copy number changes in BCL2, PIK3CA and RICTOR were the most frequently described genetic alterations in these patients. Although this may reflect the fact that there were more CTCs analysed for CNA than were WES, in the paper by Ross et al investigating clinically actionable alterations, RICTOR amplification was also the most common alteration seen [94].

Even though the overall analysis was designed just to identify alterations in clinically relevant genes the analyses performed in these patients did identify clinically actionable changes which could be used to direct therapy. For example patients with amplifications in RICTOR may respond to mTORC2 inhibitors [294]. This provides clear evidence that CTC analysis could personalise patients' therapy.

The potential to use CTCs for clinical trials raises interesting questions to consider in relation to heterogeneity in tumours. The number of CTCs that would need to be analysed from an individual patient to be confident that they did or did not have a relevant genetic alteration would need to be considered. There is evidence that mutations within tumours are heterogeneous both within the primary and between metastases [144]. As CTCs reflect the heterogeneity that occurs in tumours the presence of a predictive biomarker, such as a BRAF mutation may not be universally present amongst all the CTCs from a patient. For CTCs to be used in place of tumour tissue for entry in to clinical trials, the proportion of CTCs from a patient that would need to contain a mutation would also need to be considered. In many ways this is an issue that is not acknowledged in clinical trials currently where standard practice would be that a single biopsy was assessed for the presence of a predictive biomarker. There would be no consideration as to how this single biopsy reflected the mutation status across the patient's different tumour deposits. In many ways analysis of CTCs allow the investigation of how the demonstrated heterogeneity in a biomarker influences patients' response to specific targeted agents with more ease than tumour biopsies. In this study a patient was noted to have gain or loss of copy number in a specific gene if two or more CTCs analysed had alteration in copy number. It was, however, evident that the proportion of CTCs containing any specific CNA varied from patient to patient. It would be very interesting to see if this translated into varied degrees of response to the relevant targeted agents.

The ease of repeating blood samples in contrast to biopsies is another strength of CTC analysis. Repeating the analysis of genetic aberrations present in a patient's CTCs when they progress after receiving therapy would potentially reveal new mutations which could provide new avenues for treatment. This was demonstrated here with new mutations identified in patients' relapse CTC samples that had not been identified at baseline (table 5.10). Beyond this in clinical trials the analysis of CTCs over the course of a patient's therapy could reveal changes in the proportion of CTCs with a predictive biomarker which may be indicative of response, or alternatively the emergence of resistance biomarkers may be revealed such as the amplification of MET. Although these results demonstrate a potential clinical use for CTCs this must be tempered by

the acknowledgement that the costs and time associated with single cell analysis of CTCs do currently limit their uses. This may, however, evolve over time in a similar manner to the decreases in costs associated with NGS allowing its wider clinical use [253].

5.4B Results and Discussion: Comparison of the Genomic Profiles of CTCs Isolated at Baseline from Patients with Chemoresponsive and Chemorefractory Disease

The patients whose CTCs were molecularly analysed consisted of five patients who had chemoresponsive disease, and five patients who had chemorefractory disease to first-line therapy, having progressed within 3 months of completion of chemotherapy. The CTCs were isolated from the 10 patients at baseline, prior to treatment being given and patients were subsequently classified as chemoresponsive or chemorefractory due to their response, and the duration of that response, to first-line chemotherapy. Investigation of the genomic profiles of the CTCs in these two groups of patients was undertaken to assess if differences in these profiles could be used to elucidate the genomic basis for intrinsic resistance to chemotherapy in these patients. A footnote table detailing the chemoresponsive and chemorefractory patients' numbers is included on the pages in this section as a reminder.

5.4B.1 Comparison of CNA in CTCs Isolated from Chemoresponsive and Chemorefractory Patients

CNA analysis of 38 CTC samples isolated from the baseline blood samples of patients with chemorefractory disease and 42 CTC samples from the baseline blood samples of patients with chemoresponsive disease was performed. An initial analysis of changes in copy number in 6360 cancer-related genes (figure 5.4) and 19336 protein-coding genes (appendix 3 figure A3.11) were made across the two groups of patients. The plots of the CNA profiles in the CTCs demonstrate differences amongst these 10 patients. When examining the CNA profiles from the CTCs of an individual patient there are common regions of change seen across all the CTCs. There was, however, also intrapatient heterogeneity between their CTC samples. This is in contrast to the findings by Ni et al in which the CNA patterns from a patient were reproducible between the CTCs from a patient [229]. Given the heterogeneity reported in tumours it is not unexpected that CTCs are heterogeneous as they are likely to have come from

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

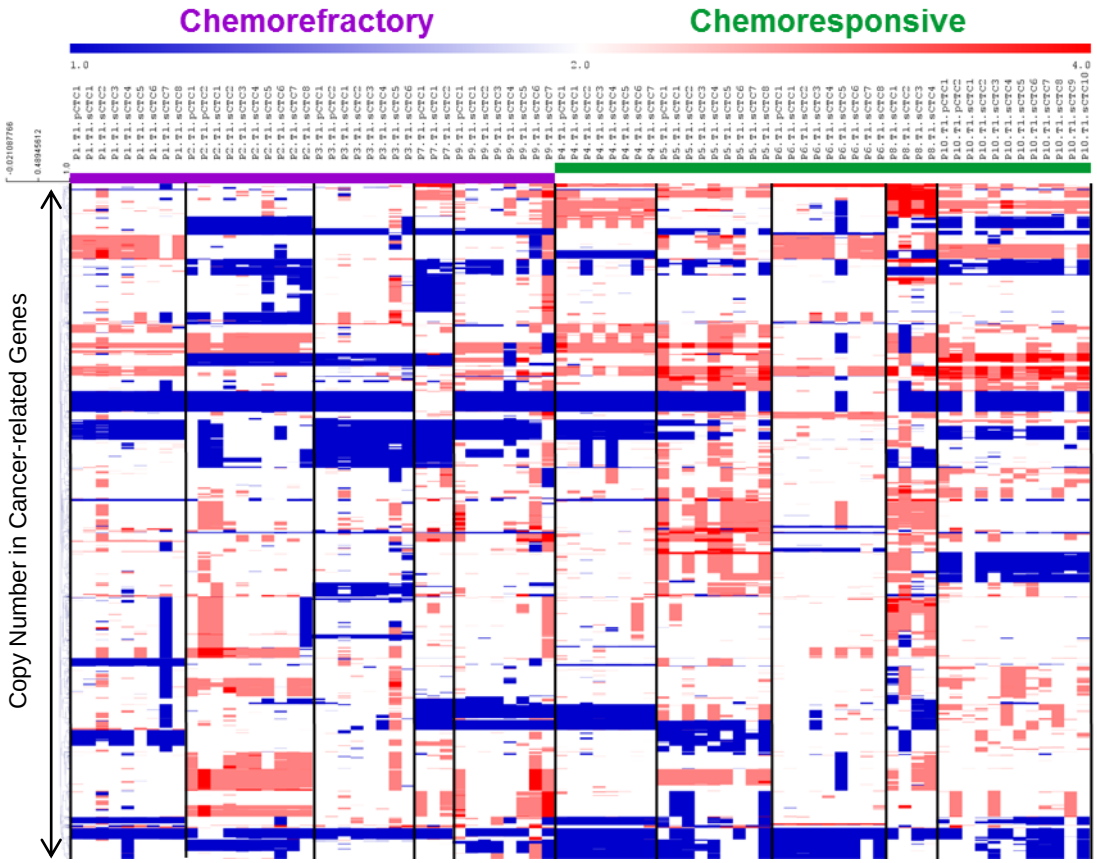


Figure 5.4 Copy number aberration analysis in cancer genes of CTCs from chemoresponsive and chemorefractory patients.
Single CTCs, and pools of CTCs were whole genome amplified and NGS of the DNA libraries performed. CNA analysis was carried out on the amplified CTCs, using sequencing data generated from 6,360 cancer-related genes. The CTCs from chemorefractory patients are in purple whilst the chemoresponsive patients' CTCs are in green. The CTCs from individual patients are divided by black lines. Gains are coloured red, losses are coloured blue and regions with no change are coloured white.

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

multiple areas of the primary tumour and/or from different metastatic sites [144]. When analysing the CDX tumours and paired CTCs in chapter 4 there was a trend of a more loss predominant pattern in the CNA of the refractory tumours whereas both losses and amplifications were seen within the chemoresponsive tumours. Visually, when analysing the CNA plots from chemoresponsive and chemorefractory patients there again appeared a trend towards more losses in the CTCs from chemorefractory patients and more amplifications in the chemoresponsive patients. To assess this further, the copy number for each of the protein-coding genes for each CTC from the chemoresponsive and chemorefractory patients was plotted on a density plot (figure 5.5). This confirmed a higher peak at a copy number of 1, indicating greater losses, in the chemorefractory patients' CTCs compared to the chemoresponsive patients' CTCs. There was also a peak above two in the chemoresponsive patients' CTCs suggesting a greater number of amplifications in this group. This therefore provides further support to the idea of more loss predominant copy number change in refractory SCLC tumours and more amplifications in the responsive SCLC tumours.

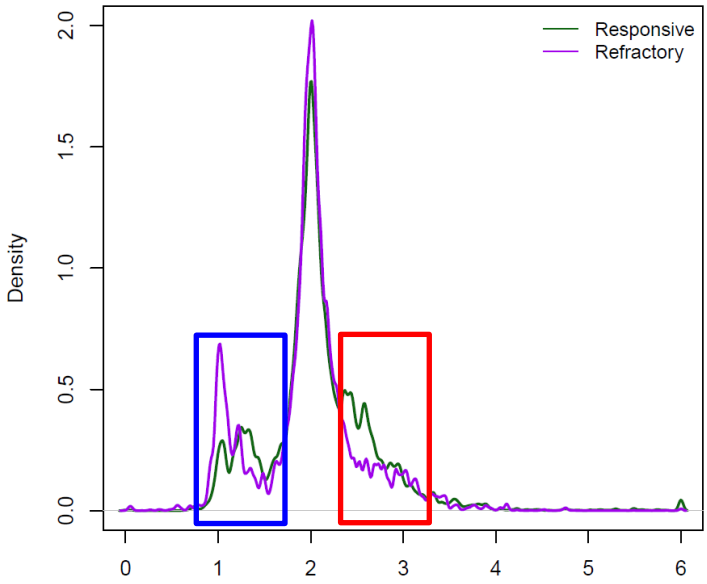


Figure 5.5 Graph of copy number changes in protein-coding genes in the CTCs from chemoresponsive and chemorefractory patients. Distribution of the copy number of 19336 protein-coding genes for the 80 CTC samples isolated from patients with chemoresponsive and chemorefractory disease was plotted. The copy number for the chemoresponsive patients is plotted in green and for the chemorefractory patients it is plotted in purple. The blue box highlights the peak below one indicating loss of copy number and the red box highlights the peak above 2 demonstrating gain in copy number.

To further investigate the differences in the CTC CNA profiles from the chemoresponsive and chemorefractory patients a PCA was carried out using the CNA

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

data generated from the 19336 protein-coding genes (figure 5.6). The CTCs from the two groups of patients, excluding patient 1, show separation on the plot in this unsupervised analysis. This confirms that there are differences in the CNA profiles between the two groups of patients. It is possible that these genomic differences may contribute to the inherent chemoresistance that occurs in the chemorefractory patients. It is again apparent from the PCA that there is heterogeneity of some patients CTC CNA profiles. Some patients, such as patient 6, have tightly grouped CTCs whilst patient 5s CTCs are widely spread. The CTCs from patient 1, who had chemorefractory disease, group with the chemoresponsive patients' CTCs on the PCA. However, analysis of the clinical history of this patient show that patient 1 responded well to the first 3 cycles of chemotherapy delivered with reduction in the volume of the thoracic disease on CXRs during treatment. The patient then developed severe neutropenic sepsis which resulted in deterioration in the patient's clinical condition and the treatment was terminated at that point. The patient went on to progress within 3 months of completing chemotherapy so according to the clinical definition used was chemorefractory. However, these results may suggest that the patient was biologically responsive to this chemotherapy and may have had a more durable response if he had completed the planned course of therapy.

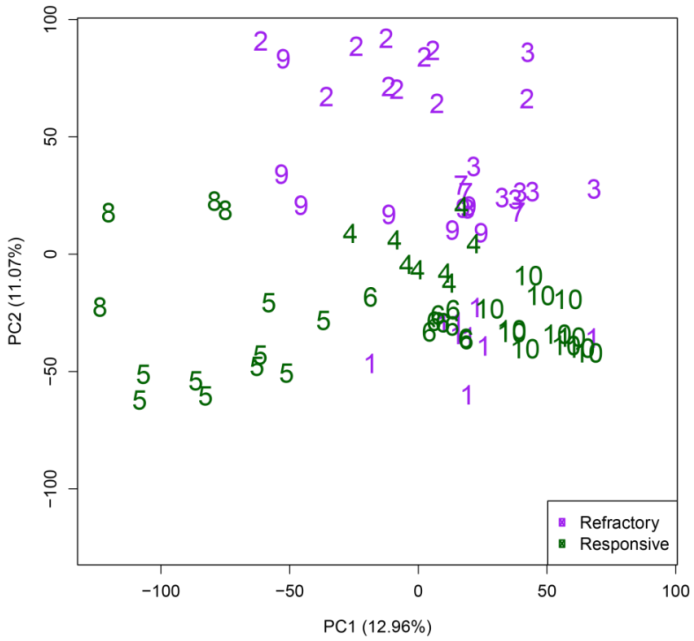


Figure 5.6 PCA of the CNA of protein-coding genes in the chemoresponsive and chemorefractory patients' CTCs. Single CTCs, and pools of CTCs were WGA and NGS of the DNA libraries performed. CNA analysis was carried out on using sequencing data generated from 19336 protein-coding genes. A PCA of the data was then performed and PC1 and 2 are plotted. The CTC samples from chemorefractory patients are coloured purple whilst the chemoresponsive patients' CTCs are coloured green. The patient number from which each

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

CTC sample comes from is indicated as the data marker. The percentage variance explained by the principal components is given in parenthesis.

To further assess the association between CTC CNA profiles and the outcomes of the patients with chemoresponsive and chemorefractory disease, a further PCA was performed using an average value for an individual patients' CTC CNA for each gene. PC1 from this analysis was then plotted against the PFS and OS of the patients (figure 5.7). The plots demonstrate a clear correlation between the increase in principal component 1 (PC1) and the increase seen in both progression free and overall survival in this group of patients.

The overall PCA plot of individual CTCs (figure 5.6) had demonstrated separation of the two groups of patients' CTCs due to differences in PC2. Interrogation of the genes contributing most highly to PC2 was therefore undertaken to reveal the key changes. The top 500 weighted genes in PC2 were identified and found to be located on a restricted number of chromosomes (chromosomes 4, 6, 11, 13, 16 and 18)(appendix 5 table A5.1). Of the top 500 genes 241 were located on chromosome 16 and 156 on chromosome 18 suggesting large scale changes in these chromosomes may influence the chemoresponsiveness of SCLC. Investigation of the direction of change in copy number in these chromosomes reveals on average that genes from chromosome 18 are amplified in the chemoresponsive CTCs, whilst the genes from chromosome 16 are lost in the chemorefractory CTCs.

Analysis of the top 500 genes in Ingenuity® Pathway Analysis linked 373 of them with cancer (appendix 6 table A6.2). However, a subset of these 500 genes have change from normal copy number in the chemoresponsive patients' CTCs and a subset have change from normal copy number in the chemorefractory patients' CTCs. Separate analysis of the genes amplified or lost in the chemoresponsive patients' CTCs, and those that altered in the chemorefractory patients' CTCs were therefore carried out in Ingenuity® Pathway Analysis, to separately analyse the genes whose alteration may be producing chemoresponsive or chemorefractory behaviour (appendix 6 table A6.3 and A6.4). When analysing the genes altered in the chemoresponsive patients' CTCs there is enrichment for genes associated with cellular function and maintenance, cell death and survival and the cell cycle which are all functions expected to be altered in SCLC. Analysis of the genes altered in the chemorefractory patients' CTCs show enrichment of genes associated with drug metabolism, small molecule biochemistry and cell to cell signalling. There is also loss of CES1, CES2 and CES3, carboxylesterase enzymes involved in the hydrolysis of xenobiotics including

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

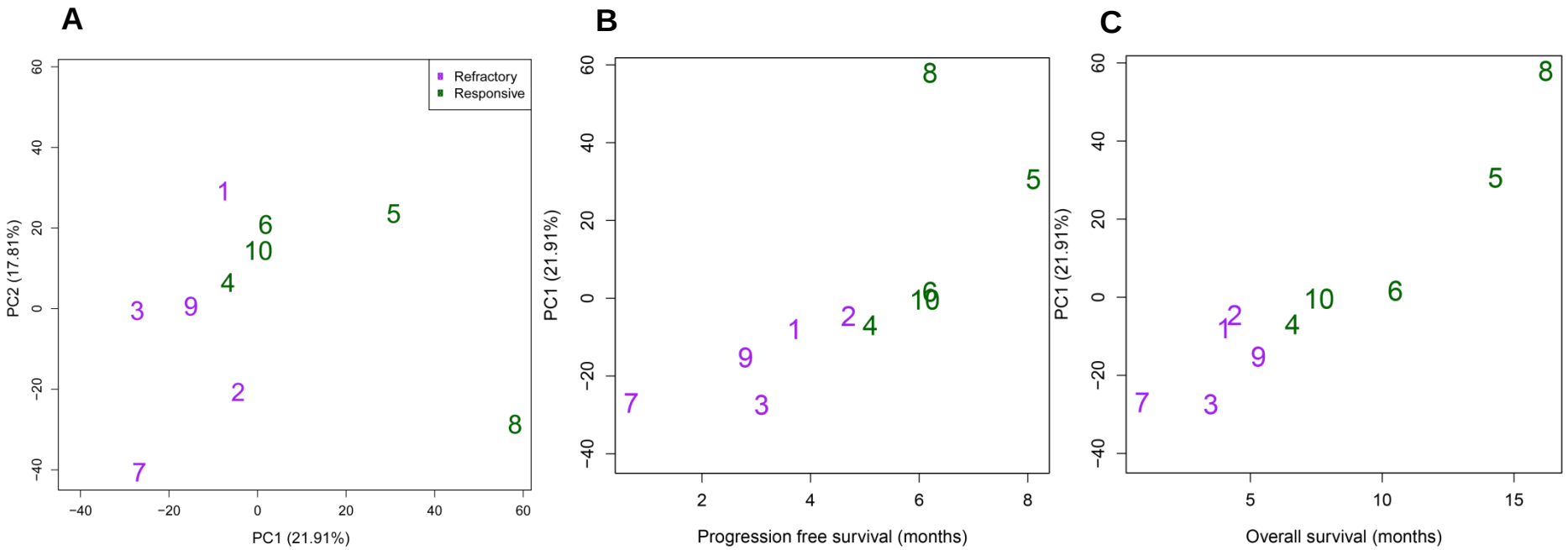


Figure 5.7 Plots of survival against PC1 generated from CNA data. A PCA of copy number changes in 19336 protein-coding genes in baseline CTC samples from the 10 SCLC patients was performed, using the average CTC CNA value per gene for each patient (A). The PFS (B) and OS (C) of the 10 SCLC patients are plotted against PC1 generated from this PCA. The chemoresponsive patients are coloured green and the chemorefractory patients are coloured purple with the patient number indicated as the data marker. The percentage variance explained by the principal components is given in parenthesis.

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

Irinotecan, in the chemorefractory patients (table 5.11). Glutathione S transferase alpha (GSTA) 1 to 5, a family of enzymes which detoxify compounds through conjugation with glutathione, are amplified in the chemorefractory CTCs in contrast to the chemoresponsive CTCs (table 5.11). Amplification of GSTAs has been linked with chemoresistance in studies including analysis of a SCLC cell line [295, 296]. Surprisingly RALBP1, an anti-apoptotic non-ABC glutathione conjugate transporter, is amplified in the chemoresponsive CTCs as it has been linked with doxorubicin resistance (table 5.11) [297, 298]. Excluding RALBP1, the alteration of genes that may promote chemoresistance were found in the chemorefractory patients' CTCs alone. The differences in the CNA profiles of CTCs may therefore be able to provide evidence of potential mechanisms associated with chemoresistance.

A further analysis of the CNA data from the chemoresponsive and chemorefractory patients' CTCs was undertaken to see if a CNA gene signature to distinguish the two groups of patients could be identified. A supervised analysis of the CNA data was undertaken using the linear models for microarray analysis programme (LIMMA) to assess differences between the two groups. LIMMA was chosen for the analysis of data as it is a very robust tool and the change in copy number can be equated to the change in gene expression usually analysed by LIMMA. The bioinformaticians with whom I worked had experience of previously analysing approximately normal data such as proteomic data using LIMMA with good results. An analysis of genes with a 1.5 fold change in copy number between the two groups of CTCs with a 5% false discovery rate (FDR) was carried out.

In total 760 genes were identified as having a significant difference in copy number between the two groups in the LIMMA analysis (appendix 5 table A5.2). The genes were located across chromosomes 3, 4, 5, 8, 11, 13, 15, 16, 18 and 19. Again, as with the unsupervised analysis, the majority of the significant genes were located on chromosomes 16 and 18, with 282 and 167 genes altered respectively on each chromosome. A heat map plot of the CNA changes in the 760 genes identified by the LIMMA analysis as being significantly different between the 80 CTCs from the chemoresponsive and chemorefractory patients was created (figure 5.8). It highlighted the differential patterns of amplification and loss seen between the two groups of CTCs. There is a large area of loss of copy number seen in the majority of the CTCs from the chemorefractory patients and regions of loss and amplification unique to the chemoresponsive patients. The CNA changes again include losses in chromosome 16 in the chemorefractory patients' CTCs and amplifications in chromosome 18 in the chemoresponsive patients' CTCs. The CTCs from patient 1 again, along with a small

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

number of CTCs from patients 2 and 9 groups with some of the CTCs from patient 4, a chemoresponsive patient, on the hierarchical clustering. Of the 760 genes identified by the supervised analysis as being significantly altered between the chemoresponsive and chemorefractory patients 424 were present amongst the top 500 genes identified by the unsupervised PCA analysis.

Gene	Copy number change in chemoresponsive patient's CTCs	Copy number change in chemorefractory patients CTCs	Function	Evidence for role in chemoresistance
CES1 - 3	Normal	Loss	Hydrolysis of xenobiotics	Hydrolysis of irinotecan to active metabolite [299, 300]
GSTA 1 – 5	Normal	Gain	Detoxification through conjugation with glutathione	Increased resistance to nitrogen mustard chemotherapies and doxorubicin [295, 296]
RALBP1	Gain	Normal	Non -ABC glutathione conjugate transporter	Resistance to doxorubicin [297, 298]
CDH1	Normal	Loss	Calcium dependent cell-cell adhesion glycoprotein	Loss promotes metastatic potential of tumour [301]
ATP7B	Normal	Loss	Copper transport protein	Overexpression associated with resistance to platinum chemotherapies [116]
CXCL genes	Loss	Normal	Chemokines with role in immune system, including leukocyte trafficking [302]	Possible mechanism for immune avoidance
PHLPP2	Gain	Loss	Phosphatase which inactivates AKT [303]	Inhibition of chemotherapy-induced apoptosis [304-306]

Table 5.11 Genes of interest identified from analysis of CNA in baseline CTCs from chemoresponsive and chemorefractory patients. Summary of genes identified through analysis of CNA data from the baseline CTCs isolated from chemoresponsive and chemorefractory patients with possible roles in chemoresistance.

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

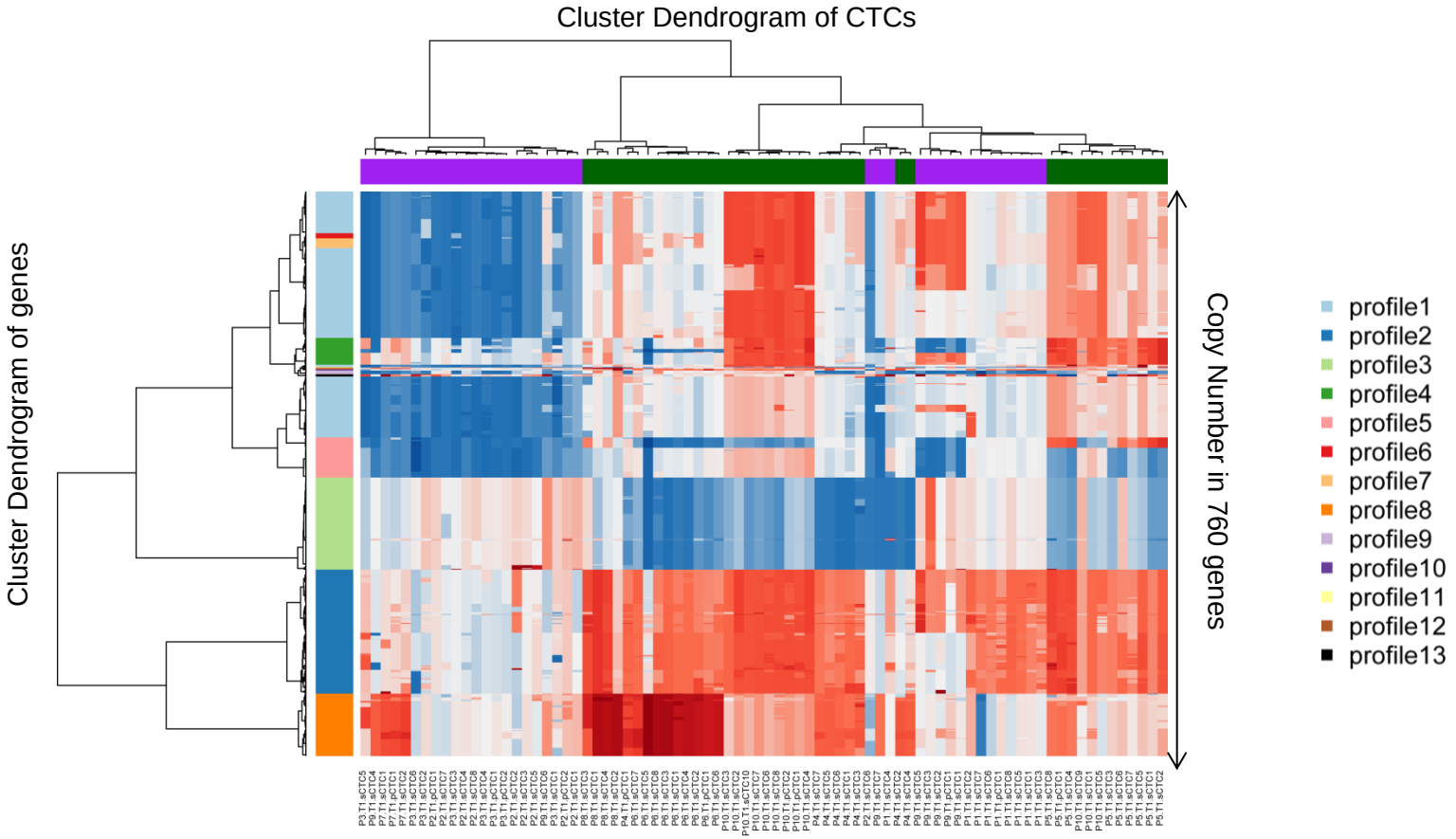


Figure 5.8 Hierarchical clustering of significant CNA differences between the chemoresponsive and chemorefractory patients' CTCs
Single CTCs, and pools of CTCs were WGA and NGS of the DNA libraries performed. CNA analysis was carried out using sequencing data from 19,336 protein-coding genes. Statistical analysis identified significant difference in the copy number data between the chemoresponsive and chemorefractory patients in 760 genes which are plotted on the heat map. Hierarchical clustering of the data was performed. The CTCs from chemorefractory patients are shown in purple whilst the chemoresponsive patients' CTCs are shown in green. Gains in copy number are coloured red, losses are coloured blue and regions with no change are coloured white.

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

Analysis of the 760 genes with significantly altered CNA in Ingenuity® Pathway Analysis confirmed that it was enriched for cancer-related genes with 574 of the genes being cancer associated (appendix 6 table A6.5). The gene set was also enriched for genes associated with cell-to-cell signalling and interaction and genes associated with cellular movement. Again the genes which change in the chemoresponsive patients were analysed separately to the genes altered in the chemorefractory patients to investigate any potential biological changes specific to the different groups of CTCs (appendix 6 table A6.6 and A6.7). Genes whose copy number is altered in the chemoresistant CTCs are enriched for genes associated with cellular growth and proliferation, drug metabolism and small molecule biochemistry. The genes altered associated with cellular growth and proliferation includes epithelial cadherin (CDH1) which has loss of copy number in the chemoresistant CTCs. As CDH1 functions as a tumour suppressor gene its loss could clearly promote the progression of the cancer in these patients and the metastatic potential of these tumours (table 5.11) [301]. Amongst the genes altered associated with drug metabolism are again the carboxylesterases which are associated with the hydrolysis of Irinotecan (table 5.11). The CTCs from the chemoresistant patients also have loss of ATP7B, a copper transporter protein (table 5.11). Overexpression of ATP7B is associated with resistance to Cisplatin and Carboplatin so this finding is at odds with the responses of the patients in this group [116]. The genes altered in the chemoresponsive patients include cell-to-cell signalling and interaction, cellular movement and cell death and survival. The loss of copy number of a cluster of CXC chemokine genes may lead to interference in the immune response through altered attraction of phagocytes and neutrophils (table 5.11) [302].

A PCA of the CNA data from the 760 genes identified as having significant changes between the two groups of patients was performed using an average value per gene from the CTCs from each patient. The PCA plot of PC1 versus PC2 highlights the separation of the chemoresponsive and chemorefractory patients with this signature (figure 5.9). Patient 1 again groups closer to the chemoresponsive patients, despite being chemorefractory.

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

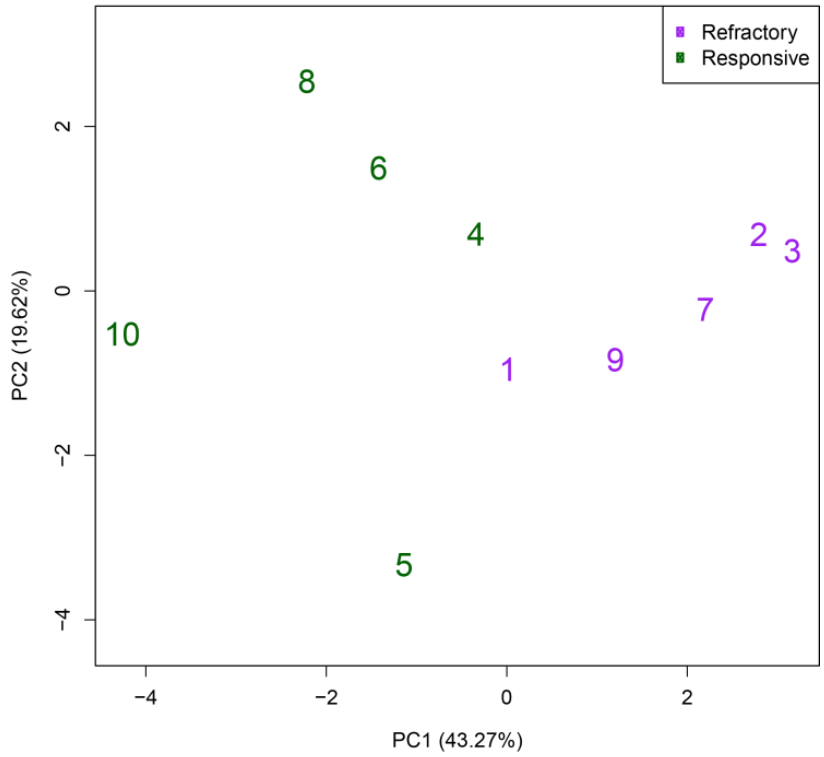


Figure 5.9 PCA of significant CNA changes between the chemoresponsive and chemorefractory patients' CTCs. Single CTCs, and pools of CTCs were whole genome amplified and NGS of the DNA libraries performed. CNA analysis was carried out on the amplified CTCs, using sequencing data from 19,336 protein-coding genes. Statistical analysis identified significant difference in the copy number data between the chemoresponsive and chemorefractory patients in 760 genes on which a PCA was performed. The chemorefractory patients are coloured purple whilst the chemoresponsive patients are coloured green with the patient number indicated. The percentage variance for the components is given in parenthesis.

As with the unsupervised PCA analysis, the correlation of the principal component data, generated from the 760 genes identified from the LIMMA analysis and patient outcomes were assessed. The patients' PFS and OS times were plotted against PC1 created from average CTC CNA values for each gene per patient. The plots again suggest a correlation between changes in the survival and changes in PC1 (figure 5.10). Patient 10, however, sits as an outlier on the trend of increased changes on PC2 being associated with shorter PFS and OS. This patient died of acute shortness of breath due to obstruction of the larynx from a tumour deposit. It therefore may be that the total volume of the disease would not have resulted in death at this point but it was the very specific location of the single area of tumour in the larynx which resulted in the patient's death. It should, however, be noted that when using the PCA data from the initial unsupervised PCA patient 10 was not an outlier.

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

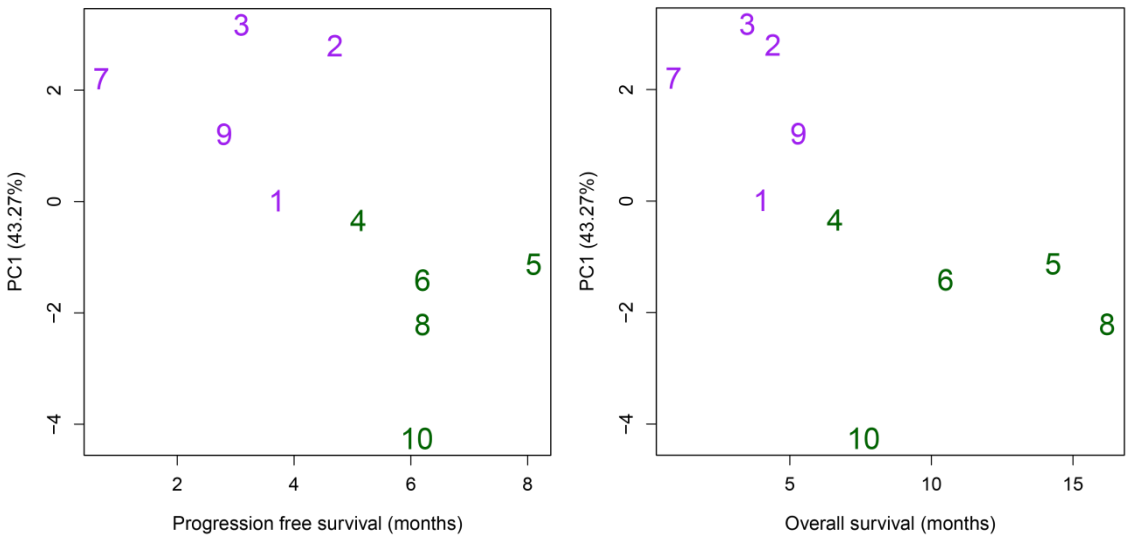


Figure 5.10 Plots of survival against PC1 generated from 760 genes with significant CNA change. The PFS and OS of the 10 SCLC patients are plotted against PC1. PC1 was generated from a PCA of CNAs in 760 protein-coding genes identified in LIMMA as having statistically significant differences between the chemoresponsive and chemorefractory patients' CTCs. The analysis was performed using CNA data from 80 CTC samples with an average copy number value for each gene calculated from all an individual patients CTCs. The chemoresponsive patients are in green and the chemorefractory patients in purple with the patient number indicated. The percentage variance accounted for by the principal components is given in parenthesis.

To further investigate the changes within the 760 gene signature, the genes were grouped into profiles according to the direction of change on average amongst the chemorefractory and the chemoresponsive patients' CTCs and the chromosome on which the genes were located. This analysis reduced the 760 genes to 13 profiles which are shown in table 5.12 with the direction and change of copy number for the profiles in the chemoresponsive or chemorefractory patients highlighted. In all but profile 7, one of the groups of patients has normal copy number for the genes whilst the other group has loss or amplification of the genes noted. In profile 7, 13 genes on chromosome 16 are amplified in the chemoresponsive patients but lost in the chemorefractory patients. This included PHLPP2 which is a negative regulator of the PI3K/AKT pathway [303]. In studies the inhibition of PHLPP2 results in decreased chemotherapy-induced apoptosis whilst its up regulation led to reduced tumour growth (table 5.11) [304]. The changes in PHLPP2 copy number noted in the CTCs are therefore consistent with the response to chemotherapy seen in the chemoresponsive and chemorefractory patients.

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

Key	Chromosome	Number of Genes	Chemoresponsive	Chemorefractory
profile1	16	262	2	1
profile2	18	167	3	2
profile3	4	124	1	2
profile4	13	36	3	2
profile5	13	54	2	1
profile6	16	7	3	2
profile7	16	13	3	1
profile8	5	85	3	2
profile9	15	5	2	1
profile10	11	2	3	2
profile11	3	1	2	1
profile12	8	1	3	2
profile13	19	3	2	3

Table 5.12 Properties of 13 profiles created from 760 genes with significant differences in copy number between the chemoresponsive and chemorefractory patients' CTCs. Analysis of amplified CTC CNA data in LIMMA identified 760 genes as having a statistically significant difference between the chemoresponsive and chemorefractory patients' CTCs. The 760 genes were reduced to 13 profiles based on the average direction of change in copy number in the CTCs from the two groups of patients and the chromosome the genes are located on. The number of genes in each profile and the average copy number of the CTCs in chemoresponsive and chemorefractory patients is noted in the table.

To assess the copy number changes in these 13 profiles further, an average value was calculated for the genes in each profile across the CTCs analysed for each individual patient. Hierarchical clustering of the copy number changes in the 13 profiles for the 10 patients was then carried out (figure 5.11). This confirms the overall trends of loss and gain in the profiles for the chemoresponsive patients and chemorefractory patients. It also demonstrates that there is some interpatient variability between the patients in the chemoresponsive and chemorefractory groups for the individual profiles. Patients 2, 3 and 7 are in a cluster whilst patients 1 and 9 are also clustered with the chemoresponsive patients 4 and 6. Patients 8, 5 and 10 with chemoresponsive disease also all group together on the hierarchical clustering. Using a variety of analysis tools consistent marked differences in copy number were noted between the chemoresponsive patients and the chemorefractory patients CTCs. These differences in copy number identified require further evaluation to see if they in part could underpin some of the mechanisms behind the divergence in the response to chemotherapy in these two groups of patients.

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

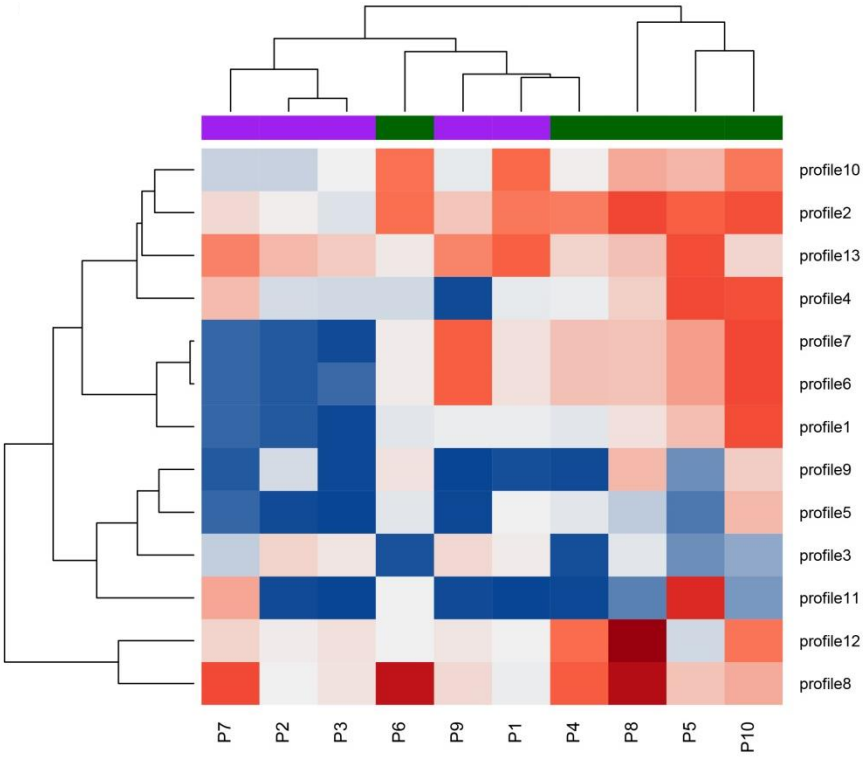


Figure 5.11 Average copy number change in the 13 gene profiles in the chemorefractory and chemoresponsive patients' CTCs. Statistically significant differences between the copy number of the chemoresponsive and chemorefractory patients' CTCs were identified in 760 genes. Thirteen profiles were created from these 760 genes, based on the average direction of change in copy number in the two patient groups' CTCs, and the chromosome the genes were located on. Hierarchical clustering of the data was performed. The chemorefractory patients are shown in purple whilst the chemoresponsive patients are shown in green. Gains in copy number are coloured red, losses are coloured blue and regions with no change are coloured white.

5.4B.2 Comparison of Mutations Present in CTCs Isolated from Chemoresponsive and Chemorefractory Patients

An analysis of differences in mutations identified in the WES data from the chemoresponsive and chemorefractory patients' CTCs was carried out in addition to evaluating differences in CNAs in each cohort of patients. WES was performed on CTC and WBC baseline samples from each of the 10 patients to allow comparison of genes mutated, either due to SNVs or Indels, in the chemoresponsive and chemorefractory patients. The WBCs were used as a control as discussed in section 2.10.6 to remove mutations, which may be due to recurrent technical errors from amplification and sequencing. For identification of differential mutations, genes mutated in three or more of the chemorefractory patients' CTCs that were not mutated in the chemoresponsive patients' CTCs, and those that were mutated in four or more of the chemorefractory patients' CTCs and mutated in just one of the chemoresponsive

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

patients' CTCs were identified. The same set of criteria was applied to identify genes mutated in the chemoresponsive patients' CTCs but not the chemorefractory patients' CTCs.

In total 34 genes were found to be differentially mutated with 5 genes identified as being mutated in the chemorefractory patients but not the chemoresponsive patients' CTCs, and 29 genes mutated in the chemoresponsive but not the chemorefractory patients' CTCs (table 5.13). The pattern of the differentially mutated genes in the chemoresponsive and chemorefractory patients is displayed in figure 5.12. Analyses of the genes mutated in the chemoresponsive patients' CTCs and those mutated in the chemorefractory patients' CTCs were carried out using Ingenuity® Pathway Analysis (appendix 6 tables A6.8 and A6.9). Investigation of the genes mutated in the chemorefractory patients linked all five with cancer. Genes of interest mutated in the CTCs from chemoresistant patients included PCDH10 and HSPG2 which have both been previously linked with chemoresistance [307-309]. Of the 29 genes mutated in the chemoresponsive patients' CTCs 28 were linked with cancer. The mutations in LRP5 in the chemoresponsive patients' CTCs are noteworthy. Cell lines with knockdown of LRP5 had decreased invasion, migration and colony formation whilst a xenograft model had smaller volume tumours in one study [310]. These examples suggest that WES of the CTCs of patients with chemorefractory disease could be used to explore the changes associated with drug resistance with validation of these changes in a larger cohort of patients being required, in addition to function testing using CDX models and SCLC cell lines.

5.4B.3 Discussion

Chemoresistance is a significant problem in oncology as its development results in treatment failure, leading to the earlier death of patients with metastatic disease, and the failure of treatment given with curative intent to those with localised disease receiving adjuvant therapy. Investigating the causes of chemoresistance and potentially revealing new therapeutic approaches is therefore a vital area of research in oncology. In SCLC the marked dichotomy of behaviour between the approximately 80% of patients with chemoresponsive disease and the 20% with disease which is refractory to first-line chemotherapy represents an interesting area of study. There is a very clear difference between these two groups of patients and one of the focuses of this research was to establish if this could be correlated with genomic alterations in CTCs. Patients with SCLC have poor survival but this is particularly true for those with

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

chemorefractory disease. The identification of targetable genomic aberrations linked to chemoresistance could lead to meaningful benefits for this group of patients. Thirty five patients with SCLC, 20 with chemoresponsive disease and 15 with chemorefractory disease were recruited to the study. Of these, molecular analyses were performed on the CTCs from 10 patients but the CTC numbers were enumerated in all the patients' blood samples using CellSearch. The median CTCs counts at baseline from the chemoresponsive patients was 412 per 7.5 ml in contrast to 2182 in the chemorefractory patients. This difference was however not statistically significant. This would imply that the poorer PFS of the chemorefractory patients cannot be explained by greater numbers of CTCs prior to chemotherapy. This would support the idea that chemorefractory patients do not just have more advanced disease with greater disease burden and that the differences in behaviour of these two groups of patients may therefore be explained by molecular alterations in their cancers. The OS was also statistically longer in the chemoresponsive than the chemorefractory patients as would be expected.

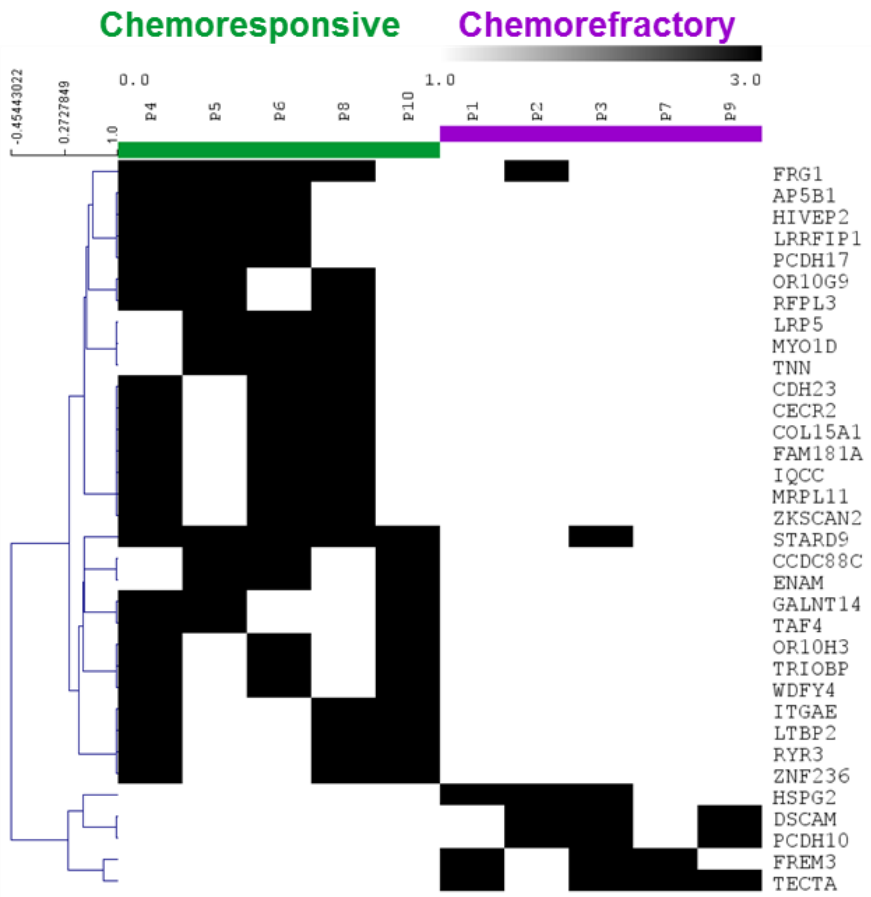


Figure 5.12 Plot of mutated genes in the chemorefractory and chemoresponsive patients' CTCs. WES was carried out on CTCs isolated from five patients with chemoresponsive SCLC

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

and five patients with chemorefractory SCLC. Genes mutated (either SNVs or Indels) in the chemorefractory but not the chemoresponsive patients' CTCs and those mutated in the chemoresponsive but not the chemorefractory patients' CTCs were identified from the sequence data. A plot of the pattern of mutated genes in each patient was created with black indicating the gene is mutated and white indicating no mutation is present. The chemoresponsive patients are in green and the chemorefractory patients in purple.

Genes mutated in the chemorefractory patients CTCs	Role	Genes mutated in the chemoresponsive patients CTCs	Role
DSCAM	Neural cell adhesion molecule	AP5B1	Protein transport
FREM3	Extracellular matrix protein	CCDC88C	Regulation of protein phosphorylation
HSPG2	Extracellular matrix component	CDH23	Cell-cell adhesion glycoprotein
PCDH10	Cadherin -related neuronal receptor	CECR2	Role in chromatin remodelling
TECTA	Tectorial membrane component	COL15A1	Alpha chain in Type XV collagen
		ENAM	Dental Enamel
		FAM181A	Protein-coding gene
		FRG1	Possible role in pre r and mRNA processing
		GALNT14	Polypeptide N-acetylgalactosaminlytransferase
		HIVEP2	Zinc finger containing transcription factor
		IQCC	Protein-coding gene
		ITGAE	Integrin protein, receptor E-Cadherin
		LRP5	Low density lipoprotein receptor
		LRRFIP1	Transcriptional repressor
		LTBP2	Extracellular Matrix Protein
		MRPL11	Ribosome structural constituent
		MYO1D	Myosin protein
		OR10G9	G protein coupled receptor, olfactory
		OR10H3	Olfactory receptor
		PCDH17	Member of cadherin superfamily
		RFPL3	Possible hTERT promoter binding protein
		RYR3	Receptor with role in calcium metabolism
		STARD9	Microtubule dependent motor protein
		TAF4	Transcription coactivator activity
		TNN	Integrin binding, role in central nervous system
		TRIOBP	Actin filament binding
		WDFY4	Protein-coding gene
		ZKSCAN2	Transcriptional regulation
		ZNF236	Transcriptional regulation

Table 5.13 Lists of genes mutated in CTCs from chemorefractory and chemoresponsive patients. WES was carried out on CTCs isolated from five patients with chemoresponsive SCLC and five patients with chemorefractory SCLC. Genes mutated (either SNVs or Indels) in the chemorefractory but not the chemoresponsive patients' CTCs and those mutated in the chemoresponsive but not the chemorefractory patients' CTCs were identified from the sequence data.

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

To investigate the genomic alterations associated with intrinsic resistance to chemotherapy a comparison of the difference in CNA and mutations between the CTCs isolated from five patients with chemorefractory disease and five patients with chemoresponsive disease was carried out. The CTCs were all isolated at baseline prior to the patients receiving any chemotherapy. CNA profiles for the 80 CTC samples isolated from the two groups of patients were created by assessing copy number in protein-coding genes from WGS data. Each of the 10 patients had unique patterns of loss and gains as would be expected but there was also heterogeneity between the CTCs from individual patients (figure 5.4). However, despite this variability there was an overall difference evident between the chemoresponsive and chemorefractory patients CTCs. The chemorefractory patients' CTCs had a more loss predominant pattern of CNA whilst those from chemoresponsive patients had more amplifications present. This visual pattern was confirmed when creating a density plot of the copy number changes in the two groups of patients (figure 5.5). It also reflects the changes seen in chapter 4 when the CDX tumours from the chemoresponsive and chemorefractory patients were profiled. Analysis at Memorial Sloane Kettering of SCLC PDX tumours derived from chemoresponsive and chemorefractory patients' tumours has also revealed similar trends in copy number between the two groups of patient-derived samples (personal correspondence CM Rudin and MC Pietanza, Memorial Sloan Kettering Cancer Centre). These results suggest that chemorefractory disease may be driven by the loss of tumour suppressor genes whilst chemoresponsive disease has both the loss of tumour suppressor genes and the amplification of oncogenes leading to its progression. It may, therefore, be that the amplifications lead to a therapeutic vulnerability in the chemoresponsive patients. The investigation of this hypothesis in the CTCs isolated from a larger cohort of chemoresponsive and chemorefractory patients would therefore be of interest.

To investigate the potential of CTC CNA profiles to discriminate the chemoresponsive and chemorefractory patients a PCA was performed (figure 5.6). This unsupervised analysis confirmed a separation of the CTCs from the two groups of patients suggesting there are changes in copy number which correlated with chemoresponsiveness. CTCs from patient 1 who had chemorefractory disease grouped with the chemoresponsive patients. However, as discussed, the patient's clinical history revealed good response to the first three cycles of chemotherapy and for non-cancer related reasons the treatment was terminated. The failure of the CTCs from patient 1 to group with the CTCs from other patients with chemorefractory disease, may reflect that although chemorefractory by the clinical definition, he was in

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

fact biologically chemoresponsive. The possibility of seeing this distinction in patients whose treatment is abandoned through deterioration in their clinical condition for non-cancer reasons would clearly require more similar patients CTCs' CNA profiles to be analysed but remains an interesting potential explanation of the data.

Comparison of the CNA changes at a patient level and PFS and OS demonstrated a clear correlation (figure 5.7). As the patients were divided into the two groups based on the length of their response to chemotherapy it is unsurprising that there was an association between the principal component and PFS. However, the marked correlation demonstrated on the plot is striking. There was also a close correlation between OS and the principal component. Of note in SCLC PFS is significantly correlated with OS [311]. The marked correlation between the PFS and the principal component generated from the unsupervised analysis of the CNA data provides support that a link with chemoresistance is being revealed. Although the patients were divided into two groups, either chemorefractory or chemoresponsive, the response to chemotherapy is in fact a spectrum. Therefore the fact that increase on PC1 was associated with increase in PFS supports the idea that the degree of response to first-line chemotherapy is associated with the degree of CNA changes present in the tumours.

Due to the link revealed between the CNA changes from the PCA and the response to chemotherapy an analysis of the cellular processes that may be altered as a result of these CNAs was carried out. An analysis of the top 500 genes contributing to PC2 from the PCA, as this caused maximum separation of the two groups of CTCs, was undertaken. Copy number changes in a number of genes with functions associated with drug resistance were seen in the CTCs from chemorefractory patients. Loss of copy number in genes associated with the hydrolysis of Irinotecan was identified in the chemorefractory patients. Irinotecan is hydrolysed by carboxylesterases to its active metabolite SN-38 [299, 300]. The loss of copy number of CES 1, 2 and 3 may therefore impair the utility of Irinotecan in these patients. Irinotecan has demonstrated single agent activity in the second-line setting in SCLC, and is also used in the first-line setting in Asia [27]. The changes in the carboxylesterases are therefore potentially very relevant when considering treatment options for these patients. One of the other interesting groups of genes with altered copy number in the chemorefractory patients' CTCs was the GSTAs. GSTAs have a role in protecting cells from oxidative stress such as lipid peroxidation [296]. Cells which overexpress GSTA1 have been shown to have increased resistance to Doxorubicin with the suggestion that this is through decreased lipid peroxidation causing damage to cell membranes [296]. The ability of

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

Doxorubicin to lead to decreased glutathione levels in cells is also reduced in GSTA1 over expressing cells, which may also contribute to the resistance seen. Bcl-2 levels were not down regulated in GSTA1 over expressing cells in response to Doxorubicin unlike those cells with normal levels of GSTA1, which again may result in protection against apoptosis. When contrasting the differences between an ovarian cell line that was resistant to nitrogen mustards to one that was sensitive, it was suggested that the behavioural difference was linked to amplification of the GSTAs in the resistance cell line [295]. This data potentially identifies resistance mechanisms related to CNAs present in the chemorefractory patients CTCs. These changes accordingly warrant further assessment in a larger cohort of patients.

The interpretation of the CNA data has some caveats. Studies have assessed the correlation between alterations in DNA copy number and changes in gene expression within tumours [312-315]. Although there is a strong link between changes in copy number and gene expression within these analyses, it is not absolute. It is therefore not possible to conclude from this data that all the copy number changes noted are linked with downstream changes in gene expression. Alterations in DNA copy number may occur at the gene level, but may also affect larger regions of chromosomes altering many genes. The amplification of a large region of a chromosome, for example, may result in the amplification of a critical oncogene that promotes tumorigenesis, with the changes in the other genes being just passenger events [316]. It is equally possible, however, that changes affecting critical processes, such as the control of cell growth, are the result of widespread alterations caused by CNAs affecting the expression of multiple genes [317, 318]. However, these limitations do not negate the value of the identification of the CNAs changes linked with chemoresistance but suggest further evidence of the downstream changes in mRNA and protein levels would be beneficial.

As differences in the CNA profiles of the chemoresponsive and chemorefractory patients' CTCs were noted, an attempt to identify a CNA signature to distinguish the two groups of patients was made. An analysis of the copy number changes in over 19000 protein-coding genes revealed statistically significant differences in 760 genes between the two patient groups CTCs. Of note this included CDH1, a tumour suppressor gene, which has loss of copy number in the chemoresistant patients' CTCs, but normal copy number in the chemoresponsive patients' CTCs. This may result in a more aggressive CTC phenotype in the chemoresistant patients with the promotion of cell motility and metastasis [301]. There were 13 genes that were amplified in the CTCs from the chemoresponsive patients, but had loss of copy number in the

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

chemoresistant patients' CTCs. This included PHLPP2 which is amplified in the chemoresponsive patients' CTCs but lost in the chemorefractory patients' CTCs. PHLPP2 acts to inactivate AKT resulting in inhibition of cell-cycle progression whilst promoting apoptosis [303]. Experiments in Mantle cell lymphoma (MCL) have showed that PHLPP2 is a negative regulator of the PI3K/AKT pathway. Overexpression of miR17~92 down regulates PHLPP2, PTEN and BIM leading to activation of the PI3K/AKT pathway and inhibition of chemotherapy-induced apoptosis [304]. Within a MCL xenograft model, inhibition of this miRNA cluster leads to inhibited growth of the tumour [304]. The loss of expression of PHLPP2 and the progression of cancers has also been noted in other studies [305] [306]. These results therefore suggest that the gene signature is identifying potentially biologically relevant changes in copy number.

The 760 genes with significant alteration of copy number between the chemoresponsive and chemorefractory patients were rationalised to changes in 13 profiles based on the direction of change of copy number in each of the groups of patients' CTCs and the chromosome the genes were located on. This provides a copy number signature that could potentially be used to identify patients at risk of having chemoresistant disease. The validation of this signature in a larger cohort of patients would be required. This however represents an interesting finding with the potential to identify patients who are unlikely to respond to platinum-based chemotherapy, which is relevant as this group of patients has a far worse prognosis than the chemoresponsive SCLC patients. Some patients are uncertain if they wish to have chemotherapy or would prefer to focus on symptom control through palliative care alone. For patients, the knowledge of whether or not they would be likely to respond to chemotherapy is of significant benefit in making this decision. The potential to avoid unnecessary toxicity of chemotherapy if a patient is unlikely to benefit would also be useful. However, many patients, particularly those who are fit, may wish to receive treatment even if the chance of benefit is small, but it may be appropriate to consider clinical trials rather than standard therapy for these patients due to the exposure to novel agents to which they may respond. As by definition patients who are chemoresponsive and chemoresistant are identified after first-line therapy it is impossible to design trials for either group of patients individually. A signature to identify chemoresistance could therefore be used to direct patients into trials designed to tackle each of these groups' different needs. The potential utility of this signature if validated in another group of patients is evident.

In addition to assessing CNA differences between the two patients, a comparison of mutations identified in the CTCs from the chemoresponsive and chemoresistant

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

patients' CTCs was also carried out. WES of CTC samples and WBC samples from each of the 10 patients was performed to identify SNVs and Indels. Mutations present in the chemoresponsive patients' CTCs, but not the chemorefractory patients' CTCs were identified and vice versa. In total 5 genes were mutated in the chemorefractory patients' CTCs but not the chemoresponsive patients' CTCs. The mutated genes included TECTA, which is a component of the tectorial membrane in the ear, and PCDH10 which are both linked with cell adhesion functions and their inactivation may therefore increase metastatic potential. Down regulation of HSPG2, which encodes a proteoglycan which is a key component in basement membranes, has been linked with chemoresistance; for example experiments in a breast cancer cell line MDA-MB-231/ADM linked chemoresistance with the downregulation of HSPG2 due to miR-663 upregulation [307]. PCDH10, a member of the cadherin superfamily, has been hypothesised to be a tumour suppressor, with its inactivation linked with the development of cervical cancer [319]. PCDH10 has also been linked with chemoresistance in studies: Silencing of PCDH10 in a leukaemia cell line reduced Imatinib-induced apoptosis [308]. In a separate study, leukaemia cell lines with inactivation of PCDH10 through methylation were less sensitive to commonly used drugs in leukaemia such as Doxorubicin and Methotrexate [309].

Twenty nine genes were mutated in the chemoresponsive but not the chemorefractory patients' CTCs. One of the genes mutated in three of the chemoresponsive patients' CTCs is LRP5 which encodes a transmembrane low density lipoprotein which internalises ligands via receptor mediated endocytosis. LRP5 has been studied in prostate cancer due to its role in skeletal metastasis. Investigating an LRP5 double knockdown cell line showed decreased cell invasion, migration and colony formation in in vitro assays [310]. In a prostate cancer xenograft model LRP5 knockdown resulted in the development of tumours of smaller volume [310]. This suggests that mutation of this gene in the CTCs of some of the chemoresponsive patients may result in less aggressive tumour growth and spread. Although all the mutations identified were nonsynonymous, the lack of function of the gene is not certain which represents a potential limitation of this data. However, the mutations identified do suggest clear biological explanations for the chemoresistant or chemoresponsive behaviour identified. The differential pattern of mutations in these two groups of patients' CTCs therefore adds to the CNA data to explain the difference in response to treatment. Investigation of the mutations identified through WES in a larger cohort of patients' CTCs, would provide further support for their potential to explain mechanisms associated with intrinsic chemoresistance.

Chemorefractory patients = 1, 2, 3, 7 and 9
Chemoresponsive patients = 4, 5, 6, 8 and 10

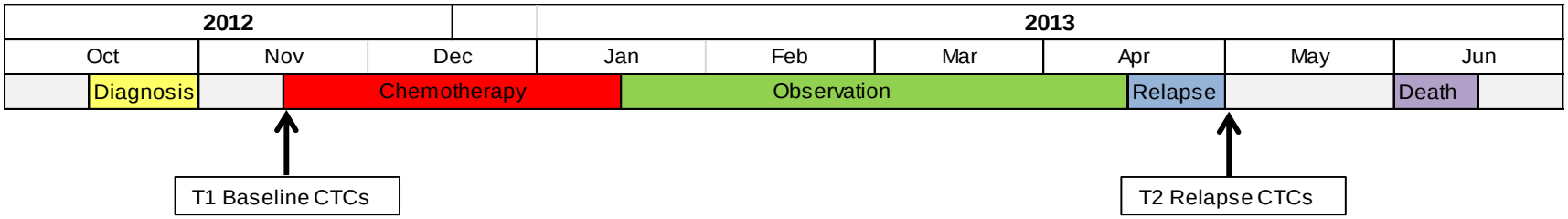
5.4C. Results and Discussion: Comparison of the Genomic Profiles of CTCs Isolated at Baseline and Relapse from Patients with Initially Chemoresponsive SCLC.

The ability of liquid biopsies such as CTCs to be used as a non-invasive method of monitoring the changes in tumours over the course of patients' therapy is postulated to be one of their strengths. Evidence has accumulated that tumours evolve over time in response to pressures such as therapy with, for example, the emergence of resistance mechanisms. The initial diagnostic biopsy will therefore fail to capture these changes in the tumour so may not be relevant in directing patients' therapy after the first-line. The analysis of CTCs isolated from longitudinal samples from patients at baseline and again at relapse was therefore undertaken to assess changes in CNAs and mutation profiles that develop in patients' CTCs. A comparison of the genomic profiles of the baseline CTC samples from the chemoresponsive patients to the CTC samples taken at relapse was undertaken, to assess changes that could be related to the development of acquired resistance. As discussed in the introduction it is acknowledged that although the majority of patients do acquire resistance to platinum chemotherapy at relapse, in some patients relapse will not equate with acquired resistance as they may respond to a rechallenge with platinum-based chemotherapy; however, those responses are generally of shorter duration than to first-line therapy [3]. The CTC and WBC samples analysed from each patient at each time point are summarised in table 5.6. The patients' treatment timelines, including when blood samples for CTC isolation were taken, are summarised in figure 5.13.

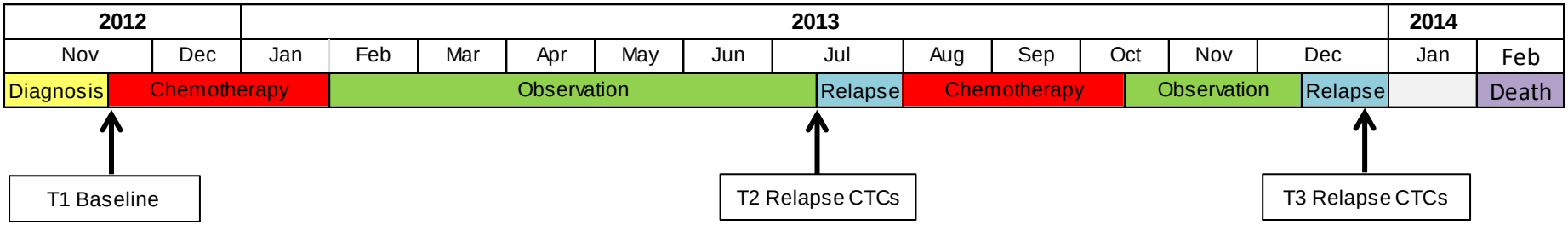
5.4C.1 Differences in CNAs between Baseline and Relapse CTC Samples from Patients with Initially Chemoresponsive SCLC.

The CNA profiles of the baseline and relapse CTC samples were assessed to see if changes between the two groups of CTCs could be identified, and what they could reveal about the mechanisms of acquired resistance. Analyses comparing the 111 CTCs isolated from the five chemoresponsive patients at baseline, relapse after first-line chemotherapy and for patients 5 and 8 relapse after second-line chemotherapy were undertaken. As an initial assessment a PCA of the CNA in 19336 protein-coding genes of all the CTCs from the 5 chemoresponsive patients' at all available time points was carried out. This was performed to see if a global difference between the baseline and relapse CTCs could be seen, irrelevant of the patient from whom they were

Patient 4



Patient 5



Patient 6

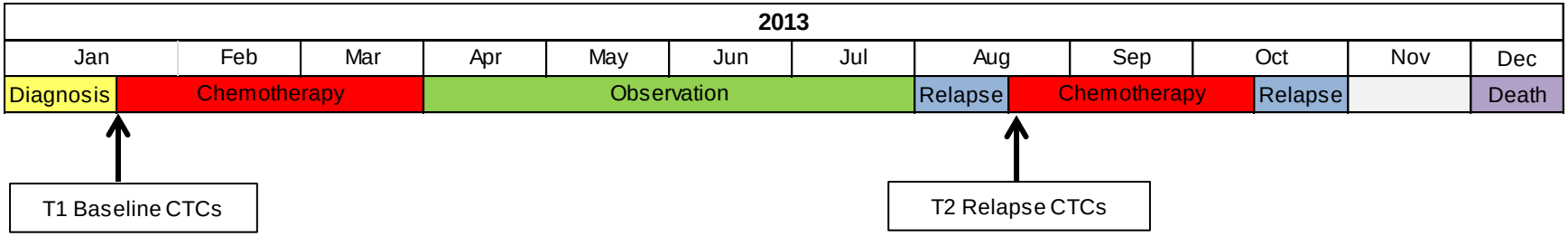
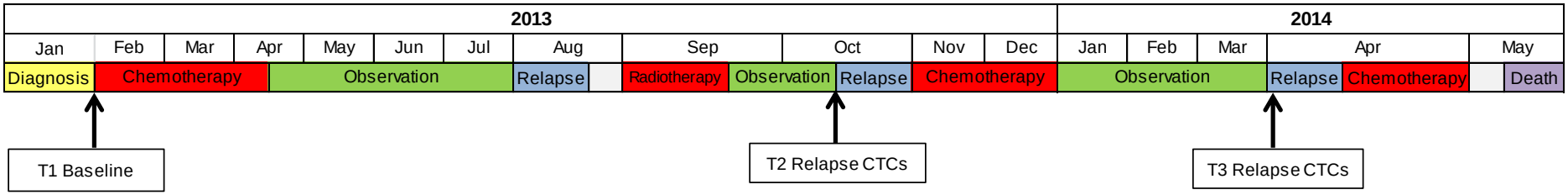


Figure 5.13 Treatment summaries of SCLC patients whose CTCs were sampled at baseline and relapse

Patient 8



Patient 10

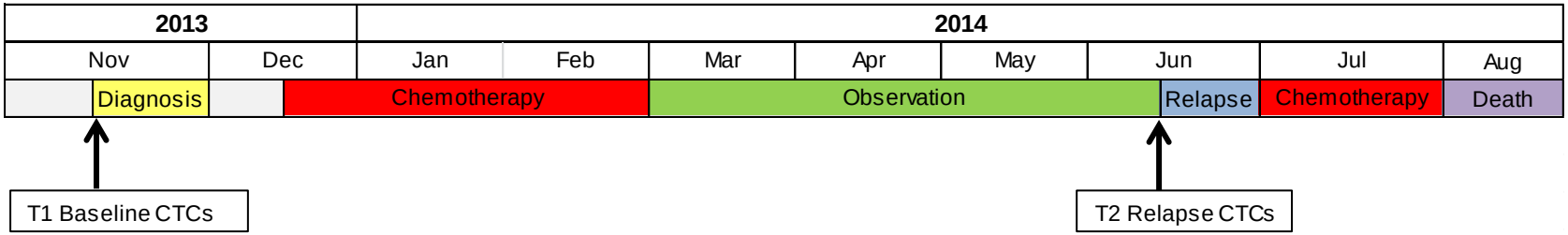


Figure 5.13 continued Treatment summaries of SCLC patients whose CTCs were sampled at baseline and relapse

isolated. On the PCA plot of PC1 and PC2 the majority of the CTCs cluster together with many of the individual CTCs from a patient clustering tightly together (figure 5.14A). There is no discernible separation of CTCs isolated at different time points evident. To investigate this further a PCA of the average CNA profile for each gene from the CTCs isolated at each time point from individual patients was performed to give a patient-specific view of each time point. The PCA plot of PC1 against PC2 for this analysis demonstrates that the CTC profile at each time point for a patient is most closely related to those at other time points from the same patient (figure 5.14B). This highlights the fact that the CTCs from an individual patient have patient-specific similarities irrelevant of the time point at which they were isolated (baseline and relapse). The patient-specific CNA changes therefore may overwhelm a pattern of subtle differences between the baseline and relapse samples across patients, preventing them from being revealed.

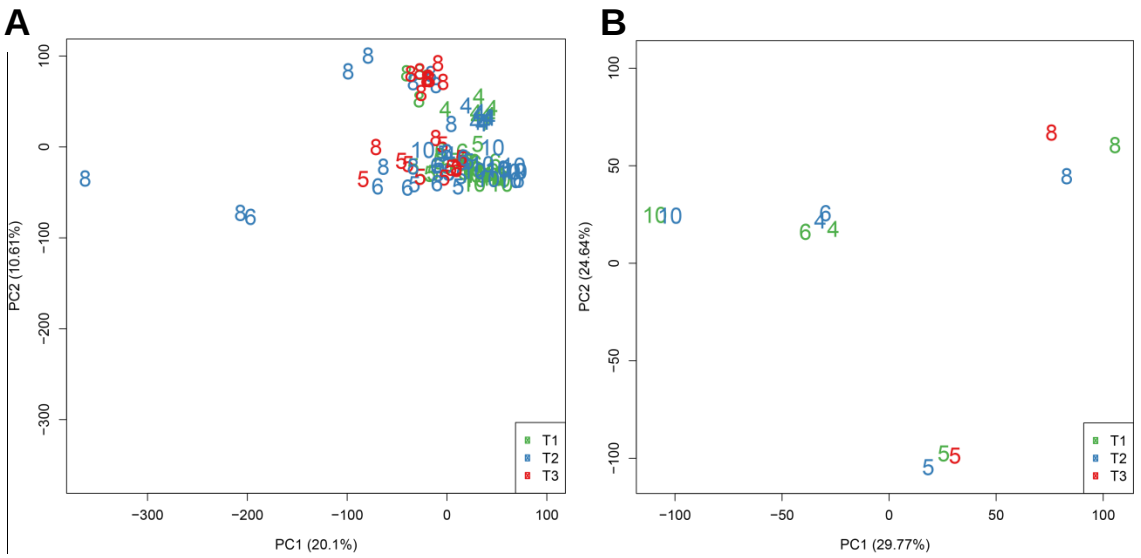


Figure 5.14 PCAs of CNA of protein-coding genes from baseline and relapse CTC samples from patients with SCLC. Single CTCs, and pools of CTCs were WGA and NGS of the DNA libraries performed. CNA analysis was carried out on the amplified CTCs, using sequencing data from 19,336 protein-coding genes. PCAs were performed either using data at an individual CTC sample level (A) or an average gene copy number across the CTCs analysed per patient at each time point (B). The percentage variance accounted for by the principal components is given in parenthesis. The samples isolated at baseline (T1) from a patient are green, those at relapse after first-line chemotherapy (T2) are blue and the samples isolated at relapse after second-line chemotherapy (T3) are red. The patient number is indicated as the data marker on the graphs.

To be able to investigate the changes that develop with relapse comparisons at an individual patient level were therefore made. LIMMA analysis of each patient's CTCs was carried out to assess significant changes between the CTCs from baseline (T1)

and relapse after first-line chemotherapy (T2). Copy number changes in protein-coding genes with a fold change between the two time points of 1.5, which were statistically significant, were identified with a FDR of 5%. The changes identified are summarised in table 5.14. In total 0 significant changes were identified in patient 4, 51 changes in patient 5, 1860 changes in patient 6, 1270 in patient 8 and 0 in patient 10. As there were no significant changes identified in patients 4 and 10 it implies copy number change across the whole population of CTCs in these two patients is not associated with the development of acquired resistance. Of note patients 6 and 8 have marked variability in CNA profiles amongst the CTCs analysed from each patient which may contribute to the higher number of significant differences seen (see appendix 3 for CTC CNA profiles). There was an overlap of just 19 changes between patient 6 and 8 and no common changes with patient 5. This implies no common changes in copy number were developing in the relapse CTCs across the five patients.

Patient	Total number of copy number changes between T1 and T2	Copy number gain in T2 compared to T1	Copy number loss in T2 compared to T1
P4	0	0	0
P5	51	0	51
P6	1860	1135	725
P8	1270	747	523
P10	0	0	0

Table 5.14 Number of significant copy number changes per patient between baseline and relapse CTCs. Copy number data was generated from amplified CTCs using NGS at baseline (T1) and when the patients relapsed (T2) from five patients with chemoresponsive disease. Comparisons in LIMMA of the copy number of 19336 genes in the baseline and relapse CTCs from the individual patients were carried out to identify significant alterations in copy number. Whether the change represents a loss or gain at T2 relative to T1 is shown in the table.

As patients 5 and 8 had responded to second-line chemotherapy, though patient 5 relapsed 2.4 months after completing second-line therapy, both patients had a third time point at which CTCs were sampled. The third time point for these two patients therefore potentially may represent the most resistant CTC profiles. Comparisons of the CTC profiles from T3 to T1 for these two patients were therefore carried out. There were 14 additional changes in patient 5 and four additional changes in patient 8 identified from these analyses. There was no commonality between the changes from patients 5 and 8 and no new changes in common between patients 6 and 8. The analyses again suggest that there are not common changes in CNA that can be

identified between the baseline and relapse samples to suggest common mechanisms for acquired resistance.

The presence of heterogeneity amongst CTCs from individual patients has already been demonstrated. Heterogeneity amongst the CTCs potentially complicates analyses as only some of the clones of CTCs may have a specific change of interest. It is possible that at baseline there are subclones of CTCs present that have a resistant genotype amongst the responsive CTCs, and at relapse there may also be heterogeneity within CTCs present, some with a resistant genotype and some with a responsive genotype. It may therefore be that the CNA profile changes associated with acquired resistance are only present in subsets of the CTCs, and that analyses of all the CTCs at baseline and relapse fail to capture these changes in subsets of CTCs amongst the whole CTC population. To investigate this hypothesis, CNA data generated from 6360 cancer-related genes from patient 4's CTCs at baseline and relapse were analysed using a PCA. A plot of PC1 versus PC3 was generated from this analysis (figure 5.15). The majority of the CTCs cluster together but three relapse CTCs cluster separately with one CTC from baseline. It is possible that this represents a resistant subclone of CTCs present at baseline and expanding at relapse. Although it is interesting to speculate that these CTCs may be a resistant subpopulation of CTCs this would be very difficult to conclude without extensive in vitro or in vivo analysis. It should also be noted that there were less obvious subpopulations of CTCs evident when PCAs of the other patients' CTCs were carried out individually (data not shown). It may therefore be very challenging to use analysis of CNAs to identify CTCs that have become resistant to chemotherapy at relapse from patients unless a predefined signature of changes associated with resistance has been developed. This could then be used to identify the proportion of CTCs present that would be resistant to chemotherapy at any time point.

5.4C.2 Comparison of CNAs in CTCs Isolated at Relapse from Initially Chemoresponsive Patients and Baseline from Chemorefractory Patients.

The results of analysing the baseline and relapse time point CTCs from the chemoresponsive patients suggest that there are no consistent changes in CNA profiles. This contrasts to the significant differences identified between the CNA profiles of the chemorefractory baseline CTC samples and the chemoresponsive baseline CTC samples (section 5.4B.1). An investigation of whether the differences in

CNA profiles identified were still present when the relapse CTC samples from the chemoresponsive patients were contrasted to the refractory patients' CTCs, in place of the baseline CTCs, was therefore made based on the hypothesis that these changes represent a chemoresistant signature. This would assess if the relapse CTCs have developed the same resistant signature as the refractory CTCs and whether the genomic changes associated with acquired resistance are the same as those associated with intrinsic resistance. Initially a PCA analysing CNA data generated from protein-coding genes in the CTCs isolated at baseline from the chemorefractory patients and the CTCs isolated at relapse from the chemoresponsive patients was carried out. A plot of PC1 and PC2 was produced from the PCA analysis (figure 5.16). There was still separation between many of the CTCs from the two groups of patients but it was not as marked as when contrasting the baseline chemorefractory and chemoresponsive patients' CTCs (figure 5.6). This suggests that although there are differences between the two groups of CTCs they are not as distinct as when the chemoresponsive versus chemorefractory patients' baseline CTCs comparison was carried out.

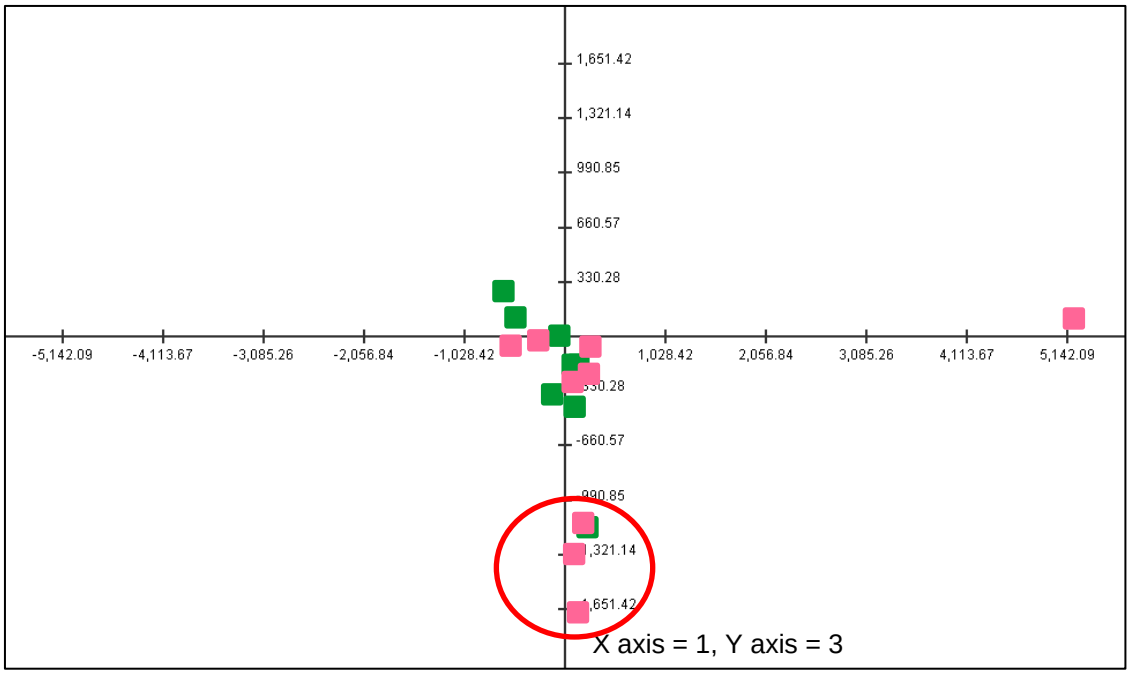


Figure 5.15 PCA of CNA data from cancer-related genes in CTCs isolated at baseline and relapse from Patient 4. CTCs isolated from patient 4's samples were amplified, DNA libraries made and WGS performed. CNA data from 6360 cancer-related genes was generated and a PCA of this data carried out. PC1 was plotted against PC3. CTCs isolated at baseline are coloured green and those isolated at relapse are coloured pink.

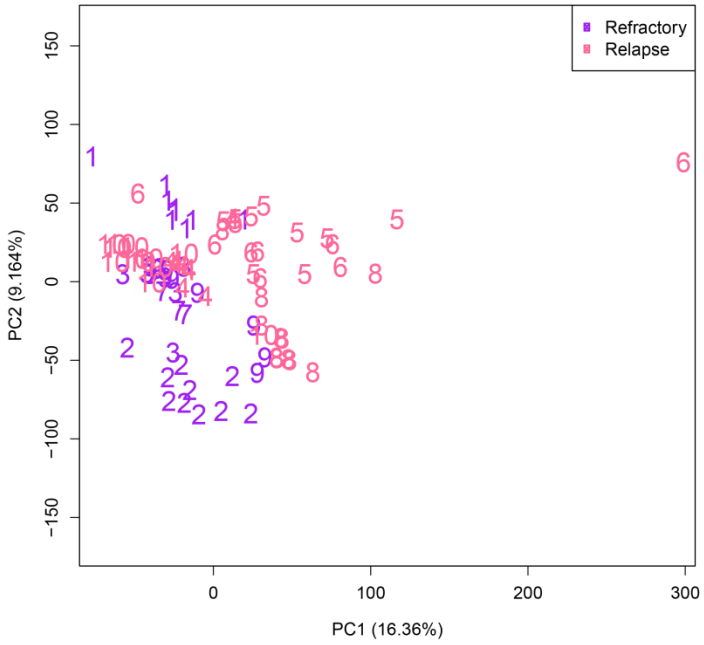


Figure 5.16 PCA of CTC CNA in protein-coding genes from baseline and relapse samples from patients with SCLC. Single CTCs, and pools of CTCs were WGA and NGS of the DNA libraries performed. CNA analysis was carried out on the amplified CTCs, using sequencing data from 19,336 protein-coding genes. A PCA analysis of the data was then performed and PC1 and 2 plotted. The percentage variance accounted for by the components is given in parenthesis. The CTCs isolated from the chemorefractory patients' baseline blood samples are in purple. The CTCs isolated from relapse blood samples from the chemoresponsive patients are in pink. The patient number is indicated as the data marker.

A signature of 760 genes copy number change differentiating the baseline chemorefractory and chemoresponsive patients' CTCs had been developed through a LIMMA analysis of CNA in protein-coding genes as previously described (section 5.4B.1). The same LIMMA analysis was performed on the baseline chemorefractory CTCs and those from the final relapse time point CTCs for the five patients with initially chemoresponsive disease (T2 for patients 4, 6 and 10 and T3 for patients 5 and 8) to see if these statistically significant differences were still present. This analysis revealed 568 genes with statistically significant difference between the two groups of CTCs (appendix 5 table A5.2). A PCA of the average copy number in the 568 genes across the CTCs from each patient from the final relapse samples for the chemoresponsive patients and the chemorefractory patients' baseline samples was carried out. PC1 was plotted against PC2 generated from this analysis (figure 5.17). The plot shows separation of the chemorefractory patients' CTCs and the chemoresponsive patients' relapse CTCs. There therefore remain significant differences between these two groups of samples suggesting there is not a single pattern of CNA associated with both intrinsic and acquired resistance.

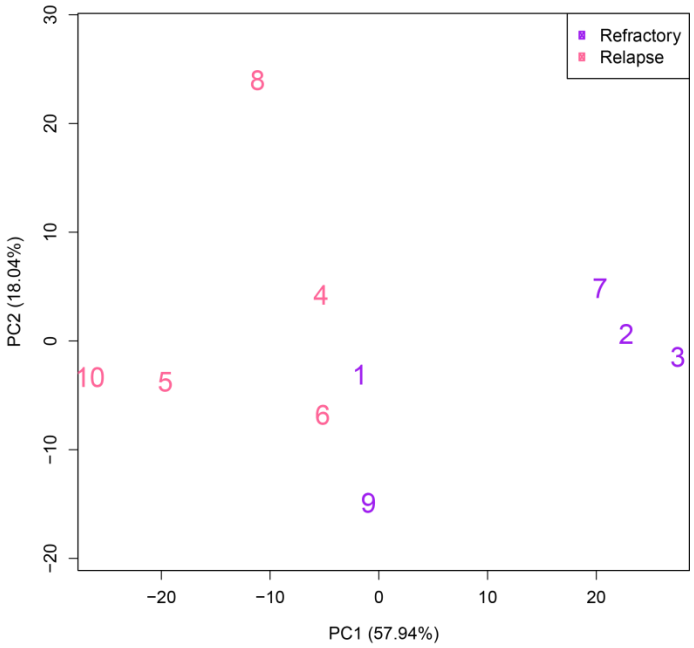


Figure 5.17 PCA of CTC CNA in protein -genes isolated from baseline and relapse blood samples from patients with SCLC. Single CTCs, and pools of CTCs were WGA and NGS of the DNA libraries performed. CNA analysis was carried out on the amplified CTCs, using sequencing data from 19,336 protein-coding genes. A subset of 568 genes had statistically significant difference in CNA between the CTCs from the chemorefractory and chemoresponsive patients when analysed in LIMMA. A PCA analysis of the data was then performed using the average copy number value across the CTCs from each patient for each of the 568 genes. PC1 and 2 generated from this analysis were plotted. The percentage variance accounted for by the components is given in parenthesis. The chemorefractory patients are in purple on the plot whilst the chemoresponsive patients (relapse samples) are in pink. The patient number is indicated as the data marker.

A comparison of the genes identified in the LIMMA analyses as being significantly altered in the baseline chemorefractory and chemoresponsive patients' CTCs analysis, and the baseline chemorefractory and chemoresponsive patients' final relapse time point CTCs analysis was performed (figure 5.18). Of the 568 genes identified when contrasting the relapse and refractory CTCs, 438 were present in the original signature of 760 genes and 130 genes were unique to this second analysis (appendix 5 table A5.2). There were now 322 copy number differences that were no longer statistically significant between the refractory and relapse time points CTCs in contrast to the chemoresponsive and chemorefractory baseline analysis. It may be possible that the changes in these genes are associated with chemoresistance. To investigate this possibility the pattern of copy number changes in these genes were assessed. The CNA patterns for each CTC from the chemorefractory and chemoresponsive patients (baseline and final relapse time point) were plotted for the 890 genes whose copy number was significantly altered in the two LIMMA analyses (figure 5.19). Of the 322 genes not identified as significant when the refractory samples were compared to the

relapse the plot highlights that the general direction of change is still similar in all the chemoresponsive patients' CTCs. However, amongst the relapse samples there is more variability in the CNA profiles than in the baseline samples evident on the plot. Analysis of the average pattern of change in copy number from the raw data from the baseline and relapse CTCs from the chemoresponsive CTCs confirms the same direction of change in all but 7 genes located on chromosome 4 cytoband q13.2. In conclusion, the direction of change in copy number for all but 7 of the 890 genes identified by the LIMMA analysis is consistent between the baseline and relapse samples from the chemoresponsive patients. The overall analyses in LIMMA would therefore imply that the copy number changes in the baseline and relapse CTCs are both markedly different to those seen in the chemorefractory CTCs and similar to each other. This would provide further evidence that unlike intrinsic resistance when there is data to support a link with CNA changes, the genomic changes associated with acquired resistance are not linked with significant CNA changes.

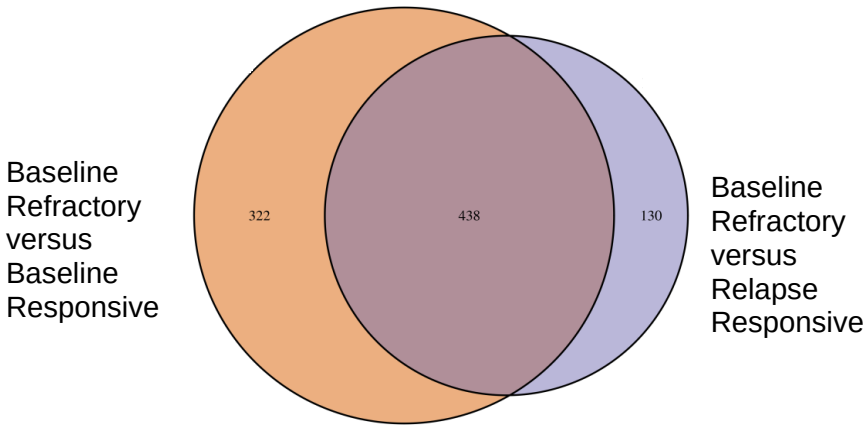


Figure 5.18 Comparison of genes with significant CNA differences identified in LIMMA analyses. CNA data was generated from WGS data from amplified CTCs isolated from patients with chemorefractory disease, and patients with chemoresponsive disease both at baseline and when they developed relapsed disease after chemotherapy. LIMMA analyses to identify genes with significant differences in copy number between the chemorefractory and chemoresponsive patients' baseline CTCs, and to identify genes with differences in copy number between the chemorefractory patients' CTCs and relapse time point CTCs from the initially chemoresponsive patients were performed. A comparison of the overlap of genes identified in the two analyses was made.

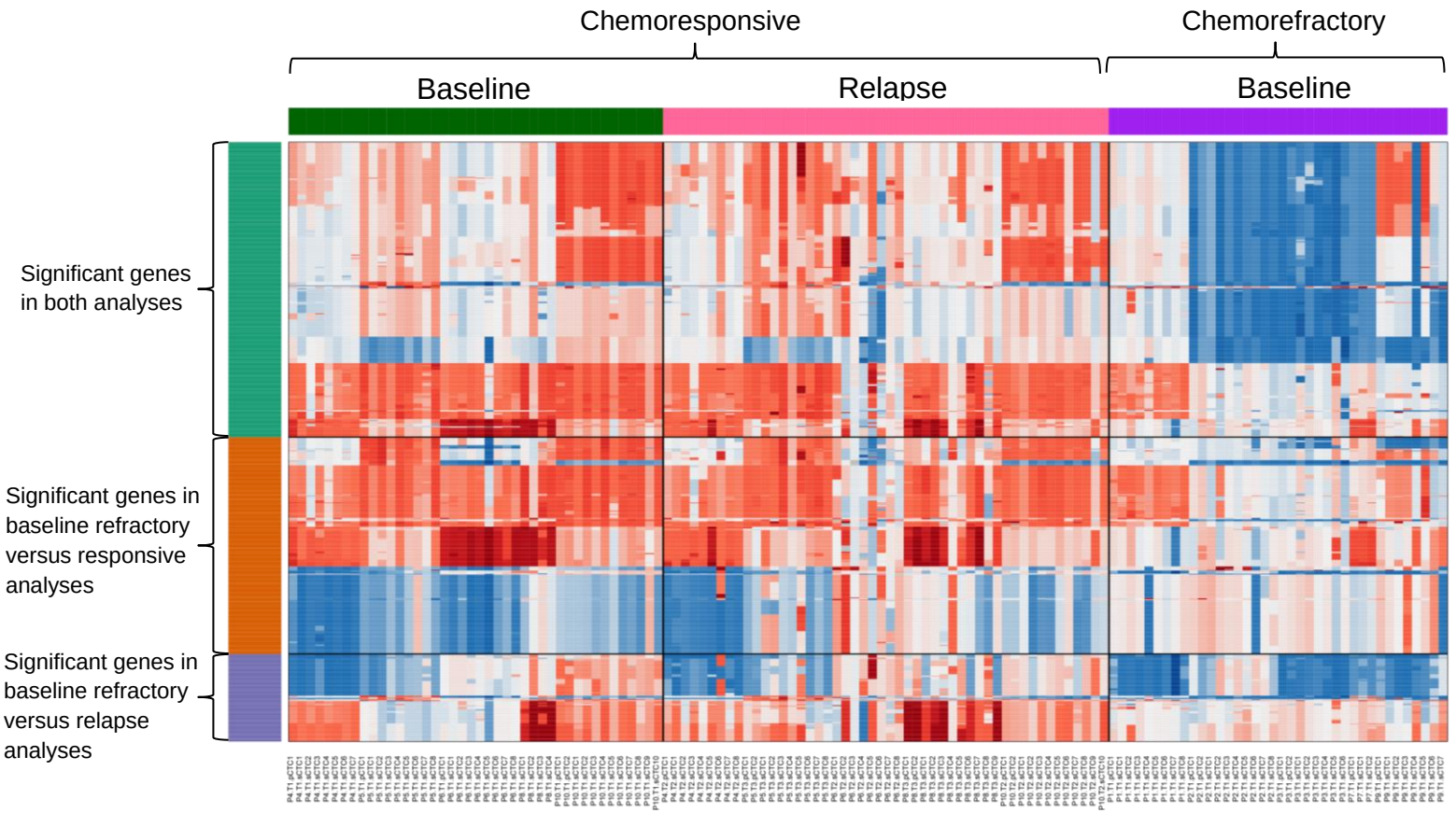


Figure 5.19 CNA profiles in genes identified as having statistically significant differences in the CTCs from chemorefractory patients and chemoresponsive patients at baseline and relapse. CNA data was generated from WGS of amplified CTCs. LIMMA analyses to identify significant differences between the chemorefractory and chemoresponsive patients' baseline CTCs, and to identify differences between the chemorefractory patients' CTCs and relapse time point CTCs isolated from initially chemoresponsive patients, were performed. A plot of the copy number of the identified genes for all CTCs analysed from the chemoresponsive (baseline and relapse) and chemorefractory patients was created. The chemoresponsive patients' baseline CTCs are green whilst their final relapse time point CTCs are pink. The chemorefractory patients' CTCs are purple. The genes whose copy number was significantly different in both analyses are green, only in the baseline chemorefractory versus chemoresponsive analysis are orange and only in the chemorefractory versus relapse analysis are purple.

5.4C.3 Differences in Mutations Identified by WES between Baseline and Relapse Samples in the CTCs of Patients with Initially Chemoresponsive SCLC

WES of CTC samples from the five initially chemoresponsive patients with both baseline and relapse samples was carried out. A search for mutated genes present in the baseline samples but not the relapse samples and vice versa was carried out. Genes with mutations, either SNVs or Indels in three or more of the baseline samples and none of the relapse samples, and four or more of the baseline samples and one of the relapse samples were identified. The same criteria were applied to identify mutations in the relapse but not the baseline samples. In total 178 genes were differentially mutated with just 7 genes mutated in the baseline samples, but not the relapse samples and 171 genes mutated in the relapse but not the baseline samples (table 5.15). The pattern of the mutated genes between the baseline and relapse samples is shown in figure 5.20.

Analyses of the mutated genes in the baseline samples, and those mutated in the relapse samples were carried out again using Ingenuity® Pathway Analysis (appendix 6 tables A6.10 and A6.11). The genes mutated in the baseline samples had all been noted to be involved in cancer with both STARD9, a kinesin-3 family member, and PTPRN2, a protein tyrosine phosphatase, having roles in the initiation and control of mitosis. Aberrant control of mitosis could be one of the methods promoting growth in the tumour.

Of the 171 genes identified as mutated in the relapse but not the baseline samples 164 had been associated with cancer. Only one gene TECTA which encodes a protein in the tectorial membrane, had also been identified in the refractory CTC samples as highly mutated in comparison to the baseline chemoresponsive samples, as well as being mutated in the relapse samples. Both KDM4E and PHF1 code proteins involved with histone modification which reflects the common occurrence of mutations in histone modifying enzymes seen in research into SCLC tumours. Of note the tumour suppressors BRCA2, KL, SASH1, SMARCA2 and TGFBR2 were found to be mutated in some of the relapse samples but not the baseline samples. Mutations in CDK4, E2F2 and PPM1J would all influence the cell cycle regulation by BTG family proteins so their mutation could promote cell growth. These mutations may result in more aggressive behaviour in the relapsed disease in contrast to that seen initially in the baseline disease. It may, therefore, be that in the CTCs from the relapse time point mutations promote increased growth as a mechanism to explain the patients'

deterioration. Mutations in genes such as TGFRβ2, a transmembrane protein in the Ser/Thr protein kinase family and ATP1A1 an ion pump, have both been linked with chemoresistance in studies and so may both contribute to a more chemoresistant phenotype at relapse. The results of the WES in the baseline and relapse CTCs therefore provide further evidence for potential mechanisms behind the development of acquired chemoresistance. Analysis of the differences in the mutations in a larger cohort of patients' baseline and relapse samples could be performed to see if the same potential mechanisms for chemoresistance are noted to develop in other patients.

Genes mutated in the baseline but not relapse CTCs	Genes mutated in the relapse but not baseline CTCs				
FAM181A	ABCA13	DMXL2	KIAA2026	PCDHGA5	TACC2
FIGN	ABCA4	DNAH10	KL	PCSK5	TCF4
LTBP2	ADAM21	DSCAML1	KRT15	PDIA6	TECTA
MYBBP1A	ADAM33	DSPP	LEPREL2	PHF1	TENM2
PTPRN2	ADAMTSL3	DSTYK	LOXHD1	PKDREJ	TGFBR2
STARD9	ADAR	E2F2	LPCAT1	PLEKHG3	TICRR
USH2A	ADCY9	EIF2B4	LPO	PNMAL1	TLR9
	AIMP2	ELF1	LRP2	PPM1J	TNRC18
	AKAP13	EPB42	LTBP1	PRUNE2	TNRC6B
	ALAS1	EPHA1	MACF1	PTPN13	TNS3
	ALS2	ERBB3	MAML2	QSOX1	TRPM1
	ANK3	ETAA1	MERTK	RBMXL1	TRPM3
	APOB	FAM205A	MGA	RC3H2	TTI1
	ARHGAP31	FILIP1	MICAL3	RNF148	TYR
	ARSG	FLRT2	MICALL1	RXRG	UHRF1BP1
	ASH1L	FOXJ2	MIPOL1	SACS	VPRBP
	ASTN1	FOXN4	MKI67	SASH1	VPS13D
	ATP1A1	FRY	MST1L	SHF	WAPAL
	BAG3	GABBR2	MYCBPAP	SIK2	XYLB
	BAZ2A	GBA2	MYH10	SIPA1L2	ZDBF2
	BAZ2B	GCN1L1	MYT1L	SLC26A11	ZFYVE26
	BRCA2	GPR158	NES	SLC6A2	ZMYND12
	C2	GRK1	NID1	SLC8A1	ZNF285
	CACNA1C	GRM5	NLRP3	SLITRK5	ZNF382
	CARD6	HECTD4	NPAT	SMARCA2	ZNF438
	CCDC129	HPX	NPC1L1	SRP68	ZNF462
	CCNI	HSPA12A	NPLOC4	STAT4	ZNF510
	CCT8L2	IGSF10	NUP205	STOX2	ZNF606
	CD22	IL16	NUP210L	STXBP5L	ZNF609
	CDK4	IQGAP1	OCA2	SULF1	ZNF687
	CEBPZ	KDM4E	OR10G2	SVEP1	ZNF780B
	CHD7	KIAA0947	OR4D5	SVIL	
	CNTRL	KIAA1199	OR5I1	SYNJ1	
	DHX9	KIAA1324L	OR6K6	SYNPO2	
	DMXL1	KIAA1683	PARP14	SZT2	

Table 5.15 Lists of genes mutated in the baseline and relapse CTCs from patients with SCLC. WES was carried out on CTCs isolated from five patients with initially chemoresponsive disease at baseline, and relapse after chemotherapy. Genes mutated (either SNVs or Indels) in the baseline but not the relapse CTCs and those mutated in the relapse but not the baseline CTCs were identified from the sequence data.

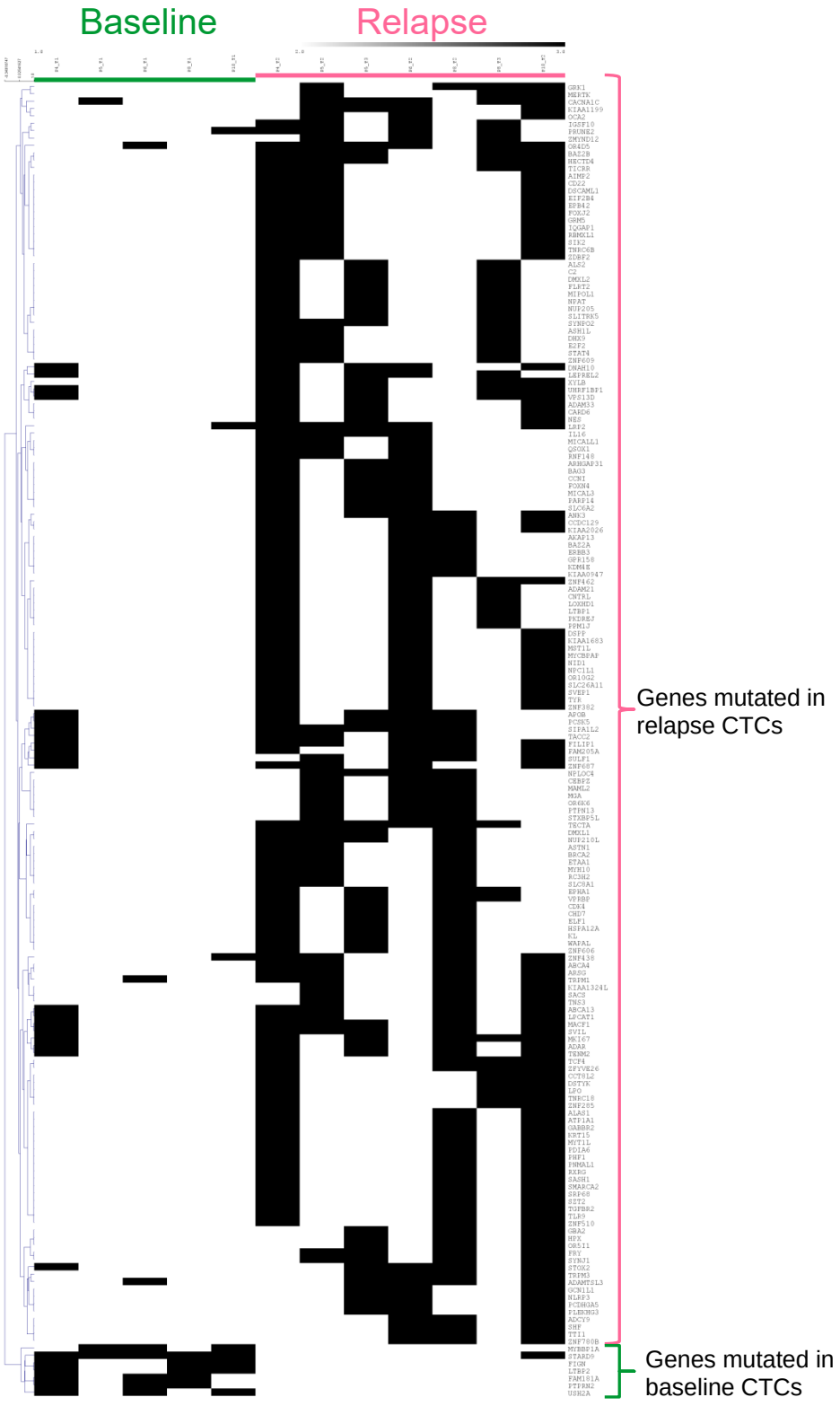


Figure 5.20 Plot of mutated genes in the chemorefractory and chemoresponsive patients' CTCs. WES was carried out on CTCs isolated from five patients with chemoresponsive disease at baseline and relapse after chemotherapy. Genes mutated (either SNVs or Indels) in the baseline but not the relapse CTCs and those mutated in the relapse but not the baseline CTCs were identified from the sequence data. A plot of the pattern of mutated genes in each patient was created with black indicating the gene is mutated and white indicating no mutation is present. The baseline samples are in green and the relapse samples in pink.

5.4C.4 Discussion

Acquired chemoresistance represents one of the greatest problems in SCLC management. The initial marked response to platinum-based chemotherapy that the majority of SCLC patients see is not sustained with virtually universal relapse noted. The short median PFS of approximately 4.5 months reflects this finding [38]. There is a marked change in behaviour from a very chemoresponsive disease to a chemoresistant one in a matter of months seen in SCLC. The genetic basis of this is therefore of significant relevance to the investigation of SCLC and potentially the development of new therapeutic options. The analysis of any changes in the genomic profiles of the CTCs isolated at baseline with those isolated once the same initially chemoresponsive patients had relapsed was made to investigate the development of acquired resistance. It is accepted that the diagnosis of relapsed disease does not always imply the development of acquired resistance as some patients respond to rechallenge with platinum-containing chemotherapy. However, from the five patients analysed three patients (6, 8 and 10) all progressed whilst receiving second- or third-line chemotherapy and patient 5 progressed at less than 3 months after receiving second-line chemotherapy. This would suggest that the clinical evidence from these four patients does support the development of acquired resistance to platinum-based chemotherapy. Patient 4 did not receive second-line chemotherapy due to deterioration in performance status so clinical confirmation of chemoresistance was not obtained in this patient. However, for the basis of these analyses as discussed in chapter 1 relapsed disease was equated with acquired resistance.

An investigation for changes in CNAs between the CTCs isolated at baseline and at the development of relapsed disease in the same initially chemoresponsive patients was carried out. PCA of the CTCs isolated from all five patients with initially chemoresponsive disease revealed no separation of CTCs according to time point, with strong commonality of the CTCs isolated from each patient, irrelevant of the time point, being revealed (figure 5.14). A comparison of changes in the CTC CNA profiles from the two time points of the five patients on an individual basis also revealed no common changes across the five patients, in fact for two of the patients there were no significant changes in copy number at all identified between baseline and relapse. These analyses suggest that there are not common changes in copy number linked with acquired resistance evident in the relapse CTCs.

An analysis of whether the CTCs at relapse had evolved to take on a signature more similar to the baseline chemorefractory patients' CTCs was therefore undertaken as an

alternate strategy for their analysis. The same LIMMA analysis was undertaken as with the baseline chemoresponsive and chemorefractory patients CTCs, but this time contrasting the CTCs isolated at relapse from the initially chemoresponsive patients with the chemorefractory patients' CTCs. Of the initially identified 760 genes, 438 still had a statistically different copy number between the two groups of patients' CTCs. The direction of change in all the other genes was the same in the chemoresponsive baseline and relapse CTCs, but the change was no longer significant in the analysis (figure 5.19). This confirms that the CTCs isolated from patients at relapse still have a very similar CNA profile to those isolated at baseline from the chemoresponsive patients, and a different CNA profile to the refractory patients' CTCs. This again suggests that there is not a common pattern of changes in copy number associated with the development of acquired resistance in the CTCs.

One hypothesis to explain these results is that there may be changes in the CNA profile of patients' CTCs in just subpopulations of cells. Resistance mechanisms to therapies may be present at a subclonal level in the initial tumour and treatments then provide a selective pressure leading to the expansion of this clone. It may follow therefore that in the baseline samples there are a small number of CTCs with a resistant genotype that become more dominant with treatment. Although a potential subpopulation was present in the PCA of the CTCs of patient 4 (figure 5.15), clear subpopulations were not evident in the analyses of the other four patients CTCs' to support this hypothesis. The results of the different analysis therefore do not support a correlation between changes in patterns of CNAs and acquired resistance.

Although the results in this chapter highlight the links between CNA profiles and intrinsic resistance when contrasting the baseline chemoresponsive and chemorefractory patients, it does not follow that the same processes account for acquired resistance. The genomic profile that leads to chemorefractory as opposed to chemoresponsive behaviour is produced over the entire time course that the tumour develops. Changes may evolve that provide growth advantages to the tumour but would also potentially result in chemoresistance. In acquired resistance the changes in the tumour such that it is no longer chemoresponsive occur in a potentially shorter window, under the direct evolutionary pressure of treatment. It therefore follows that there are different pressures potentially driving the development of the two forms of resistance. Although the data suggest strong links between CNA patterns and intrinsic resistance, other mechanisms such as mutations or epigenetic changes may be responsible for acquired resistance.

To investigate if specific mutations are associated with the development of acquired resistance, a comparison of the mutations in the baseline and relapse CTCs isolated from the five patients with initially chemoresponsive disease was undertaken. There were 178 genes identified of which just 7 genes were mutated in the baseline but not relapse samples and 171 genes developed mutations in the relapse samples. In the relapse samples there were mutations in 5 tumour suppressor genes, BRCA2, KL, SASH1, SMARCA2 and TGFBR2, identified. Mutations in CDK4, E2F2 and PPM1J which all influence the control of the cell cycle were also noted. These mutations may contribute to a more aggressive pattern of growth at the relapse time point. If tumours are very rapidly increasing patients may deteriorate too quickly to achieve benefit from a course of chemotherapy.

Possible mechanisms for the chemoresistant behaviour at relapse were also identified including mutations in TGFRβ2, a transmembrane protein in the Ser/Thr protein kinase family. A study analysed the gene expression in two FOLFOX-resistant colon cancer cell lines which were enriched for cancer stem cells [320]. The authors noted up regulation of miR-21within these cell lines. They demonstrated that the overexpression of miR-21 leads to down regulation of TGFRβ2 and PDCD4 and hypothesised that TGFRβ2 targeting by miR-21 alters the stemness of the cell through effects on the Wnt/β-catenin signalling. Stem cells have been proposed to have properties which may increase their resistance to chemotherapy such as increased drug efflux and DNA damage repair [138]. If the relapse CTC population were enriched for cells with stem cell like properties due, for example, to mutations in TGFRβ2 this could lead to the chemoresistance seen in the patients' tumours at relapse.

Mutations in ATP1A1, an ion pump, could also potentially promote the chemoresistant phenotype of the patients' tumours at relapse. Analysis of a Cisplatin resistant ovarian cell line revealed one of the genes with most significantly altered expression was ATP1A1 with decreased expression seen in contrast to a Cisplatin sensitive line [321]. Platinum accumulation defects have also been demonstrated through decreased expression of ATP1A1 in rat hepatoma [322]. One could therefore speculate that the mutations in ATP1A1 may contribute to decreased platinum sensitivity in the mutated SCLC tumours through decreased drug accumulation. As platinum-based chemotherapy is the backbone of treatment in SCLC this may represent a very significant change.

TECTA, a gene which encodes a protein in the tectorial membrane of the ear, was noted to be frequently mutated in both the CTCs isolated at baseline from the

chemorefractory patients and also the CTCs isolated at relapse from the chemoresponsive patients. TECTA has been previously noted to be mutated in WES of breast cancer tumours [323]. There is, however, no evidence for a clear role in the development of chemoresistance in the literature. The findings in this research suggest it may warrant further investigation in SCLC to elucidate a possible mechanism by which it could promote chemoresistance. The analysis of the mutations in the relapse samples in contrast to those in the baseline samples has provided evidence for potential mechanisms for the development of a more resistant phenotype. It suggests that acquired resistance may be driven by the accumulation of mutations in critical genes. This again warrants investigation in a larger population of patients to see if these results can be replicated.

5.5 Conclusions

The results of the molecular analysis of CTCs isolated from 10 patients with SCLC, five with chemoresponsive disease and five with chemorefractory disease were reported in this chapter. Initial comparisons of the data generated from CNA analysis and WES of the CTCs to published data from SCLC biopsies were performed. The CNA data generated from the CTCs captured the commonly noted abnormalities in copy number, both at the chromosome and gene level, that have been previously reported. The WES data also demonstrated mutations in frequently altered genes with the exception of RB1. Evidence of frequent mutations in LRP1B in CTCs addition to those reported in tumour biopsies suggests a potential role as a tumour suppressor in SCLC, which therefore represents an area for further research. The techniques used for CNA analysis and WES in this chapter were able to demonstrate genomic aberrations in potentially clinically actionable genes. This provides the initial evidence of a potential role of CTCs in identifying biomarkers for targeted therapies in SCLC.

In a cohort of 35 patients no statistical difference in CTC number was noted between the chemorefractory or chemoresponsive patients, or at baseline and relapse after chemotherapy. This suggests the differences in response to chemotherapy noted in the chemorefractory patients and when the initially chemoresponsive patients relapse cannot be explained by the burden of disease alone. Molecular analysis of the CTCs isolated from the chemorefractory and chemoresponsive patients confirmed different CNA patterns, with a more loss predominant pattern in the chemorefractory patients' CTCs and more amplifications in the chemoresponsive patients' CTCs. PCA of the

CNA data generated from the CTCs from the chemoresponsive and chemorefractory patients confirmed the marked differences in the CNA profiles, and demonstrated a correlation between the principal components and OS and PFS. The CNA data from the over 19000 protein-coding genes was distilled into a 760 gene signature to distinguish chemorefractory and chemoresponsive patients' CTC's. If validated in a larger cohort of patients the clinical potential of this signature would be of interest. In contrast there was no difference in the CNA patterns in the CTCs isolated from initially chemoresponsive patients at baseline and when they developed relapsed disease. This suggests that although a link was demonstrated between CNA patterns and intrinsic resistance, the same mechanisms do not account for the development of acquired resistance. When analysing the WES data 171 genes had acquired new mutations in the relapse time points CTCs in contrast to the baseline CTCs. Analysis of the genes involved revealed potential mechanisms for the development of acquired resistance within these patients' tumours. This suggests that acquired resistance may result from the accumulation of mutations in key genes.

The results presented in this chapter have demonstrated a number of ways in which CTCs could be used in research, and even as clinical tools despite the challenges associated with their processing. Biopsies are potentially going to become limiting features in recruitment to lung cancer trials due to patients' reluctance and the logistical issues in their arrangement, making liquid biopsies such as CTCs appealing. Beyond using CTCs for the analysis of predictive biomarkers it is the possibility of using CTCs to explore the biology of cancers that is a most striking prospect. In SCLC chemoresistance represents perhaps the single greatest challenge to overcome to improve patients' outcomes, excluding earlier diagnosis. The data presented in this chapter suggest that larger scale studies investigating the links between changes in copy number and mutations with intrinsic and acquired resistance in SCLC using CTCs are warranted.

Chapter 6: General Discussion

6.1 The Increasing Profile of SCLC Research

There remains a marked need to improve the management of SCLC given the very poor prognosis of patients with this disease, with a 5 year survival rate of just 5% [4, 5]. Over the past 20 years clinical trials have failed to deliver the improvements in SCLC that have been seen in other cancers. With this in mind, the National Cancer Institute (NCI) identified SCLC as one of its research priorities, following the Recalcitrant Cancer Research Act being passed in 2013 in the USA [324]. There were five research strategies developed by the NCI which were felt to have the potential to improve SCLC patients' outcomes (figure 6.1). Amongst the priorities identified by the NCI was the need to develop better research tools, including optimising the collection of tumour specimens and creating new tumour models. They also highlighted the need for comprehensive genomic profiling of SCLC, the development of new therapeutic strategies based on the specific molecular vulnerabilities of SCLC and research on the mechanisms responsible for chemoresistance. Cancer Research UK (CRUK) identified lung cancer as one of the cancers with the most unmet need in its research strategy for 2009 to 2014, along with oesophageal and pancreatic cancer, noting the poor survival in these common cancers [325]. The importance of further research in lung cancer was again highlighted in CRUKs 2014 strategy with the aim to have a 2- to 3-fold increase in lung cancer research, as it remains an area of significant unmet need [326]. During the course of this thesis the Manchester Cancer Research Centre was accredited as a CRUK Lung Cancer Centre of Excellence in partnership with University College London, reflecting the quality of the research into lung cancer being carried out at the centre. As noted in the introduction, although the rates of smoking and therefore SCLC have declined in Western Europe and the USA, this pattern is not uniform [1, 7]. For example, smoking rates remain high in the North West of England where this research was carried out and so research in to SCLC is particularly relevant to this population [10]. Rates of smoking are also increasing in regions such as South East Asia and so in addition to public health measures to encourage smoking cessation, research in to SCLC diagnosis and management remains critical [1]. This thesis aimed to explore the use of CTCs as research tools and to consider the potential clinical roles that they could fulfil in SCLC, particularly in relation to chemoresistance thus addressing key issues in SCLC. The overall findings of the thesis will be summarised in this general discussion with a focus on potential applications of the approaches developed to meet areas of unmet need in SCLC research and treatment.

NCI Priorities for SCLC Research

1) Develop better research tools for the study of SCLC by
    a) Optimising tumour tissue collection representing the distinct phases of SCLC e.g. at diagnosis and at recurrence post treatment.
    b) Developing new tumour models that reflect the phases of SCLC found in the clinic such as PDX and GEMMs.
2) Expand comprehensive genomic profiling studies of SCLC specimens, with associated clinical information, to improve knowledge of the molecular abnormalities that exist both at diagnosis and following therapeutic relapse.
3) Investigate new diagnostic approaches for populations at high risk of developing SCLC.
4) Focus therapeutic development efforts on specific molecular vulnerabilities identified in SCLC.
5) Investigate the mechanisms underlying both the initial high response rate of SCLC to first-line therapy and the rapid emergence of resistance following treatment.

Figure 6.1 Key areas for SCLC research identified by the NCI [324].

6.2 Summary of Findings from the Thesis

This thesis set out to develop a workflow for the molecular characterisation of CTCs, and to apply these methods to clinical samples from SCLC patients. The aim in doing so was to explore common molecular aberrations present in SCLC CTCs and to investigate whether CTCs could be used to investigate chemoresistance in SCLC. The key findings from this thesis are summarised below.

1) A workflow for the isolation and genomic profiling of CTCs from SCLC patients was optimised (chapters 3, 4 and 5). Different CNA profiles were identified from CTCs and WBCs that were isolated and then amplified from the blood samples of patients with SCLC using these methods, providing validation for the workflow developed (chapter 4). Methodology for a number of downstream molecular processes was also developed, including WES of CTCs with methods for the reamplification of DNA libraries and multiplexing of samples for the in-solution target enrichment established (chapter 3).

2) The genomic profiling of CDX tumours confirmed they had the same common genetic aberrations as SCLC tumours, supporting the hypothesis they have been generated from SCLC CTCs (chapter 4). CTCs isolated from blood samples and CDX tumours created using a paired blood sample from the same patient had the same CNA profiles and TP53 and RB1 mutations. This helps substantiate the hypothesis that the CTCs used to create the CDX models originate from the same initial pool of CTCs as those isolated for genomic profiling using CellSearch as the initial enrichment step. It also adds validity to the methodology used for the single-cell genomic analysis, as the results from single cell profiling reflected those from bulk tumour profiling (chapter 4).
3) It was demonstrated that a method for focussed NGS, TAm-Seq, could be used to sequence and identify mutations in a panel of genes in SCLC CTCs. TP53 mutations were identified in each of the six patients' CTCs that were analysed in this manner (chapter 4).
4) Baseline CTC numbers, as enumerated by CellSearch in 35 patients, were not statistically different between patients with chemoresponsive or chemorefractory disease. The number of CTCs isolated at baseline and when the same patient progressed were also not significantly different (chapter 5).
5) Genomic profiling of SCLC CTCs confirms they have the same common genomic aberrations that have previously been identified in SCLC tumour biopsies. Common CNA changes at both the chromosome and gene level noted in tumour biopsies were identified in the SCLC CTCs analysed. Mutations in commonly altered genes and pathways were identified in the CTCs including TP53 mutations. RB1 mutations were not identified in the CTCs but this was felt to be due to the technical challenges with sequencing single cells rather than representing a biological difference (chapter 5).
6) WGS to identify potentially druggable genetic aberrations, including changes in copy number and mutations, was carried out on CTCs isolated from patients with SCLC. Longitudinal sampling of CTCs to monitor potentially clinically actionable genes and to identify new mutations developing was performed (chapter 5).
7) Differences in copy number profiles between the CTCs isolated from patients with chemorefractory and chemoresponsive patients were demonstrated in an unsupervised analysis. Correlation between changes in the CTC CNA profile and both the PFS and OS of 10 patients with SCLC were highlighted, generating hypotheses that the pattern of changes in CNAs is directly linked with chemoresponsiveness and therefore survival (chapter 5). 760 genes had

significant alteration in copy number between the CTCs isolated from patients with chemoresponsive and chemorefractory disease in an analysis. This was refined to changes in 13 profiles based on the direction of change in copy number in the CTCs from each group of patients and the chromosome the genes were located on. If validated this potential signature could be used to distinguish patients likely to have intrinsic resistance to platinum-based chemotherapy (chapter 5).

8) No common changes in copy number were identified when contrasting CTCs isolated from chemoresponsive patients at baseline and when they were diagnosed with relapsed disease after receiving platinum-based chemotherapy. Genes that were frequently mutated in CTC samples isolated at relapse but not baseline were identified from WES data. The results suggested that CNA changes do not form the basis of acquired resistance unlike intrinsic resistance to platinum-based chemotherapy in SCLC (as suggested by the baseline chemoresponsive versus chemorefractory patients' CTC analysis). However, the results of the WES suggest that there may be commonly mutated genes associated with the development of progressive disease and these should be investigated in a larger cohort of patients (chapter 5).

6.3 Challenges to CTC Research – Successes and Future Hurdles to Overcome

To enable the utility of CTCs in SCLC research to be developed and potentially to enable improved outcomes for patients, reliable methods for their identification, isolation, amplification and molecular analysis are required. The results presented in this thesis highlight the fact that the workflow that I have developed during my PhD project is robust, consistent and can be used to generate biologically meaningful results (figure 2.1). The approach developed by our team is also flexible and a wide range of downstream analyses can, and have, been performed in this thesis. Sanger sequencing, focussed NGS of a panel of genes, WGS to analyse copy number and WES were all used to generate CTC data. When my research was initiated the use of many of these techniques to analyse CTCs had not been published, particularly with reference to SCLC and so the results presented represent the emergence of a new area of research.

One of the challenges of the analysis of single or small numbers of cells is the very limited amount of nucleic acids that are present [211]. WGA methods are therefore

employed but the amount of DNA created is still limited. This is particularly relevant when considering WES as large quantities of DNA are recommended as an input for the target enrichment. It is also important from a method validation standpoint to be able to perform multiple downstream analyses from a single cell to compare the results generated. Consequently comparisons of reamplifying either the WGA DNA or DNA libraries prepared from amplified cells was carried out. The results of the reamplification of the DNA library in contrast to the reamplification of the WGA DNA most closely reflected the original sample CNA output. This method was therefore adopted and optimised to maximise the efficiency and minimise the number of further cycles of PCR required in an attempt to reduce the incorporation of errors. An assessment of the use of WGA DNA for the in-solution target enrichment and subsequent WES, in place of unamplified DNA, was also made to assess its impact. The results confirmed the success of the exonic enrichment and the sequencing. However, areas in the target exonic region were not sequenced when using amplified DNA in contrast to unamplified DNA. This was due most likely to problems resulting from the digestion of the DNA with the Mse1 restriction endonuclease at TTAA sites during the WGA process. The amplicons resulting from this may not have been compatible with the baits used in the target enrichment process. The areas of the genome that were sequenced were consistent between the amplified samples analysed, allowing conclusions relating to the mutations present in these regions to be drawn when comparing, for example, the CTCs from chemoresponsive and chemorefractory patients. Multiplexing samples prior to the in-solution capture was also investigated in contrast to enriching the samples singly, to reduce the costs associated with this process. It was shown not to affect the success of the target enrichment or sequencing, and there was also no evidence of contamination across the samples processed in this manner. It was therefore adopted for the analysis of clinical samples.

One of the challenges of creating a larger role for CTC analysis in research and the clinic is the associated costs compared to bulk tumour analysis. When analysing tumour samples, in general a single tumour biopsy is analysed per patient with a paired normal tissue sample. However, when analysing CTCs it would be normal, given the risks of failure of isolation and amplification, and the heterogeneity noted, to analyse more than one CTC increasing the associated costs dramatically. There are also costs associated with the enrichment and isolation of CTCs that are not present when analysing bulk tumour biopsies. When developing the methods in this thesis, attempts to reduce the costs of processing the samples were considered. The development of a

reliable method for reamplification of the DNA libraries allows multiple analyses to be performed on individual CTCs, maximising the data that can be generated from this finite source. It also means that the costs of enrichment, isolation and amplification are then spread over several downstream analyses. Multiplexing of samples for in-solution capture was also optimised, reducing the costs associated with this step of the processing. The overall costs of single CTCs analyses, however, still remain considerable and so clear advantages of CTC analysis will need to be demonstrated for it to be adopted in the clinic. To be able to expand the research role or clinical role for the analysis of CTCs further methods of reducing costs will need to be considered. Over the past 10 years there has been a marked reduction in the costs of NGS allowing it to be used more widely [253]. It is therefore hoped that in the same manner there will be reductions in the costs of single-cell analysis in the future, which will potentially enable more widespread adoption of these methods.

Another area of concern that was addressed when processing and analysing the samples was the risk of attributing biological significance to changes resulting from technical errors. WBCs were isolated from patients' blood samples alongside CTCs and processed in the same manner. The confirmation that CTCs and WBCs have different CNA profiles and group separately on PCAs reinforces the conclusion that the two types of cells are being reliably identified. It also suggests that the results generated from the molecular analyses can be used to distinguish biologically distinct cell types. The demonstration that CTCs have the same genomic aberrations as both matched CDX tumours and published results from SCLC tumour biopsies also supports the validity of the processes used for the molecular analyses, as they are able to identify expected alterations. The WBCs were used as controls to the CTCs and their results were used to identify regions of the genome with greater risk of technical error. The results generated from these regions, such as the cytobands and genes excluded from CNA analysis detailed in tables 2.3 and 2.4, were then excluded from the analyses performed on the data produced from the CTCs. This approach was adopted to reduce conclusions being drawn from data potentially resulting from recurrent technical errors. The majority of the analyses were also made by contrasting multiple CTCs rather than single CTCs which reduced the possibility of drawing conclusions from stochastic errors (with PCR error rates estimated at 8×10^{6} per base per replication [264]). Although these approaches may have resulted in the loss of some genuine data, the overall aim was to increase confidence in the remaining data analysed and the conclusions drawn from it.

CellSearch was used as the initial method to enrich CTCs in this project. It enriches cells based on EpCAM expression, as described in chapter 1, and therefore CTCs which are EpCAM negative, such as those which have undergone EMT, are missed [153]. The genomic profiles of this subset of CTCs have not been explored in this research. Although, as demonstrated in this thesis, interesting and potentially clinically relevant results have been obtained from the EpCAM positive CTCs in SCLC, the additional significance of the EpCAM negative CTC population is as yet uncertain. Consideration of whether marker independent methods of CTC enrichment are required to ensure representation of all the CTCs from patients for future clinical analysis of CTCs are required.

The feasibility of analysis of CTCs isolated from serial blood sampling from individual patients was demonstrated in this thesis. CTCs isolated at baseline and then at relapse with progressive disease from individual patients were processed together in a batch allowing a comparison of the results generated. This confirms the practicability of longitudinal sampling and analysis of CTCs from patients which is proposed to be one of the key benefits of liquid biopsies. In practice when trying to obtain relapse samples from SCLC patients there are significant challenges which led to relapse samples being obtained from only 24 of the 35 patients in this study (69%). Patients with SCLC often deteriorate very quickly when they progress after chemotherapy, so there may be a limited timescale over which relapse samples can be obtained. This problem was particularly acute as the study was run through a tertiary cancer centre and when many patients deteriorate they may be managed through their local hospital or community palliative care services, rather than returning to the tertiary centre, if they are not going to receive further treatment. These problems, however, are not unique to obtaining CTC samples from SCLC patients but would apply equally, if not more so, to obtaining repeat biopsy specimens. It just reinforces the challenges of longitudinally sampling in SCLC which is essential if the evolution of this disease over the course of patients' treatment is to be studied. Programmes of analysing autopsy samples for patients who have died of SCLC would also have the potential to provide the much needed samples after patients have developed progressive disease. However, this would also not be without its difficulties, as it can be challenging to raise the possibility of involvement in such studies with some patients and their families due to their reluctance to discuss their limited prognosis. As this study was run through a tertiary cancer centre, as with blood tests, if patients die in the community or their local hospital it may be challenging to coordinate their transfer to the tertiary centre for autopsy.

The methods developed and applied in this thesis confirm that CTCs represent a new tool for the investigation of SCLC biology and treatment response. The expansion of their use in the future would require a number of further developments. The potential limitations associated with the costs of CTCs have been discussed above. Currently there are numerous approaches for the enrichment and isolation of CTCs used by the research community. CellSearch was selected as the initial enrichment technique in this study as it is semi-automated and robust, and sample processing can be centralised given the 96 hour window for processing after blood sampling which would all facilitate its adoption in clinical practice. For CTCs to gain a wider clinical role all the techniques for enriching, isolating and analysing CTCs would need to go through the stringent process of clinical biomarker evaluation [176]. Beyond this, although the techniques used in this thesis have been applied to the research of SCLC, further validation would be required before this could be routinely applied to any clinical role.

6.4 The Molecular Profiling of SCLC CTCs

Over the past 3 years there has been increased interest in mapping the genomic landscape of SCLC. In 2012 two large scale studies, performing NGS of SCLC tumour biopsies, were published [70, 91]. They provided new insights in to the common genomic aberrations found in SCLC. The data presented in this thesis is the first evaluation of whether the commonly identified changes in copy number and mutations noted in analyses of SCLC tumours are also found in CTCs. The analysis of CNA changes in 149 CTC samples isolated from 10 patients with SCLC was carried out. CTCs reflected the changes noted in copy number in tumours at both the chromosome and gene level (figure 5.3). For example in SCLC CTCs, in addition to SCLC tumour biopsies, common aberrations such as the loss of 3p and 17p and the amplification of 3q are identified in the majority of cases. The loss of critical tumour suppressors such as TP53, RB1 and FHIT are also frequently noted. This again provides further evidence of the ubiquity of these genomic alterations in SCLC and in particular in the invasive subset of SCLC cells. The amplification of specific genes is not as universally noted in SCLC as the loss of tumour suppressor genes. In both tumour biopsies reported in the literature and the CTCs analysed in this thesis, amplifications of oncogenes such as members of the MYC family are noted in some samples. However, these changes are not as common as the loss of specific tumour suppressor genes. This data suggests that the loss of certain tumour suppressor genes are critical for the development of SCLC but that there is not a specific oncogene that needs to be

amplified to promote its pathogenesis but a range of amplifications of oncogenes are seen.

WES of CTCs isolated from the same 10 SCLC patients was also performed to assess the pattern of mutations seen. Mutations identified in critical pathways and affecting specific cellular functions such as the PI3K pathway and histone modification again reflected the changes previously reported in tumour biopsies (table 5.8) [70, 91]. As in the tumours analysed these mutations only occurred in small numbers of patients, as in SCLC only TP53 and RB1 have very high rates of mutations [70, 91]. Although TP53 mutations were noted frequently no RB1 mutations were identified in the CTCs sequenced. An RB1 mutation was identified by Sanger sequencing in the CTCs isolated from the same patients whose samples generated CDX 2 confirming that RB1 mutations do occur in SCLC CTCs. Given the ubiquity of RB1 mutations in SCLC the lack of RB1 mutations identified from the WES of CTCs is likely to represent a technical limitation of NGS when analysing CTCs. It is not felt to be due to a biological difference considering how closely CTCs represent tumours in other genetic alterations. Overall, both the CNA changes and mutations noted in the SCLC CTCs closely reflected the changes that had been noted in tumours.

The common pattern of CNAs and mutations noted in SCLC CTCs and tumour biopsies provides evidence for the idea that CTCs could potentially be used in place of tumour biopsies for the research of SCLC. One of the key areas identified by the NCI which could make an impact in SCLC was to have better research tools for the study of SCLC through the optimisation of tumour specimen collection. However, rather than focussing on increasing patient biopsies, the results presented in this thesis suggest that CTCs potentially could be used in place of tumour biopsies to study SCLC, although a direct comparison of CTCs and tumour biopsies was not performed in this research. This would overcome a potential limiting issue in research without exposing patients to additional biopsies. As CTCs also have the common genomic alterations present in tumour biopsies, the investigation of the biological significance of these aberrations and their evolution in the face of selective pressures such as chemotherapies can be studied. CTCs could be used to investigate the frequency, heterogeneity and range of genetic aberrations that occur in SCLC. The expansion of comprehensive genomic profiling is an additional area of research required to improve outcomes in SCLC through increasing knowledge about this disease. The results of this thesis suggest that CTCs, in addition to tumour biopsies, could be used to meet this aim.

6.5 Utilising CTCs to Investigate Chemoresistance Mechanisms in SCLC

Chemoresistance contributes significantly to the poor outcomes noted in SCLC and has been highlighted by the NCI as a key area for future research in SCLC. The majority of cancer patients' deaths can be linked to the failure of treatment due to chemoresistance in both extensive and early stage disease [112]. This is particularly relevant in SCLC as up to 80% of patients will respond to first-line chemotherapy, demonstrating a higher response rate than many other solid tumours [31, 33-35]. Twenty per cent of patients have intrinsic resistance to chemotherapy, manifested as chemorefractory disease progressing within 3 months of completion of chemotherapy, and do not benefit from first-line therapy and are unlikely to respond to second-line treatment options [2, 3]. This therefore results in a very poor survival for this group of patients, even when considering the overall poor survival seen in SCLC [39]. Despite the initial high response rate to first-line therapy seen in the majority of SCLC patients, this does not translate to sustained benefit for patients, with almost universal relapses with progressive disease seen and usually in just a small number of months. At this stage there is a marked reduction in the chemosensitivity of tumours with poor responses to second-line therapies seen [3]. Due to the rapid progression of SCLC with the associated decline in patients' performance status and the inherent toxicity of chemotherapy not all patients are suitable for further chemotherapy, which in addition to the development of acquired resistance to chemotherapy seen contributes to the poor survival in SCLC [39]. Intrinsic resistance and particularly acquired resistance therefore represent significant hurdles to overcome to improve the outcomes for patients with SCLC, which is why they were a major focus in this thesis. The data presented in this thesis using CTCs to explore chemoresistance in SCLC is very novel. To date no work has been published using CTCs to investigate intrinsic resistance in SCLC and there is data from just one patient published to address changes in relapsed CTCs [229].

6.5.1 The Potential of CTCs to Investigate Intrinsic Chemoresistance in SCLC

To investigate intrinsic resistance CTCs were enumerated by CellSearch at baseline, prior to receiving chemotherapy, in SCLC patients who went on to respond to chemotherapy and also from those who were refractory to treatment. This confirmed that there was no significant difference between the baseline CTC numbers of these

two groups of patients. This would suggest that chemorefractory behaviour is not just the result of a greater disease burden which overwhelms patients before they can respond to chemotherapy. Analysis of molecular changes in CTCs isolated from patients who responded to and those who were refractory to first-line therapy was therefore undertaken to investigate this further. Visual differences in the CNA profiles of CDX tumours created from patients with chemoresponsive disease and chemorefractory disease had been noted, with a trend to greater numbers of losses in the chemorefractory tumours and amplifications in the chemoresponsive tumours (figure 4.2). This trend was also seen in the analysis of CTCs isolated from the chemoresponsive and chemorefractory patients (figure 5.5). Clear differences between the two groups of patients' CTCs were highlighted through an unsupervised statistical analysis of the CNA data generated from protein-coding genes (figure 5.6). This suggests there is an association between the differences in CNA and the behaviour noted in these two groups of patients. The strong correlation noted between both PFS and OS of the patients and the principal component generated from the PCA of the CNA data provides further credence to this assertion (figure 5.7).

An analysis of the genes with most marked differences in copy number between the two groups of patients provided plausible biological explanations for the responses to chemotherapy noted. Examples included the amplification of GSTA genes, which are involved in the detoxification of compounds through conjugation with glutathione, and the loss of carboxylesterases, which are involved in the hydrolysis of xenobiotics, seen in the CTCs isolated from the chemorefractory patients that could both result in resistance to chemotherapy [295, 296, 299, 300]. The loss of carboxylesterases may impair the conversion of Irinotecan, which is used in both the first and second-line in SCLC treatment, to its active metabolite SN-38 [299, 300]. The amplification of GSTA genes has been linked with resistance to both nitrogen mustard chemotherapies and Doxorubicin [295, 296]. An analysis of the different patterns of mutations identified through WES of CTCs from these two patient groups highlighted further differences that may contribute to the different responses to chemotherapy seen. Mutations in HSPG2 and PCDH10 were noted in the chemorefractory patients' CTCs and these genes have both been linked with chemoresistance when downregulated in other studies [307-309]. HSPG2, which encodes a component of basement membranes, has been noted to be downregulated in chemoresistant breast cancer tissue and cell lines. Mutations in PCDH10, a member of the cadherin superfamily, have been linked with resistance to Imatinib, Doxorubicin and Methotrexate in studies, for example through impaired apoptosis [308, 309]. The results generated from both the CNA and WES

data have therefore highlighted potential mechanisms for the intrinsic resistance seen in these patients which could be explored further in larger trials. Confirmation that the genomic aberrations lead to changes in mRNA and protein levels in the CTCs would also be required. This group has recently published a single-cell mRNA workflow that could be used to further investigate changes in mRNA expression of key genes of interest in SCLC CTCs [327]. The methods used for CNA analyses and WES could also be used to personalise individual patient's therapy for example avoiding the use of Irinotecan in patients with loss of copy number of carboxylesterases.

6.5.2 A CTC-Based Biomarker of Intrinsic Resistance?

To further investigate the marked differences noted in CNAs between the CTCs from chemoresponsive and chemorefractory patients, a further analysis to identify significant differences in copy number was performed identifying 760 affected genes (figure 5.8). This potential CNA signature again produced clear separation of the two groups of patients CTCs. The changes in copy number were further refined into changes in just 13 profiles based on the chromosome the genes were located on and the direction of change in copy number in the two groups of patients' CTCs. This potential signature provides a method of distinguishing these two groups of patients from their baseline CTC samples in contrast to the clinical criteria, which by definition, requires the patients to be monitored for a minimum of 3 months post completion of chemotherapy. There are currently no molecular signatures developed from SCLC tumours which can be used to distinguish these two groups of patient. This signature has been created from the analysis of 10 patients CTCs and clearly would need to be validated through the analysis of CTCs from a larger cohort of patients. It, however, represents an exciting finding from this thesis.

As a clinician I believe that there are clear potential clinical and research uses of a signature which could distinguish chemoresponsive and chemorefractory patients. This information could be used to assist patients in making informed choices about treatment, though there would clearly need to be further investigation done to establish the sensitivity and specificity of the signature. As there are clear differences in behaviour between these groups of patients it would be appropriate to identify them prior to entry to clinical trials for a number of reasons. Trials assessing treatment with new agents in addition to chemotherapy in the first-line setting ensure the arm receiving standard therapy and the arm with the novel therapy are balanced for a number of factors, such as the performance status of patients, age and gender. This is

carried out to avoid biasing the results through differences in these clinical factors. It would also be useful to balance arms for patients who are likely to be chemorefractory. Failure to do this could result in failing to identify benefit from a new agent or falsely attributing benefit to a new agent when in fact it just reflected the patient mix in the two arms. Given the marked differences in behaviour of tumours from patients who do, or do not respond to first-line therapy, it may also be beneficial to design trials for each group independently to meet their specific needs. For example, given the low response rate seen in patients with chemorefractory disease to standard chemotherapy, these patients may be ideal for first-line trials of rational novel agent combinations. Patients who are likely to respond to first-line chemotherapy may be better served by trials of novel agents in addition to chemotherapy or maintenance therapies to maintain the response they achieve. A signature to identify the two groups of patients in advance of receiving therapy would be essential for this.

6.5.2 Using CTCs to Interrogate Mechanisms of Acquired Resistance in SCLC

Acquired resistance represents in many ways a greater issue in SCLC than intrinsic resistance, as it results in treatment failure in greater numbers of patients (approximately 80%). As the investigation of acquired resistance requires the identification of the development of new genomic aberrations over time serial sampling is required, for which CTCs are ideally suited. The same approach as that which was used to investigate intrinsic resistance was applied to the investigation of acquired resistance which was equated with the development of progressive disease more than 3 months after demonstrating an initial response to chemotherapy. CTCs isolated at baseline from chemoresponsive patients, and in the same patients when they relapsed after receiving chemotherapy were analysed to investigate CNAs and mutations. PCA of the copy number profiles of the CTCs failed to show any separation between the CTCs from the two time points and just highlighted the similarities of CTCs from individual patients irrelevant of when they were isolated. This contrasts to the data presented when comparing the baseline CTCs isolated from chemoresponsive and chemorefractory patients. When comparing CNA between the two groups of CTCs from baseline and relapse there were no significant common changes identified across the patients. The CTCs isolated when patients relapsed also still demonstrated marked differences from those isolated from patients with chemorefractory disease, as they still retained a very similar CNA profile to the baseline CTCs. This suggests that

the development of acquired resistance has different underlying mechanisms than seen in intrinsic resistance, as opposed to it just representing the same process developing at a later time point.

The results from the analyses presented in this thesis of CTC CNA profiles from baseline and relapse samples demonstrated no common changes. This reflects the findings of Ni et al who examined the CNA profiles of CTCs before and after chemotherapy from one patient with SCLC [229]. There are a number of potential explanations for these results. It is possible that at baseline there are a small number of CTCs with a resistant genotype which become more prominent at relapse, though the analysis of the longitudinal patients' CTCs failed to demonstrate these subgroups clearly. The difference in profiles across the CTCs at both time points would be diluted by the presence of the resistant CTCs amongst the baseline CTCs. If there are changes in the proportions of CTCs with a specific signature between the two time points the failure to capture this may represent an issue of insufficient sampling of CTCs. Patients may have up to several thousand CTCs enriched from their blood sample from which up to 10 singles and 2 pools of 10 CTCs were analysed at any time point. It is feasible that the range of CTCs from a patient was not adequately represented. One of the significant limiting factors when analysing CTCs from a sample is the DEPArray as there are a defined number of cell recoveries that are possible. Although the ability of the DEPArray to recover more samples has been improved with recent software and hardware upgrades it still represents a constraint. Despite these potential sampling issues clear differences in the profiles of the baseline chemoresponsive and chemorefractory patients' CTCs were still identified which may mean this theory is less relevant.

The changes seen in response to chemotherapy may not be accounted for by the accumulation of new changes in copy number in the relapse CTCs. It is possible the change in behaviour is driven by other factors such as differences in mutations, epigenetic changes, protein expression and post translational changes. Examination of mutations, both SNVs and Indels, in the baseline and relapse CTCs revealed a large number of genes that were exclusively mutated at relapse. This included five tumour suppressor genes whose mutation may contribute to a more aggressive growth pattern in tumours when patients relapse. This included the identification of mutations in TGFRβ2, a tumour suppressor gene, whose mutation may promote more 'stem' like and therefore potentially chemoresistant behaviour [320]. Mutations in ATP1A1, an ion pump, were also noted amongst the relapse CTCs. ATP1A1 mutations have been linked with resistance to Cisplatin in studies, through platinum accumulation defects

[321, 322]. As platinum-based treatment is the cornerstone of SCLC, this potentially represents a very critical change. These data suggest that there may be mutations associated with the development of acquired resistance accumulating within these patients CTCs. Much of the research investigating chemoresistance has been performed using cell lines and the results of such studies do not always translate into the clinic. The mutations identified in CTCs in this thesis have come from changes in patients' samples in response to treatment and so represent a novel alternative method of exploring chemoresistance.

6.5.3 Different Mechanisms Produce Intrinsic and Acquired Resistance in SCLC?

The data presented from this research provides evidence that there are different mechanisms associated with intrinsic and acquired resistance. There is strong evidence from the data presented of a link between intrinsic resistance and specific changes in copy number. However, the results did not demonstrate a clear change in CNA profiles at the development of acquired resistance, although the acquisition of new mutations was demonstrated through WES of CTCs. A possible explanation of the data presented is that the development of the two forms of resistance occurs in response to different selection pressures and over different time scales. The genomic aberrations associated with intrinsic resistance can arise during tumorigenesis and may occur to promote tumour growth and metastasis formation whilst also coincidentally resulting in chemoresistance. The alterations associated with acquired resistance potentially appear in a shorter timescale under the direct selective pressure of treatment. It is therefore feasible that due to the different pressures and timescales present during the evolution of the two forms of resistance they involve contrasting mechanisms.

The results presented in this thesis have revealed potential resistance mechanisms occurring in patients' tumours through analysing CTCs. This represents a novel approach to the investigation of chemoresistance in SCLC in contrast to using for example cell lines or tumour biopsies. It also clearly demonstrates the strengths of CTCs, and the methods developed to analyse them, in the research of chemoresistance. Given the marked problems associated with chemoresistance in SCLC and the high number of CTCs present in this disease, SCLC represents an ideal candidate for the further investigation of CTCs as a research tool for chemoresistance.

6.6 Potential Utility of CTC Isolation and Analysis for the Treatment of SCLC Patients – What We Know and Unanswered Questions

In SCLC evidence for CellSearch detected CTCs as prognostic biomarkers has previously been established [189, 204, 205]. However, as demonstrated in this thesis CTCs could potentially have far wider roles. The entry into clinical trials is increasingly dependent on the identification of specific genetic predictive biomarkers (such as EGFR mutations in NSCLC). This necessitates analysis of tumour biopsy samples beyond establishing the histological diagnosis. The ability to demonstrate mutations and CNAs in potentially clinically actionable genes in CTCs in this thesis suggests that CTCs could act as surrogates for tumour biopsies. This is particularly relevant in lung cancer where biopsies are not always technically possible and are not without risk for patients. One of the key benefits for using CTCs as opposed to biopsies in clinical trials is the ease of obtaining serial samples. This means it is feasible to track changes in the prevalence and molecular composition of CTCs containing the predictive biomarker(s) of interest, or to track the development of resistance mechanisms over the course of a patient's treatment. Again within this thesis the ability to monitor the evolution of mutations at two time points was demonstrated confirming the practicability of this approach (table 5.10). It should, however, be stressed that although it is easier to obtain serial samples of CTCs than biopsies, their analysis presents increased challenges.

CTCs can also be used to interrogate the biology of SCLC, with increased knowledge of this disease likely to be very important when trying to improve patients' outcomes. Critical issues for the research of SCLC such as chemoresistance can be investigated, as highlighted in this thesis, using CTCs. The results of this sort of research can then be used to generate new hypotheses for investigation in clinical research and trials such as the investigation of the prevalence of specific mechanisms of resistance. It is possible that initial small-scale analyses of single CTCs could be carried out to generate relevant genomic profiles such as the potential CNA or mutation profiles identified from the chemoresponsive and chemorefractory patients' CTCs. The presence of these profiles in for example enriched CTC samples or cfDNA could then be investigated. This would reduce the complexity of further analyses and their costs and enable larger numbers of patients to be screened. The potential future clinical roles for CTCs in SCLC are summarised in figure 6.2.

6.6.1 CTCs and cfDNA – Strengths and Weaknesses

CTCs and cfDNA are often discussed together as potential liquid biopsies, though each has its own advantages and disadvantages. The benefits of cfDNA analysis, in contrast to CTCs, include the ease of obtaining cfDNA from plasma samples without the need for enrichment and then potentially isolation which is required for CTC analysis [328]. The cost associated with cfDNA isolation, in contrast to CTCs, also represents a positive aspect of its analysis. Genotyping tumours and monitoring therapy response represent areas of potential clinical application for cfDNA. For example in NSCLC the identification of activating EGFR mutations and the emergence of T790M resistance mutations whilst patients received an EGFR TKI have been demonstrated from cfDNA analysis [329, 330]. By its very nature though cfDNA analysis is restricted to the analysis of DNA alone with possible analysis including point mutations, structural rearrangements, copy number change and DNA methylation.

Single-cell analysis of CTCs allows the analysis of pure tumour DNA whilst cfDNA consists of both wild-type DNA from normal cells and variable concentrations of tumour derived DNA. As analysis of the DNA from single CTCs can be carried out it is possible to examine coexisting mutations which enables the investigation of heterogeneity and evolution in tumours [152]. The analysis of CTCs also allows the assessment of a whole cell meaning that DNA, RNA and protein analyses are possible, unlike with cfDNA. Studies have also demonstrated the ability to combine DNA and mRNA analysis from single cells [328]. One of the most striking benefits of CTC analysis is the ability to carry out in vivo functional studies through the development of CDX [24, 273]. This will allow interrogation of therapy responses, the genetics underlying these responses and investigation of the biology of cancers. It is therefore likely that CTCs and cfDNA will have complementary roles in the future based on the unique strengths of each. The use of cfDNA is likely to be focussed on genotyping and monitoring changes in a biomarker or the evolution of a resistance mechanism with treatment whilst CTCs will be used to explore the biology of cancer metastasis and investigate the use of existing and novel treatments.

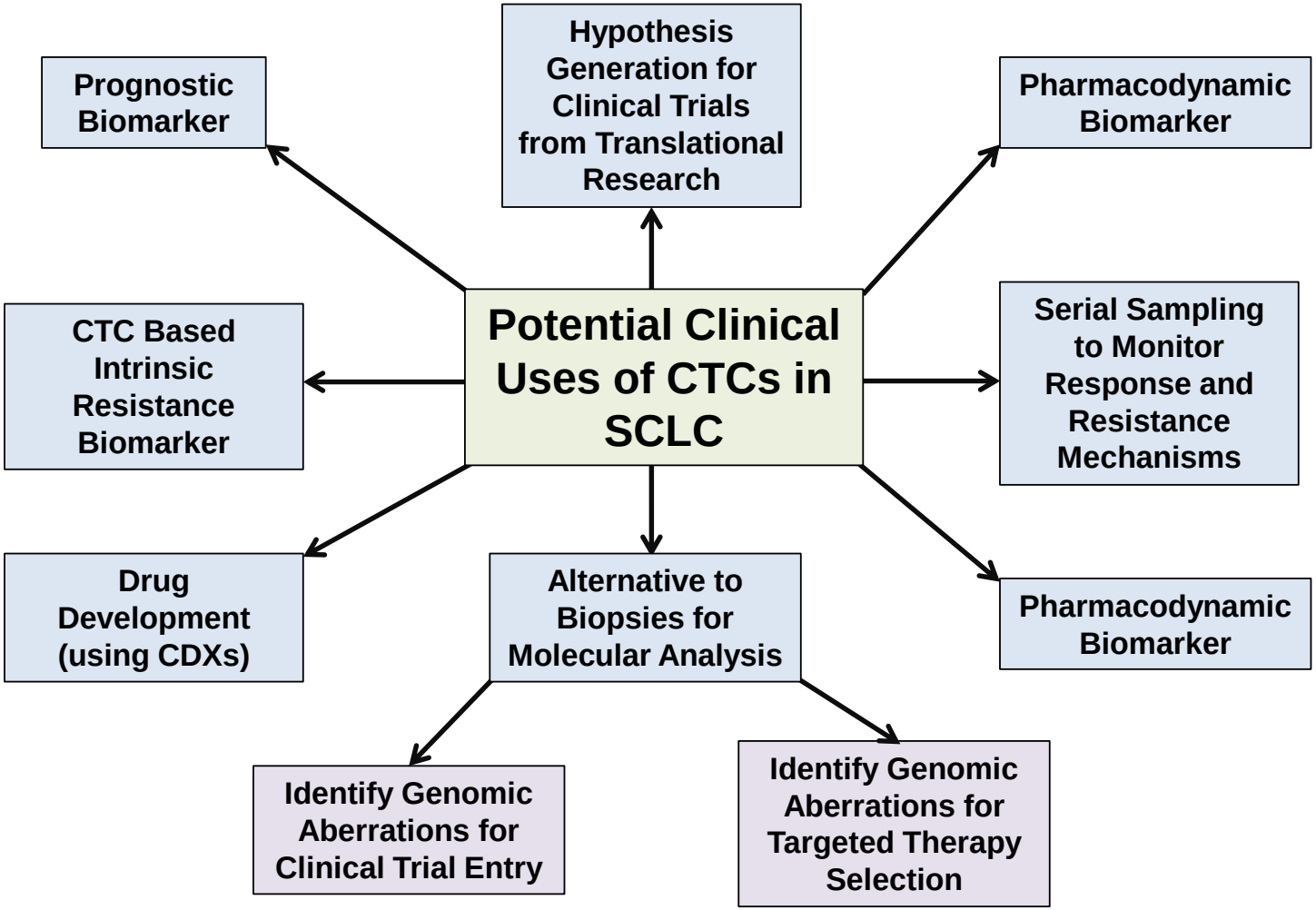


Figure 6.2 Potential clinical uses of CTCs in SCLC

6.6.2 CDX Models – A New Approach to the Exploration of SCLC Biology and Drug Development

The CDX models represent an important way in which CTCs could be used to advance the understanding and optimisation of treatment in SCLC. The demonstration that the response of patients to Cisplatin and Etoposide chemotherapy could be recapitulated in the CDX tumours generated from their CTCs suggests that CDX could have a meaningful role in drug development. Pre-clinical investigation of novel therapies could be carried out using the CDX models. The molecular profiling of the CDX models would enable the identification of potential biomarkers associated with responders and non-responders. The paired CTC sample, taken at the same time as the sample used to create the CDX model, enriched using the CellSearch platform could also be analysed to confirm the presence of the same biomarkers. If patients were subsequently enrolled in clinical trials of the investigational agent their CTCs could again be analysed for the presence of this biomarker. Given the failure of targeted therapies in SCLC this could present a rational approach for trial design, potentially improving the chance of success in future trials. Within the pre-clinical team in CEP studies assessing DNA damage repair inhibitors in combination with chemotherapy in the CDX are being carried out, with promising initial data emerging, highlighting the potential of CDX analyses.

6.6.3 CTCs: Progression from Research Tool to Clinical Implementation

Despite the positive results in this thesis it should be acknowledged that the adoption of single-cell analysis of CTCs would require a number of issues to be addressed. Key questions such as the number of CTCs that need to be analysed per patient to ascertain whether or not they can be considered to have a predictive biomarker is as yet unclear. As it is evident that CTCs are heterogeneous, consideration of the proportion of CTCs that need to contain the biomarkers to conclude that a patients may respond to a targeted agent also needs to be examined. Larger scale studies would clearly be needed to address these issues. The significant costs associated with the isolation of CTCs and the analysis of single cells has already been discussed and may inhibit the adoption of these techniques in the near future.

As with NGS of tumour biopsies, the sequencing of CTCs generates large quantities of data which requires complex bioinformatic analysis. In CTCs the bioinformatic analysis of NGS data is further complicated by the use of the amplified DNA. This was

exemplified by the need to develop a different approach when analysing the WES data generated from the amplified DNA, as opposed to unamplified DNA, due to the poor mapping of the reads initially. The errors and bias from the amplification process also need to be addressed when examining single cells as opposed to bulk tumour biopsies. During this study the expertise of the RNA Biology/Computational Biology group, led by Crispin Miller in this institute, have been invaluable in establishing robust methodology for the bioinformatic analysis of CTC NGS data. If the use of NGS of CTCs was to become widespread clinically, the speed with which results could be analysed and final reports for clinicians created would need to be considered. This is particularly true in SCLC where due to the aggressive nature of the disease patients deteriorate quickly, necessitating rapid clinical decision making. WGS and WES of CTCs may be most appropriate for the discovery of new biomarkers whilst focussed NGS may be most suited for the analysis of clinical samples, given the reduced bioinformatic analysis required. Although CTCs have a clear role in research and potentially in the clinical management of SCLC further work to tackle these methodological and bioinformatic issues will be required. The data presented in the thesis, however, would suggest this would be a very worthwhile focus which could lead to improvements in outcomes given the promise already shown.

6.7 Future Work and Perspectives

The results generated in this thesis suggest the molecular analysis of CTCs is a fruitful area for future research in SCLC. Further studies addressing the potential mechanisms of chemoresistance highlighted would be of interest. The CNA data generated suggest that there are changes in copy number which could be contributing to the chemoresistant behaviour seen in the patients' tumours. However, the generation of gene expression data to correlate with the CNA data would strengthen the argument that the CNAs are producing downstream changes. The potential CTC CNA signature to discriminate the chemoresponsive and chemorefractory patients also represents a promising area for future research. Initially a repeat analysis in a second cohort of SCLC patients would be essential to validate the potential signature. If validated an assessment of whether the signature could be detected in either enriched samples of CTCs, as opposed to isolated CTCs or cfDNA would be of interest. This is because due to the decreased processing required for enriched CTCs or cfDNA it may enable the signature to be more widely used. Within CEP methodology for the isolation and analysis of cfDNA has already been optimised. CNA profiles have been generated

from CTCs and cfDNA isolated from two patients with SCLC demonstrating similar profiles (see figure 6.3 and Appendix 7). This methodology could be applied to investigating the presence of the potential resistance CNA signature in cfDNA isolated from SCLC patients' blood, as opposed to CTCs.

In the investigation of changes between CTCs isolated at baseline and at relapse from patients with SCLC there were different patterns of mutations noted. The expansion of this analysis to a further group of patients would also be of interest to address if the same potential mechanisms for resistance are identified again. The analysis of a core set of mutations associated with acquired resistance could then be carried out in a larger cohort of patients using a focussed NGS approach such as TAm-Seq. The same approach could be applied to confirming the presence of the differential mutations identified between the CTCs from the chemoresponsive and chemorefractory patients.

The assessment of the degree of heterogeneity amongst an individual patient's CTCs, and how this may be of significance for treatment is an additional area of interest for further research. To investigate heterogeneity large numbers of CTCs from individual patients would need to be profiled. The DEPArray remains a limiting factor for this with the maximum number of CTCs it is possible to isolate currently being 32. To maximise the analysis of CTCs from a patient a number of options could be considered. These include dividing the enriched CTC sample from CellSearch between a number of DEPArray runs. Other potential options for single cell isolation that could be used in place of the DEPArray could be explored such as the CellCelector [331]. It may also be relevant to optimise methods for the long term storage of enriched CTC samples. Single CTCs could then be isolated from an aliquot of the enriched sample and analysed. If genomic aberrations of interest were identified the other aliquots of the cells could then be processed. This would increase the potential numbers of CTCs which could be analysed from each sample, but would ensure the expense associated with the additional processing was only generated when genetic abnormalities of interest were identified. Initial studies investigating freezing the enriched CTCs in glycerol prior to isolation, enrichment and amplification are promising.

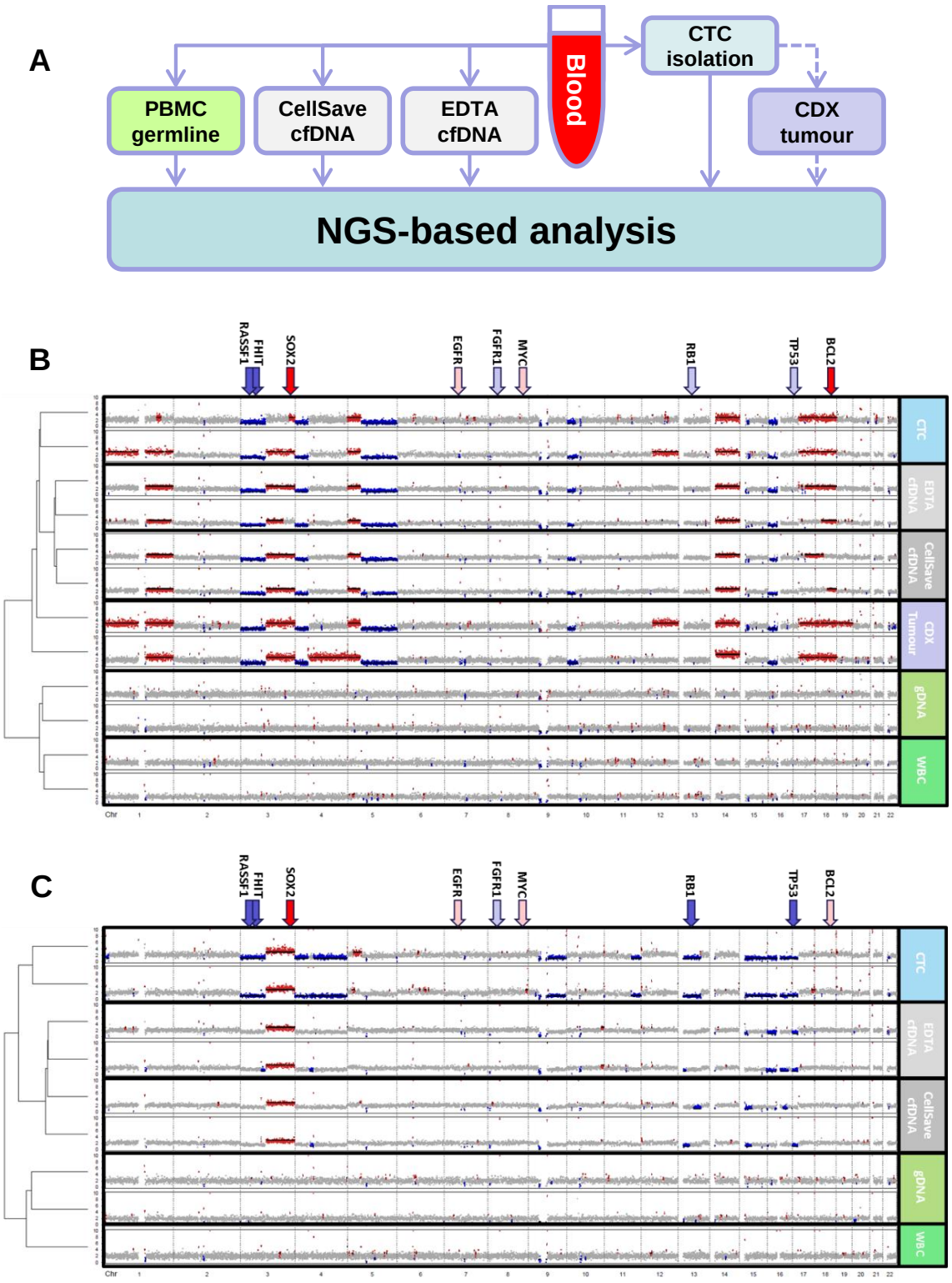


Figure 6.3. Comparison of CNA profiles generated from CTCs, cfDNA, WBCs and gDNA isolated from two patients with SCLC. (Reproduced from Rothwell et al 2015, In Review). **A.** Schematic showing procedure for processing of a single blood sample to give NGS analysis of cfDNA and CTCs. **B & C.** Unsupervised, hierarchical clustering of CNA profiles in two SCLC patients. CNA profiles were generated from isolated CTCs, EDTA cfDNA, CellSave cfDNA, two CDX tumours (A only), germline gDNA and isolated WBC. Matching patterns of gain (regions of red) and loss (regions of blue) were seen across all tumour material and were absent from germline controls. Arrows indicate location of common copy number aberrations found in SCLC

with red indicating gain and blue loss. Dark filled arrows indicate loci altered in the patient sample.

Suggested further studies are listed below.

1) Assessment of gene expression from CTCs isolated at baseline from chemoresponsive and chemorefractory patients to correlate with CNAs identified in the two groups' CTCs.
2) Validation of the ability of the potential CTC CNA signature to discriminate between chemoresponsive and chemorefractory patients using samples from a larger cohort of SCLC patients.
3) Investigation of whether the potential CTC CNA signature used to discriminate chemoresponsive and chemorefractory patients can be identified in the post CellSearch enriched CTC sample as opposed to using single CTCs.
4) Assessment of whether the differential mutations identified by WES between CTCs isolated at baseline and when patients relapse after first-line therapy can be identified in a larger cohort of patients' CTCs, to confirm if there are common mechanisms associated with acquired resistance.
5) Investigation of the heterogeneity of genomic aberrations in patients' CTCs through profiling larger numbers of CTCs from individual SCLC patients, and assess its relevance to patients' outcomes.

6.8 Concluding Remarks

There is significant unmet need in SCLC and research leading to improvements in patient outcomes is long overdue. This thesis has provided novel evidence of the potential of CTCs to be used to address key areas of research in this disease, adding to the published literature in this field. Results demonstrating that CTCs contain the same common genomic aberrations as SCLC tumours have been presented, suggesting they could be used as a surrogate for tumour biopsies to explore the genomic landscape of this complex disease. The capability of CTCs to investigate chemoresistance in SCLC has also been established. Potential mechanisms associated with both intrinsic and acquired resistance have been revealed through analysis of patterns of CNAs and mutations in CTCs. Results from this thesis have been presented at numerous national and international meetings and generated great interest. There are also several projects utilising SCLC CTCs continuing within this laboratory.

In England and Wales 3000 people are diagnosed every year with SCLC with the majority dying within a year of diagnosis [332]. This starkly highlights the need to improve the treatments available for these patients. Collaborative approaches to SCLC research are required with scientists, bioinformaticians, oncologists and radiologists working together to translate scientific advances in to meaningful improvements for patients. It is hoped that CTCs, as the results in this thesis suggest, could be a valuable research tool and eventually clinical tool to help drive forward these improvements.

References

1. Jemal, A., et al., *Global cancer statistics.* CA Cancer J Clin, 2011. **61**(2): p. 69-90.
2. Byers, L.A. and C.M. Rudin, *Small cell lung cancer: Where do we go from here?* Cancer, 2014.
3. Fruh, M., et al., *Small-cell lung cancer (SCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Ann Oncol, 2013. **24 Suppl 6**: p. vi99-105.
4. Worden, F.P. and G.P. Kalemkerian, *Therapeutic advances in small cell lung cancer.* Expert Opin Investig Drugs, 2000. **9**(3): p. 565-79.
5. Cooper, S. and S.G. Spiro, *Small cell lung cancer: treatment review.* Respirology, 2006. **11**(3): p. 241-8.
6. Zelen, M., *Keynote address on biostatistics and data retrieval.* Cancer Chemother Rep 3, 1973. **4**(2): p. 31-42.
7. van Meerbeeck, J.P., D.A. Fennell, and D.K. De Ruysscher, *Small-cell lung cancer.* Lancet, 2011. **378**(9804): p. 1741-55.
8. Thatcher, N., C. Faivre-Finn, and P. Lorigan, *Management of small-cell lung cancer.* Ann Oncol, 2005. **16 Suppl 2**: p. ii235-9.
9. Chan, B.A. and J.I. Coward, *Chemotherapy advances in small-cell lung cancer.* J Thorac Dis, 2013. **5 Suppl 5**: p. S565-78.
10. *Integrated Household Survey: January 2012 to December 2012 (Experimental Statistics)*, O.N. Statistics, Editor 2013.
11. Kalemkerian, G.P., et al., *Small cell lung cancer.* J Natl Compr Canc Netw, 2013. **11**(1): p. 78-98.
12. Sakurai, H., et al., *Isolated solitary brain metastasis as a relapse of small cell lung cancer.* Oncol Lett, 2013. **6**(4): p. 1108-1110.
13. Ries, L.A.G., et al., *SEER Survival Monograph: Cancer Survival Among Adults: U.S. SEER Program, 1988-2001, Patient and Tumour Characteristics*, N.C. Institute, Editor 2007: Bethesda, MD.
14. Park, K.S., et al., *Characterization of the cell of origin for small cell lung cancer.* Cell Cycle, 2011. **10**(16): p. 2806-15.
15. Jackman, D.M. and B.E. Johnson, *Small-cell lung cancer.* Lancet, 2005. **366**(9494): p. 1385-96.
16. Brambilla, E., et al., *The new World Health Organization classification of lung tumours.* Eur Respir J, 2001. **18**(6): p. 1059-68.
17. Hirsch, F.R., et al., *Histopathologic classification of small cell lung cancer. Changing concepts and terminology.* Cancer, 1988. **62**(5): p. 973-7.
18. Tan, D. and D.S. Zander, *Immunohistochemistry for assessment of pulmonary and pleural neoplasms: a review and update.* Int J Clin Exp Pathol, 2008. **1**(1): p. 19-31.
19. Davidson, M.R., A.F. Gazdar, and B.E. Clarke, *The pivotal role of pathology in the management of lung cancer.* J Thorac Dis, 2013. **5 Suppl 5**: p. S463-78.
20. Thunnissen, E., et al., *The challenge of NSCLC diagnosis and predictive analysis on small samples. Practical approach of a working group.* Lung Cancer, 2012. **76**(1): p. 1-18.
21. Daniel, V.C., et al., *A primary xenograft model of small-cell lung cancer reveals irreversible changes in gene expression imposed by culture in vitro.* Cancer Res, 2009. **69**(8): p. 3364-73.
22. Iwakawa, R., et al., *Genome-wide identification of genes with amplification and/or fusion in small cell lung cancer.* Genes Chromosomes Cancer, 2013. **52**(9): p. 802-16.

23. Manhire, A., et al., *Guidelines for radiologically guided lung biopsy.* Thorax, 2003. **58**(11): p. 920-36.
24. Hodgkinson, C.L., et al., *Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer.* Nat Med, 2014. **20**(8): p. 897-903.
25. Rossi, A., O. Martelli, and M. Di Maio, *Treatment of patients with small-cell lung cancer: from meta-analyses to clinical practice.* Cancer Treat Rev, 2013. **39**(5): p. 498-506.
26. Simon, G.R. and A. Turrisi, *Management of small cell lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition).* Chest, 2007. **132**(3 Suppl): p. 324S-339S.
27. Califano, R., et al., *Management of small cell lung cancer: recent developments for optimal care.* Drugs, 2012. **72**(4): p. 471-90.
28. Socinski, M.A. and J.A. Bogart, *Limited-stage small-cell lung cancer: the current status of combined-modality therapy.* J Clin Oncol, 2007. **25**(26): p. 4137-45.
29. Auperin, A., et al., *Prophylactic cranial irradiation for patients with small-cell lung cancer in complete remission. Prophylactic Cranial Irradiation Overview Collaborative Group.* N Engl J Med, 1999. **341**(7): p. 476-84.
30. Rodriguez, E. and R.C. Lilenbaum, *Small cell lung cancer: past, present, and future.* Curr Oncol Rep, 2010. **12**(5): p. 327-34.
31. Demedts, I.K., K.Y. Vermaelen, and J.P. van Meerbeeck, *Treatment of extensive-stage small cell lung carcinoma: current status and future prospects.* Eur Respir J, 2010. **35**(1): p. 202-15.
32. Slotman, B., et al., *Prophylactic cranial irradiation in extensive small-cell lung cancer.* N Engl J Med, 2007. **357**(7): p. 664-72.
33. Fukuoka, M., et al., *Randomized trial of cyclophosphamide, doxorubicin, and vincristine versus cisplatin and etoposide versus alternation of these regimens in small-cell lung cancer.* J Natl Cancer Inst, 1991. **83**(12): p. 855-61.
34. Skarlos, D.V., et al., *Randomized comparison of etoposide-cisplatin vs. etoposide-carboplatin and irradiation in small-cell lung cancer. A Hellenic Co-operative Oncology Group study.* Ann Oncol, 1994. **5**(7): p. 601-7.
35. Loehrer, P.J., Sr., et al., *Cisplatin plus etoposide with and without ifosfamide in extensive small-cell lung cancer: a Hoosier Oncology Group study.* J Clin Oncol, 1995. **13**(10): p. 2594-9.
36. Schiller, J.H., et al., *Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer.* N Engl J Med, 2002. **346**(2): p. 92-8.
37. Scagliotti, G.V., et al., *Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer.* J Clin Oncol, 2008. **26**(21): p. 3543-51.
38. Hanna, N., et al., *Randomized phase III trial comparing irinotecan/cisplatin with etoposide/cisplatin in patients with previously untreated extensive-stage disease small-cell lung cancer.* J Clin Oncol, 2006. **24**(13): p. 2038-43.
39. Cheng, S., et al., *Chemotherapy for relapsed small cell lung cancer: a systematic review and practice guideline.* J Thorac Oncol, 2007. **2**(4): p. 348-54.
40. Masuda, N., et al., *CPT-11: a new derivative of camptothecin for the treatment of refractory or relapsed small-cell lung cancer.* J Clin Oncol, 1992. **10**(8): p. 1225-9.
41. Asai, N., et al., *Relapsed small cell lung cancer: treatment options and latest developments.* Ther Adv Med Oncol, 2014. **6**(2): p. 69-82.
42. von Pawel, J., et al., *Topotecan versus cyclophosphamide, doxorubicin, and vincristine for the treatment of recurrent small-cell lung cancer.* J Clin Oncol, 1999. **17**(2): p. 658-67.
43. O'Brien, M.E., et al., *Phase III trial comparing supportive care alone with supportive care with oral topotecan in patients with relapsed small-cell lung cancer.* J Clin Oncol, 2006. **24**(34): p. 5441-7.

44. Oze, I., et al., *Twenty-seven years of phase III trials for patients with extensive disease small-cell lung cancer: disappointing results.* PLoS One, 2009. **4**(11): p. e7835.
45. Berge, E.M. and R.C. Doebele, *Targeted therapies in non-small cell lung cancer: emerging oncogene targets following the success of epidermal growth factor receptor.* Semin Oncol, 2014. **41**(1): p. 110-25.
46. Savas, P., B. Hughes, and B. Solomon, *Targeted therapy in lung cancer: IPASS and beyond, keeping abreast of the explosion of targeted therapies for lung cancer.* J Thorac Dis, 2013. **5 Suppl 5**: p. S579-92.
47. Thatcher, N., et al., *Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer).* Lancet, 2005. **366**(9496): p. 1527-37.
48. Mok, T.S., et al., *Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma.* N Engl J Med, 2009. **361**(10): p. 947-57.
49. Wong, D.W., et al., *The EML4-ALK fusion gene is involved in various histologic types of lung cancers from nonsmokers with wild-type EGFR and KRAS.* Cancer, 2009. **115**(8): p. 1723-33.
50. Soda, M., et al., *Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer.* Nature, 2007. **448**(7153): p. 561-6.
51. Takeuchi, K., et al., *Multiplex reverse transcription-PCR screening for EML4-ALK fusion transcripts.* Clin Cancer Res, 2008. **14**(20): p. 6618-24.
52. Perner, S., et al., *EML4-ALK fusion lung cancer: a rare acquired event.* Neoplasia, 2008. **10**(3): p. 298-302.
53. Kwak, E.L., et al., *Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer.* N Engl J Med, 2010. **363**(18): p. 1693-703.
54. Abidin, A.Z., et al., *Targeted therapies in small cell lung cancer: a review.* Ther Adv Med Oncol, 2010. **2**(1): p. 25-37.
55. Lu, H.Y., X.J. Wang, and W.M. Mao, *Targeted therapies in small cell lung cancer.* Oncol Lett, 2013. **5**(1): p. 3-11.
56. Pillai, R.N. and T.K. Owonikoko, *Small cell lung cancer: therapies and targets.* Semin Oncol, 2014. **41**(1): p. 133-42.
57. Rudin, C.M., A. Drilon, and J.T. Poirier, *RET mutations in neuroendocrine tumors: including small-cell lung cancer.* J Thorac Oncol, 2014. **9**(9): p. 1240-2.
58. Gazdar, A.F., et al., *Lung cancer cell lines as tools for biomedical discovery and research.* J Natl Cancer Inst, 2010. **102**(17): p. 1310-21.
59. Pleasance, E.D., et al., *A small-cell lung cancer genome with complex signatures of tobacco exposure.* Nature, 2010. **463**(7278): p. 184-90.
60. Gazdar, A.F., B. Gao, and J.D. Minna, *Lung cancer cell lines: Useless artifacts or invaluable tools for medical science?* Lung Cancer, 2010. **68**(3): p. 309-18.
61. Moro, M., et al., *Patient-derived xenografts of non small cell lung cancer: resurgence of an old model for investigation of modern concepts of tailored therapy and cancer stem cells.* J Biomed Biotechnol, 2012. **2012**: p. 568567.
62. Kwon, M.C. and A. Berns, *Mouse models for lung cancer.* Mol Oncol, 2013. **7**(2): p. 165-77.
63. Poupon, M.F., et al., *Response of small-cell lung cancer xenografts to chemotherapy: multidrug resistance and direct clinical correlates.* J Natl Cancer Inst, 1993. **85**(24): p. 2023-9.
64. Jacks, T., et al., *Effects of an Rb mutation in the mouse.* Nature, 1992. **359**(6393): p. 295-300.
65. Maandag, E.C., et al., *Developmental rescue of an embryonic-lethal mutation in the retinoblastoma gene in chimeric mice.* EMBO J, 1994. **13**(18): p. 4260-8.
66. Williams, B.O., et al., *Cooperative tumorigenic effects of germline mutations in Rb and p53.* Nat Genet, 1994. **7**(4): p. 480-4.

67. Meuwissen, R., et al., *Induction of small cell lung cancer by somatic inactivation of both Trp53 and Rb1 in a conditional mouse model.* Cancer Cell, 2003. **4**(3): p. 181-9.
68. Minna, J.D., J.M. Kurie, and T. Jacks, *A big step in the study of small cell lung cancer.* Cancer Cell, 2003. **4**(3): p. 163-6.
69. Vogelstein, B., et al., *Cancer genome landscapes.* Science, 2013. **339**(6127): p. 1546-58.
70. Peifer, M., et al., *Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer.* Nat Genet, 2012. **44**(10): p. 1104-10.
71. Pfeifer, G.P., et al., *Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers.* Oncogene, 2002. **21**(48): p. 7435-51.
72. Hecht, S.S., *Progress and challenges in selected areas of tobacco carcinogenesis.* Chem Res Toxicol, 2008. **21**(1): p. 160-71.
73. Karro, J.E., et al., *Exponential decay of GC content detected by strand-symmetric substitution rates influences the evolution of isochore structure.* Mol Biol Evol, 2008. **25**(2): p. 362-74.
74. Wistuba, II, A.F. Gazdar, and J.D. Minna, *Molecular genetics of small cell lung carcinoma.* Semin Oncol, 2001. **28**(2 Suppl 4): p. 3-13.
75. Lane, D.P., *Cancer. p53, guardian of the genome.* Nature, 1992. **358**(6381): p. 15-6.
76. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
77. Kandoth, C., et al., *Mutational landscape and significance across 12 major cancer types.* Nature, 2013. **502**(7471): p. 333-9.
78. Olivier, M., M. Hollstein, and P. Hainaut, *TP53 mutations in human cancers: origins, consequences, and clinical use.* Cold Spring Harb Perspect Biol, 2010. **2**(1): p. a001008.
79. Wistuba, II, et al., *Molecular changes in the bronchial epithelium of patients with small cell lung cancer.* Clin Cancer Res, 2000. **6**(7): p. 2604-10.
80. Arriola, E., et al., *Genetic changes in small cell lung carcinoma.* Clin Transl Oncol, 2008. **10**(4): p. 189-97.
81. Onuki, N., et al., *Genetic changes in the spectrum of neuroendocrine lung tumors.* Cancer, 1999. **85**(3): p. 600-7.
82. Kitamura, H., et al., *Small cell lung cancer: significance of RB alterations and TTF-1 expression in its carcinogenesis, phenotype, and biology.* Endocr Pathol, 2009. **20**(2): p. 101-7.
83. Cagle, P.T., et al., *Differential retinoblastoma protein expression in neuroendocrine tumors of the lung. Potential diagnostic implications.* Am J Pathol, 1997. **150**(2): p. 393-400.
84. Harbour, J.W., et al., *Abnormalities in structure and expression of the human retinoblastoma gene in SCLC.* Science, 1988. **241**(4863): p. 353-7.
85. Mori, N., et al., *Variable mutations of the RB gene in small-cell lung carcinoma.* Oncogene, 1990. **5**(11): p. 1713-7.
86. Joerger, A.C. and A.R. Fersht, *The tumor suppressor p53: from structures to drug discovery.* Cold Spring Harb Perspect Biol, 2010. **2**(6): p. a000919.
87. Petitjean, A., et al., *Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database.* Hum Mutat, 2007. **28**(6): p. 622-9.
88. Han, J.Y., et al., *Whole-genome analysis of a patient with early-stage small-cell lung cancer.* Pharmacogenomics J, 2014. **14**(6): p. 503-8.
89. McFadden, D.G., et al., *Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing.* Cell, 2014. **156**(6): p. 1298-311.
90. Whang-Peng, J., et al., *Specific chromosome defect associated with human small-cell lung cancer; deletion 3p(14-23).* Science, 1982. **215**(4529): p. 181-2.

91. Rudin, C.M., et al., *Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer.* Nat Genet, 2012. **44**(10): p. 1111-6.
92. Sos, M.L., et al., *A framework for identification of actionable cancer genome dependencies in small cell lung cancer.* Proc Natl Acad Sci U S A, 2012. **109**(42): p. 17034-9.
93. Sher, T., G.K. Dy, and A.A. Adjei, *Small cell lung cancer.* Mayo Clin Proc, 2008. **83**(3): p. 355-67.
94. Ross, J.S., et al., *Next-generation sequencing reveals frequent consistent genomic alterations in small cell undifferentiated lung cancer.* J Clin Pathol, 2014.
95. Takahashi, T., et al., *p53: a frequent target for genetic abnormalities in lung cancer.* Science, 1989. **246**(4929): p. 491-4.
96. Yokomizo, A., et al., *PTEN/MMAC1 mutations identified in small cell, but not in non-small cell lung cancers.* Oncogene, 1998. **17**(4): p. 475-9.
97. Shibata, T., et al., *Oncogenic mutation of PIK3CA in small cell lung carcinoma: a potential therapeutic target pathway for chemotherapy-resistant lung cancer.* Cancer Lett, 2009. **283**(2): p. 203-11.
98. Tatematsu, A., et al., *Epidermal growth factor receptor mutations in small cell lung cancer.* Clin Cancer Res, 2008. **14**(19): p. 6092-6.
99. Umemura, S., et al., *Therapeutic Priority of the PI3K/AKT/mTOR Pathway in Small Cell Lung Cancers as Revealed by a Comprehensive Genomic Analysis.* J Thorac Oncol, 2014. **9**(9): p. 1324-31.
100. Shiao, T.H., et al., *Epidermal growth factor receptor mutations in small cell lung cancer: a brief report.* J Thorac Oncol, 2011. **6**(1): p. 195-8.
101. Moore, A.M., et al., *Gefitinib in patients with chemo-sensitive and chemo-refractory relapsed small cell cancers: a Hoosier Oncology Group phase II trial.* Lung Cancer, 2006. **52**(1): p. 93-7.
102. Zakowski, M.F., M. Ladanyi, and M.G. Kris, *EGFR mutations in small-cell lung cancers in patients who have never smoked.* N Engl J Med, 2006. **355**(2): p. 213-5.
103. Okamoto, I., et al., *EGFR mutation in gefitinib-responsive small-cell lung cancer.* Ann Oncol, 2006. **17**(6): p. 1028-9.
104. Joshi, M., A. Ayoola, and C.P. Belani, *Small-cell lung cancer: an update on targeted therapies.* Adv Exp Med Biol, 2013. **779**: p. 385-404.
105. Tarhini, A., et al., *Phase II study of everolimus (RAD001) in previously treated small cell lung cancer.* Clin Cancer Res, 2010. **16**(23): p. 5900-7.
106. Pandya, K.J., et al., *A randomized, phase II trial of two dose levels of temsirolimus (CCI-779) in patients with extensive-stage small-cell lung cancer who have responding or stable disease after induction chemotherapy: a trial of the Eastern Cooperative Oncology Group (E1500).* J Thorac Oncol, 2007. **2**(11): p. 1036-41.
107. Byers, L.A., et al., *A Phase 2 Study of Amuvatinib (MP-470), the first RAD51 inhibitor in combination with Platinum-Etoposide (PE) in Refractory or Relapsed Small Cell Lung Cancer (ESCAPE)*, in *AACR Annual Meeting*2013: Washington.
108. Lu, Y., et al., *Evidence that SOX2 overexpression is oncogenic in the lung.* PLoS One, 2010. **5**(6): p. e11022.
109. Byers, L.A., et al., *Proteomic profiling identifies dysregulated pathways in small cell lung cancer and novel therapeutic targets including PARP1.* Cancer Discov, 2012. **2**(9): p. 798-811.
110. Kalari, S., et al., *The DNA methylation landscape of small cell lung cancer suggests a differentiation defect of neuroendocrine cells.* Oncogene, 2013. **32**(30): p. 3559-68.

111. Wilson, T.R., D.B. Longley, and P.G. Johnston, *Chemoresistance in solid tumours.* Ann Oncol, 2006. **17 Suppl 10**: p. x315-24.
112. Longley, D.B. and P.G. Johnston, *Molecular mechanisms of drug resistance.* J Pathol, 2005. **205**(2): p. 275-92.
113. Thomas, H. and H.M. Coley, *Overcoming multidrug resistance in cancer: an update on the clinical strategy of inhibiting p-glycoprotein.* Cancer Control, 2003. **10**(2): p. 159-65.
114. Wang, Y., R. Zhao, and I.D. Goldman, *Decreased expression of the reduced folate carrier and folypolyglutamate synthetase is the basis for acquired resistance to the pemetrexed antifolate (LY231514) in an L1210 murine leukemia cell line.* Biochem Pharmacol, 2003. **65**(7): p. 1163-70.
115. Meijer, C., et al., *Relationship of cellular glutathione to the cytotoxicity and resistance of seven platinum compounds.* Cancer Res, 1992. **52**(24): p. 6885-9.
116. Stewart, D.J., *Mechanisms of resistance to cisplatin and carboplatin.* Crit Rev Oncol Hematol, 2007. **63**(1): p. 12-31.
117. Stewart, D.J., *Tumor and host factors that may limit efficacy of chemotherapy in non-small cell and small cell lung cancer.* Crit Rev Oncol Hematol, 2010. **75**(3): p. 173-234.
118. Maheswaran, S., et al., *Detection of mutations in EGFR in circulating lung-cancer cells.* N Engl J Med, 2008. **359**(4): p. 366-77.
119. Bean, J., et al., *MET amplification occurs with or without T790M mutations in EGFR mutant lung tumors with acquired resistance to gefitinib or erlotinib.* Proc Natl Acad Sci U S A, 2007. **104**(52): p. 20932-7.
120. Evans, W.K., et al., *VP-16 alone and in combination with cisplatin in previously treated patients with small cell lung cancer.* Cancer, 1984. **53**(7): p. 1461-6.
121. Giaccone, G., et al., *Teniposide in the treatment of small-cell lung cancer: the influence of prior chemotherapy.* J Clin Oncol, 1988. **6**(8): p. 1264-70.
122. Shen, D.W., et al., *Cisplatin resistance: a cellular self-defense mechanism resulting from multiple epigenetic and genetic changes.* Pharmacol Rev, 2012. **64**(3): p. 706-21.
123. Stordal, B. and M. Davey, *Understanding cisplatin resistance using cellular models.* IUBMB Life, 2007. **59**(11): p. 696-9.
124. O'Grady, S., et al., *The role of DNA repair pathways in cisplatin resistant lung cancer.* Cancer Treat Rev, 2014. **40**(10): p. 1161-1170.
125. Giaccone, G., et al., *MRP is frequently expressed in human lung-cancer cell lines, in non-small-cell lung cancer and in normal lungs.* Int J Cancer, 1996. **66**(6): p. 760-7.
126. Triller, N., et al., *Multidrug resistance in small cell lung cancer: expression of P-glycoprotein, multidrug resistance protein 1 and lung resistance protein in chemo-naive patients and in relapsed disease.* Lung Cancer, 2006. **54**(2): p. 235-40.
127. Ushijima, R., et al., *Immunohistochemical expression of MRP2 and clinical resistance to platinum-based chemotherapy in small cell lung cancer.* Anticancer Res, 2007. **27**(6C): p. 4351-8.
128. Oguri, T., et al., *MRP7/ABCC10 expression is a predictive biomarker for the resistance to paclitaxel in non-small cell lung cancer.* Mol Cancer Ther, 2008. **7**(5): p. 1150-5.
129. Bergman, A.M., et al., *Collateral sensitivity to gemcitabine (2',2'-difluorodeoxycytidine) and cytosine arabinoside of daunorubicin- and VM-26-resistant variants of human small cell lung cancer cell lines.* Biochem Pharmacol, 2001. **61**(11): p. 1401-8.
130. Glisson, B.S. and M.D. Alpeter, *Multidrug resistance in a small cell lung cancer line: rapid selection with etoposide and differential chemosensitization with cyclosporin A.* Anticancer Drugs, 1992. **3**(4): p. 359-66.

131. Takigawa, N., et al., *Establishment and characterization of an etoposide-resistant human small cell lung cancer cell line.* Acta Med Okayama, 1992. **46**(3): p. 203-12.
132. Campling, B.G., et al., *Expression of the MRP and MDR1 multidrug resistance genes in small cell lung cancer.* Clin Cancer Res, 1997. **3**(1): p. 115-22.
133. Milroy, R., *A randomised clinical study of verapamil in addition to combination chemotherapy in small cell lung cancer. West of Scotland Lung Cancer Research Group, and the Aberdeen Oncology Group.* Br J Cancer, 1993. **68**(4): p. 813-8.
134. Wood, L., et al., *Results of a phase III, double-blind, placebo-controlled trial of megestrol acetate modulation of P-glycoprotein-mediated drug resistance in the first-line management of small-cell lung carcinoma.* Br J Cancer, 1998. **77**(4): p. 627-31.
135. Adams, J.M. and S. Cory, *The Bcl-2 apoptotic switch in cancer development and therapy.* Oncogene, 2007. **26**(9): p. 1324-37.
136. Rudin, C.M., et al., *Phase II study of single-agent navitoclax (ABT-263) and biomarker correlates in patients with relapsed small cell lung cancer.* Clin Cancer Res, 2012. **18**(11): p. 3163-9.
137. Sartorius, U.A. and P.H. Krammer, *Upregulation of Bcl-2 is involved in the mediation of chemotherapy resistance in human small cell lung cancer cell lines.* Int J Cancer, 2002. **97**(5): p. 584-92.
138. Eyler, C.E. and J.N. Rich, *Survival of the fittest: cancer stem cells in therapeutic resistance and angiogenesis.* J Clin Oncol, 2008. **26**(17): p. 2839-45.
139. Navin, N. and J. Hicks, *Future medical applications of single-cell sequencing in cancer.* Genome Med, 2011. **3**(5): p. 31.
140. Schilsky, R.L., *Personalized medicine in oncology: the future is now.* Nat Rev Drug Discov, 2010. **9**(5): p. 363-6.
141. McDermott, U. and J. Settleman, *Personalized cancer therapy with selective kinase inhibitors: an emerging paradigm in medical oncology.* J Clin Oncol, 2009. **27**(33): p. 5650-9.
142. Dienstmann, R., J. Rodon, and J. Tabernero, *Biomarker-driven patient selection for early clinical trials.* Curr Opin Oncol, 2013. **25**(3): p. 305-12.
143. Gonzalez de Castro, D., et al., *Personalized cancer medicine: molecular diagnostics, predictive biomarkers, and drug resistance.* Clin Pharmacol Ther, 2013. **93**(3): p. 252-9.
144. Gerlinger, M., et al., *Intratumor heterogeneity and branched evolution revealed by multiregion sequencing.* N Engl J Med, 2012. **366**(10): p. 883-92.
145. Pantel, K. and C. Alix-Panabieres, *Real-time liquid biopsy in cancer patients: fact or fiction?* Cancer Res, 2013. **73**(21): p. 6384-8.
146. Alix-Panabieres, C. and K. Pantel, *Challenges in circulating tumour cell research.* Nat Rev Cancer, 2014. **14**(9): p. 623-31.
147. Ashworth, T., *A case of cancer in which cells similar tho those in the tumors were seen in the blood after death.* Aust Med J., 1869. **14**: p. 146-149.
148. Wittekind, C. and M. Neid, *Cancer invasion and metastasis.* Oncology, 2005. **69 Suppl 1**: p. 14-6.
149. Pantel, K. and R.H. Brakenhoff, *Dissecting the metastatic cascade.* Nat Rev Cancer, 2004. **4**(6): p. 448-56.
150. Kim, M.Y., et al., *Tumor self-seeding by circulating cancer cells.* Cell, 2009. **139**(7): p. 1315-26.
151. Alix-Panabieres, C., H. Schwarzenbach, and K. Pantel, *Circulating tumor cells and circulating tumor DNA.* Annu Rev Med, 2012. **63**: p. 199-215.
152. Krebs, M.G., et al., *Molecular analysis of circulating tumour cells-biology and biomarkers.* Nat Rev Clin Oncol, 2014. **11**(3): p. 129-44.
153. Joosse, S.A., T.M. Gorges, and K. Pantel, *Biology, detection, and clinical implications of circulating tumor cells.* EMBO Mol Med, 2014.

154. Vona, G., et al., *Isolation by size of epithelial tumor cells : a new method for the immunomorphological and molecular characterization of circulatingtumor cells.* Am J Pathol, 2000. **156**(1): p. 57-63.
155. Muller, V., et al., *Circulating tumor cells in breast cancer: correlation to bone marrow micrometastases, heterogeneous response to systemic therapy and low proliferative activity.* Clin Cancer Res, 2005. **11**(10): p. 3678-85.
156. Gascoyne, P.R., et al., *Isolation of rare cells from cell mixtures by dielectrophoresis.* Electrophoresis, 2009. **30**(8): p. 1388-98.
157. Allard, W.J., et al., *Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases.* Clin Cancer Res, 2004. **10**(20): p. 6897-904.
158. Riethdorf, S., et al., *Detection of circulating tumor cells in peripheral blood of patients with metastatic breast cancer: a validation study of the CellSearch system.* Clin Cancer Res, 2007. **13**(3): p. 920-8.
159. Talasaz, A.H., et al., *Isolating highly enriched populations of circulating epithelial cells and other rare cells from blood using a magnetic sweeper device.* Proc Natl Acad Sci U S A, 2009. **106**(10): p. 3970-5.
160. Saucedo-Zeni, N., et al., *A novel method for the in vivo isolation of circulating tumor cells from peripheral blood of cancer patients using a functionalized and structured medical wire.* Int J Oncol, 2012. **41**(4): p. 1241-50.
161. Nagrath, S., et al., *Isolation of rare circulating tumour cells in cancer patients by microchip technology.* Nature, 2007. **450**(7173): p. 1235-9.
162. Stott, S.L., et al., *Isolation of circulating tumor cells using a microvortex-generating herringbone-chip.* Proc Natl Acad Sci U S A, 2010. **107**(43): p. 18392-7.
163. Ozkumur, E., et al., *Inertial focusing for tumor antigen-dependent and -independent sorting of rare circulating tumor cells.* Sci Transl Med, 2013. **5**(179): p. 179ra47.
164. Saliba, A.E., et al., *Microfluidic sorting and multimodal typing of cancer cells in self-assembled magnetic arrays.* Proc Natl Acad Sci U S A, 2010. **107**(33): p. 14524-9.
165. Zieglschmid, V., et al., *Combination of immunomagnetic enrichment with multiplex RT-PCR analysis for the detection of disseminated tumor cells.* Anticancer Res, 2005. **25**(3A): p. 1803-10.
166. Hou, H.W., et al., *Isolation and retrieval of circulating tumor cells using centrifugal forces.* Sci Rep, 2013. **3**: p. 1259.
167. Riahi, R., et al., *A novel microchannel-based device to capture and analyze circulating tumor cells (CTCs) of breast cancer.* Int J Oncol, 2014. **44**(6): p. 1870-8.
168. Lu, J., et al., *Isolation of circulating epithelial and tumor progenitor cells with an invasive phenotype from breast cancer patients.* Int J Cancer, 2010. **126**(3): p. 669-83.
169. Lopez-Riquelme, N., et al., *Imaging cytometry for counting circulating tumor cells: comparative analysis of the CellSearch vs ImageStream systems.* APMIS, 2013.
170. Marrinucci, D., et al., *Fluid biopsy in patients with metastatic prostate, pancreatic and breast cancers.* Phys Biol, 2012. **9**(1): p. 016003.
171. Alix-Panabieres, C. and K. Pantel, *Circulating tumor cells: liquid biopsy of cancer.* Clin Chem, 2013. **59**(1): p. 110-8.
172. Cristofanilli, M., et al., *Circulating tumor cells, disease progression, and survival in metastatic breast cancer.* N Engl J Med, 2004. **351**(8): p. 781-91.
173. de Bono, J.S., et al., *Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer.* Clin Cancer Res, 2008. **14**(19): p. 6302-9.

174. Cohen, S.J., et al., *Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer.* J Clin Oncol, 2008. **26**(19): p. 3213-21.
175. Harb, W., et al., *Mutational Analysis of Circulating Tumor Cells Using a Novel Microfluidic Collection Device and qPCR Assay.* Transl Oncol, 2013. **6**(5): p. 528-38.
176. Parkinson, D.R., et al., *Considerations in the development of circulating tumor cell technology for clinical use.* J Transl Med, 2012. **10**: p. 138.
177. Thiery, J.P., *Epithelial-mesenchymal transitions in development and pathologies.* Curr Opin Cell Biol, 2003. **15**(6): p. 740-6.
178. Gorges, T.M., et al., *Circulating tumour cells escape from EpCAM-based detection due to epithelial-to-mesenchymal transition.* BMC Cancer, 2012. **12**: p. 178.
179. Hou, J.M., et al., *Circulating tumor cells as a window on metastasis biology in lung cancer.* Am J Pathol, 2011. **178**(3): p. 989-96.
180. Liu, Z., et al., *Negative enrichment by immunomagnetic nanobeads for unbiased characterization of circulating tumor cells from peripheral blood of cancer patients.* J Transl Med, 2011. **9**: p. 70.
181. Deneve, E., et al., *Capture of viable circulating tumor cells in the liver of colorectal cancer patients.* Clin Chem, 2013. **59**(9): p. 1384-92.
182. Chen, Y.Y. and G.B. Xu, *Effect of circulating tumor cells combined with negative enrichment and CD45-FISH identification in diagnosis, therapy monitoring and prognosis of primary lung cancer.* Med Oncol, 2014. **31**(12): p. 240.
183. Stathopoulou, A., et al., *Real-time quantification of CK-19 mRNA-positive cells in peripheral blood of breast cancer patients using the lightcycler system.* Clin Cancer Res, 2003. **9**(14): p. 5145-51.
184. Chen, C.L., et al., *Single-cell analysis of circulating tumor cells identifies cumulative expression patterns of EMT-related genes in metastatic prostate cancer.* Prostate, 2013. **73**(8): p. 813-26.
185. Fabbri, F., et al., *Detection and recovery of circulating colon cancer cells using a dielectrophoresis-based device: KRAS mutation status in pure CTCs.* Cancer Lett, 2013. **335**(1): p. 225-31.
186. Peeters, D.J., et al., *Semiautomated isolation and molecular characterisation of single or highly purified tumour cells from CellSearch enriched blood samples using dielectrophoretic cell sorting.* Br J Cancer, 2013. **108**(6): p. 1358-67.
187. Krebs, M.G., et al., *Evaluation and prognostic significance of circulating tumor cells in patients with non-small-cell lung cancer.* J Clin Oncol, 2011. **29**(12): p. 1556-63.
188. Muinelo-Romay, L., et al., *Evaluation of Circulating Tumor Cells and Related Events as Prognostic Factors and Surrogate Biomarkers in Advanced NSCLC Patients Receiving First-Line Systemic Treatment.* Cancers (Basel), 2014. **6**(1): p. 153-65.
189. Hiltermann, T.J., et al., *Circulating tumor cells in small-cell lung cancer: a predictive and prognostic factor.* Ann Oncol, 2012. **23**(11): p. 2937-42.
190. Tanaka, F., et al., *Circulating tumor cell as a diagnostic marker in primary lung cancer.* Clin Cancer Res, 2009. **15**(22): p. 6980-6.
191. Ilie, M., et al., *"Sentinel" circulating tumor cells allow early diagnosis of lung cancer in patients with chronic obstructive pulmonary disease.* PLoS One, 2014. **9**(10): p. e111597.
192. Carlsson, A., et al., *Circulating tumor microemboli diagnostics for patients with non-small-cell lung cancer.* J Thorac Oncol, 2014. **9**(8): p. 1111-9.
193. Rosell, R., et al., *Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell*

lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol, 2012. **13**(3): p. 239-46.
194. Solomon, B.J., et al., *First-line crizotinib versus chemotherapy in ALK-positive lung cancer.* N Engl J Med, 2014. **371**(23): p. 2167-77.
195. Pailler, E., et al., *Detection of circulating tumor cells harboring a unique ALK rearrangement in ALK-positive non-small-cell lung cancer.* J Clin Oncol, 2013. **31**(18): p. 2273-81.
196. Ilie, M., et al., *ALK-gene rearrangement: a comparative analysis on circulating tumour cells and tumour tissue from patients with lung adenocarcinoma.* Ann Oncol, 2012. **23**(11): p. 2907-13.
197. Liu, Y., et al., *Circulating tumor cells in HER2-positive metastatic breast cancer patients: a valuable prognostic and predictive biomarker.* BMC Cancer, 2013. **13**: p. 202.
198. Hirose, T., et al., *Relationship of circulating tumor cells to the effectiveness of cytotoxic chemotherapy in patients with metastatic non-small-cell lung cancer.* Oncol Res, 2012. **20**(2-3): p. 131-7.
199. Punnoose, E.A., et al., *Evaluation of circulating tumor cells and circulating tumor DNA in non-small cell lung cancer: association with clinical endpoints in a phase II clinical trial of pertuzumab and erlotinib.* Clin Cancer Res, 2012. **18**(8): p. 2391-401.
200. Yu, K.H., et al., *Pharmacogenomic Modeling of Circulating Tumor and Invasive Cells for Prediction of Chemotherapy Response and Resistance in Pancreatic Cancer.* Clin Cancer Res, 2014.
201. Yu, M., et al., *Cancer therapy. Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility.* Science, 2014. **345**(6193): p. 216-20.
202. Gazzaniga, P., et al., *Chemosensitivity profile assay of circulating cancer cells: prognostic and predictive value in epithelial tumors.* Int J Cancer, 2010. **126**(10): p. 2437-47.
203. Gradilone, A., et al., *Circulating tumor cells (CTCs) in metastatic breast cancer (MBC): prognosis, drug resistance and phenotypic characterization.* Ann Oncol, 2011. **22**(1): p. 86-92.
204. Hou, J.M., et al., *Clinical significance and molecular characteristics of circulating tumor cells and circulating tumor microemboli in patients with small-cell lung cancer.* J Clin Oncol, 2012. **30**(5): p. 525-32.
205. Naito, T., et al., *Prognostic impact of circulating tumor cells in patients with small cell lung cancer.* J Thorac Oncol, 2012. **7**(3): p. 512-9.
206. Normanno, N., et al., *Prognostic value of circulating tumor cells' reduction in patients with extensive small-cell lung cancer.* Lung Cancer, 2014.
207. Shi, W.L., et al., *CK-19 mRNA-positive cells in peripheral blood predict treatment efficacy and survival in small-cell lung cancer patients.* Med Oncol, 2013. **30**(4): p. 755.
208. Fu, L., et al., *Circulating Tumor Cells Correlate with Recurrence in Stage III Small-cell Lung Cancer after Systemic Chemoradiotherapy and Prophylactic Cranial Irradiation.* Jpn J Clin Oncol, 2014.
209. Navin, N., et al., *Tumour evolution inferred by single-cell sequencing.* Nature, 2011. **472**(7341): p. 90-4.
210. Swanton, C., *Intratumor heterogeneity: evolution through space and time.* Cancer Res, 2012. **72**(19): p. 4875-82.
211. Van Loo, P. and T. Voet, *Single cell analysis of cancer genomes.* Curr Opin Genet Dev, 2014. **24**: p. 82-91.
212. Dean, F.B., et al., *Comprehensive human genome amplification using multiple displacement amplification.* Proc Natl Acad Sci U S A, 2002. **99**(8): p. 5261-6.
213. Spits, C., et al., *Whole-genome multiple displacement amplification from single cells.* Nat Protoc, 2006. **1**(4): p. 1965-70.

214. Zhang, L., et al., *Whole genome amplification from a single cell: implications for genetic analysis.* Proc Natl Acad Sci U S A, 1992. **89**(13): p. 5847-51.
215. Klein, C.A., et al., *Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells.* Proc Natl Acad Sci U S A, 1999. **96**(8): p. 4494-9.
216. Zong, C., et al., *Genome-wide detection of single-nucleotide and copy-number variations of a single human cell.* Science, 2012. **338**(6114): p. 1622-6.
217. Tang, F., et al., *RNA-Seq analysis to capture the transcriptome landscape of a single cell.* Nat Protoc, 2010. **5**(3): p. 516-35.
218. Ramskold, D., et al., *Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells.* Nat Biotechnol, 2012.
219. Pan, X., et al., *Two methods for full-length RNA sequencing for low quantities of cells and single cells.* Proc Natl Acad Sci U S A, 2013. **110**(2): p. 594-9.
220. Hashimshony, T., et al., *CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification.* Cell Rep, 2012. **2**(3): p. 666-73.
221. Magbanua, M.J. and J.W. Park, *Advances in genomic characterization of circulating tumor cells.* Cancer Metastasis Rev, 2014.
222. Smirnov, D.A., et al., *Global gene expression profiling of circulating tumor cells.* Cancer Res, 2005. **65**(12): p. 4993-7.
223. Magbanua, M.J., et al., *Isolation and genomic analysis of circulating tumor cells from castration resistant metastatic prostate cancer.* BMC Cancer, 2012. **12**: p. 78.
224. Magbanua, M.J., et al., *Genomic profiling of isolated circulating tumor cells from metastatic breast cancer patients.* Cancer Res, 2013. **73**(1): p. 30-40.
225. Gasch, C., et al., *Heterogeneity of Epidermal Growth Factor Receptor Status and Mutations of KRAS/PIK3CA in Circulating Tumor Cells of Patients with Colorectal Cancer.* Clin Chem, 2013. **59**(1): p. 252-60.
226. Heitzer, E., et al., *Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing.* Cancer Res, 2013. **73**(10): p. 2965-75.
227. Steinert, G., et al., *Immune escape and survival mechanisms in circulating tumor cells of colorectal cancer.* Cancer Res, 2014. **74**(6): p. 1694-704.
228. Swennenhuis, J.F., et al., *Efficiency of whole genome amplification of single circulating tumor cells enriched by CellSearch and sorted by FACS.* Genome Med, 2013. **5**(11): p. 106.
229. Ni, X., et al., *Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients.* Proc Natl Acad Sci U S A, 2013. **110**(52): p. 21083-8.
230. Klein, C.A., et al., *Combined transcriptome and genome analysis of single micrometastatic cells.* Nat Biotechnol, 2002. **20**(4): p. 387-92.
231. Ulmer, A., et al., *Immunomagnetic enrichment, genomic characterization, and prognostic impact of circulating melanoma cells.* Clin Cancer Res, 2004. **10**(2): p. 531-7.
232. Hannemann, J., et al., *Quantitative high-resolution genomic analysis of single cancer cells.* PLoS One, 2011. **6**(11): p. e26362.
233. Zhao, L., et al., *High-purity prostate circulating tumor cell isolation by a polymer nanofiber-embedded microchip for whole exome sequencing.* Adv Mater, 2013. **25**(21): p. 2897-902.
234. Deng, G., et al., *Single cell mutational analysis of PIK3CA in circulating tumor cells and metastases in breast cancer reveals heterogeneity, discordance, and mutation persistence in cultured disseminated tumor cells from bone marrow.* BMC Cancer, 2014. **14**: p. 456.
235. Dago, A.E., et al., *Rapid phenotypic and genomic change in response to therapeutic pressure in prostate cancer inferred by high content analysis of single circulating tumor cells.* PLoS One, 2014. **9**(8): p. e101777.

236. Neves, R.P., et al., *Genomic high-resolution profiling of single CKpos/CD45neg flow-sorting purified circulating tumor cells from patients with metastatic breast cancer.* Clin Chem, 2014. **60**(10): p. 1290-7.
237. Polzer, B., et al., *Molecular profiling of single circulating tumor cells with diagnostic intention.* EMBO Mol Med, 2014. **6**(11): p. 1371-86.
238. Powell, A.A., et al., *Single cell profiling of circulating tumor cells: transcriptional heterogeneity and diversity from breast cancer cell lines.* PLoS One, 2012. **7**(5): p. e33788.
239. Cann, G.M., et al., *mRNA-Seq of single prostate cancer circulating tumor cells reveals recapitulation of gene expression and pathways found in prostate cancer.* PLoS One, 2012. **7**(11): p. e49144.
240. Lohr, J.G., et al., *Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer.* Nat Biotechnol, 2014.
241. Ting, D.T., et al., *Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells.* Cell Rep, 2014. **8**(6): p. 1905-18.
242. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.
243. DePristo, M.A., et al., *A framework for variation discovery and genotyping using next-generation DNA sequencing data.* Nat Genet, 2011. **43**(5): p. 491-8.
244. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): p. e164.
245. Boeva, V., et al., *Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization.* Bioinformatics, 2011. **27**(2): p. 268-9.
246. Krzywinski, M., et al., *Circos: an information aesthetic for comparative genomics.* Genome Res, 2009. **19**(9): p. 1639-45.
247. Saeed, A.I., et al., *TM4: a free, open-source system for microarray data management and analysis.* Biotechniques, 2003. **34**(2): p. 374-8.
248. Yates, T., M.J. Okoniewski, and C.J. Miller, *X:Map: annotation and visualization of genome structure for Affymetrix exon array analysis.* Nucleic Acids Res, 2008. **36**(Database issue): p. D780-6.
249. Smyth, G.K., *Linear models and empirical bayes methods for assessing differential expression in microarray experiments.* Stat Appl Genet Mol Biol, 2004. **3**: p. Article3.
250. Benjamini, Y. and Y. Hochberg, *Controlling the False Discover Rate: A Practical and Powerful Approach to Multiple Testing.* Journal of Royal Statistical Society: Series B (Statistical Methodology), 1995. **57**(1).
251. Voelkerding, K.V., S.A. Dames, and J.D. Durtschi, *Next-generation sequencing: from basic research to diagnostics.* Clin Chem, 2009. **55**(4): p. 641-58.
252. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
253. Xuan, J., et al., *Next-generation sequencing in the clinic: promises and challenges.* Cancer Lett, 2013. **340**(2): p. 284-95.
254. Johansen Taber, K.A., B.D. Dickinson, and M. Wilson, *The promise and challenges of next-generation genome sequencing for clinical care.* JAMA Intern Med, 2014. **174**(2): p. 275-80.
255. Campbell, P.J., et al., *Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing.* Nat Genet, 2008. **40**(6): p. 722-9.
256. Bamshad, M.J., et al., *Exome sequencing as a tool for Mendelian disease gene discovery.* Nat Rev Genet, 2011. **12**(11): p. 745-55.

257. Ahn, J.W., et al., *Identification of somatic mutations in EGFR/KRAS/ALK-negative lung adenocarcinoma in never-smokers.* Genome Med, 2014. **6**(2): p. 18.
258. Yin, S., et al., *Exome sequencing identifies frequent mutation of MLL2 in non-small cell lung carcinoma from Chinese patients.* Sci Rep, 2014. **4**: p. 6036.
259. Kim, Y., et al., *Integrative and comparative genomic analysis of lung squamous cell carcinomas in East Asian patients.* J Clin Oncol, 2014. **32**(2): p. 121-8.
260. Zheng, C.X., et al., *Whole-exome sequencing to identify novel somatic mutations in squamous cell lung cancers.* Int J Oncol, 2013. **43**(3): p. 755-64.
261. Nguyen, K.S., J.W. Neal, and H. Wakelee, *Review of the current targeted therapies for non-small-cell lung cancer.* World J Clin Oncol, 2014. **5**(4): p. 576-87.
262. Mamanova, L., et al., *Target-enrichment strategies for next-generation sequencing.* Nat Methods, 2010. **7**(2): p. 111-8.
263. Mertes, F., et al., *Targeted enrichment of genomic DNA regions for next-generation sequencing.* Brief Funct Genomics, 2011. **10**(6): p. 374-86.
264. Cline, J., J.C. Braman, and H.H. Hogrefe, *PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases.* Nucleic Acids Res, 1996. **24**(18): p. 3546-51.
265. Davenport, R.D., *Diagnostic value of crush artifact in cytologic specimens. Occurrence in small cell carcinoma of the lung.* Acta Cytol, 1990. **34**(4): p. 502-4.
266. Ashman, J.N., et al., *Chromosomal alterations in small cell lung cancer revealed by multicolour fluorescence in situ hybridization.* Int J Cancer, 2002. **102**(3): p. 230-6.
267. Ocak, S., et al., *DNA copy number aberrations in small-cell lung cancer reveal activation of the focal adhesion pathway.* Oncogene, 2010. **29**(48): p. 6331-42.
268. Hasmats, J., et al., *Validation of whole genome amplification for analysis of the p53 tumor suppressor gene in limited amounts of tumor samples.* Biochem Biophys Res Commun, 2012. **425**(2): p. 379-83.
269. Taub, M.A., H. Corrada Bravo, and R.A. Irizarry, *Overcoming bias and systematic errors in next generation sequencing data.* Genome Med, 2010. **2**(12): p. 87.
270. Holt, R.A. and S.J. Jones, *The new paradigm of flow cell sequencing.* Genome Res, 2008. **18**(6): p. 839-46.
271. Forshew, T., et al., *Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA.* Sci Transl Med, 2012. **4**(136): p. 136ra68.
272. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
273. Baccelli, I., et al., *Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay.* Nat Biotechnol, 2013. **31**(6): p. 539-44.
274. De Witt Hamer, P.C., et al., *The genomic profile of human malignant glioma is altered early in primary cell culture and preserved in spheroids.* Oncogene, 2008. **27**(14): p. 2091-6.
275. Sasai, K., et al., *Shh pathway activity is down-regulated in cultured medulloblastoma cells: implications for preclinical studies.* Cancer Res, 2006. **66**(8): p. 4215-22.
276. Hudson, A.M., et al., *Discrepancies in cancer genomic sequencing highlight opportunities for driver mutation discovery.* Cancer Res, 2014. **74**(22): p. 6390-6.
277. Fumagalli, D., et al., *A rapid, sensitive, reproducible and cost-effective method for mutation profiling of colon cancer and metastatic lymph nodes.* BMC Cancer, 2010. **10**: p. 101.

278. Camidge, D.R., W. Pao, and L.V. Sequist, *Acquired resistance to TKIs in solid tumours: learning from lung cancer.* Nat Rev Clin Oncol, 2014. **11**(8): p. 473-81.
279. Liu, C.X., et al., *Genomic organization of a new candidate tumor suppressor gene, LRP1B.* Genomics, 2000. **69**(2): p. 271-4.
280. Liu, C.X., et al., *The putative tumor suppressor LRP1B, a novel member of the low density lipoprotein (LDL) receptor family, exhibits both overlapping and distinct properties with the LDL receptor-related protein.* J Biol Chem, 2001. **276**(31): p. 28889-96.
281. Jardim, D.L., et al., *Analysis of 1,115 Patients Tested for MET Amplification and Therapy Response in the MD Anderson Phase I Clinic.* Clin Cancer Res, 2014. **20**(24): p. 6336-45.
282. Salgia, R., *Fibroblast growth factor signaling and inhibition in non-small cell lung cancer and their role in squamous cell tumors.* Cancer Med, 2014. **3**(3): p. 681-92.
283. Swanton, C., A. Futreal, and T. Eisen, *Her2-targeted therapies in non-small cell lung cancer.* Clin Cancer Res, 2006. **12**(14 Pt 2): p. 4377s-4383s.
284. Prim, N., et al., *Germ-line exon 21 EGFR mutations, V843I and P848L, in nonsmall cell lung cancer patients.* Eur Respir Rev, 2014. **23**(133): p. 390-2.
285. Sekulic, A., et al., *Efficacy and safety of vismodegib in advanced basal-cell carcinoma.* N Engl J Med, 2012. **366**(23): p. 2171-9.
286. Tang, J.Y., et al., *Inhibiting the hedgehog pathway in patients with the basal-cell nevus syndrome.* N Engl J Med, 2012. **366**(23): p. 2180-8.
287. Hou, J.M., et al., *Evaluation of circulating tumor cells and serological cell death biomarkers in small cell lung cancer patients undergoing chemotherapy.* Am J Pathol, 2009. **175**(2): p. 808-16.
288. Dehainault, C., et al., *A deep intronic mutation in the RB1 gene leads to intronic sequence exonisation.* Eur J Hum Genet, 2007. **15**(4): p. 473-7.
289. Ni, S., et al., *Down expression of LRP1B promotes cell migration via RhoA/Cdc42 pathway and actin cytoskeleton remodeling in renal cell cancer.* Cancer Sci, 2013. **104**(7): p. 817-25.
290. Ding, L., et al., *Somatic mutations affect key pathways in lung adenocarcinoma.* Nature, 2008. **455**(7216): p. 1069-75.
291. Craig, D.W., et al., *Genome and transcriptome sequencing in prospective metastatic triple-negative breast cancer uncovers therapeutic vulnerabilities.* Mol Cancer Ther, 2013. **12**(1): p. 104-16.
292. Cowin, P.A., et al., *LRP1B deletion in high-grade serous ovarian cancers is associated with acquired chemotherapy resistance to liposomal doxorubicin.* Cancer Res, 2012. **72**(16): p. 4060-73.
293. Liu, C.X., et al., *LRP-DIT, a putative endocytic receptor gene, is frequently inactivated in non-small cell lung cancer cell lines.* Cancer Res, 2000. **60**(7): p. 1961-7.
294. Sparks, C.A. and D.A. Guertin, *Targeting mTOR: prospects for mTOR complex 2 inhibitors in cancer therapy.* Oncogene, 2010. **29**(26): p. 3733-44.
295. Lewis, A.D., et al., *Amplification and increased expression of alpha class glutathione S-transferase-encoding genes associated with resistance to nitrogen mustards.* Proc Natl Acad Sci U S A, 1988. **85**(22): p. 8511-5.
296. Sharma, A., et al., *Glutathione S-transferases as antioxidant enzymes: small cell lung cancer (H69) cells transfected with hGSTA1 resist doxorubicin-induced apoptosis.* Arch Biochem Biophys, 2006. **452**(2): p. 165-73.
297. Singhal, S.S., et al., *Depletion of RLIP76 sensitizes lung cancer cells to doxorubicin.* Biochem Pharmacol, 2005. **70**(3): p. 481-8.
298. Vatsyayan, R., et al., *Role of RLIP76 in doxorubicin resistance in lung cancer.* Int J Oncol, 2009. **34**(6): p. 1505-11.

299. Sanghani, S.P., et al., *Carboxylesterases expressed in human colon tumor tissue and their role in CPT-11 hydrolysis.* Clin Cancer Res, 2003. **9**(13): p. 4983-91.
300. Sanghani, S.P., et al., *Hydrolysis of irinotecan and its oxidative metabolites, 7-ethyl-10-[4-N-(5-aminopentanoic acid)-1-piperidino] carbonyloxycamptothecin and 7-ethyl-10-[4-(1-piperidino)-1-amino]-carbonyloxycamptothecin, by human carboxylesterases CES1A1, CES2, and a newly expressed carboxylesterase isoenzyme, CES3.* Drug Metab Dispos, 2004. **32**(5): p. 505-11.
301. van Roy, F., *Beyond E-cadherin: roles of other cadherin superfamily members in cancer.* Nat Rev Cancer, 2014. **14**(2): p. 121-34.
302. Buenestado, A., et al., *The role of adenosine receptors in regulating production of tumour necrosis factor-alpha and chemokines by human lung macrophages.* Br J Pharmacol, 2010. **159**(6): p. 1304-11.
303. Brognard, J., et al., *PHLPP and a second isoform, PHLPP2, differentially attenuate the amplitude of Akt signaling by regulating distinct Akt isoforms.* Mol Cell, 2007. **25**(6): p. 917-31.
304. Rao, E., et al., *The miRNA-17 approximately 92 cluster mediates chemoresistance and enhances tumor growth in mantle cell lymphoma via PI3K/AKT pathway activation.* Leukemia, 2012. **26**(5): p. 1064-72.
305. Liao, W.T., et al., *microRNA-224 promotes cell proliferation and tumor growth in human colorectal cancer by repressing PHLPP1 and PHLPP2.* Clin Cancer Res, 2013. **19**(17): p. 4662-72.
306. Wen, Y.A., et al., *Downregulation of PHLPP expression contributes to hypoxia-induced resistance to chemotherapy in colon cancer cells.* Mol Cell Biol, 2013. **33**(22): p. 4594-605.
307. Hu, H., et al., *The overexpression of hypomethylated miR-663 induces chemotherapy resistance in human breast cancer cells by targeting heparin sulfate proteoglycan 2 (HSPG2).* J Biol Chem, 2013. **288**(16): p. 10973-85.
308. Ding, K., et al., *Inhibition of apoptosis by downregulation of hBex1, a novel mechanism, contributes to the chemoresistance of Bcr/Abl+ leukemic cells.* Carcinogenesis, 2009. **30**(1): p. 35-42.
309. Narayan, G., et al., *Promoter methylation-mediated inactivation of PCDH10 in acute lymphoblastic leukemia contributes to chemotherapy resistance.* Genes Chromosomes Cancer, 2011. **50**(12): p. 1043-53.
310. Rabbani, S.A., A. Arakelian, and R. Farookhi, *LRP5 knockdown: effect on prostate cancer invasion growth and skeletal metastasis in vitro and in vivo.* Cancer Med, 2013. **2**(5): p. 625-35.
311. Nickolich, M., et al., *Clinical trial design in small cell lung cancer: surrogate end points and statistical evolution.* Clin Lung Cancer, 2014. **15**(3): p. 207-12.
312. Pollack, J.R., et al., *Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors.* Proc Natl Acad Sci U S A, 2002. **99**(20): p. 12963-8.
313. Hyman, E., et al., *Impact of DNA amplification on gene expression patterns in breast cancer.* Cancer Res, 2002. **62**(21): p. 6240-5.
314. Tsafrir, D., et al., *Relationship of gene expression and chromosomal abnormalities in colorectal cancer.* Cancer Res, 2006. **66**(4): p. 2129-37.
315. Platzer, P., et al., *Silence of chromosomal amplifications in colon cancer.* Cancer Res, 2002. **62**(4): p. 1134-8.
316. Knudson, A.G., Jr., *Mutation and cancer: statistical study of retinoblastoma.* Proc Natl Acad Sci U S A, 1971. **68**(4): p. 820-3.
317. Solimini, N.L., et al., *Recurrent hemizygous deletions in cancers may optimize proliferative potential.* Science, 2012. **337**(6090): p. 104-9.
318. Xue, W., et al., *A cluster of cooperating tumor-suppressor gene candidates in chromosomal deletions.* Proc Natl Acad Sci U S A, 2012. **109**(21): p. 8212-7.

319. Narayan, G., et al., *Protocadherin PCDH10, involved in tumor progression, is a frequent and early target of promoter hypermethylation in cervical cancer.* Genes Chromosomes Cancer, 2009. **48**(11): p. 983-92.
320. Yu, Y., et al., *MicroRNA-21 induces stemness by downregulating transforming growth factor beta receptor 2 (TGFbetaR2) in colon cancer cells.* Carcinogenesis, 2012. **33**(1): p. 68-76.
321. Stordal, B., et al., *Resistance to paclitaxel in a cisplatin-resistant ovarian cancer cell line is mediated by P-glycoprotein.* PLoS One, 2012. **7**(7): p. e40717.
322. Kishimoto, S., et al., *Role of Na+, K+-ATPase alpha1 subunit in the intracellular accumulation of cisplatin.* Cancer Chemother Pharmacol, 2006. **57**(1): p. 84-90.
323. Wood, L.D., et al., *The genomic landscapes of human breast and colorectal cancers.* Science, 2007. **318**(5853): p. 1108-13.
324. *Scientific Framework for Small Cell Lung Cancer (SCLC)*, 2014, National Cancer Institute.
325. *Cancer Research UK's Strategy 2009 - 2014*, 2008, Cancer Research UK.
326. *Beating Cancer Sooner: Our Research Strategy*, 2014, Cancer Research UK.
327. Rothwell, D.G., et al., *Evaluation and validation of a robust single cell RNA-amplification protocol through transcriptional profiling of enriched lung cancer initiating cells.* BMC Genomics, 2014. **15**(1): p. 1129.
328. Ignatiadis, M. and S.J. Dawson, *Circulating Tumor Cells and Circulating tumor DNA for precision medicine: Dream or reality?* Ann Oncol, 2014.
329. Taniguchi, K., et al., *Quantitative detection of EGFR mutations in circulating tumor DNA derived from lung adenocarcinomas.* Clin Cancer Res, 2011. **17**(24): p. 7808-15.
330. Murtaza, M., et al., *Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA.* Nature, 2013. **497**(7447): p. 108-12.
331. Yao, X., et al., *Tumor cells are dislodged into the pulmonary vein during lobectomy.* J Thorac Cardiovasc Surg, 2014. **148**(6): p. 3224-31 e1-5.
332. National Collaborating Centre for Cancer, U.K., *The Diagnosis and Treatment of Lung Cancer (Update)*, 2011: Cardiff, UK.

Appendix 1: CNA and WES Read Numbers

Table A1.1 Number of uniquely mapped reads used to generate CNA data in section 4A.4

Sample	Genome Integrity Index	Number of mapped reads
CDX1L	NA	1,464,585,606
CDX1R	NA	1,456,181,156
CDX2	NA	1,287,175,558
CDX3L	NA	2,790,256
CDX3R	NA	8,624,078
CDX4	NA	7,533,784
CDX1LA	4	3,025,068
CDX1RA	4	3,554,070
CDX2A	4	2,454,638
CDX3LA	4	2,668,108
CDX3RA	4	2,014,106
CDX4A	4	2,177,222
P2CTC1	1	2,757,448
P2CTC2	4	2,005,038
P2CTC3	1	1,777,978
P2CTC4	2	2,253,328
P2CTC5	4	498,760
P2CTC6	2	1,773,938
P2CTCP1	4	1,767,944
P2CTCP2	4	2,065,400
P2WBC	4	1,694,098
P4CTC1	4	1,605,976
P4CTC2	4	3,667,812
P4CTCP	4	1,095,910
P4WBC	4	3,278,072
23R11-sWBC	3	2,056,516
97R14-sWBC	4	1,301,998
B12-sWBC	4	1,148,410
B13-sWBC	4	2,005,434
23R13-pWBC	4	1,335,576
68R15-pWBC	4	1,139,980
B4-sCTC	3	2,465,300
A1-sCTC	4	973,500
A5-sCTC	2	1,405,322
A8-sCTC	1	2,502,424
A9-pCTC	4	4,856,342
B9-pCTC	4	981,160

Table A1.2 Genomic integrity index score and the number of uniquely mapped reads used to generate CNA data for section 4.B.4.

Sample Name	Genome Integrity Index	Number of uniquely mapped reads
P1-SCTC1	3	1202260
P1-SCTC2	2	467912
P1-SCTC3	3	557536
P1-SCTC4	3	534936
P1-SCTC5	3	998774
P1-SCTC6	3	553854
P1-SCTC7	4	638296
P1-SCTC8	1	477334
P1-pCTC1	4	935916
P1-pWBC1	4	797972
P1-pWBC2	4	904174
P2-sCTC1	2	249720
P2-sCTC2	1	647026
P2-sCTC3	1	182992
P2-sCTC4	3	409296
P2-sCTC5	3	387142
P2-sCTC6	3	803522
P2-sCTC7	0	n/a
P2-sCTC8	1	640990
P2-pCTC1	1	456470
P2-pCTC2	4	792410
P2-pWBC1	4	755172
P3-sCTC1	1	384760
P3-sCTC2	2	256782
P3-sCTC3	4	617630
P3-sCTC4	4	1007072
P3-sCTC5	2	350046
P3-sCTC6	4	927490
P3-sCTC7	3	392678
P3-sCTC8	3	663346
P3-pCTC1	1	581890
P3-pCTC2	3	294384
P3-pWBC1	3	826336
P3-pWBC2	4	735122
P4-sCTC1	1	2757448
P4-sCTC2	4	2005038
P4-sCTC3	1	1777978
P4-sCTC4	2	2253328
P4-sCTC5	4	498760

Table A1.2 continued

Sample Name	Genome Integrity Index	Number of uniquely mapped reads
P4-sCTC6	2	1773938
P4-pCTC1	4	1767944
P4-pCTC2	4	2065400
P4-pWBC1	4	1694098
P5-sCTC1	4	973500
P5-sCTC2	4	711146
P5-sCTC3	2	904642
P5-sCTC4	3	1405322
P5-sCTC5	4	756740
P5-sCTC6	4	1055564
P5-sCTC7	1	2502424
P5-pCTC1	4	4856342
P5-pCTC2	4	n/a
P5-pWBC1	4	924806
P5-pWBC2	4	934730
P6-sCTC1	4	556908
P6-sCTC2	4	1160754
P6-sCTC3	3	237386
P6-sCTC4	4	561424
P6-sCTC5	4	878814
P6-sCTC6	4	455578
P6-sCTC7	3	479288
P6-sCTC8	4	363714
P6-pCTC1	4	388744
P6-sWBC1	3	650314
P6-sWBC2	3	124296
P6-pWBC1	4	1335576

Table A1.3 Genomic integrity index score of cells and the number of uniquely mapped reads used to generate CNA data for chapter 5.

Sample Name	Genome Integrity Index	Uniquely Mapped Reads
P1-T1-sCTC1	3	6,773,084
P1-T1-sCTC2	2	2,663,396
P1-T1-sCTC3	3	3,316,540
P1-T1-sCTC4	3	2,824,412
P1-T1-sCTC5	3	5,259,736
P1-T1-sCTC6	3	3,155,294
P1-T1-sCTC7	4	3,602,000
P1-T1-sCTC8	1	2,746,038
P1-T1-pCTC1	4	5,013,912
P1-T1-pWBC1	4	4,677,220
P1-T1-pWBC2	4	5,242,892
P2-T1-sCTC1	1	700,602
P2-T1-sCTC2	2	463,716
P2-T1-sCTC3	4	1,124,658
P2-T1-sCTC4	4	1,819,560
P2-T1-sCTC5	3	624,200
P2-T1-sCTC6	4	1,719,744
P2-T1-sCTC7	3	723,150
P2-T1-sCTC8	3	1,224,088
P2-T1-pCTC1	1	1,065,696
P2-T1-pCTC2	2	524,542
P2-T1-pWBC2	2	1,496,768
P2-T1-pWBC2	4	1,310,676
P3-T1-sCTC1	1	2,757,448
P3-T1-sCTC2	4	2,005,038
P3-T1-sCTC3	1	1,777,978
P3-T1-sCTC4	2	2,253,328
P3-T1-sCTC5	4	498,760
P3-T1-sCTC6	2	1,773,938
P3-T1-pCTC1	4	1,767,944
P3-T1-pCTC2	4	2,065,400
P3-T1-pWBC1	4	1,694,098
P4-T1-sCTC1	4	5,669,166
P4-T1-sCTC2	4	4,219,410
P4-T1-sCTC3	2	5,027,320
P4-T1-sCTC4	2	5,094,792
P4-T1-sCTC5	4	7,871,908
P4-T1-sCTC6	4	6,141,118
P4-T1-sCTC7	1	15,327,248
P4-T1-pCTC1	4	28,426,024
P4-T1-pWBC1	4	5,619,602

Table A1.3 continued

New Sample Name	Genome Integrity Index	Uniquely Mapped Reads
P4-T1-pWBC2	4	5,446,820
P4-T2-sCTC1	3	1,762,706
P4-T2-sCTC2	3	4,431,338
P4-T2-sCTC3	3	4,589,116
P4-T2-sCTC4	3	14,875,030
P4-T2-sCTC5	3	3,639,842
P4-T2-sCTC6	2	475,586
P4-T2-sCTC7	3	4,577,138
P4-T2-sCTC8	3	2,139,854
P4-T2-pCTC1	4	5,521,742
P4-T2-sWBC1	4	3,976,830
P4-T2-sWBC2	4	1,660,194
P4-T2-sWBC3	4	4,548,416
P4-T2-sWBC4	4	12,076,506
P4-T2-pWBC1	4	2,924,650
P5-T1-sCTC1	4	2,758,934
P5-T1-sCTC2	4	5,564,320
P5-T1-sCTC3	3	1,086,420
P5-T1-sCTC4	4	2,647,782
P5-T1-sCTC5	4	4,160,806
P5-T1-sCTC6	4	2,056,556
P5-T1-sCTC7	4	2,233,050
P5-T1-sCTC8	4	1,687,546
P5-T1-pCTC1	4	1,757,934
P5-T1-sWBC1	3	3,027,784
P5-T1-sWBC2	3	560,134
P5-T1-pWBC1	4	6,383,086
P5-T1-gDNA1	4	5,174,508
P5-T2-sCTC1	4	2,240,002
P5-T2-sCTC2	2	1,403,810
P5-T2-sCTC3	4	2,304,690
P5-T2-sCTC4	2	4,885,862
P5-T2-sCTC5	4	4,024,656
P5-T2-sCTC6	4	6,075,950
P5-T2-sCTC7	3	6,298,868
P5-T2-pCTC1	3	6,361,002
P5-T2-pCTC2	4	8,239,850
P5-T2-sWBC1	4	3,510,392
P5-T2-pWBC1	4	7,067,618
P5-T3-sCTC1	4	1,158,610
P5-T3-sCTC2	4	1,475,528
P5-T3-sCTC3	4	1,313,592

Table A1.3 continued

New Sample Name	Genome Integrity Index	Uniquely Mapped Reads
P5-T3-sCTC4	4	1,303,060
P5-T3-sCTC5	4	1,703,574
P5-T3-sCTC6	4	1,968,704
P5-T3-sCTC7	4	1,191,684
P5-T3-sCTC8	4	1,520,160
P5-T3-pCTC1	4	1,732,500
P5-T3-pCTC2	4	1,465,614
P5-T3-sWBC1	4	1,304,766
P5-T3-pWBC1	4	1,314,258
P6-T1-sCTC1	1	1,887,940
P6-T1-sCTC2	4	1,749,798
P6-T1-sCTC3	4	6,779,308
P6-T1-sCTC4	4	1,959,726
P6-T1-sCTC5	4	3,713,322
P6-T1-sCTC6	4	4,022,006
P6-T1-sCTC7	4	9,874,186
P6-T1-sCTC8	2	5,566,562
P6-T1-pCTC1	4	8,230,750
P6-T1-sWBC1	4	7,702,048
P6-T1-sWBC2	4	1,811,766
P6-T1-sWBC3	4	9,217,696
P6-T2-sCTC1	1	3,278,974
P6-T2-sCTC2	2	1,271,140
P6-T2-sCTC3	3	925,280
P6-T2-sCTC4	4	3,844,480
P6-T2-sCTC5	1	2,795,608
P6-T2-sCTC6	3	4,106,644
P6-T2-sCTC7	4	7,562,246
P6-T2-sCTC8	2	3,021,836
P6-T2-sWBC1	3	611,328
P6-T2-sWBC2	3	1,624,722
P6-T2-pWBC1	4	21,383,848
P6-T1-gDNA1	4	15,125,682
P7-T1-sCTC1	4	1,605,976
P7-T1-sCTC2	4	3,667,812
P7-T1-pCTC1	4	1,095,910
P7-T1-pWBC1	4	3,278,072

Table A1.3 continued

New Sample Name	Genome Integrity Index	Uniquely Mapped Reads
P8-T1-sCTC1	3	3,514,960
P8-T1-sCTC2	4	5,511,374
P8-T1-sCTC3	4	4,231,534
P8-T1-sCTC4	4	6,152,236
P8-T1-sWBC1	4	4,425,820
P8-T1-sWBC2	4	8,004,616
P8-T1-sWBC3	4	3,314,454
P8-T1-sWBC4	4	4,542,326
P8-T1-pWBC1	4	6,214,474
P8-T1-gDNA1	4	6,108,462
P8-T2-sCTC1	2	1,423,366
P8-T2-sCTC2	1	382,308
P8-T2-sCTC3	2	1,963,376
P8-T2-sCTC4	2	1,080,060
P8-T2-sCTC5	2	1,525,854
P8-T2-sCTC6	2	1,163,860
P8-T2-sCTC7	1	297,598
P8-T2-pCTC1	4	5,260,990
P8-T2-pCTC2	3	2,652,148
P8-T2-pCTC3	3	2,709,198
P8-T2-sWBC1	2	1,506,982
P8-T2-pWBC1	4	3,647,350
P8-T3-sCTC1	4	6,468,566
P8-T3-sCTC2	4	5,223,922
P8-T3-sCTC3	3	7,132,320
P8-T3-sCTC4	1	1,637,022
P8-T3-sCTC5	4	6,244,950
P8-T3-sCTC6	4	5,517,126
P8-T3-sCTC7	4	5,507,134
P8-T3-sCTC8	3	10,749,918
P8-T3-sCTC9	3	4,951,988
P8-T3-pCTC1	1	5,912,816
P8-T3-pCTC2	4	6,916,582
P8-T3-sWBC1	4	6,735,604
P8-T3-sWBC2	4	4,262,056
P8-T3-pWBC1	4	14,433,680
P9-T1-sCTC1	4	1,670,050
P9-T1-sCTC2	3	941,802
P9-T1-sCTC3	3	1,221,208
P9-T1-sCTC4	4	2,063,192
P9-T1-sCTC5	4	1,033,730
P9-T1-sCTC6	4	1,980,576

Table A1.3 continued

New Sample Name	Genome Integrity Index	Uniquely Mapped Reads
P9-T1-sCTC7	4	829,916
P9-T1-pCTC1	4	2,087,812
P9-T1-sWBC1	4	1,207,334
P9-T1-sWBC2	4	515,538
P9-T1-pWBC1	4	1,621,510
P10-T1-sCTC1	2	1,744,092
P10-T1-sCTC2	4	1,250,014
P10-T1-sCTC3	4	1,982,686
P10-T1-sCTC4	2	1,923,520
P10-T1-sCTC5	4	2,062,154
P10-T1-sCTC6	3	2,191,946
P10-T1-sCTC7	3	1,799,438
P10-T1-sCTC8	3	2,085,044
P10-T1-sCTC9	4	2,017,746
P10-T1-sCTC10	3	2,146,900
P10-T1-pCTC1	4	2,018,366
P10-T1-pCTC2	4	1,913,792
P10-T1-sWBC1	4	1,570,354
P10-T1-pWBC1	4	1,722,690
P10-T1-gDNA1	4	1,997,812
P10-T2-sCTC1	3	1,936,538
P10-T2-sCTC2	4	1,905,692
P10-T2-sCTC3	1	1,961,768
P10-T2-sCTC4	3	1,847,574
P10-T2-sCTC5	2	1,654,302
P10-T2-sCTC6	4	1,880,000
P10-T2-sCTC7	4	1,744,204
P10-T2-sCTC8	4	1,683,196
P10-T2-sCTC9	2	1,458,860
P10-T2-sCTC10	4	2,098,670
P10-T2-pCTC1	4	1,611,840
P10-T2-pCTC2	4	1,540,482
P10-T2-sWBC1	4	1,204,480
P10-T2-sWBC2	4	1,938,610
P10-T2-pWBC1	4	2,064,334

Table A1.4 Percentage of reads covering the exonic regions and the number of uniquely mapped reads for the WES samples used in chapter 5.

Sample Name	Percentage of Reads Covering Exon	Uniquely Mapped Reads	Percentage of Exome with 5X coverage
P1-T1-pCTC1	37.81%	12720010	41.65%
P1-T1-pWBC2	54.18%	10423968	45.22%
P2-T1-pCTC2	53.85%	2823702	20.23%
P2-T1-pWBC2	38.24%	9484610	32.35%
P3-T1-pCTC1	43.92%	14088692	47.23%
P3-T1-pWBC1	43.36%	17414388	55.79%
P4-T1-sCTC1	25.49%	35549626	54.94%
P4-T1-sCTC4	64.93%	28157266	57.28%
P4-T1-pCTC1	58.38%	49857442	67.4%
P4-T1-pWBC1	62.01%	69226810	72.69%
P4-T2-sCTC2	23.62%	30528442	45.63%
P4-T2-sCTC7	29.43%	21862840	41.24%
P4-T2-pCTC1	55.17%	43266144	61.72%
P4-T2-sWBC2	27.45%	6601392	21.51%
P5-T1-pCTC1	58.77%	16005486	55.61%
P5-T1-pWBC1	62.00%	50101538	69.47%
P5-T2-pCTC1	59.74%	48035450	58.52%
P5-T3-pCTC2	60.84%	44301684	58.77%
P6-T1-pCTC1	29.29%	16688778	45.64%
P6-T2-sCTC4	34.26%	3886812	17.25%
P6-T2-sCTC7	33.18%	9390054	26.57%
P6-T2-pWBC1	38.82%	17615478	46.64%
P7-T1-pCTC1	60.11%	44041308	68.45%
P7-T1-pWBC1	62.27%	58432864	69.15%
P8-T1-sCTC3	29.58%	14159620	39.16%
P8-T2-pCTC3	30.29%	8524716	26.82%
P8-T2-pWBC1	25.87%	25729488	26.95%
P8-T3-pCTC1	31.24%	6785774	26.57%
P9-T1-pCTC1	27.54%	16131692	39%
P9-T1-pWBC1	24.85%	16312134	40.47%
P10-T1-pCTC1	24.49%	32088922	52.38%
P10-T2-pCTC1	33.35%	21879708	49.39%
P10-T2-pCTC2	29.09%	41513390	58.39%
P10-T2-pWBC1	37.27%	44216846	64.32%

Appendix 2: TAm-Seq Primers

Table A2.1 Table of primers used in TAm-Seq analysis in chapter 4

Gene	Chromosome	Primer Name	Amplicon Start Position	Amplicon End Position	Forward Primer Sequence	Reverse Primer Sequence
BRAF	7	BRAF_AA_HS	140453108	140453256	TCATAATGCTTGCTCTGATAGGA	CTGATGGGACCCACTCCAT
CTNNB1	3	CTNNB1_D0008_001	41266060	41266179	AGCGGCTGTTAGTCACTGG	GTATCCACATCCTCTTCCTCAGG
EGFR	7	EGFR_D0008_008	55241589	55241709	GTCTCTGTGTTCTTGTCCCCC	GCCCAGCACTTTGATCTTTTTGAA
EGFR	7	EGFR_D0008_009	55241658	55241746	TCTCTTGAGGATCTTGAAGGAAAC	GGGACCTTACCTTATACACCGT
EGFR	7	EGFR_E00001601336_1	55248902	55249111	GCGTCTTCACCTGGAAGGG	CCGGACATAGTCCAGGAGG
EGFR	7	EGFR_E00001601336_2	55249005	55249213	GCGTGGACAACCCCCAC	GGCTCCTTATCTCCCCTCC
EGFR	7	EGFR_E00001681524_1	55259352	55259542	GGATGCAGAGCTTCTTCCCA	TTCTCTTCCGCACCCAG
EGFR	7	EGFR_E00001681524_2	55259395	55259591	GGTCTTCTCTGTTTCAGGGCAT	GCTGACCTAAAGCCACCTCC
EGFR	7	EGFR_E00001779947_1	55266362	55266571	TGTTCATTCATGATCCCACTGC	CCACCAGTCACTCACACTTG
EGFR	7	EGFR_E00001779947_2	55266461	55266641	TCCCTGCCAGCGAGAT	AGGGATGCAAAGGCCTCA
EGFR	7	EGFR_E00001801208_1	55268806	55268987	CCCCTGCTCCTATAGCCAA	ATGAGGTACTCGTCGGCATC
EGFR	7	EGFR_E00001801208_2	55268921	55269101	ACTTCTACCGTGCCCTGA	GTTCAAATGAGTAGACACAGCTT
EGFR	7	EGFR_E00001773562_1	55269336	55269516	TACCCTCCATGAGGCACAC	GGAGAGCTGTAAATTCTGGCTT
EGFR	7	EGFR_D0008_005	55227942	55228041	GCTATGCAAATACAATAAACTGGAAA	GGTGACTTACTGCAGCTGTTTT
KRAS	12	KRAS_D0008_001	25378518	25378613	CAGATCTGTATTTATTTCAGTGTTACTTACCT	CAGGCTCAGGACTTAGCAAGAA
KRAS	12	KRAS_D0008_002	25380216	25380337	ACACAAAGAAAGCCCTCCCC	AGGAAGCAAGTAGTAATTGATGGAGA
KRAS	12	KRAS_D0009_002	25398211	25398303	TATTGTTGGATCATATTCGTCCACAAAA	CTTGTGGTAGTTGGAGCTGGT
KRAS	12	KRAS_D0008_003	25398246	25398337	GAATTAGCTGTATCGTCAAGGCAC	ATTATAAGGCCTGCTGAAAATGACTG

Table A.2.1 continued

Gene	Chromosome	Primer Name	Amplicon Start Position	Amplicon End Position	Forward Primer Sequence	Reverse Primer Sequence
NRAS	1	NRAS_D0008_001	115256498	115256617	GTATTGGTCTCTCATGGCACTGT	TACCCTCCACACCCCCAG
NRAS	1	NRAS_D0008_002	115258699	115258807	AAGTGGTTCTGGATTAGCTGGATT	TTCCAACAGGTTCTTGCTGGT
PIK3CA	3	PIK3CA_D0008_001	178936028	178936135	AGGGAAAATGACAAAGAACAGC	TTTTAGCACTTACCTGTGACTCCA
PIK3CA	3	PIK3CA_D0018_001	178952038	178952128	TGAGCAAGAGGCTTTGGAGT	TGTGTGGAAGATCCAATCCA
PTEN	10	PTEN_D0023_001	89624175	89624374	GCAGCTTCTGCCATCTCTCT	CATCCGTCTACTCCCACGTT
PTEN	10	PTEN_D0023_002	89653778	89653958	TCAGATATTTATCCAAACATTATTGC	TCTTTTTCTGTGGCTTAGAAATCTT
PTEN	10	PTEN_D0023_005	89692956	89693048	AGGCACAAGAGGCCCTAGAT	TCCAGGAAGAGGAAAGGAAAA
PTEN	10	PTEN_D0023_006	89711954	89712056	TGGCACTGTTGTTTCACAAG	TGTTCCAATACATGGAAGGATG
PTEN	10	PTEN_D0023_007	89717605	89717813	TGCAGATCCTCAGTTTGTGG	TTTTGGATATTTCTCCCAATGAA
TP53	17	TP53_E00001757276_1	7572850	7573030	GACCCAAAACCCAAAATGGC	TCCCTGCTTCTGTCTCCTAC
TP53	17	TP53_E00001728015_1	7573859	7574054	GGAATCCTATGGCTTTCCAACC	CCCCCTCCTCTGTTGCTG
TP53	17	TP53_E00001789298_1	7576786	7576983	AGAAAACGGCATTTTGAGTGT	AAGGGTGCAGTTATGCCTCA
TP53	17	TP53_E00001789298_2	7576908	7577075	CTGGTGTTGTTGGGCAGT	ATCTCCGCAAGAAAGGGGAG
TP53	17	TP53_E00001789298_3	7577003	7577187	TGTCCTGCTTGCTTACCTCG	GCCTCTTGCTTCTCTTTTCCT
TP53	17	TP53_E00001665758_1	7577432	7577631	GGGGTCAGAGGCAAGCAG	CTTGGGCCTGTGTTATCTCC
TP53	17	TP53_D0018_01	7578140	7578274	AACCCCTCCTCCCAGAGAC	AGCATCTTATCCGAGTGGAAGG
TP53	17	TP53_E00001255919_3	7578229	7578406	TCCAAATACTCCACACGCAAA	GCTGCCCCCACCATGAG
TP53	17	TP53_E00001255919_5	7578361	7578525	AGCTGCTCACCATCGCTA	CCAACTGGCCAAGACCT
TP53	17	TP53_E00001255919_6	7578425	7578594	TGTGCTGTGACTGCTTGTAG	TGCCCTGACTTTCAACTCTGT
TP53	17	TP53_E00001612188_1	7579260	7579421	ATACGGCCAGGCATTGAAGT	CCTCCTGGCCCCTGTC
TP53	17	TP53_E00001612188_2	7579359	7579520	GGAAACCGTAGCTGCCCTG	AAGACCCAGGTCCAGATGAA

Table A.2.1 continued

Gene	Chromosome	Primer Name	Amplicon Start Position	Amplicon End Position	Forward Primer Sequence	Reverse Primer Sequence
TP53	17	EXP0116_TP53_E4	7579479	7579626	CAGCCTCTGGCATTCTGG	CCTGGTCCTCTGACTGCTCT
TP53	17	EXP0116_TP53_E3	7579557	7579754	TCAAATCATCCATTGCTTGG	CCATGGGACTGACTTTCTGC
TP53	17	TP53_E00001596491_1	7579758	7579940	TTTCGCTTCCCACAGGTCTC	CAGCCAGACTGCCTTCCG

Appendix 3: Hierarchical Clustering of CTC CNA Data from Chapter 5

Hierarchical clustering of CNA data generated from the each of the 10 patients CTCs from chapter 5 is included in this appendix.

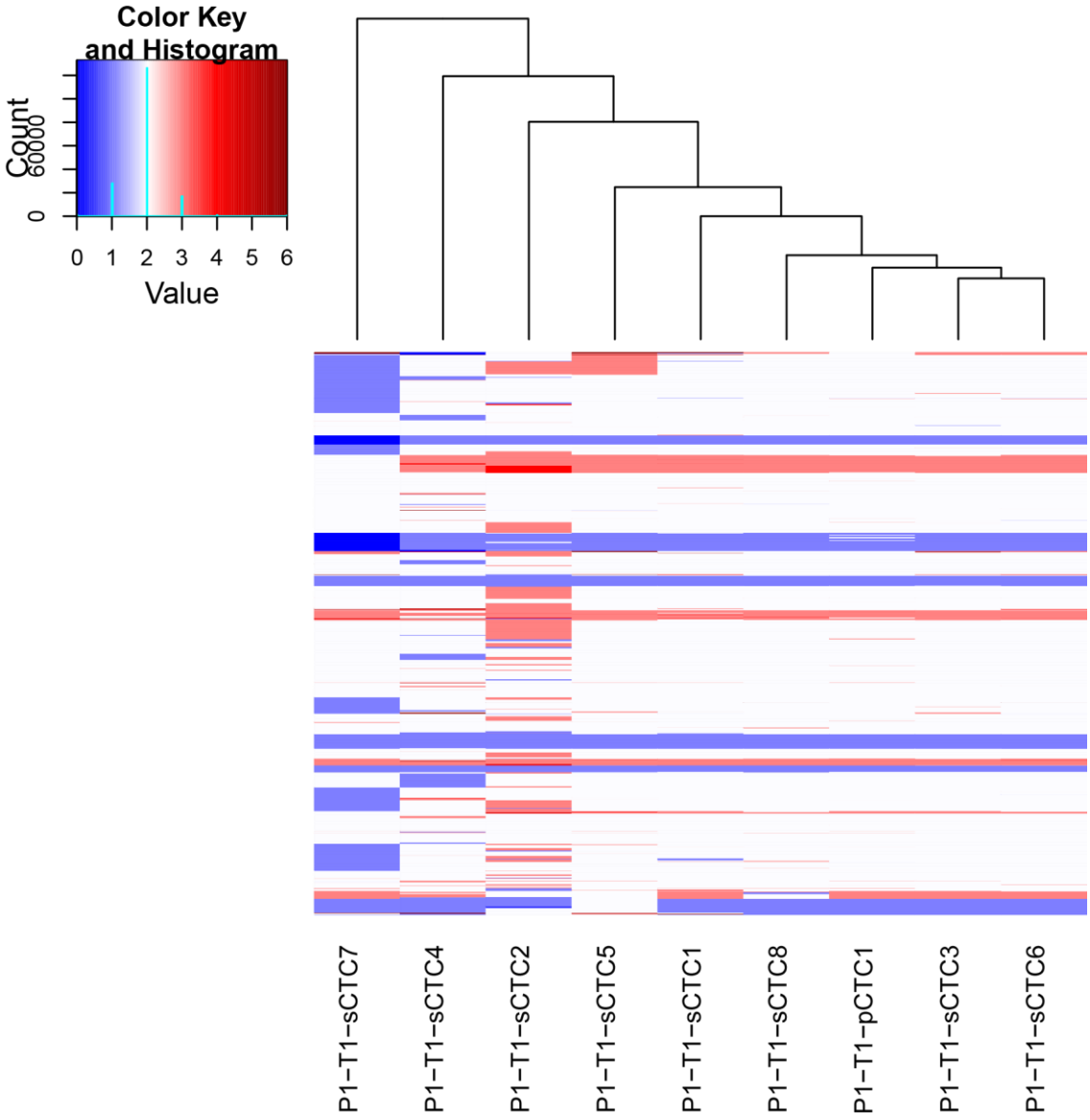


Figure A3.1 Hierarchical clustering of copy number data generated from CTCs isolated from patient 1. CTCs were isolated from a blood sample from patient 1 taken at baseline, prior to the patient receiving chemotherapy (T1). The samples were either single CTCs (sCTC) or pools of CTCs (pCTC). The amplified CTCs were subject to WGA and DNA libraries then created. The DNA libraries were WGS and the data used to calculate the copy number of 19336 protein-coding genes. Hierarchical clustering of the CNA data from the CTCs was then carried out. Regions of gain of copy number are red, regions of loss are blue and regions with no change are white.

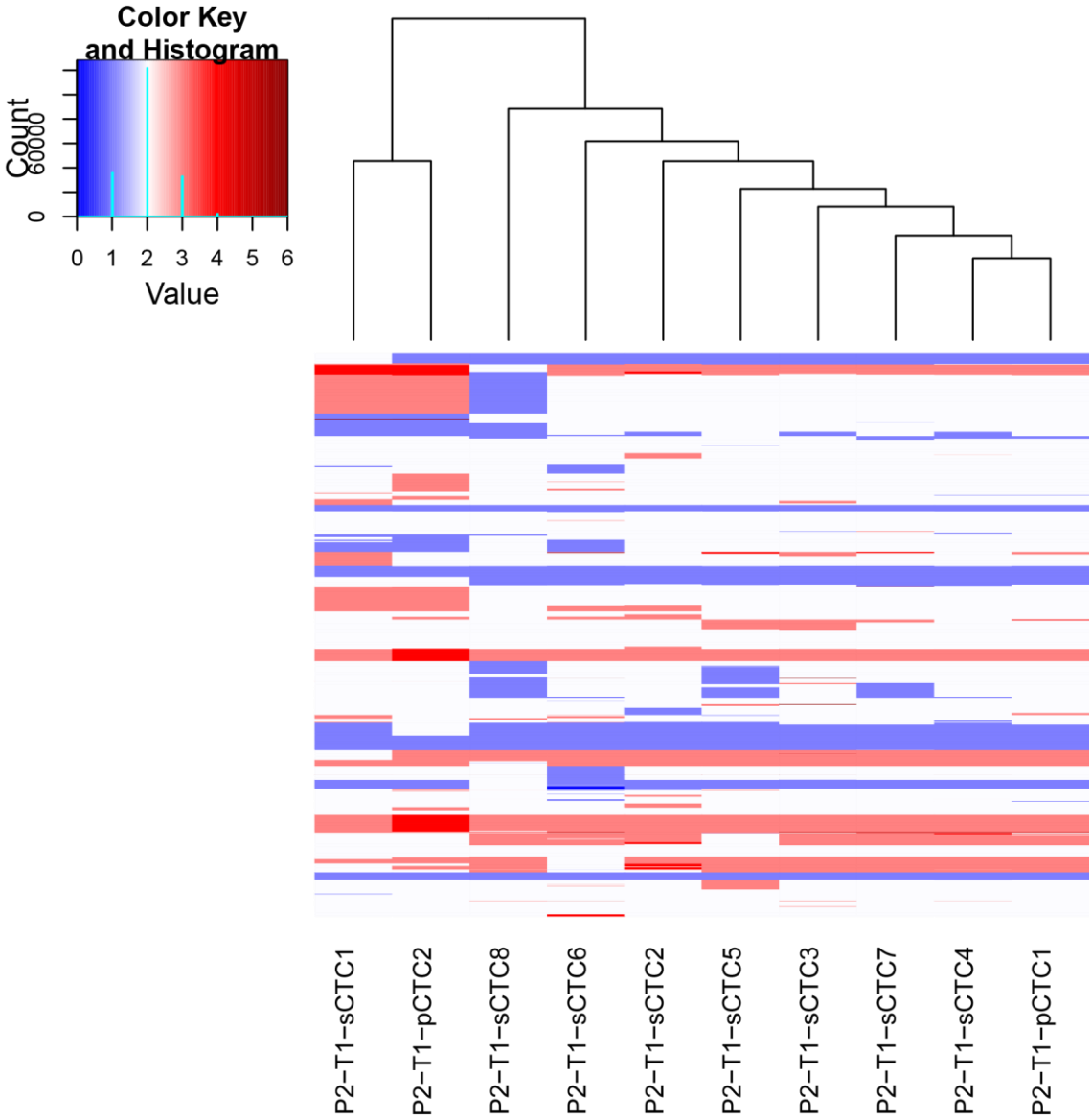


Figure A3.2 Hierarchical clustering of copy number data generated from CTCs isolated from patient 2. CTCs were isolated from a blood sample from patient 2 taken at baseline, prior to the patient receiving chemotherapy (T1). The samples were either single CTCs (sCTC) or pools of CTCs (pCTC). The amplified CTCs were subject to WGA and DNA libraries then created. The DNA libraries were WGS and the data used to calculate the copy number of 19336 protein-coding genes. Hierarchical clustering of the CNA data from the CTCs was then carried out. Regions of gain of copy number are red, regions of loss are blue and regions with no change are white.

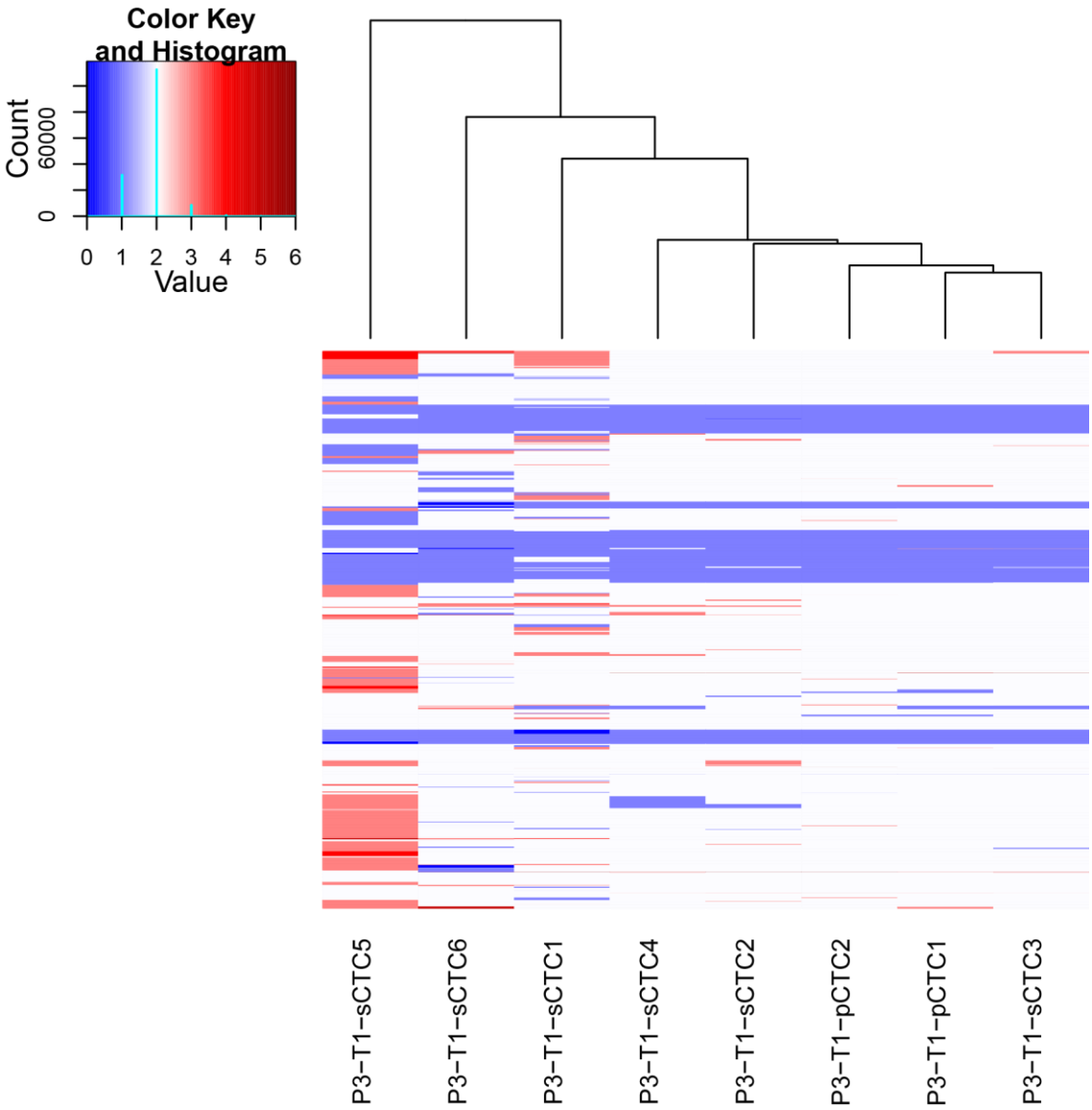


Figure A3.3 Hierarchical clustering of copy number data generated from CTCs isolated from patient 3. CTCs were isolated from a blood sample from patient 3 taken at baseline, prior to the patient receiving chemotherapy (T1). The samples were either single CTCs (sCTC) or pools of CTCs (pCTC). The amplified CTCs were subject to WGA and DNA libraries then created. The DNA libraries were WGS and the data used to calculate the copy number of 19336 protein-coding genes. Hierarchical clustering of the CNA data from the CTCs was then carried out. Regions of gain of copy number are red, regions of loss are blue and regions with no change are white.

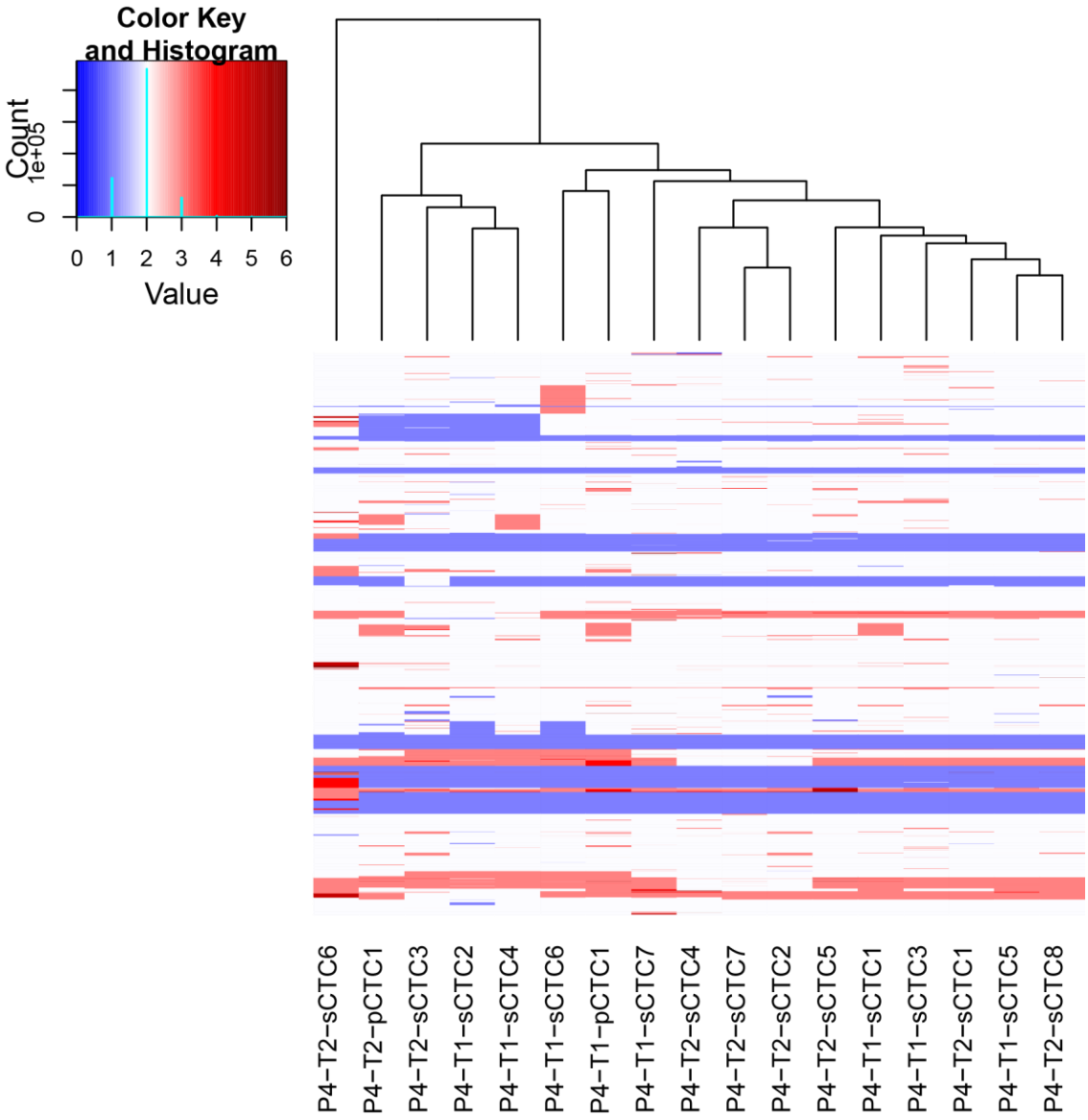


Figure A3.4 Hierarchical clustering of copy number data generated from CTCs isolated from patient 4. CTCs were isolated from blood samples from patient 4 taken at baseline, prior to the patient receiving chemotherapy (T1) and when the patient relapsed after first-line chemotherapy (T2). The samples were either single CTCs (sCTC) or pools of CTCs (pCTC). The amplified CTCs were subject to WGA and DNA libraries then created. The DNA libraries were WGS and the data used to calculate the copy number of 19336 protein-coding genes. Hierarchical clustering of the CNA data from the CTCs was then carried out. Regions of gain of copy number are red, regions of loss are blue and regions with no change are white.

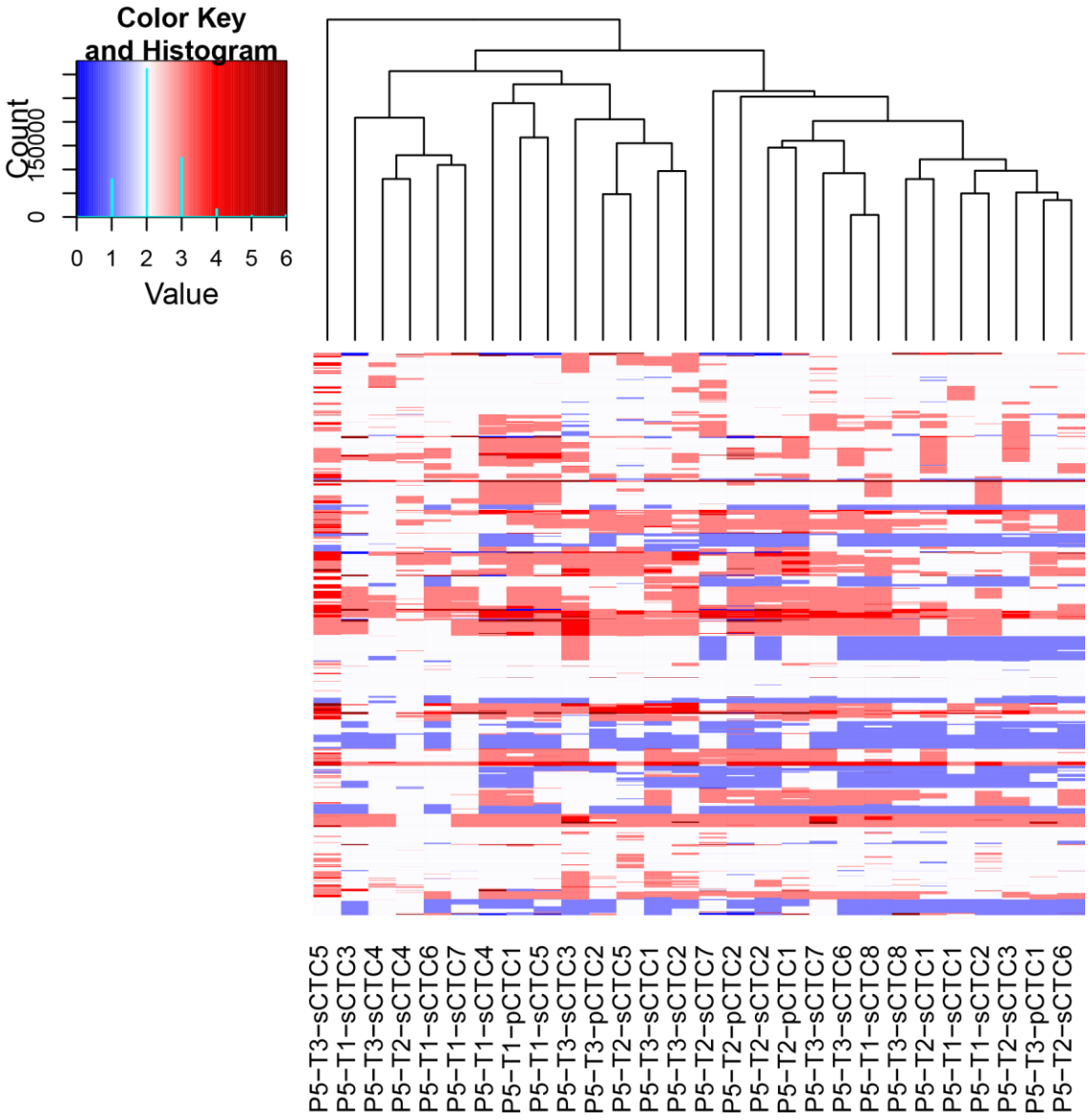


Figure A3.5 Hierarchical clustering of copy number data generated from CTCs isolated from patient 5. CTCs were isolated from blood samples from patient 5 taken at baseline, prior to the patient receiving chemotherapy (T1), when the patient relapsed after first-line chemotherapy (T2) and when the patient relapsed after second-line chemotherapy (T3). The samples were either single CTCs (sCTC) or pools of CTCs (pCTC). The amplified CTCs were subject to WGA and DNA libraries then created. The DNA libraries were WGS and the data used to calculate the copy number of 19336 protein-coding genes. Hierarchical clustering of the CNA data from the CTCs was then carried out. Regions of gain of copy number are red, regions of loss are blue and regions with no change are white.

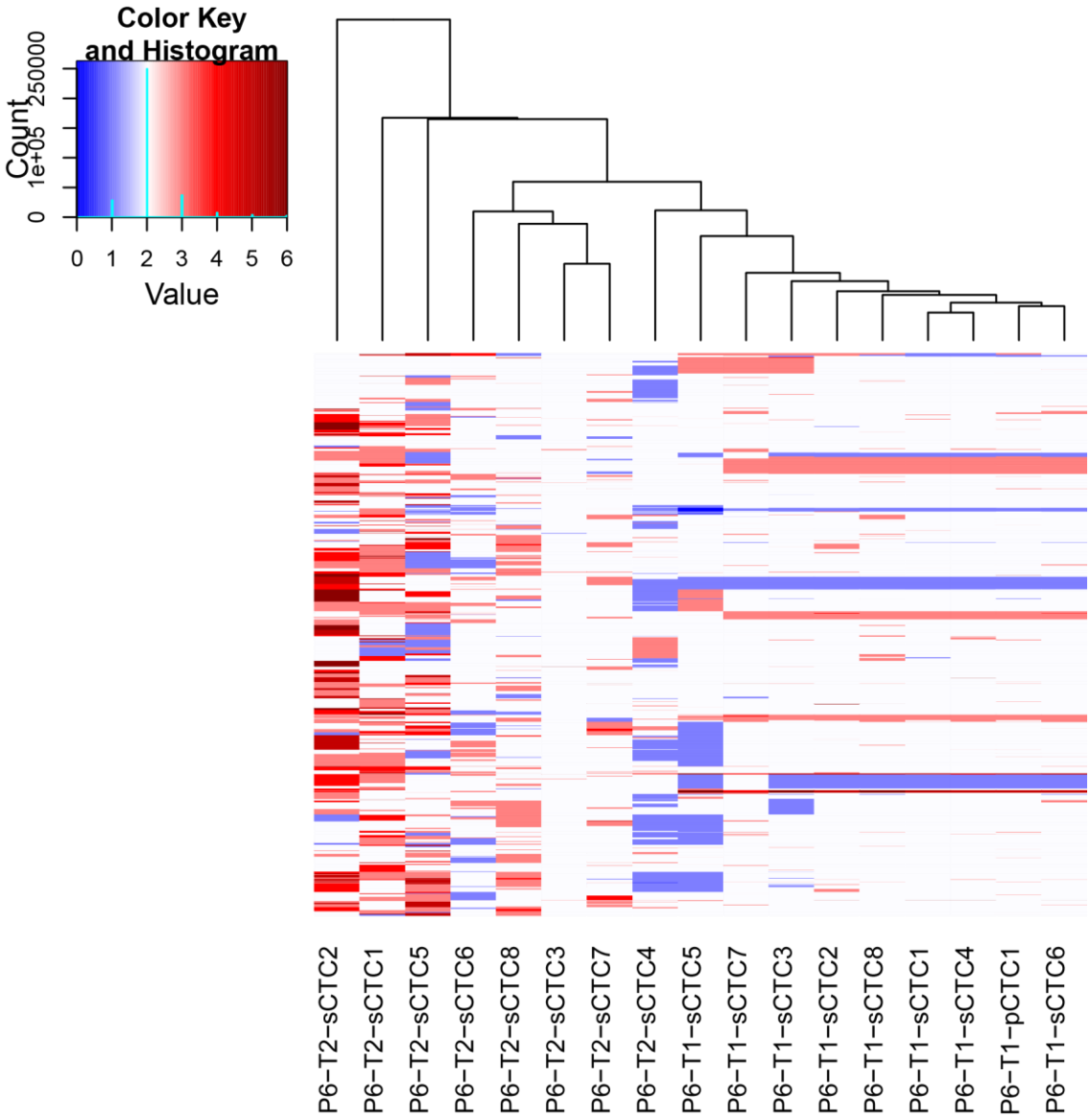


Figure A3.6 Hierarchical clustering of copy number data generated from CTCs isolated from patient 6. CTCs were isolated from blood samples from patient 6 taken at baseline, prior to the patient receiving chemotherapy (T1) and when the patient relapsed after first-line chemotherapy (T2). The samples were either single CTCs (sCTC) or pools of CTCs (pCTC). The amplified CTCs were subject to WGA and DNA libraries then created. The DNA libraries were WGS and the data used to calculate the copy number of 19336 protein-coding genes. Hierarchical clustering of the CNA data from the CTCs was then carried out. Regions of gain of copy number are red, regions of loss are blue and regions with no change are white.

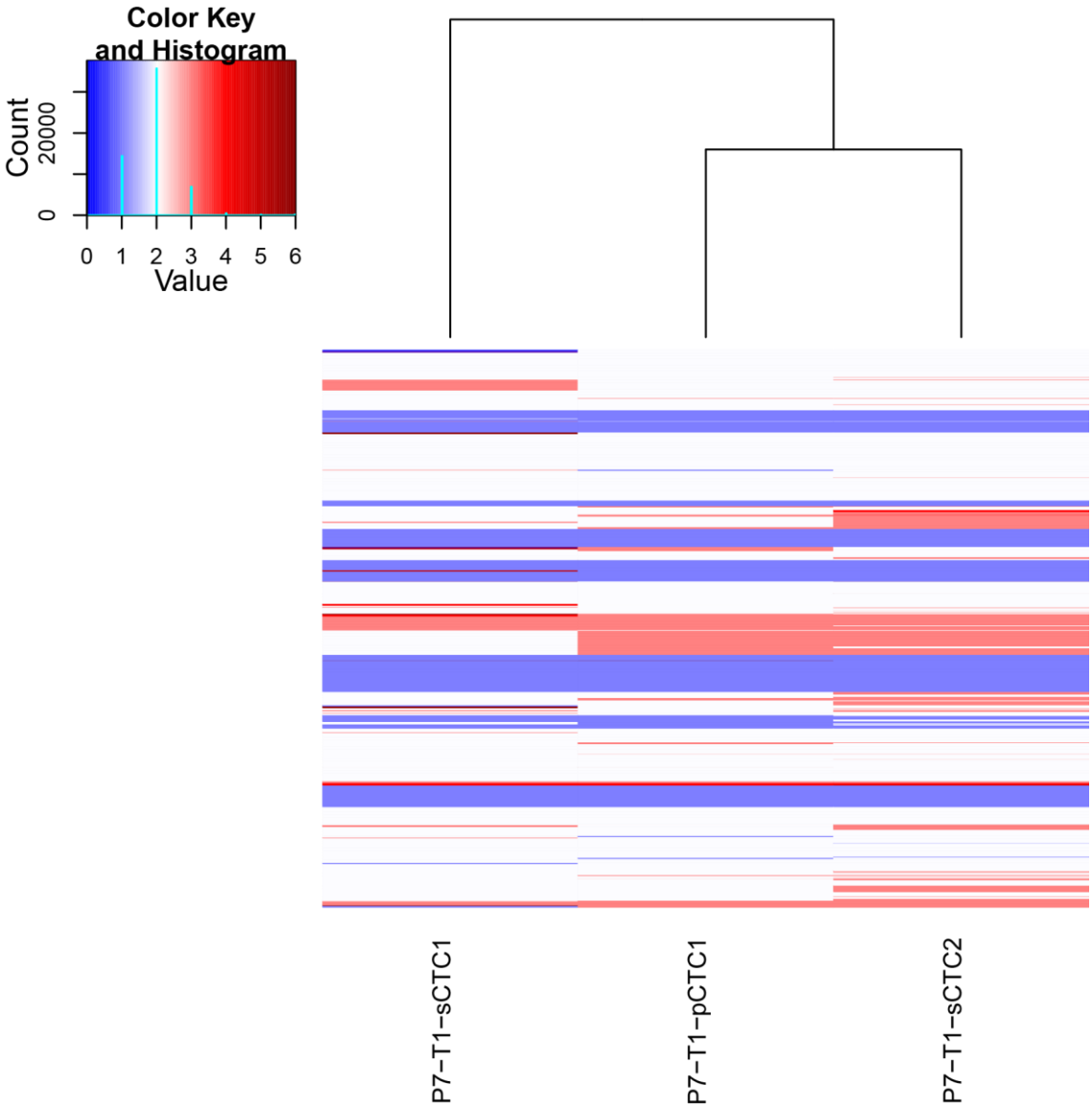


Figure A3.7 Hierarchical clustering of copy number data generated from CTCs isolated from patient 7. CTCs were isolated from a blood sample from patient 7 taken at baseline, prior to the patient receiving chemotherapy (T1). The samples were either single CTCs (sCTC) or pools of CTCs (pCTC). The amplified CTCs were subject to WGA and DNA libraries then created. The DNA libraries were WGS and the data used to calculate the copy number of 19336 protein-coding genes. Hierarchical clustering of the CNA data from the CTCs was then carried out. Regions of gain of copy number are red, regions of loss are blue and regions with no change are white.

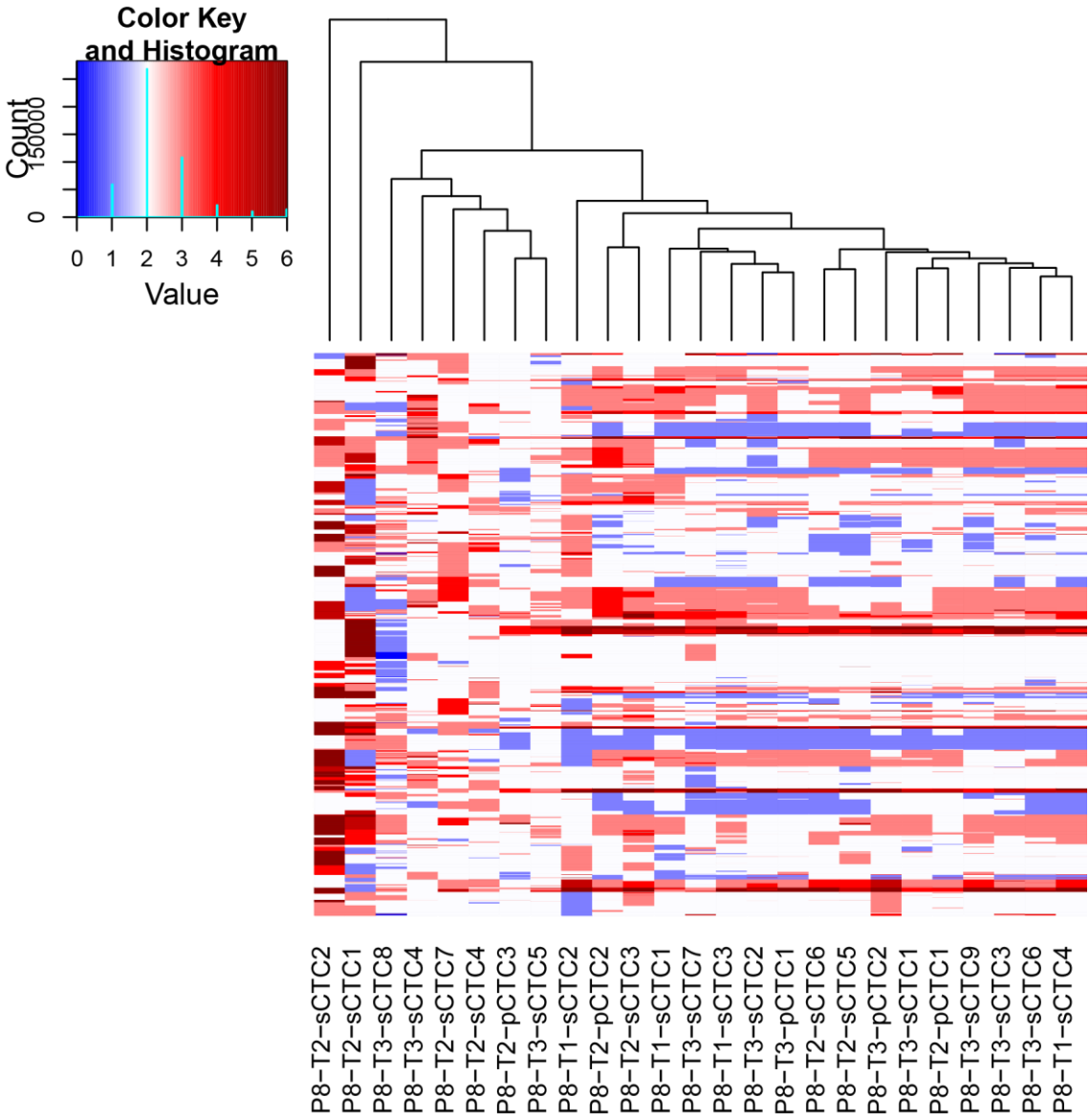


Figure A3.8 Hierarchical clustering of copy number data generated from CTCs isolated from patient 8. CTCs were isolated from blood samples from patient 8 taken at baseline, prior to the patient receiving chemotherapy (T1), when the patient relapsed after first-line chemotherapy (T2) and when the patient relapsed after second-line chemotherapy (T3). The samples were either single CTCs (sCTC) or pools of CTCs (pCTC). The amplified CTCs were subject to WGA and DNA libraries then created. The DNA libraries were WGS and the data used to calculate the copy number of 19336 protein-coding genes. Hierarchical clustering of the CNA data from the CTCs was then carried out. Regions of gain of copy number are red, regions of loss are blue and regions with no change are white.

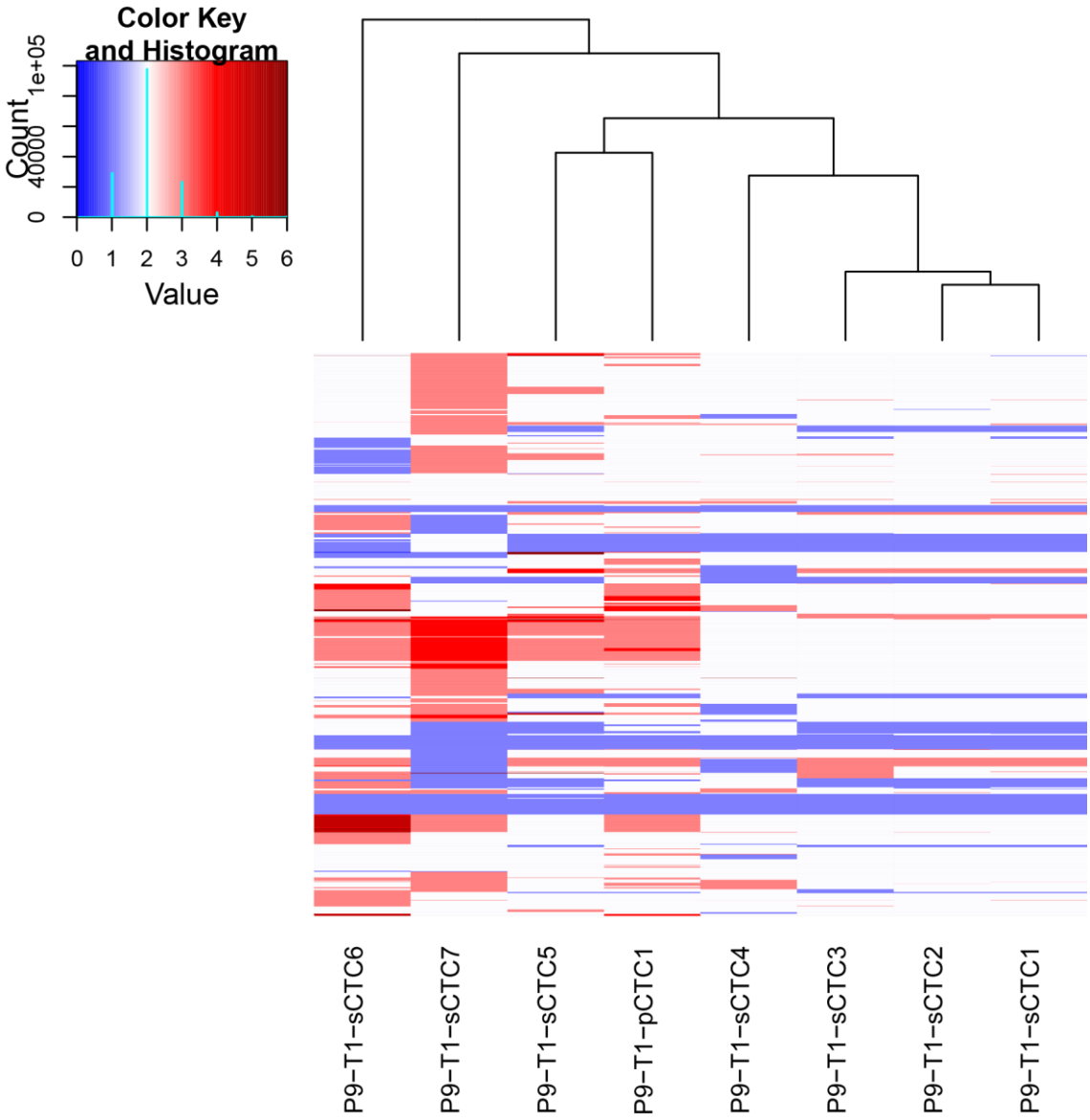


Figure A3.9 Hierarchical clustering of copy number data generated from CTCs isolated from patient 9. CTCs were isolated from a blood sample from patient 9 taken at baseline, prior to the patient receiving chemotherapy (T1). The samples were either single CTCs (sCTC) or pools of CTCs (pCTC). The amplified CTCs were subject to WGA and DNA libraries then created. The DNA libraries were WGS and the data used to calculate the copy number of 19336 protein-coding genes. Hierarchical clustering of the CNA data from the CTCs was then carried out. Regions of gain of copy number are red, regions of loss are blue and regions with no change are white.

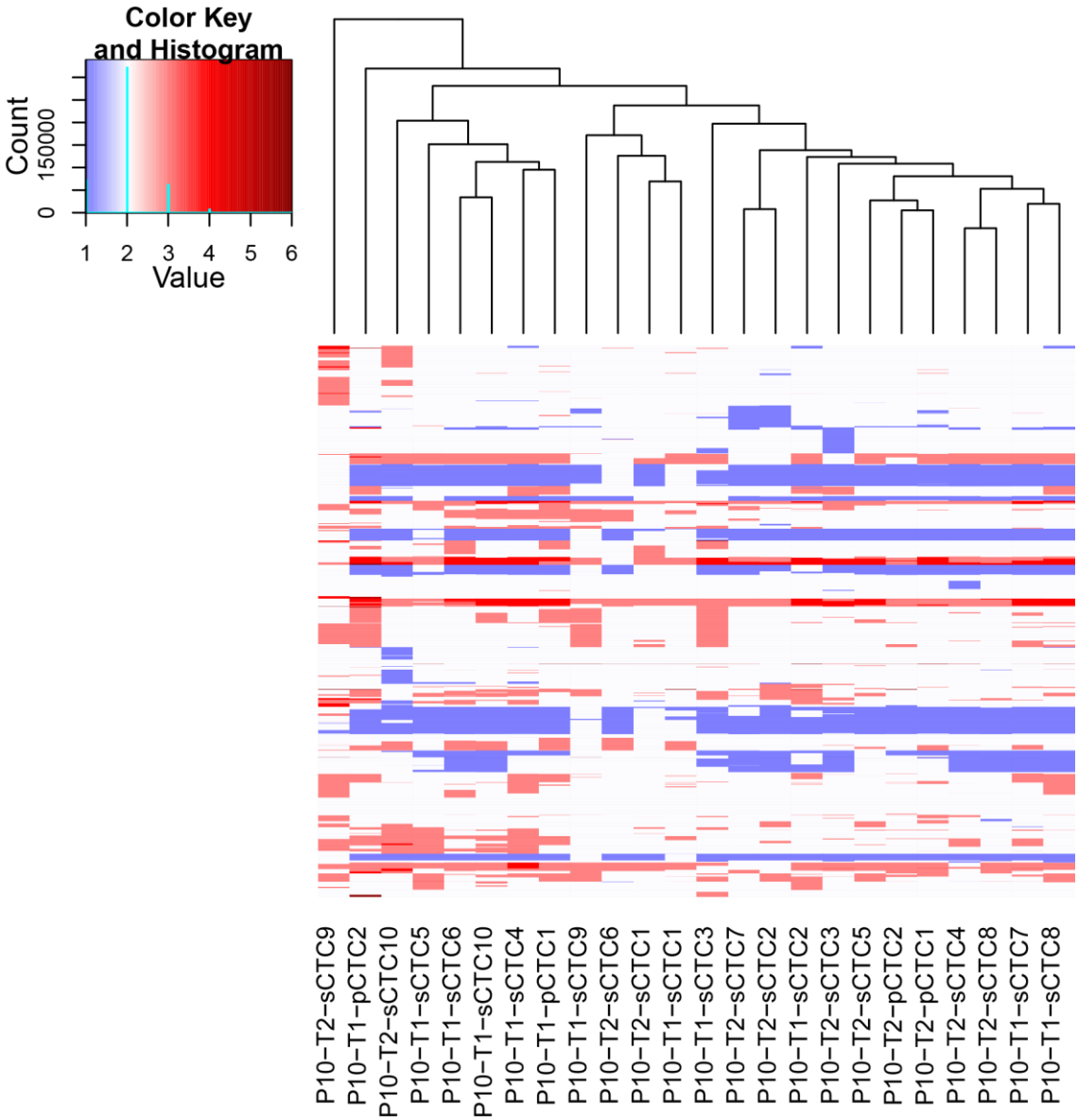


Figure A3.10 Hierarchical clustering of copy number data generated from CTCs isolated from patient 10. CTCs were isolated from blood samples from patient 10 taken at baseline, prior to the patient receiving chemotherapy (T1) and when the patient relapsed after first-line chemotherapy (T2). The samples were either single CTCs (sCTC) or pools of CTCs (pCTC). The amplified CTCs were subject to WGA and DNA libraries then created. The DNA libraries were WGS and the data used to calculate the copy number of 19336 protein-coding genes. Hierarchical clustering of the CNA data from the CTCs was then carried out. Regions of gain of copy number are red, regions of loss are blue and regions with no change are white.

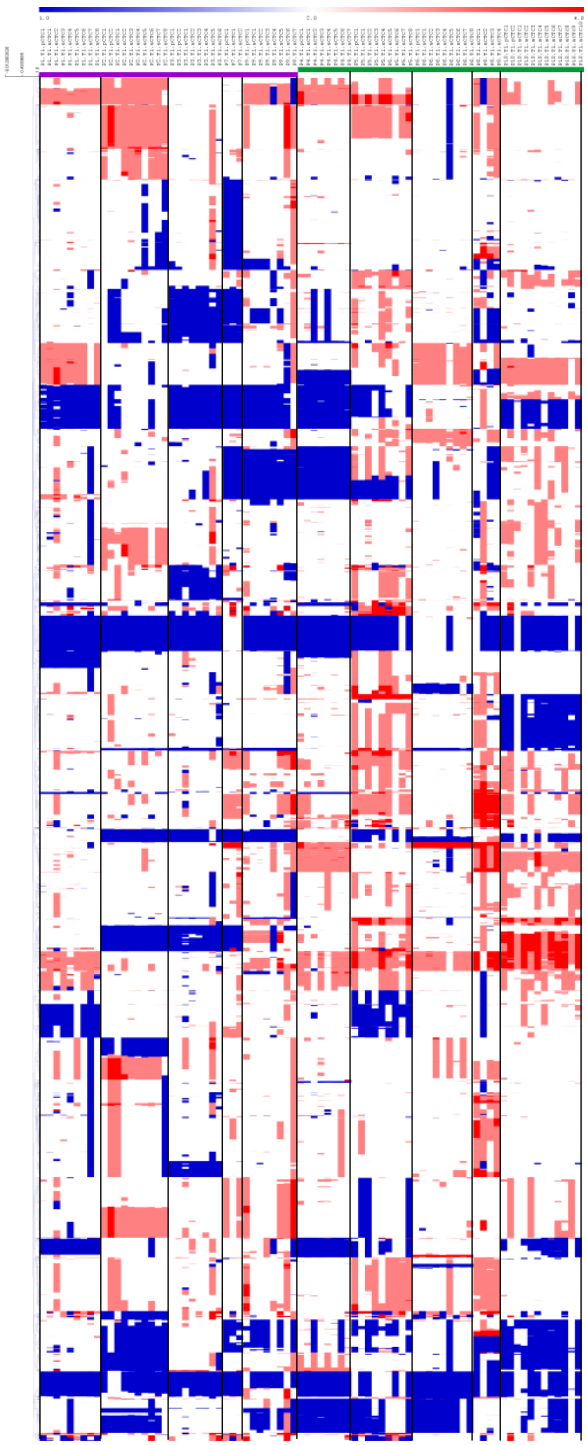


Figure A3.11 CNA analysis in protein-coding genes of CTCs from chemoresponsive and chemorefractory patients. Single CTCs and pools of CTCs were whole genome amplified and NGS of the DNA libraries performed. CNA analysis was carried out on the amplified CTCs, using sequencing data generated from 19336 protein-coding genes. The CTCs from chemorefractory patients are in purple whilst the chemoresponsive CTCs are in green. The CTCs from individual patients are divided by black lines. Gains are coloured red, losses are coloured blue and regions with no change are coloured white.

Appendix 4: CTC Confirmatory Sanger Sequencing from Chapter 5

Gene	Chromosome	Position	Reference	Report	Forward Primer Sequence	Reverse Primer Sequence	Patient with Mutation	Patient with Wild-Type Sequence
BRAF	7	140534539	G	T	AATCTCTGGGGAACGGAACT	TTTGTTGGGCAGGAAGACTC	P8 Baseline	P8 Relapse
LRP1B	2	141135833	T	A	TTTTATCCCCAGAGCCACAG	CTTGCCGCACTTATTGGACT	P1	P9
LRP1B	2	141777669	G	T	TGAGAGAGGGCCTATTATTCACA	TTCCTATGGCCATCATTGGT	P7	P9
LRP5	11	68174189	G	A	TGCTGGGCTGTTGATGTTTA	CTTTGAGGCAGGAACAGAGG	P6	P4
LRP5	11	68174189	G	A	TGCTGGGCTGTTGATGTTTA	CTTTGAGGCAGGAACAGAGG	P8	P4
PCDH10	4	134072249	T	A	ACGGTGAGGTCGTGTACTCC	GAAAGGCACGTCTCCCAGTA	P9	P1
TP53	17	7578536	T	C	GTTTCTTTGCTGCCGTCTTC	ACACGCAAATTTCCTTCCAC	P4	P10
TP53	17	7578461	C	A	GTTTCTTTGCTGCCGTCTTC	ACACGCAAATTTCCTTCCAC	P1	P10
TP53	17	7578275	G	A	GTTTCTTTGCTGCCGTCTTC	ACACGCAAATTTCCTTCCAC	P6	P10

Table A4.1 Confirmatory Sanger sequencing from WES in chapter 5. Confirmatory Sanger sequencing of 9 mutated and 9 wild-type amplicons identified from the WES of CTCs in chapter 5 was performed as described in section 2.7. Results that were concordant between the WES and Sanger sequencing are highlighted in green. The result that was mutated in the WES but wild-type in the Sanger sequencing is highlighted in blue.

Appendix 5: Gene Lists from Chapter 5

Table A5.1 Top 500 Genes from PC2 from the PCA analysis of the chemoresponsive and chemorefractory patients' CTCs. CTCs were isolated from blood samples of five chemoresponsive and five chemorefractory patients CTCs. DNA libraries created from amplified CTCs were WGS and the data used to calculate the copy number of 19336 protein-coding genes. A PCA was performed and PC2 led to the maximum separation of the two groups of patients CTCs. The top 500 genes contributing to this principal component are detailed in the table.

Gene Symbol				
SMCHD1	RAB31	NPC1	MAPRE2	PIAS2
EMILIN2	TXNDC2	ANKRD29	ZNF397	KATNAL2
LPIN2	VAPA	LAMA3	ZSCAN30	TCEB3CL2
MYOM1	APCDD1	TTC39C	ZNF24	TCEB3CL
MYL12A	NAPG	IMPACT	ZNF396	TCEB3C
MYL12B	PIEZO2	HRH4	RP11-322E11.6	TCEB3B
TGIF1	GNAL	ZNF521	INO80C	HDHD2
DLGAP1	CHMP1B	SS18	GALNT1	SKOR2
CTAGE1	MPPE1	PSMA8	C18orf21	SMAD2
CTAGE1	IMPA2	TAF4B	RPRD1A	ZBTB7C
CABLES1	SLMO1	KCTD1	SLC39A6	AC091150.1
TMEM241	SPIRE1	AQP4	ELP2	CTIF
CABYR	AP005482.1	CHST9	MOCOS	SMAD7
OSBPL1A	CEP76	CDH2	FHOD3	DYM
IER3IP1	PSMG2	DSC3	TPGS2	C18orf32
IER3IP1	PTPN2	DSC2	KIAA1328	RPL17-C18ORF32
C18orf42	SEH1L	DSC1	CELF4	RPL17
TMEM200C	CEP192	DSG1	RIT2	LIPG
RP11-865B13.1	LDLRAD4	DSG4	SYT4	ACAA2
ARHGAP28	FAM210A	DSG3	SETBP1	RP11-886H22.1
LAMA1	RNMT	DSG2	SLC14A2	MYO5B
LRRC30	MC5R	TTR	SLC14A1	CCDC11
PTPRM	ESCO1	B4GALT6	SIGLEC15	MBD1
AP001094.1	SNRPD1	SLC25A52	EPG5	CXXC1
RAB12	ABHD3	TRAPPC8	PSTPIP2	SKA1
SOGA2	MIB1	RNF125	ATP5A1	MAPK4
NDUFV2	RP11-595B24.2	RNF138	HAUS1	MRO
ANKRD12	GATA6	FAM59A	C18orf25	ME2
TWSG1	RBBP8	MEP1B	RNF165	ELAC1
RALBP1	RIOK3	NOL4	LOXHD1	RP11-729L2.2
PPP4R1	C18orf8	DTNA	ST8SIA5	SMAD4

Table A5.1 continued

Gene Symbol				
MEX3C	C16orf86	RP11-343C2.7	RP11-77K12.7	HSDL1
NAE1	GFOD2	NIP7	CHST5	DNAAF1
CA7	RANBP10	TMED6	TMEM231	TAF1C
PDP2	TSNAXIP1	TERF2	GABARAPL2	ADAD2
CDH16	CENPT	CYB5B	ADAT1	KCNG4
RRAD	THAP11	FTSJD1	KARS	WFDC1
FAM96B	NUTF2	CALB2	TERF2IP	ATP2C2
CES2	EDC4	ZNF23	AC025287.1	KIAA1609
CES3	NRN1L	ZNF19	CNTNAP4	COTL1
CES4A	PSKH1	ZNF19	RP11-58C22.1	KLHL36
CBFB	CTRL	CHST4	MON1B	USP10
C16orf70	CTC-479C5.12	TAT	SYCE1L	CRISPLD2
B3GNT9	PSMB10	MARVELD3	ADAMTS18	ZDHHC7
TRADD	LCAT	PHLPP2	NUDT7	KIAA0513
FBXL8	SLC12A4	AP1G1	VAT1L	FAM92B
RP11-5A19.5	DPEP3	ATXN1L	CLEC3A	RP11-680G10.1
HSF4	DPEP2	IST1	WWOX	GSE1
NOL3	DUS2L	ZNF821	PIH1	GINS2
KIAA0895L	DDX28	IST1	MAF	C16orf74
EXOC3L1	NFATC3	DHODH	DYNLRB2	EMC8
E2F4	ESRP2	TXNL4B	CDYL2	COX4I1
ELMO3	PLA2G15	HP	CMC2	IRF8
LRRC29	SLC7A6	HPR	CENPN	FOXF1
AC040160.1	SLC7A6OS	DHX38	ATMIN	MTHFSD
TMEM208	PRMT7	PMFBP1	C16orf46	FOXC2
FHOD1	SMPD3	ZFHX3	GCSH	FOXL1
SLC9A5	ZFP90	C16orf47	PKD1L2	C16orf95
PLEKHG4	CDH3	RFWD3	BCMO1	RP11-178L8.4
KCTD19	CDH1	MLKL	GAN	FBXO31
LRRC36	TMCO7	FA2H	CMIP	MAP1LC3B
TPPP3	HAS3	WDR59	PLCG2	ZCCHC14
ZDHHC1	CHTF8	ZNRF1	SDR42E1	JPH3
HSD11B2	CHTF8	LDHD	HSD17B2	AC010536.1
ATP6V0D1	CIRH1A	ZFP1	MPHOSPH6	FLJ00104
AGRP	SNTB2	CTRB2	CDH13	IRX3
FAM65A	VPS4A	CTRB1	HSBP1	IRX5
CTCF	RP11-343C2.3	BCAR1	MLYCD	IRX6
RLTPR	COG8	CFDP1	OSGIN1	MMP2
ACD	PDF	RP11-77K12.1	NECAB2	LPCAT2
PARD6A	RP11-343C2.8	TMEM170A	SLC38A8	CAPNS2
ENKD1	COG8	CHST6	MBTPS1	SLC6A2

Table A5.1 continued

Gene Symbol			
CES1	FBXL3	FDX1	GNMT
CES5A	MYCBP2	ARHGAP20	PEX6
GNAO1	SCEL	IL17A	PPP2R5D
AMFR	SLAIN1	IL17F	MEA1
NUDT21	EDNRB	MCM3	KLHDC3
OGFOD1	POU4F1	PAQR8	RRP36
BBS2	RNF219	EFHC1	CUL7
MT4	RBM26	TRAM2	KLC4
MT3	NDFIP2	TMEM14A	MRPL2
MT2A	SPRY2	GSTA2	PTK7
MT1E	SLITRK1	GSTA1	PKHD1
AC026461.1	SLITRK6	GSTA5	TMPRSS11F
MT1M	SLITRK5	GSTA3	TMPRSS11D
MT1A	TBC1D4	GSTA4	TMPRSS11BNL
MT1B	COMMD6	ICK	TMPRSS11B
MT1F	UCHL3	FBXO9	YTHDC1
MT1G	LMO7	GCM1	TMPRSS11E
MT1H	PANX1	ELOVL5	
MT1X	FOLR4	GCLC	
NUP93	GPR83	KLHL31	
SLC12A3	CNTN5	LRRC1	
HERPUD1	ARHGAP42	MLIP	
CETP	TMEM133	TINAG	
NLRC5	PGR	FAM83B	
CPNE2	DCUN1D5	BYSL	
CCDC113	DYNC2H1	CCND3	
PRSS54	CASP12	TAF8	
GINS3	CASP4	C6orf132	
NDRG4	CASP5	GUCA1A	
SETD6	CASP1	RP1-139D8.6	
CNOT1	CARD16	GUCA1B	
SLC38A7	CARD17	MRPS10	
GOT2	CARD18	TRERF1	
RP11-105C20.2	GRIA4	UBR2	
CMTM4	MSANTD4	PRPH2	
DYNC1LI2	KBTBD3	TBCC	
CCDC79	AASDHPPT	KIAA0240	
C13orf45	GUCY1A2	RPL7L1	
KCTD12	C11orf87	C6orf226	
IRG1	ZC3H12C	PTCRA	
CLN5	RDX	CNPY3	

Table A5.2 Genes with significant differences in copy number identified in LIMMA analyses. CTCs were isolated from blood samples of five chemorefractory patients CTCs and five chemoresponsive patients CTCs (at both baseline and relapse after chemotherapy). DNA libraries created from amplified CTCs were WGS and the data used to calculate the copy number of 19336 protein-coding genes. LIMMA analyses were performed to identify significant differences in copy number between the baseline CTCs from the chemorefractory and chemoresponsive patients and separately between the baseline CTCs from the chemorefractory patients and the relapse CTCs from the chemoresponsive patients. Genes identified in both analyses are indicated in green. Genes identified just in the comparison of the baseline CTCs from the chemorefractory and chemoresponsive patients are in blue. Genes identified just in the comparison of the baseline CTCs from the chemorefractory patients and the relapse CTCs from the chemoresponsive patients are indicated in red.

		Gene Symbol		
FAM192A	GINS3	HSDL1	GINS2	BBS2
RSPRY1	NDRG4	USP10	NLRC5	MT4
ARL2BP	SETD6	CRISPLD2	NUDT21	AMFR
PLLP	CNOT1	RPGRIP1L	MT1E	IRX3
CCL22	SLC38A7	SLC38A8	AC026461.1	WWOX
CX3CL1	GOT2	MBTPS1	MT1M	JPH3
CCL17	RP11-105C20.2	RP11-680G10.1	MT1A	PIH1
CIAPIN1	ATMIN	DYNLRB2	MT1B	MAF
COQ9	C16orf46	CDYL2	MT1F	LPCAT2
POLR2C	GCSH	HSBP1	MT1G	CAPNS2
DOK4	PKD1L2	C16orf95	MT1H	SLC6A2
CCDC102A	BCMO1	MLYCD	MT1X	CES1
GPR114	GAN	CMC2	NUP93	CES5A
GPR56	CMIP	EMC8	SLC12A3	PLCG2
GPR97	CENPN	COX4I1	HERPUD1	SDR42E1
CCDC135	FTO	IRF8	CETP	HSD17B2
KATNB1	DNAAF1	FOXF1	RP11-178L8.4	MMP2
KIFC3	TAF1C	MTHFSD	FBXO31	IRX5
CTD-2600O9.1	ADAD2	FOXC2	MAP1LC3B	IRX6
CNGB1	KCNG4	FOXL1	ZCCHC14	RBBP8
TEPP	WFDC1	CPNE2	MON1B	RP11-324D17.1
ZNF319	ZDHHC7	AC010536.1	SYCE1L	CDH13
USB1	KIAA0513	FLJ00104	ADAMTS18	CNTNAP4
MMP15	FAM92B	C16orf74	NUDT7	MPHOSPH6
C16orf80	ATP2C2	NECAB2	VAT1L	LATS2
CSNK2A2	KIAA1609	GNAO1	CLEC3A	RP11-58C22.1
CCDC113	COTL1	MT3	OSGIN1	APCDD1
PRSS54	KLHL36	MT2A	OGFOD1	NAPG

Table A5.2 continued

Gene Symbol				
PIEZO2	C1QTNF9	ZNF19	TXNDC2	NIP7
CABLES1	C1QTNF9B-AS1	ZNF19	CHMP1B	TMED6
SPRY2	PARP4	CHST4	C16orf47	TERF2
TMEM241	ATP12A	TAT	L3MBTL4	CYB5B
RIOK3	RNF17	GATA6	EPB41L3	TPPP3
AP1G1	CENPJ	CEP192	AP001094.1	ZDHHC1
CABYR	LSP1	ANKRD29	CDK8	C16orf70
OSBPL1A	PABPC3	RP11-77K12.1	WASF3	B3GNT9
ATXN1L	FAM123A	ADAT1	GPR12	TRADD
TXNL4B	MTMR6	WDR59	CTAGE1	FBXL8
TTC39C	NUPL1	DLGAP1	CTAGE1	RP11-5A19.5
ZNF521	ATP8A2	RALBP1	RAB12	HSF4
SS18	AL138815.1	GNAL	PTPN2	NOL3
PSMA8	AL138815.2	LAMA1	PDP2	KIAA0895L
AQP4	SHISA2	KARS	CBFB	EXOC3L1
CHST9	CDH8	TERF2IP	HSD11B2	E2F4
ZFP161	PHLPP2	AC025287.1	SMCHD1	ELMO3
TAF4B	CES4A	PPP4R1	NAE1	LRRC29
HPR	MARVELD3	RAB31	CA7	AC040160.1
LAMA3	VAPA	CFDP1	ATP6V0D1	TMEM208
SAP18	HRH4	RRAD	ANKRD12	FHOD1
SKA3	IMPACT	FTSJD1	FAM65A	SLC9A5
MRP63	DHX38	CDH16	LRRC36	PLEKHG4
ZDHHC20	PMFBP1	C18orf42	AGRP	TWSG1
ZDHHC20	NPC1	RNF6	DYNC1LI2	RP11-865B13.1
EFHA1	CHST6	FAM96B	KCTD19	ARHGAP28
GSE1	TMEM170A	CES2	CCDC79	CMTM4
SGCG	RP11-77K12.7	CES3	CDH2	SPIRE1
SACS	CHST5	EMILIN2	SOGA2	AP005482.1
TNFRSF19	TMEM231	LPIN2	TMCO7	CEP76
FGF9	GABARAPL2	MPPE1	HAS3	PSMG2
IST1	MYL12A	IMPA2	CHTF8	TMEM200C
KCTD1	MYL12B	ZNRF1	CHTF8	CDH1
C18orf8	TGIF1	LDHD	CIRH1A	FAM124A
ZNF821	MYOM1	ZFP1	SNTB2	SERPINE3
IST1	USP12	CTRB2	VPS4A	INTS6
DHODH	RPL21	CTRB1	RP11-343C2.3	CTCF
HP	RASL11A	BCAR1	COG8	PTPRM
MIPEP	GTF3A	SEH1L	PDF	CDH3
AL139080.1	MTIF3	LRRC30	RP11-343C2.8	CMTM2
C1QTNF9B	CALB2	NDUFV2	COG8	CMTM3
SPATA13	ZNF23	ZFHX3	RP11-343C2.7	CDH5

Table A5.2 continued

Gene Symbol				
RLTPR	DSC2	HCN1	SDAD1	G3BP2
ACD	DSC1	MKNK2	AC110615.1	USO1
PARD6A	MIB1	ATP8B3	CXCL9	PPEF2
ENKD1	RP11-595B24.2	REXO1	ART3	SLC4A4
C16orf86	DLEU7	LDLRAD4	CXCL10	GC
GFOD2	SLITRK6	FAM210A	CXCL11	NPFFR2
CMTM1	C18orf25	RNMT	NUP54	ADAMTS3
RANBP10	METTL4	MC5R	SCARB2	ARHGAP24
TSNAXIP1	NDC80	SLMO1	FAM47E	UGT2B15
CENPT	SLC25A52	LPHN3	STBD1	UGT2B10
THAP11	PSTPIP2	ANXA3	CCDC158	UGT2A3
NUTF2	ATP5A1	BMP2K	SHROOM3	UGT2B7
EDC4	HAUS1	PAQR3	SOWAHB	UGT2B11
NRN1L	RNF125	NAA11	SEPT11	UGT2B28
ESRP2	RNF138	GK2	CCNI	UGT2B4
PLA2G15	FAM84B	ANTXR2	CCNG2	UGT2A1
SLC7A6	PTGER4	PRDM8	CXCL13	SULT1B1
SLC7A6OS	TTC33	FGF5	CNOT6L	SULT1E1
CDH11	PRKAA1	C4orf22	MRPL1	CSN1S1
ZFP90	RPL37	BMP3	AFM	CSN2
CKLF-CMTM1	CARD6	PRKG2	AFP	STATH
NFATC3	C7	RASGEF1B	RASSF6	HTN3
SMPD3	HEATR7B2	HNRNPD	IL8	HTN1
PRMT7	C6	HNRPDL	CXCL6	C4orf40
CKLF	PLCXD3	ENOPH1	PF4V1	ODAM
TK2	C5orf51	TMEM150C	CXCL1	FDCSP
PSKH1	FBXO4	SCD5	PF4	CSN3
RNASEH2B	OXCT1	SEC31A	PPBP	CABS1
OLFM4	GHR	THAP9	CXCL5	SMR3A
RP11-403P17.5	CCDC152	LIN54	CXCL3	SMR3B
AC132186.1	PDZD2	COPS4	CXCL2	PROL1
CTRL	GOLPH3	PLAC8	AC093677.1	MUC7
CTC-479C5.12	CDH6	COQ2	COX18	AMTN
PSMB10	DROSHA	HPSE	ANKRD17	AMBN
LCAT	C5orf22	HELQ	ALB	ENAM
SLC12A4	MRPS30	MRPS18C	NAAA	IGJ
DPEP3	CDH10	FAM175A	BTC	UTP3
DPEP2	C5orf17	AGPAT9	PARM1	RUFY3
DUS2L	PRDM9	NKX6-1	RCHY1	GRSF1
DDX28	CDH9	CDS1	THAP6	MOB1B
BEAN1	FGF10	WDFY3	C4orf26	DCK
DSC3	HTR1F	FRAS1	CDKL2	KCTD12

Table A5.2 continued

Gene Symbol				
IRG1	MEP1B	FA2H	SKP2	SLITRK5
CLN5	ACAA2	CELF4	NADKD1	COMMD6
FBXL3	RP11-886H22.1	SEPP1	RANBP3L	UCHL3
MYCBP2	PIK3C3	EPHA5	HMGCS1	TBC1D4
SCEL	MYO5B	C13orf45	CCL28	CBLN2
SLAIN1	CCDC11	ANXA2R	C5orf28	TMPRSS11F
EDNRB	MBD1	GDNF	C5orf34	NALCN
SMAD2	CXXC1	EGFLAM	PAIP1	TM9SF2
POU4F1	SKA1	LIFR	RP11-322E11.6	CLYBL
RNF219	C18orf32	AC091435.1	INO80C	ITGBL1
TECRL	RPL17-C18ORF32	AC091435.2	TARS	ZIC5
DSG1	RPL17	OSMR	ADAMTS12	ZIC2
SLC14A1	LIPG	RICTOR	RXFP3	UBAC2
SIGLEC15	MAPK4	FYB	SLC45A2	GPR18
EPG5	SETBP1	AC008964.1	AMACR	GPR183
RNF165	FAM59A	C9	C1QTNF3	GPC5
LOXHD1	DYM	DAB2	RAI14	WDFY2
ST8SIA5	MRO	ZNF131	TTC23L	DHRS12
PIAS2	ME2	C5orf42	RAD1	CCDC70
KATNAL2	ELAC1	AVEN	BRIX1	ATP7B
TCEB3CL2	RP11-729L2.2	SNURF	DNAJC21	ALG11
TCEB3CL	SMAD4	ZFR	AGXT2	UTP14C
TCEB3C	MEX3C	NIPBL	FAM134B	IPO5
TCEB3B	SYT4	TPGS2	ESCO1	FARP1
HDHD2	KIAA1328	MARCH11	SNRPD1	RNF113B
IER3IP1	NOL4	UGT2B17	ABHD3	STK24
IER3IP1	KLHL14	SLC1A3	GALNT1	SLC15A1
SKOR2	AC012123.1	NNT	PRLR	DOCK9
RBM26	CCDC178	CHRM5	AC010368.2	LECT1
NDFIP2	LNX2	EMC7	FHOD3	PCDH8
SLC14A2	NUP155	PGBD4	ZNF622	FOLR4
DSG4	WDR70	NIM1	C18orf21	GPR83
DSG3	ASXL3	FBXL7	RPRD1A	EXD1
ZBTB7C	DTNA	SUB1	TMPRSS11BNL	CHP1
AC091150.1	MAPRE2	NPR3	TMPRSS11B	OIP5
CTIF	ZNF397	AC026703.1	YTHDC1	NUSAP1
SMAD7	ZSCAN30	SPEF2	TMPRSS11E	C15orf57
RIT2	ZNF24	IL7R	MOCOS	SLITRK1
DSG2	ZNF396	CAPSL	ELP2	ITPKA
TTR	MTMR12	UGT3A1	MYO10	LTK
B4GALT6	RFWD3	UGT3A2	SLC39A6	RPAP1
TRAPPC8	MLKL	LMBRD2	LMO7	RTF1

Table A5.2 continued

Gene Symbol		
NDUFAF1	WDR67	DUOXA1
CASC5	FAM83A	DUOX1
RAD51	C8orf76	SHF
FAM82A2	ZHX1-C8ORF76	DUOX2
INO80	ZHX1	SQRDL
TDRD3	ATAD2	SPATA5L1
DIAPH3	WDYHV1	C15orf48
TYRO3	FBXO32	SLC30A4
GCHFR	KLHL38	BLOC1S6
DNAJC17	ANXA13	SQRDL
C15orf62	FAM91A1	ST3GAL1
ZFYVE19	FER1L6	EFR3A
PPP1R14D	HAS2	OC90
SPINT1	MEIS2	OC90
RHOV	SPG11	HHLA1
VPS18	ZHX2	KCNQ3
DLL4	MTSS1	LRRC6
CHAC1	TMCO5A	TMEM71
UBE3A	TMEM65	PHF20L1
TMEM87A	TRMT12	TG
STARD9	RNF139	ZFAT
CDAN1	TATDN1	KHDRBS3
TTBK2	NDUFB9	PTK2
UBR1	TRIM69	SLA
CASC4	SLC45A4	POU5F1B
DLEU1	SNTB1	MYC
HAUS2	ZNF572	TMEM75
CTDSPL2	SQLE	WISP1
LRRC57	KIAA0196	NDRG1
SNAP23	ADCY8	TRAPPC9
ZFP106	PATL2	CHRAC1
CAPN3	B2M	EIF2C2
PREX1	NSMCE2	C8orf17
RASGRP1	FAM135B	MOB3A
EIF3J	COL22A1	CGGBP1
GANC	KCNK9	RP11-159G9.5
DENND3	TRIB1	
SPRED1	SLC28A2	
FAM98B	GSDMC	
C15orf41	FAM49B	
MTBP	ASAP1	
DERL1	DUOXA2	

Appendix 6: Ingenuity Pathway Analyses Summaries

Summaries of the analyses run using QIAGEN's Ingenuity® Pathway Analysis (IPA®) (www.qiagen.com/ingenuity) which are presented in chapter 5 are detailed in this appendix. The analyses run were all the core analysis with standard settings.

Table A6.1 Summary of the results of the Ingenuity Pathway Analysis of the 19 most frequently mutated genes identified from the WES data of the SCLC CTCs.

Top Diseases and Bio Functions

Diseases and Disorders

Name	p-value	# Molecules
Cancer	3.55E-08 - 6.66E-03	19
Organismal Injury and Abnormalities	3.55E-08 - 5.71E-03	17
Renal and Urological Disease	3.55E-08 - 5.71E-03	10
Reproductive System Disease	3.55E-08 - 4.76E-03	16
Gastrointestinal Disease	2.63E-06 - 6.66E-03	17

Molecular and Cellular Functions

Name	p-value	# Molecules
Cell Cycle	2.15E-04 - 6.66E-03	3
Cell Death and Survival	3.94E-04 - 6.66E-03	3
DNA Replication, Recombination, and Repair	5.03E-04 - 5.71E-03	3
Cellular Assembly and Organization	7.25E-04 - 4.76E-03	4
Carbohydrate Metabolism	9.54E-04 - 4.76E-03	1

Physiological System Development and Function

Name	p-value	# Molecules
Humoral Immune Response	2.15E-04 - 2.86E-03	2
Hematological System Development and Function	6.25E-04 - 6.62E-03	2
Hematopoiesis	6.25E-04 - 4.76E-03	2
Lymphoid Tissue Structure and Development	6.25E-04 - 4.76E-03	2
Tissue Morphology	6.25E-04 - 6.62E-03	3

Table A6.2 Summary of the results of the Ingenuity Pathway Analysis of the top 500 weighted genes from PC2 from the PCA of the CNA data protein-coding genes generated from the chemoresponsive and chemorefractory patients' baseline CTCs.

Top Diseases and Bio Functions

Diseases and Disorders

Name	p-value	# Molecules
Cancer	2.59E-09 - 2.14E-02	373
Organismal Injury and Abnormalities	4.59E-07 - 2.14E-02	189
Reproductive System Disease	4.59E-07 - 2.14E-02	165
Dermatological Diseases and Conditions	6.88E-06 - 2.14E-02	20
Hereditary Disorder	6.88E-06 - 2.14E-02	75

Molecular and Cellular Functions

Name	p-value	# Molecules
Drug Metabolism	9.73E-06 - 2.14E-02	12
Small Molecule Biochemistry	9.73E-06 - 2.14E-02	48
Cell Morphology	4.57E-04 - 2.14E-02	22
Cell-To-Cell Signaling and Interaction	4.57E-04 - 2.14E-02	25
Cellular Compromise	4.57E-04 - 2.14E-02	10

Physiological System Development and Function

Name	p-value	# Molecules
Embryonic Development	9.73E-06 - 2.14E-02	40
Organ Development	9.73E-06 - 2.14E-02	27
Organ Morphology	9.73E-06 - 2.14E-02	22
Organismal Development	9.73E-06 - 2.14E-02	43
Renal and Urological System Development and Function	9.73E-06 - 2.14E-02	11

Table A6.3 Summary of the results of the Ingenuity Pathway Analysis of the genes which changed in the chemoresponsive patients' CTCs from the top 500 weighted genes from PC2 from the PCA of the CNA data protein-coding genes generated from the chemoresponsive and chemorefractory patients' baseline CTCs.

Top Diseases and Bio Functions

Diseases and Disorders

Name	p-value	# Molecules
Cancer	1.10E-07 - 3.53E-02	160
Dermatological Diseases and Conditions	2.13E-07 - 2.66E-02	11
Hereditary Disorder	2.13E-07 - 3.54E-02	34
Cardiovascular Disease	1.36E-05 - 3.26E-02	25
Developmental Disorder	1.36E-05 - 3.54E-02	36

Molecular and Cellular Functions

Name	p-value	# Molecules
Cellular Function and Maintenance	7.98E-05 - 3.32E-02	41
Cell Death and Survival	4.73E-04 - 3.54E-02	11
Cell Morphology	4.73E-04 - 3.54E-02	39
Molecular Transport	7.20E-04 - 3.54E-02	15
Cell Cycle	7.84E-04 - 3.54E-02	12

Physiological System Development and Function

Name	p-value	# Molecules
Skeletal and Muscular System Development and Function	7.98E-05 - 2.66E-02	14
Tissue Development	7.98E-05 - 3.54E-02	45
Embryonic Development	2.08E-04 - 3.54E-02	28
Tissue Morphology	2.08E-04 - 3.54E-02	19
Organ Development	2.38E-04 - 2.66E-02	19

Table A6.4 Summary of the results of the Ingenuity Pathway Analysis of the genes which changed in the chemorefractory patients' CTCs from the top 500 weighted genes from PC2 from the PCA of the CNA data protein-coding genes generated from the chemoresponsive and chemorefractory patients' baseline CTCs.

Top Diseases and Bio Functions

Diseases and Disorders

Name	p-value	# Molecules
Cancer	5.79E-09 - 3.83E-02	231
Organismal Injury and Abnormalities	5.79E-09 - 3.78E-02	72
Reproductive System Disease	5.79E-09 - 3.66E-02	50
Psychological Disorders	4.24E-05 - 2.67E-02	8
Inflammatory Response	1.80E-04 - 3.98E-02	10

Molecular and Cellular Functions

Name	p-value	# Molecules
Drug Metabolism	2.40E-06 - 2.67E-02	12
Small Molecule Biochemistry	2.40E-06 - 3.42E-02	40
Cell-To-Cell Signaling and Interaction	1.80E-04 - 3.19E-02	16
Lipid Metabolism	1.80E-04 - 3.19E-02	21
Nucleic Acid Metabolism	1.80E-04 - 2.67E-02	5

Physiological System Development and Function

Name	p-value	# Molecules
Hematological System Development and Function	1.80E-04 - 3.98E-02	11
Tissue Development	1.80E-04 - 3.98E-02	46
Hair and Skin Development and Function	5.35E-04 - 3.78E-02	6
Respiratory System Development and Function	5.35E-04 - 3.98E-02	6
Skeletal and Muscular System Development and Function	5.35E-04 - 3.98E-02	12

Table A6.5 Summary of the results of the Ingenuity Pathway Analysis of the 760 genes identified by LIMMA as having significant differences in copy number between the chemorefractory and chemoresponsive patients' CTCs.

Top Diseases and Bio Functions

Diseases and Disorders

Name	p-value	# Molecules
Cancer	2.50E-11 - 2.94E-02	574
Inflammatory Response	1.70E-05 - 3.04E-02	36
Dermatological Diseases and Conditions	3.36E-05 - 1.74E-02	54
Hereditary Disorder	3.86E-05 - 3.09E-02	65
Gastrointestinal Disease	6.92E-05 - 2.55E-02	403

Molecular and Cellular Functions

Name	p-value	# Molecules
Cell-To-Cell Signaling and Interaction	5.84E-09 - 2.98E-02	67
Cellular Movement	5.84E-09 - 2.77E-02	110
Lipid Metabolism	1.01E-06 - 3.04E-02	69
Small Molecule Biochemistry	1.01E-06 - 3.04E-02	97
Cellular Development	1.70E-05 - 2.35E-02	74

Physiological System Development and Function

Name	p-value	# Molecules
Immune Cell Trafficking	5.84E-09 - 3.01E-02	43
Hematological System Development and Function	6.39E-07 - 3.01E-02	45
Embryonic Development	1.70E-05 - 3.05E-02	50
Endocrine System Development and Function	1.70E-05 - 2.73E-02	31
Hematopoiesis	1.70E-05 - 2.06E-02	10

Table A6.6 Summary of the results of the Ingenuity Pathway Analysis of the genes with altered copy number in the chemoresponsive patients' CTCs from the 760 genes identified by LIMMA as having significant differences in copy number between the chemorefractory and chemoresponsive patients' CTCs.

Top Diseases and Bio Functions

Diseases and Disorders

Name	p-value	# Molecules
Cancer	3.89E-10 - 1.96E-02	344
Gastrointestinal Disease	2.13E-07 - 1.96E-02	258
Organismal Injury and Abnormalities	6.91E-07 - 1.96E-02	185
Reproductive System Disease	6.91E-07 - 1.88E-02	147
Inflammatory Response	8.36E-07 - 1.96E-02	30

Molecular and Cellular Functions

Name	p-value	# Molecules
Cell-To-Cell Signaling and Interaction	2.72E-09 - 1.96E-02	37
Cellular Movement	2.72E-09 - 1.96E-02	76
Lipid Metabolism	4.60E-08 - 1.96E-02	32
Small Molecule Biochemistry	4.60E-08 - 1.96E-02	43
Cell Death and Survival	1.44E-04 - 1.96E-02	51

Physiological System Development and Function

Name	p-value	# Molecules
Immune Cell Trafficking	2.72E-09 - 1.96E-02	29
Hematological System Development and Function	1.14E-08 - 1.96E-02	30
Endocrine System Development and Function	2.12E-06 - 1.96E-02	20
Hematopoiesis	2.12E-06 - 1.96E-02	7
Digestive System Development and Function	2.34E-06 - 1.96E-02	26

Table A6.7 Summary of the results of the Ingenuity Pathway Analysis of the genes with altered copy number in the chemorefractory patients' CTCs from the 760 genes identified by LIMMA as having significant differences in copy number between the chemorefractory and chemoresponsive patients' CTCs.

Top Diseases and Bio Functions

Diseases and Disorders

Name	p-value	# Molecules
Cancer	6.46E-08 - 2.78E-02	237
Organismal Injury and Abnormalities	6.46E-08 - 4.14E-02	70
Reproductive System Disease	6.46E-08 - 2.78E-02	50
Psychological Disorders	4.87E-05 - 2.78E-02	13
Ophthalmic Disease	5.81E-04 - 2.78E-02	10

Molecular and Cellular Functions

Name	p-value	# Molecules
Cellular Development	5.52E-07 - 2.78E-02	19
Cellular Growth and Proliferation	5.52E-07 - 2.78E-02	22
Drug Metabolism	2.72E-06 - 2.78E-02	8
Small Molecule Biochemistry	2.72E-06 - 4.13E-02	40
Cell-To-Cell Signaling and Interaction	1.95E-04 - 4.14E-02	24

Physiological System Development and Function

Name	p-value	# Molecules
Embryonic Development	5.52E-07 - 4.14E-02	28
Nervous System Development and Function	5.52E-07 - 2.89E-02	19
Organ Morphology	5.52E-07 - 4.14E-02	29
Organismal Development	5.52E-07 - 4.14E-02	46
Tissue Development	5.52E-07 - 4.14E-02	40

Table A6.8 Summary of the results of the Ingenuity Pathway Analysis of the genes identified as mutated in the chemoresponsive patients' CTCs but not the chemorefractory patients' CTCs from the WES data.

Top Diseases and Bio Functions

Diseases and Disorders

Name	p-value	# Molecules
Cancer	1.30E-04 - 4.27E-02	28
Gastrointestinal Disease	1.30E-04 - 4.79E-03	24
Hepatic System Disease	1.30E-04 - 1.30E-04	19
Hereditary Disorder	1.23E-03 - 3.41E-02	6
Connective Tissue Disorders	1.44E-03 - 3.41E-02	4

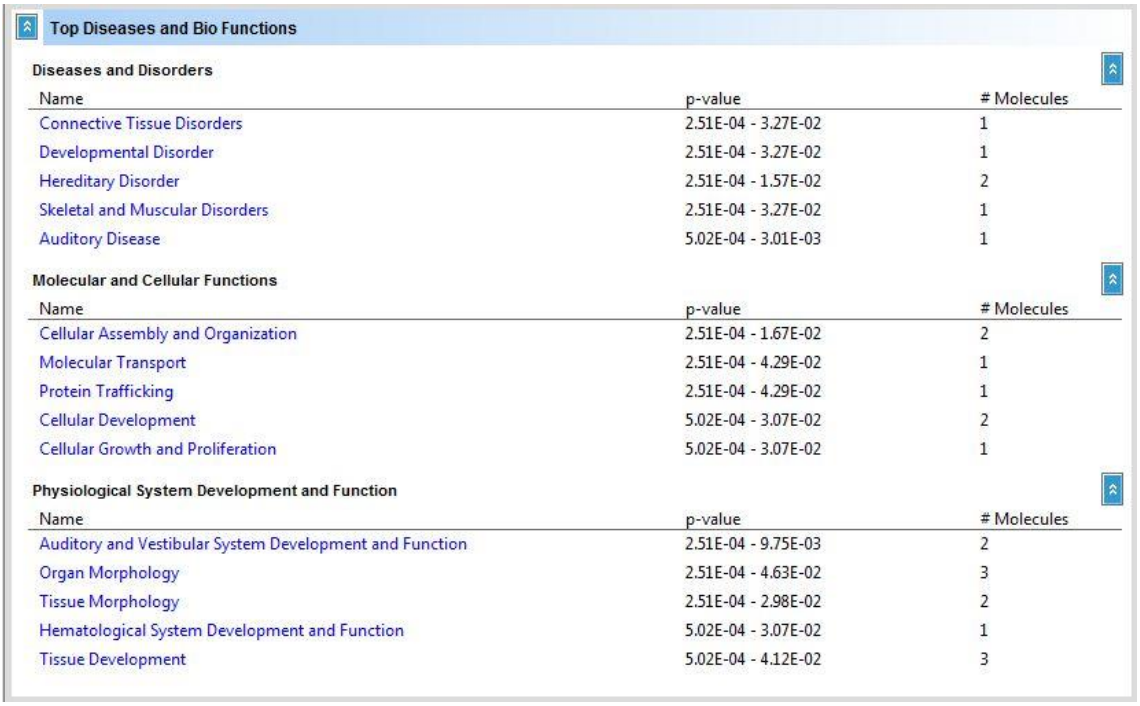
Molecular and Cellular Functions

Name	p-value	# Molecules
Cell Morphology	1.44E-03 - 3.69E-02	8
Carbohydrate Metabolism	1.44E-03 - 7.20E-03	1
Cell-To-Cell Signaling and Interaction	1.44E-03 - 4.38E-02	5
Cellular Assembly and Organization	1.44E-03 - 3.83E-02	2
Cellular Function and Maintenance	1.44E-03 - 4.38E-02	5

Physiological System Development and Function

Name	p-value	# Molecules
Connective Tissue Development and Function	1.44E-03 - 3.97E-02	5
Tissue Morphology	1.44E-03 - 4.52E-02	6
Cell-mediated Immune Response	1.44E-03 - 7.20E-03	2
Hematological System Development and Function	1.44E-03 - 4.38E-02	3
Immune Cell Trafficking	1.44E-03 - 7.20E-03	2

Table A6.9 Summary of the results of the Ingenuity Pathway Analysis of the genes identified as mutated in the chemorefractory patients' CTCs but not the chemoresponsive patients' CTCs from the WES data. The changes in cancer-related genes are included in a separate panel as for this analysis they were not included in the main summary produced by the ingenuity software but were discussed in the text of chapter 5.



Top Diseases and Bio Functions

Diseases and Disorders

Name	p-value	# Molecules
Connective Tissue Disorders	2.51E-04 - 3.27E-02	1
Developmental Disorder	2.51E-04 - 3.27E-02	1
Hereditary Disorder	2.51E-04 - 1.57E-02	2
Skeletal and Muscular Disorders	2.51E-04 - 3.27E-02	1
Auditory Disease	5.02E-04 - 3.01E-03	1

Molecular and Cellular Functions

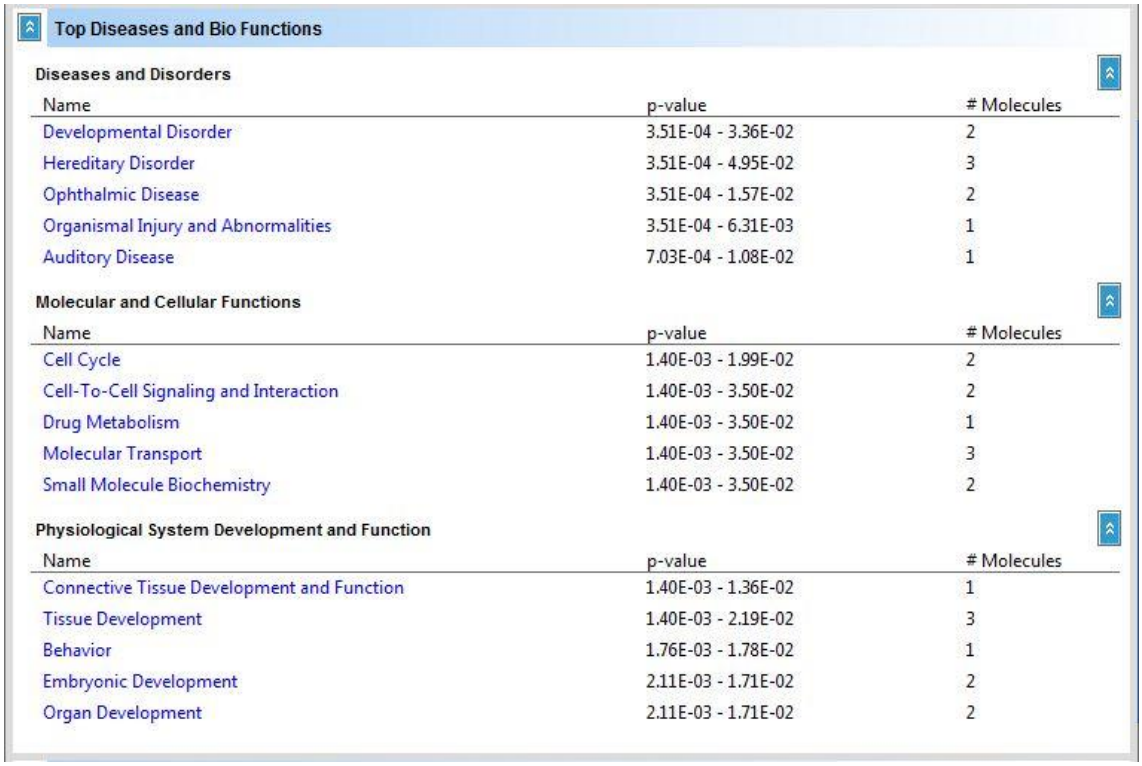
Name	p-value	# Molecules
Cellular Assembly and Organization	2.51E-04 - 1.67E-02	2
Molecular Transport	2.51E-04 - 4.29E-02	1
Protein Trafficking	2.51E-04 - 4.29E-02	1
Cellular Development	5.02E-04 - 3.07E-02	2
Cellular Growth and Proliferation	5.02E-04 - 3.07E-02	1

Physiological System Development and Function

Name	p-value	# Molecules
Auditory and Vestibular System Development and Function	2.51E-04 - 9.75E-03	2
Organ Morphology	2.51E-04 - 4.63E-02	3
Tissue Morphology	2.51E-04 - 2.98E-02	2
Hematological System Development and Function	5.02E-04 - 3.07E-02	1
Tissue Development	5.02E-04 - 4.12E-02	3

Categories	Diseases or Functions Annotation	p-Value	Molecules	# Molecules
Cancer, Cellular Development, Cellular Growth and Proliferation, Neurological Disease, Tumor Morphology	growth of adrenal carcinoma cells	5.02E-04	HSPG2 all 1	1
Cancer, Hematological Disease, Immunological Disease, Organismal Injury and Abnormalities	hairy-cell leukemia	2.54E-03	DSCAM, PCDH10 all 2	2
Cancer	growth of tumor	2.10E-02	HSPG2, PCDH10 all 2	2
Cancer	squamous-cell carcinoma	2.25E-02	DSCAM, PCDH10 all 2	2
Cancer, Gastrointestinal Disease	gastrointestinal carcinoma	2.51E-02	DSCAM, HSPG2, PCDH10, TECTA all 4	4
Cancer	pelvic cancer	2.93E-02	DSCAM, FREM3, HSPG2, TECTA all 4	4
Cancer, Organismal Injury and Abnormalities	urogenital cancer	3.02E-02	DSCAM, FREM3, HSPG2, TECTA all 4	4
Cancer, Gastrointestinal Disease	oral squamous cell carcinoma	4.65E-02	PCDH10 all 1	1
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	adenocarcinoma in endometrium	4.75E-02	DSCAM, FREM3, TECTA all 3	3
Cancer	endometrioid carcinoma	4.80E-02	DSCAM, FREM3, TECTA all 3	3

Table A6.10 Summary of the results of the Ingenuity Pathway Analysis of the genes identified as mutated in the baseline time point CTCs but not the relapse time point CTCs from the WES data. The changes in cancer-related genes are included in a separate panel as for this analysis they were not included in the main summary produced by the ingenuity software but were discussed in the text of chapter 5.



Top Diseases and Bio Functions

Diseases and Disorders

Name	p-value	# Molecules
Developmental Disorder	3.51E-04 - 3.36E-02	2
Hereditary Disorder	3.51E-04 - 4.95E-02	3
Ophthalmic Disease	3.51E-04 - 1.57E-02	2
Organismal Injury and Abnormalities	3.51E-04 - 6.31E-03	1
Auditory Disease	7.03E-04 - 1.08E-02	1

Molecular and Cellular Functions

Name	p-value	# Molecules
Cell Cycle	1.40E-03 - 1.99E-02	2
Cell-To-Cell Signaling and Interaction	1.40E-03 - 3.50E-02	2
Drug Metabolism	1.40E-03 - 3.50E-02	1
Molecular Transport	1.40E-03 - 3.50E-02	3
Small Molecule Biochemistry	1.40E-03 - 3.50E-02	2

Physiological System Development and Function

Name	p-value	# Molecules
Connective Tissue Development and Function	1.40E-03 - 1.36E-02	1
Tissue Development	1.40E-03 - 2.19E-02	3
Behavior	1.76E-03 - 1.78E-02	1
Embryonic Development	2.11E-03 - 1.71E-02	2
Organ Development	2.11E-03 - 1.71E-02	2

Categories	Diseases or Functions Annotation	p-Value	Molecules	# Molecules
Cancer	abdominal cancer	3.21E-02	FAM181A, FIGN, LTBP2, MYBBP1A, PTPRN2, STARD9, USH2A all 7	7
Cancer	carcinoma	3.36E-02	FAM181A, FIGN, LTBP2, MYBBP1A, PTPRN2, STARD9, USH2A all 7	7
Cancer, Gastrointestinal Disease, Hepatic System Disease	liver cancer	3.37E-02	LTBP2, MYBBP1A, PTPRN2, STARD9, USH2A all 5	5

Table A6.11 Summary of the results of the Ingenuity Pathway Analysis of the genes identified as mutated in the relapse time point CTCs but not the baseline time point CTCs from the WES data.

Top Diseases and Bio Functions

Diseases and Disorders

Name	p-value	# Molecules
Cancer	2.60E-24 - 1.67E-02	164
Gastrointestinal Disease	1.06E-19 - 1.67E-02	133
Hepatic System Disease	1.06E-19 - 1.67E-02	110
Organismal Injury and Abnormalities	1.16E-17 - 1.70E-02	113
Reproductive System Disease	1.16E-17 - 1.67E-02	92

Molecular and Cellular Functions

Name	p-value	# Molecules
Cell Morphology	6.95E-05 - 1.67E-02	18
Cell-To-Cell Signaling and Interaction	4.12E-04 - 1.67E-02	11
Cellular Compromise	4.12E-04 - 1.67E-02	12
Cellular Function and Maintenance	4.12E-04 - 1.67E-02	20
Cell Signaling	6.83E-04 - 1.67E-02	5

Physiological System Development and Function

Name	p-value	# Molecules
Hair and Skin Development and Function	6.95E-05 - 8.36E-03	7
Embryonic Development	8.45E-05 - 1.67E-02	27
Nervous System Development and Function	8.45E-05 - 1.67E-02	31
Organ Development	8.45E-05 - 1.67E-02	25
Organ Morphology	8.45E-05 - 1.67E-02	21

Appendix 7 Reliable and Reproducible Isolation of Both Circulating Free DNA and Circulating Tumour Cells from the Same Preserved Whole Blood Sample - Paper Draft

Draft of Publication In Review at Cancer Epidemiology, Biomarkers and Prevention.

Reliable and reproducible isolation of both circulating free DNA and Circulating Tumour Cells from the same preserved whole blood sample

Dominic G. Rothwell[1], Nigel Smith[1], Daniel Morris[1], Hui Sun Leong[2], Yaoyong Li[2, 3], Antoine Hollebecque[1], Mahmood Ayub[1], Louise Carter[1], Lynsey Franklin[1], Fiona Blackhall[3,] Crispin Miller[2, 4], Caroline Dive[1] and Ged Brady[1]

[1]Clinical and Experimental Pharmacology Group, [2]Computational Biology Support and [4]RNA Biology Group, CR-UK Manchester Institute, University of Manchester, M20 4BX, UK. [3]Faculty of Life Sciences, University of Manchester, Michael Smith Building, Oxford Road, Manchester, M13 9PT, UK. [5]Christie NHS Foundation Trust, Institute of Cancer Sciences, University of Manchester, M20 4BX, UK.

Running title: Stable whole blood processing from a single blood sample delivering CTCs and cfDNA for up to 4 days post blood draw.

Financial support: This work was supported by core funding to CR-UK Manchester Institute (C5759/A12328), and via Manchester CR-UK Centre Award (A12197).

Corresponding author:
Dr Ged Brady
Phone +44 161 4467052
Fax +44 161 4463109
ged.brady@cruk,mancehster.ac.uk

Conflicts of interest: no conflicts of interest

ABSTRACT

Molecular information obtained from cancer patients' blood is an emerging and powerful research tool whose potential as a companion diagnostic for patient stratification and monitoring is being realised. Blood, sampled simply and routinely, provides a means of inferring the current genetic status of patients' tumours via analysis of circulating tumour cells (CTCs) or circulating tumour DNA (ctDNA). Accurate assessment of both CTCs and ctDNA requires all blood cells be maintained intact until samples are processed, particularly when analytes present are at very low concentrations. Here we describe a blood collection protocol that does not require on-site processing, and which is amenable for analysis of both CTCs and ctDNA following transport and storage at ambient temperature in CellSave vacutainers for up to four days after blood collection. We demonstrate that yields of circulating free DNA (cfDNA) obtained from whole blood CellSave samples are equivalent to those obtained from conventional EDTA plasma processed within 4 hours of blood draw. Targeted and genome wide next generation sequencing revealed comparable DNA quality and resultant sequence information from cfDNA within CellSave and EDTA samples. We also demonstrate that CTCs and ctDNA can be isolated from the same patient blood sample, allowing direct comparison of the genetic status of patients' tumours. Our results demonstrate the utility of a simple approach enabling blood collection combined with controlled processing at specialised central laboratories. This avoids site-to-site variability and extends robust CTC and cfDNA analyses to multi-site clinical trials and to clinical sites lacking standardised equipment for blood processing.

INTRODUCTION

Technological advances in blood borne cancer biomarkers now make it possible to routinely analyse RNA and DNA from single cells (1, 2, 3) including isolated circulating tumour cells (CTC)s and the minute amounts of tumour derived DNA present in patient blood samples (reviewed in 4, 5). Circulating free DNA (cfDNA) analysis is emerging as a relatively simple yet powerful biomarker for monitoring disease status and reporting mechanisms of treatment resistance in cancer patients, with the important advantage of being minimally invasive and suitable for longitudinal sampling (6). CTCs have also been shown to be clinically informative with CTC enumeration recognised as a prognostic biomarker by the FDA in metastatic breast, prostate and colorectal cancers (7, 8, 9). More recently, CTCs have been expanded *in vitro* and *in vivo* providing valuable insights into tumour biology (10). For accurate and sensitive analysis of both

CTCs and cfDNA, it is important to ensure that blood collection, transport and processing do not result in cell damage or lysis resulting in loss of CTCs or dilution of cfDNA by lysed white blood cell (WBC) contents. Dilution of ctDNA due to WBC lysis may hinder the ability to detect clinically important tumour mutations, or lead to misleading estimates of the mutant fraction of cfDNA thereby impairing studies of residual disease and emergent mechanisms of treatment resistance. In standard cfDNA protocols, WBC lysis is minimized by preparation of plasma within a short time from the blood draw (typically 1-4h), which may be challenging in non-specialized sites and busy clinics. Recently, the use of dedicated blood collection tubes containing a preservative which allows transport of whole blood at ambient temperature for several days prior to cfDNA isolation has been shown to extend the window within which samples can be used for cfDNA extraction (11). For CTC analysis, the gradual loss of cell integrity with prolonged storage of a standard EDTA blood sample is overcome by using a CellSearch® CellSave Preservation tube. This preserves cells in whole blood for up to 4 days at room temperature and allows international transport of blood samples and a standardised workflow without the need for sample processing at collection sites. Using the CellSearch CellSave system, CTCs can be fluorescently labeled and enumerated (12), isolated and characterized by whole genome sequencing (WGS) (10). Analysis of CTCs and cfDNA from the same whole blood sample would extend the molecular information extracted from a single blood sample and enable a direct comparison of CTC and cfDNA readouts. Here we describe the isolation of both CTCs and cfDNA from CellSave blood samples, followed by genome wide and focused next generation sequencing (NGS) to establish reliable and effective analysis of both CTCs and cfDNA from whole blood transported up to 4 days at ambient temperature. This will allow participation of non-specialised clinical sites to ship blood samples to central laboratories for processing and expert analysis of ‘liquid biopsies’, reducing time required for blood processing in busy clinics and minimizing variability in the resultant molecular data obtained

MATERIAL AND METHODS

cfDNA preparation and quantification

Plasma was separated from whole blood by means of two sequential centrifugations (2,000g, 10 minutes) and stored at -80°C in 1 ml aliquots. cfDNA was isolated from 1ml of double spun plasma using the QIAamp Circulating Nucleic Acid Kit (Qiagen) as per

manufacturer's instructions. Following isolation the cfDNA yield was quantified using the TaqMan® RNase P Detection Kit (Life Technologies).

NGS library preparation and sequencing

Whole genome sequencing (WGS) of CTC derived explant tumours (CDX), CTCs and WBCs was carried out as previously described (10). Focused NGS of samples with a minimum of 8 ng cfDNA was performed using the Qiagen GeneRead system as described by the manufacturer except input was reduced to as low as 8 ng DNA (ensuring ≥2 ng input into each of the 4 Qiagen GeneRead multiplex PCR reactions). WGS of cfDNA was carried out using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® kit using 5 ng DNA input. NGS for both focused GeneRead libraries and WGS cfDNA libraries was carried out using an Illumina® MiSeq desktop sequencer.

Targeted NGS analysis

Analysis of the GeneRead NGS data was performed on the Qiagen Cloud-Based DNAseq Sequence Variant Analysis software according to the manufacturer's instructions. For WGS, analysis paired-end sequence reads were aligned to the human reference genome GRCh37/hg19 using the Burrows-Wheeler alignment tool (BWA, version 0.7.4) with default parameters and the BWA-MEM algorithm. The alignments were sorted and indexed by chromosome coordinates using SAMtools (version 0.1.19), followed by PCR duplicates removal using Picard tools MarkDuplicates function (version 1.96) (http://picard.sourceforge.net). Single nucleotide variants (SNVs) were identified using VarScan2 (version 2.3.7) with the following settings: min-coverage=8, min-reads2=2, min-avg.qual=15, min-var-freq=0.01, p-value=0.01.

Copy Number Aberration analysis from WGS data

Paired-end sequence reads were aligned to the human reference genome GRCh37/hg19 using SMALT aligner (version 0.7.1, http://www.sanger.ac.uk/resources/software/smalt/). SMALT index was built by setting k=20 and s=13. The alignments were sorted and indexed by chromosome coordinates using SAMtools (version 0.1.18). Copy number variations were predicted by using Control-FREEC(version 6.4) with the following settings: coefficientOfVariation = 0.1, ploidy = 2, mateOrientation = FR. The samples were clustered hierarchically by their

copy number profiles based on the Euclidean distance and the Ward linkage method in R.

Evaluation of NGS error rates

Two metrics were used to infer mutation rate in the CellSave and EDTA samples: the first was calculated as the number of SNV detected divided by total number of bases in the pileup file; the second metric was calculated by dividing the number of SNV detected by number of bases with at least 8x coverage in the pileup file.

To account for variation in sequencing depth between samples, we performed 100 down-sampling of the aligned data, keeping 1 million read pairs in each iteration. We re-calculated the mutation rates by averaging the output from all iterations. A two-tailed *t*-test was performed to assess if the mutation rate is significantly different between CellSave and EDTA samples.

RESULTS & DISCUSSION

Isolation of cfDNA from EDTA and CellSave HNV blood samples

Our objective was to evaluate the 'real life' utility of CellSave preserved whole blood collection for analysis of cfDNA and CTCs as applied to blood samples obtained in multiple sites and shipped to a centralised laboratory for analysis. To determine the effect of WBC lysis on cfDNA yields following long term storage (> 24 hours) of whole blood in EDTA, we isolated plasma from blood within 1 hour of collection in a standard EDTA vacutainers tube and then at 24, 48, 72 and 96 hours post-draw. Following isolation, the cfDNA yield was determined using the RNAseP real-time PCR assay (Figure 1A). Increasing amounts of cfDNA were detected over time, with almost a 3-fold increase seen by 24-hours post-draw, increasing to over 60 fold by 96 hours, which could reduce the ability to analyse the ctDNA fraction within clinical samples.

To evaluate the suitability of using CellSave to reduce WBC lysis and facilitate cfDNA analysis, we undertook a 20 healthy normal volunteers (HNV) study where each HNV donated two EDTA and two CellSave samples. For each HNV donor cfDNA was isolated from one EDTA and one CellSave tube within 4 hours post blood draw (isolation range 2.0 – 3.3 hours, mean = 2.8 hours). The remaining EDTA and CellSave tubes were sent through the British postal system back to the host institute using a Royal Mail Safe Box™, then maintained at ambient temperature storage for up to 96 hours post-draw (isolation range 93.3 – 95.3 hours, mean = 94.5 hours) (Figure 1B). The yield of cfDNA from all samples was determined using an RNAseP real-time

PCR assay, and showed no significance difference between the 4 hour EDTA, 4 hour CellSave and 96 hour CellSave samples (Figure 1C). As expected, a significant increase in cfDNA was seen in the 96 hour EDTA sample compared to both the CellSave samples and the 4 hour EDTA sample, reflecting extensive WBC lysis. These results are consistent with both an effective reduction in WBC lysis and efficient cfDNA isolation from whole blood CellSave samples kept at ambient temperature for up to 96 hours, allowing time for shipment of samples in a multi-site setting to the analysis laboratory.

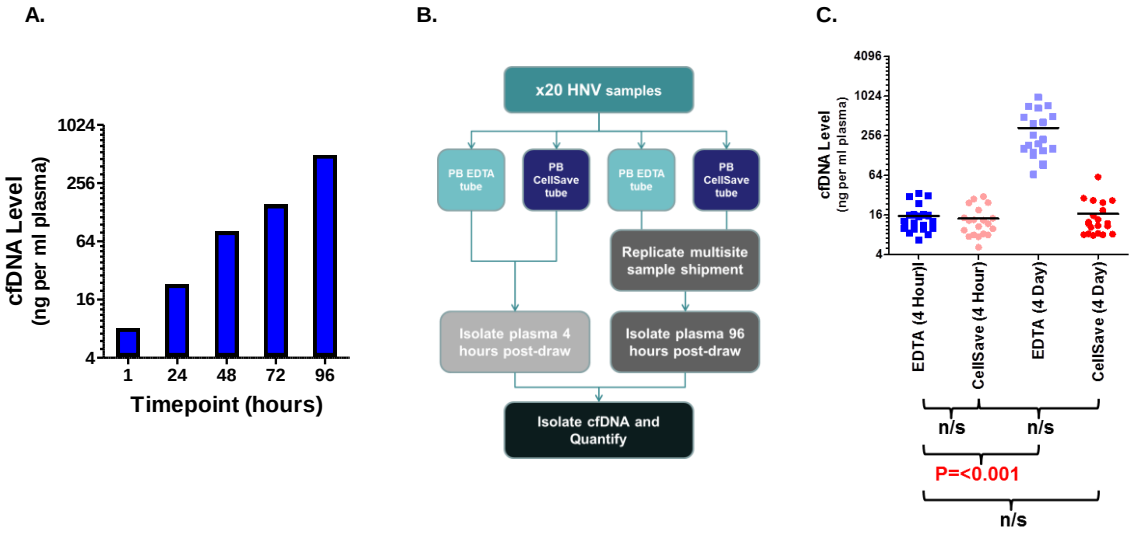


Figure 1. A. Graph showing increase in cfDNA levels in plasma from EDTA blood left at room temperature for up to 96 hours post-draw. **B.** Schematic of EDTA and CellSave cfDNA stability study. **C.** cfDNA yields from 20 HNV blood samples collected in EDTA or CellSave and processed either 4 hours or 96 hours post-draw. No significant difference in overall yields between the 4 hour EDTA, 4 hour CellSave and 96 hour CellSave samples with a highly significant increase in cfDNA yield following 96 hours in EDTA

Evaluation of EDTA and CellSave cfDNA NGS error rates

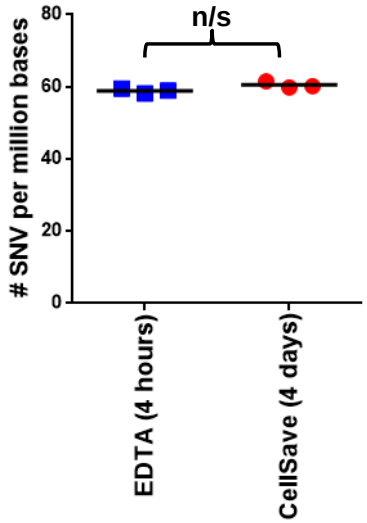
Although the CellSave preservative significantly reduced the level of WBC lysis, thereby maintaining the ctDNA fraction within samples, it is possible that the components of the CellSave tube could act as a DNA damaging agent and effectively increase background sequencing errors. To test, this standard EDTA and 96 hour CellSave cfDNA samples from the 20 HNV were pooled and subjected to WGS. To estimate the overall mutation burden low pass WGS Illumina MiSeq sequencing data were generated from three technical replicates of each pool. Over $1.5x10^8$ bases were sequenced for each library with approximately $9.5x10^3$ single nucleotide variants (SNVs) identified per sample when analysed against the Hg19 genome (Figure 2A). No significant difference was found between the mutation rates of the CellSave (60.4 SNV per million bases) compared to the EDTA samples (58.9 SNV per million bases)

indicating CellSave cfDNA is compatible with extended NGS strategies (Figure 2B). Analysis of the types of SNV detected within the cfDNA in each collection tubes was also performed, with similar frequencies of transitions and transversions seen in both sample type suggesting no effect of CellSave preservative on cfDNA integrity (Figure 2C).

A.

Sample	Bases in pileup	SNP predicted value (p-value<0.01)	# SNV per Million bases
EDTA 1	157057344	9345	59.501
EDTA 2	158379699	9210	58.152
EDTA 3	157007320	9249	58.908
CellSave 1	161532233	9725	60.205
CellSave 2	159561656	9816	61.519
CellSave 3	161275644	9649	59.829

B.



C.

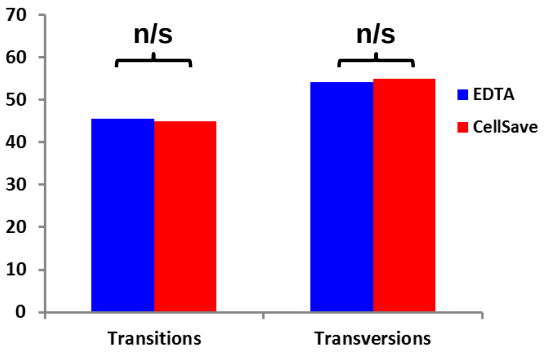


Figure 2. A. Bases sequenced, SNVs identified and the number of SNVs per million bases read for three EDTA and three CellSave cfDNA whole genome NGS libraries. **B.** Number of single nucleotide variations identified in a pool of HNV cfDNA prepared from either EDTA processed up to 4 hours post blood draw and CellSave processed 96 hours post blood draw. There was no significant difference in SNVs per million bases for the EDTA and CellSave cfDNA samples. **C.** Repertoire of mutations detected in each collection with equal frequencies of transistions and transversions seen in both EDTA and CellSave samples.

Isolation of cfDNA from EDTA and CellSave patient blood samples

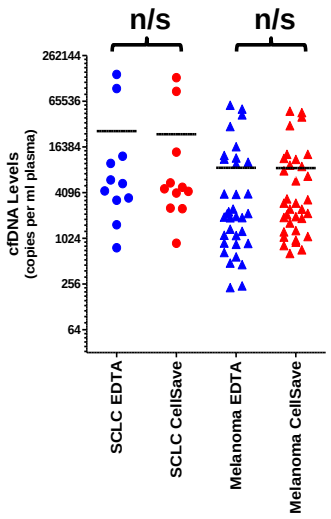
CellSave vacutainers are routinely used for CTC enumeration using the CellSearch® platform and molecular analysis of CTCs retrieved from CellSearch® cartridges can be achieved using both focused and genome wide NGS (10, 13, 14). Since the CellSearch® system requires 7.5 ml blood input and the CellSave vacutainer can hold up to 10 ml there is often surplus blood, which can be used for additional analyses. To test the suitability of CellSave for cfDNA analysis of clinical samples, we compared

yields of cfDNA obtained from surplus CellSave blood to yields of cfDNA obtained from sample obtained from a parallel EDTA blood sample processed to plasma within 4 hours. Analysis of 11 SCLC and 34 melanoma patient samples showed comparable yields of patient cfDNA from 4 hour EDTA plasma (hereafter referred to as standard EDTA) to cfDNA isolated from CellSave blood kept at room temperature for up to 96 hours (Figure 3A).

Targeted NGS of EDTA and CellSave cfDNA

To test the suitability of CellSave cfDNA for targeted NGS analysis of clinical samples, 5 of the 11 SCLC patients with above 8ng of cfDNA available for both standard EDTA and CellSave cfDNAs were analysed using the Qiagen GeneRead Lung Cancer Panel. Analysis of the NGS data was carried out and compared to a corresponding germline sample from each patient for each EDTA and CellSave sample following cfDNA isolation and quantification (Figure 3B). In keeping with the high frequency of TP53 mutations in SCLC (15, 16), somatic TP53 mutations were identified in 4 of the 5 SCLC patients analysed with essentially identical results observed for both EDTA and CellSave matched samples (Figure 3B). For 1 patient (SCLC-03) who did not harbour a detectable TP53 mutation, an ALK mutation was detected, again with similar levels seen for both EDTA and CellSave matched samples. Although ALK mutations at this locus have not been previously reported, low frequency ALK translocations have been observed in SCLC (17) raising the possibility that the detected mutation is involved in the pathology of the disease. For patient SCLC-05, in addition to a TP53 mutation, a second lower frequency mutation in ERBB2 was also consistently identified in both EDTA and CellSave samples suggesting possible tumour heterogeneity within this patient. In summary, targeted NGS of matched EDTA and CellSave patient cfDNA samples identified the same mutations with similar tumour allele frequencies (Figure 3B) supporting the suitability of CellSave whole blood collection for cfDNA molecular analysis of clinical samples.

A.



B.

			TP53 7578212 G>A Stop gain		TP53 7577022 G>A Stop gain		TP53 7577574 A>G Y236C		TP53 7578281 C>A Stop P190T		ERBB2 37884019 G>T G1164C		ALK 29416481 A>G K1491R	
Patient Sample	Sample type	cfDNA input (ng/rxn)	Mutation Detected	Mutation frequency	Mutation Detected	Mutation frequency	Mutation Detected	Mutation frequency	Mutation Detected	Mutation frequency	Mutation Detected	Mutation frequency	Mutation Detected	Mutation frequency
SCLC-01	WT gDNA	20.0	WT	NA	WT	NA	WT	NA	WT	NA	WT	NA	WT	NA
	EDTA cfDNA	10.4	Mut	0.47	WT	NA	WT	NA	WT	NA	WT	NA	WT	NA
	CellSave cfDNA	14.7	Mut	0.48	WT	NA	WT	NA	WT	NA	WT	NA	WT	NA
SCLC-02	WT gDNA	2.0	WT	NA	WT	NA	WT	NA	WT	NA	WT	NA	WT	NA
	EDTA cfDNA	2.0	WT	NA	WT	NA	Mut	0.26	WT	NA	WT	NA	WT	NA
	CellSave cfDNA	2.0	WT	NA	WT	NA	Mut	0.31	WT	NA	WT	NA	WT	NA
SCLC-03	WT gDNA	2.0	WT	NA	WT	NA	WT	NA	WT	NA	WT	NA	WT	NA
	EDTA cfDNA	2.0	WT	NA	WT	NA	WT	NA	WT	NA	WT	NA	Mut	0.34
	CellSave cfDNA	2.0	WT	NA	WT	NA	WT	NA	WT	NA	WT	NA	Mut	0.42
SCLC-04	WT gDNA	20.0	WT	NA	WT	NA	WT	NA	WT	NA	WT	NA	WT	NA
	EDTA cfDNA	22.5	WT	NA	Mut	0.63	WT	NA	WT	NA	WT	NA	WT	NA
	CellSave cfDNA	19.5	WT	NA	Mut	0.65	WT	NA	WT	NA	WT	NA	WT	NA
SCLC-05	WT gDNA	2.0	WT	NA	WT	NA	WT	NA	WT	NA	WT	NA	WT	NA
	EDTA cfDNA	2.0	WT	NA	WT	NA	WT	NA	Mut	0.79	Mut	0.08	WT	NA
	CellSave cfDNA	2.0	WT	NA	WT	NA	WT	NA	Mut	0.71	Mut	0.09	WT	NA

Figure 3. A. Yields of cfDNA from duplicate clinical samples collected in EDTA and CellSave bloods from a cohort of 11 SCLC and 34 melanoma patients. No significant difference was found between each collection type. **B.** Mutations identified in five SCLC patient samples using a targeted NGS approach. Germline gDNA, EDTA cfDNA and CellSave cfDNA was analysed for each patient. Mutations were called with read counts >200 and frequency >10%. Mutated samples are indicated by red fill with WT alleles indicated by green fill.

Whole Genome Copy Number Alteration (CNA) of Matched cfDNA and CTCs

As well as the identification of tumour associated mutations, low depth whole genome sequencing (WGS) of cfDNA can be used to characterise CNA patterns arising from the circulating tumour DNA (ctDNA) present in the total cfDNA (18, 19). Since we and others have shown CNA analysis can be readily applied to CTCs isolated following CellSearch enrichment (10, 13, 14), use of CellSave for cfDNA isolation would enable combined CTC and cfDNA analysis from the same collection tube (Figure 4A), maximizing the potential clinical information that can be elucidated, facilitating direct comparison of the two as potential liquid biopsies from a single blood sample. This also allows an evaluation of the importance of determining whether genetic alterations

picked up by ctDNA assessment are co-expressed in single CTCs. To establish combined CTC and cfDNA analysis, 7.5 ml of a CellSave whole blood sample was used for CTC isolation via CellSearch and DEParray as previously described (10) and the remaining CellSave blood (typically 1-2.5 mL) was used to prepare cfDNA. WBCs were used as a germline control for CTC CNA analysis and WGS of whole blood DNA served as a germline control for the cfDNA samples. For one patient, we were also previously able to generate CDX tumours in an immune-compromised mouse following CTC enrichment of a parallel EDTA blood sample (10), enabling us to compare CNA patterns from both CTCs and cfDNA obtained from the corresponding CellSave blood sample. Figure 4 shows the comparison of CNA profiles generated from isolated CTCs, EDTA cfDNA, CellSave cfDNA, two CDX tumours, germline gDNA and isolated WBC DNA from 2 SCLC patients. The results show a clear tumour related CNA patterns in matched CTC, CDX and cfDNA with similar patterns seen for both CellSave and EDTA cfDNA. The pattern of gain and loss in the two CDX tumours in patient 1 (Figure 4B) are consistent with previously published studies on CNA in SCLC (15, 16) with regions containing RASSF1 and FHIT being lost and regions containing SOX2 and BCL2 showing amplification. The CDX tumours also show amplification of regions of chromosomes 2 and 14, with this pattern also observed in both CTCs and all cfDNA samples. In patient 2 (Figure 4C) there was no CDX tumour available, but regions of loss and gain in the CTCs correspond well with published data, including loss of chromosome 17 (TP53) and amplification of chromosome 3 (SOX2). A similar pattern of loss and gain is also seen in the CNA of the cfDNA samples, with good correlation between the EDTA and CellSave samples showing CellSave cfDNA to be suitable for NGS CNA and compatible with combined CTC collection and analysis.

In summary, we have demonstrated the suitability of whole blood CellSave samples for both CTC and cfDNA molecular analysis. The ability to generate informative molecular profiles of both CTCs and cfDNA from a simple whole blood sample shipped at ambient temperature for up to 4 days represents a significant methodological improvement for clinical benefit. The ability to process samples at a single recipient site avoids site-to-site variability, a major confounding issue in cfDNA analysis (20). Furthermore, the use of a simple blood collection protocol does not require specialised equipment, such as centrifuges or even refrigeration, extending the number of clinical sites that can participate in patient evaluation via liquid biopsies to anywhere where a blood draw is taken. For example, following initial cancer therapy, patients in remission can be monitored via a blood draw at a local medical practice rather than necessitating often lengthy/expensive trips to a specialised oncology centre. Furthermore, use of whole

blood transport for cfDNA and CTC analysis would be of particular benefit in underdeveloped countries where local blood processing will be impractical. In September 2015, the first ctDNA companion diagnostic assessing EGFR mutation for patient stratification was approved by the EDA (21). We posit that as minimally invasive, liquid biopsies become increasingly employed for cancer patient management, the ability to routinely and simply draw blood and ship samples to accredited biomarker assessment laboratories will facilitate the dawn of this new development in the delivery of personalised cancer medicines.

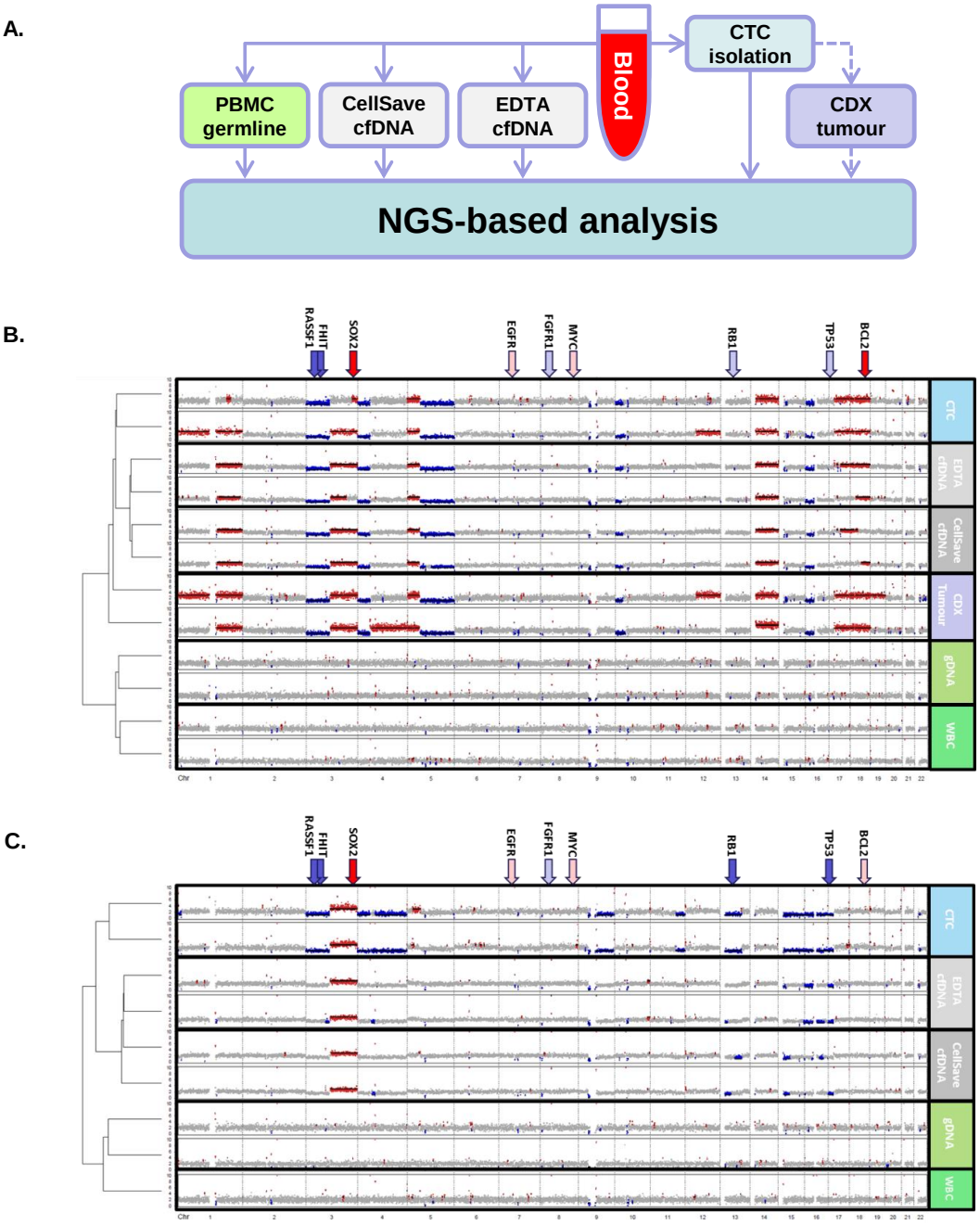


Figure 4. A. Schematic showing procedure for processing of a single blood sample to give NGS analysis of cfDNA and CTCs**. B & C.** Unsupervised, hierarchical clustering of CNA profiles in

two SCLC patients. CNA profiles were generated from isolated CTCs, EDTA cfDNA, CellSave cfDNA, two CDX tumours (A only), germline gDNA and isolated WBC. Matching patterns of gain (regions of red) and loss (regions of blue) were seen across all tumour material and were absent from germline controls. Arrows indicate location of common copy number aberrations found in SCLC with red indicating gain and blue loss. Dark filled arrows indicate loci altered in the patient sample.

ACKNOWLEDGMENTS

We would like to acknowledge the help of the CR-UK Molecular Biology Core Facilities and members of the Clinical and Experimental Pharmacology Group for their support of this study. We also thank the patients and healthy volunteers who provided their blood samples for this study.

AUTHORS' CONTRIBUTIONS

GB devised the study and co-wrote the manuscript. DGR co-wrote the manuscript, planned and ran cfDNA NGS experiments. YL and HS performed the bioinformatic analysis. NS devised and ran the NHV study. DM isolated the patient cfDNA. AH and MA generated and ran some of the cfDNA NGS experiments. LC collected clinical samples and analysed data. FHB and LK oversaw ethical permission and patient consent for blood samples. CM and CD designed and supervised the project. All authors read and approved the final manuscript

REFERENCES

1. Rothwell DG, Li Y, Ayub M, Tate C, Newton G, Hey Y, *et al.* Evaluation and validation of a robust single cell RNA-amplification protocol through transcriptional profiling of enriched lung cancer initiating cells. BMC Genomics. 2014;15:1129. doi: 10.1186/1471-2164-15-1129.

2. Ramsköld D1, Luo S, Wang YC, Li R, Deng Q, Faridani OR, *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol. 2012 Aug;30(8):777-82.

3. Gužvić M, Braun B, Ganzer R, Burger M, Nerlich M, Winkler S, *et al.* Combined genome and transcriptome analysis of single disseminated cancer cells from bone marrow of prostate cancer patients reveals unexpected transcriptomes. Cancer Res. 2014 Dec 15;74(24):7383-94. doi: 10.1158/0008-5472.CAN-14-0934. Epub 2014 Oct 15.

4. Krebs MG, Metcalf RL, Carter L, Brady G, Blackhall FH, Dive C. Molecular analysis of circulating tumour cells-biology and biomarkers. Nat Rev Clin Oncol. 2014 Mar;11(3):129-44. doi: 10.1038/nrclinonc.2013.253. Epub 2014 Jan 21.

5. Diaz LA Jr1, Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. J Clin Oncol. 2014 Feb 20;32(6):579-86. doi: 10.1200/JCO.2012.45.2011. Epub 2014 Jan 21.

6. Murtaza M, Dawson SJ, Tsui DW, Gale D, Forshew T, Piskorz AM, *et al.* Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature. 2013 May 2;497(7447):108-12. doi: 10.1038/nature12065. Epub 2013 Apr 7.

7. Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Matera J, Miller MC, *et al.* Circulating tumor cells, disease progression, and survival in metastatic breast cancer. N Engl J Med. 2004 Aug 19;351(8):781-91.

8. de Bono JS, Scher HI, Montgomery RB, Parker C, Miller MC, Tissing H, *et al.* Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. Clin Cancer Res. 2008 Oct 1;14(19):6302-9.

9. Cohen SJ, Punt CJ, Iannotti N, Saidman BH, Sabbath KD, Gabrail NY, *et al.* Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. J Clin Oncol. 2008;26(19):3213-3221.

10. Hodgkinson CL, Morrow CJ, Li Y, Metcalf RL, Rothwell DG, Trapani F, *et al.* Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. Nat Med. 2014 Aug;20(8):897-903. doi: 10.1038/nm.3600. Epub 2014 Jun 1.

11. Norton SE, Luna KK, Lechner JM, Qin J, Fernando MR. A new blood collection device minimizes cellular DNA release during sample storage and shipping when compared to a standard device. J Clin Lab Anal. 2013 Jul;27(4):305-11. doi: 10.1002/jcla.21603.

12. Hou JM, Krebs MG, Lancashire L, Sloane R, Backen A, Swain RK, *et al.* Clinical significance and molecular characteristics of circulating tumor cells and circulating tumor microemboli in patients with small-cell lung cancer. J Clin Oncol. 2012 Feb 10;30(5):525-32. doi: 10.1200/JCO.2010.33.3716. Epub 2012 Jan 17

13. Gasch C, Bauernhofer T, Pichler M, Langer-Freitag S, Reeh M, Seifert AM, *et al.* Heterogeneity of epidermal growth factor receptor status and mutations of KRAS/PIK3CA in circulating tumor cells of patients with colorectal cancer. Clin Chem. 2013 Jan;59(1):252-60. doi: 10.1373/clinchem.2012.188557. Epub 2012 Nov 7

14. Heitzer E, Auer M, Gasch C, Pichler M, Ulz P, Hoffmann EM, *et al.* Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing. Cancer Res. 2013 May 15;73(10):2965-75. doi: 10.1158/0008-5472.CAN-12-4140. Epub 2013 Mar

15. Peifer M, Fernández-Cuesta L, Sos ML, George J, Seidel D, Kasper LH, *et al.* Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. Nat Genet. 2012 Oct;44(10):1104-10. doi: 10.1038/ng.2396. Epub 2012 Sep 2.

16. Rudin CM, Durinck S, Stawiski EW, Poirier JT, Modrusan Z, Shames DS, *et al.* Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. Nat Genet. 2012 Oct;44(10):1111-6. doi: 10.1038/ng.2405. Epub 2012 Sep 2.

17. Toyokawa G, Takenoyama M, Taguchi K, Toyozawa R, Inamasu E, Kojo M, *et al.* An extremely rare case of small-cell lung cancer harboring variant 2 of the EML4-ALK fusion gene. Lung Cancer. 2013 Sep;81(3):487-90.

18. Leary RJ, Sausen M, Kinde I, Papadopoulos N, Carpten JD, Craig D, *et al.* Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. Sci Transl Med. 2012 Nov 28;4(162):162ra154. doi: 10.1126/scitranslmed.3004742.

19. Mohan S, Heitzer E, Ulz P, Lafer I, Lax S, Auer M, *et al.* Changes in colorectal carcinoma genomes under anti-EGFR therapy identified by whole-genome plasma DNA sequencing. PLoS Genet. 2014 Mar 27;10(3):e1004271. doi: 10.1371/journal.pgen.1004271. eCollection 2014.

20. Gormally E, Hainaut P, Caboux E, Airoldi L, Autrup H, Malaveille C, *et al.* Amount of DNA in plasma and cancer risk: a prospective study. Int J Cancer. 2004 Sep 20;111(5):746-9. Erratum in: Int J Cancer. 2006 Jul 15;119(2):476.

21. Douillard JY, Ostoros G, Cobo M, Ciuleanu T, McCormack R, Webster A, *et al.* First-line gefitinib in Caucasian EGFR mutation-positive NSCLC patients: a phase-IV, open-label, single-arm study. Br J Cancer. 2014 Jan 7;110(1):55-62.

Appendix 8 Tumorigenicity and Genetic Profiling of Circulating Tumor Cells in Small Cell Lung Cancer - Publication Reprint

Publication Reprint

Hodgkinson C, Morrow C, Li Y, Metcalf R, Rothwell D, Trapani F, Polanski R, Burt D, Simpson K, Morris K, Pepper S, Nonaka D, Greystoke A, Kelly P, Bola B, Krebs M, Antonello J, Faulkner S, Priest L, **Carter L**, Tate C, Miller C, Blackhall F, Brady G, Dive C. Tumorigenicity and Genetic Profiling of Circulating Tumor Cells in Small Cell Lung Cancer. Nature Medicine Aug 2014.

This article covers pages 306 to 314 of this thesis.

nature medicine

Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer

Cassandra L Hodgkinson[1,7], Christopher J Morrow[1,7], Yaoyong Li[2], Robert L Metcalf[1], Dominic G Rothwell[1], Francesca Trapani[1], Radoslaw Polanski[1], Deborah J Burt[1], Kathryn L Simpson[1], Karen Morris[1], Stuart D Pepper[3], Daisuke Nonaka[4], Alastair Greystoke[1,4,5], Paul Kelly[1], Becky Bola[1], Matthew G Krebs[1], Jenny Antonello[1], Mahmood Ayub[1], Suzanne Faulkner[1], Lynsey Priest[1], Louise Carter[1], Catriona Tate[1], Crispin J Miller[2,6], Fiona Blackhall[4,5,8], Ged Brady[1,8] & Caroline Dive[1,8]

Small-cell lung cancer (SCLC), an aggressive neuroendocrine tumor with early dissemination and dismal prognosis, accounts for 15–20% of lung cancer cases and ~200,000 deaths each year. Most cases are inoperable, and biopsies to investigate SCLC biology are rarely obtainable. Circulating tumor cells (CTCs), which are prevalent in SCLC, present a readily accessible 'liquid biopsy'. Here we show that CTCs from patients with either chemosensitive or chemorefractory SCLC are tumorigenic in immune-compromised mice, and the resultant CTC-derived explants (CDXs) mirror the donor patient's response to platinum and etoposide chemotherapy. Genomic analysis of isolated CTCs revealed considerable similarity to the corresponding CDX. Most marked differences were observed between CDXs from patients with different clinical outcomes. These data demonstrate that CTC molecular analysis via serial blood sampling could facilitate delivery of personalized medicine for SCLC. CDXs are readily passaged, and these unique mouse models provide tractable systems for therapy testing and understanding drug resistance mechanisms.

Improved treatment outcomes for patients with SCLC require new approaches to interrogate the biology and genetics of this disease, appropriate methods to investigate resistance to current chemotherapy and tractable, patient-derived, clinically relevant models to test new therapies. Moreover, minimally invasive monitoring of patients with SCLC is needed to optimize therapy selection. We sought to address these issues by developing unique patient-derived mouse models exploiting the abundant CTCs in patients with SCLC and testing their response to standard platinum and etoposide chemotherapy. In parallel, we validated CTC profiling for patient monitoring.

The high response rates, including complete responses, to platinum-based chemotherapy regimens for SCLC in the 1970s and 1980s[1–3] led to the belief that cures might soon follow. Four decades later, 5-year survival rates for patients with SCLC remain at 5% owing to inherent or, most commonly, acquired treatment resistance. SCLC cell lines were amongst the first cancer cell lines developed and used for drug testing[4,5], and the NCI-H209 line was amongst the first cancer cell lines to be deep sequenced[6], accelerating knowledge of SCLC biology. However, although many hypotheses were generated using cell lines, they were not upheld in the clinic[7]. Trials of targeted therapies in SCLC have proved universally disappointing, with no major advances since the advent of cisplatin and etoposide treatment[8]. The frequent, rapid and marked biological transition from chemotherapy-sensitive to chemotherapy-resistant disease suggests that much is unknown regarding drivers of acquired chemotherapy resistance in SCLC. A genetically engineered mouse model of SCLC developed via conditional deletion of the tumor suppressor genes *Trp53* and *Rb1* allowed new insights into SCLC progression[9], but it has not proven amenable to pharmacology-based studies[10].

A major barrier to comprehensive understanding of human SCLC biology and discovery of 'druggable' targets is that access to fresh, sufficient tumor tissue for research is rare. This most aggressive neuroendocrine tumor has a short doubling time and high growth fraction and disseminates early such that surgery is rarely performed. Explant models derived from patients with SCLC (patient derived xenografts, PDXs) exist[11,12], but the poor take rate and low frequency with which tumor biopsies are obtained, along with their typically small size and high necrosis content[13], make this approach challenging. In the absence of sequential biopsies, the molecular basis of acquired drug resistance in SCLC has yet to be interrogated comprehensively. Comparison of SCLC PDXs and derived cell lines indicates that swift and irreversible changes in gene expression occur in the latter[11]. Next-generation

[1]Clinical and Experimental Pharmacology Group, Cancer Research UK Manchester Institute, University of Manchester, Manchester, UK. [2]Computational Biology Support Group, Cancer Research UK Manchester Institute, University of Manchester, Manchester, UK. [3]Molecular Biology Core Facility, Cancer Research UK Manchester Institute, University of Manchester, Manchester, UK. [4]The Christie NHS Foundation Trust, Manchester, UK. [5]Institute of Cancer Sciences, University of Manchester, Manchester, UK. [6]RNA Biology Group, Cancer Research UK Manchester Institute, University of Manchester, Manchester, UK. [7]These authors contributed equally to this work. [8]These authors jointly supervised this work. Correspondence should be addressed to C.D. (caroline.dive@cruk.manchester.ac.uk).

Received 17 March; accepted 16 May; published online 1 June 2014; doi:10.1038/nm.3600

Table 1 Clinical characteristics of donors with SCLC and subsequent generation of CTC tumors in immunocompromised mice

Patient	CTC count per 7.5 ml[a]	Metastatic sites	Chemosensitive/refractory	Patient survival (months)[b]	Time from CTC implantation to palpable tumor (months)
1	458	Bone, lung, lymph node	Sensitive	7.3	4.4
2	1,625	Bone, brain, meningeal	Refractory	3.5	2.4
3	507	Lymph node	Sensitive	9.7	4.3
4	1,376	Liver, lung, lymph node	Refractory	0.9	3.2
5	222	Liver, lymph node, pancreas	Refractory	1.7	No tumor at 13.7 (mouse culled)
6	20	Lymph node, pleura	Sensitive	13.4[c]	No tumor at 12.3 (mouse culled)

[a]CTC count performed on CellSearch platform. [b]From date of CTC sample blood draw. [c]Patient alive at last follow-up.

sequencing (NGS) technology recently expanded the genomic landscape of SCLCs beyond the nearly ubiquitous inactivation of the tumor suppressor genes *TP53* and *RB1*, revealing high mutation rates and frequent alterations in regulators of histone modification that underpin genomic instability and tumor heterogeneity[6,14,15].

With a new approach, we sought to develop patient-derived *in vivo* models of SCLC that would inform our understanding of metastatic disease and that would be sufficiently tractable for therapy testing. We demonstrated CTCs (expressing epithelial cell adhesion molecule (EpCAM) and cytokeratins) detected using the CellSearch platform were highly prevalent in patients with SCLC compared to other cancers and that CTC number was of prognostic significance[16–21]. We reasoned that tumor-initiating cells must be present within the CTC population. Here, we present the first formal demonstration, to our knowledge, that CTCs from chemotherapy-naive patients with extensive-stage metastatic SCLC are tumorigenic in immunocompromised mice and show that CDXs faithfully recapitulate response to cisplatin and etoposide treatment of donor patients. We also report the first direct genomic comparison of single CTCs directly isolated from patient blood and the resultant matched CDXs obtained following transplant into mice.

RESULTS

CTCs from patients with SCLC are tumorigenic

Blood samples were obtained from six patients with chemotherapy-naive, extensive-stage SCLC (two males, four females) who presented between August 2012 and February 2013. Details on patient selection are in the Online Methods. All patients had a tobacco smoking history (mean 47 pack-years; s.d. 24). Median age was 69 years (range 56–78 years), and patients were performance status 1–3. Three patients (1, 3 and 6) were subsequently chemotherapy sensitive, and three patients (2, 4 and 5) had progressive disease within 3 months of completion of chemotherapy (**Table 1**), defined as refractory disease[22].

To establish whether patients' CTCs could form tumors in immunocompromised mice, we enriched the blood from each patient (10 ml) for CTCs and injected it into one or both flanks of a non-obese diabetic (NOD) severe combined immunodeficient (SCID) interleukin-2 receptor γ–deficient (NSG) mouse. The number of epithelial CTCs ($EpCAM^+cytokeratin^+$) implanted was estimated in a paired 7.5-ml blood sample by CellSearch (**Table 1**). CTCs from patients 1–4 generated tumors in mice (termed CDXs 1–4, respectively). We detected palpable tumors within 4 months of implantation with doubling times ranging from 5 to 21 d (**Fig. 1a–c**). CTC number in the paired blood sample correlated with time to palpable tumor (**Supplementary Fig. 1**). CTC numbers were higher in chemorefractory as compared to chemosensitive patients whose samples gave rise to CDX, and the resulting CDX grew faster (**Supplementary Fig. 1**). CellSearch CTC number in patients whose blood samples gave rise to CDX were all >400 CTCs per 7.5 ml (**Table 1**). In contrast, the CellSearch CTC counts for patients 5 and 6, whose blood samples have not generated CDXs, were 222 and 20 in 7.5 ml, respectively.

CDXs represent clinical SCLC

We assessed the histopathology and immunohistochemistry (IHC) of CDXs in comparison to their corresponding clinical specimens (patient 1: fine needle aspirate from the subcarinal lymph node; patient 2: pleural fluid cytology; patient 3: tracheal biopsy; patient 4: bronchial biopsy). We observed typical SCLC morphology[23] in both diagnostic specimens and CDXs, with clusters and sheets of densely packed small round or oval cells with scant cytoplasm, enlarged hyperchromatic

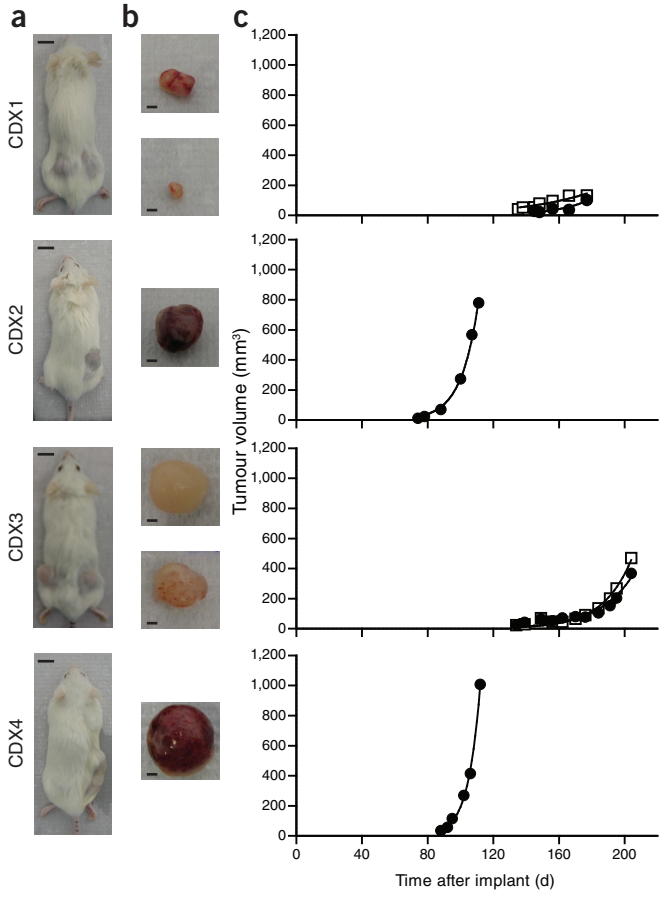


Figure 1 SCLC CTCs are tumorigenic. CTCs enriched from patients with SCLC were injected into mice. Mice carrying CDX1 and CDX3 were injected on both flanks and mice carrying CDX2 and CDX4 on the right flank. (**a**) Tumor-bearing mice. Scale bar, 1 cm. (**b**) Tumors at death. Scale bar, 2 mm. (**c**) Tumor volume over time after implant. Black circles, right tumor; white squares, left tumor; solid line, exponential growth line of best fit. Data show the four passage-1 mice and resultant six CDX.

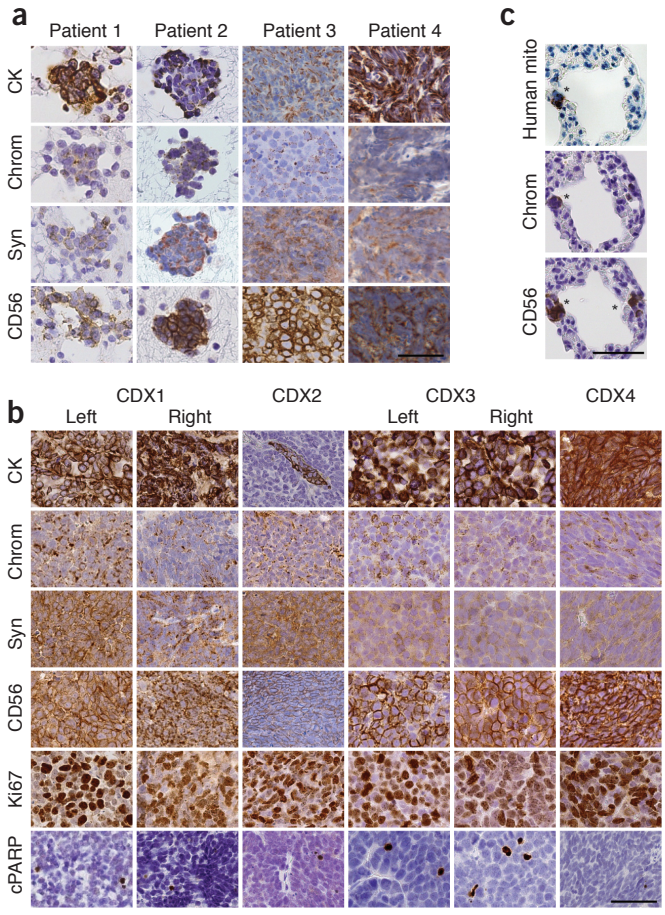


Figure 2 CDXs and mouse micrometastases are representative of patient specimens. (**a**,**b**) Patient specimens (**a**) and CDXs (**b**) stained for cytokeratins (CK), chromogranin A (Chrom), synaptophysin (Syn) and CD56. CDXs were also stained for Ki67 and cleaved PARP (cPARP). (**c**) Lungs from a mouse bearing CDX2 stained for human mitochondria (Human mito), chromogranin A and CD56. Asterisks indicate SCLC micrometastases. Scale bars, 50 μm. Data show the six passage-1 CDXs and images of lungs from mice bearing passage-1 CDX2.

nuclei, inconspicuous nucleoli, speckled chromatin and focal nuclear molding (**Fig. 2a**,**b**). CDXs demonstrated minimal stroma, expressed at least one cytokeratin (detected using a pan-cytokeratin antibody) and neuroendocrine markers synaptophysin, chromogranin A and CD56. We frequently observed mitotic and apoptotic cells (Ki67 and cleaved poly ADP ribose polymerase (PARP) indices ~75% and 3%, respectively), which are typical of SCLC. Crush artifact[13] and necrotic foci were also frequent.

To determine whether metastases were present in mice bearing CDXs, we harvested internal organs from mice bearing CDX1 and CDX2 but did not note macrometastases on visual inspection. However, we detected micrometastases (indicated by human DNA detection using quantitative PCR (qPCR)) in the lungs of both mice and brain of the mouse bearing CDX1 (**Supplementary Fig. 2**). Subsequent detailed histological examination of serial lung sections revealed small clusters of tumor cells in the alveolar wall (**Fig. 2c**) and scattered single cells infiltrating pulmonary parenchyma (data not shown). Metastatic foci were composed of <20 cells (three times larger than lymphocytes) with scant cytoplasm, dispersed chromatin and irregular nuclei, consistent with SCLC. We confirmed their human origin and SCLC histology by staining with an anti-human mitochondrial antibody and expression of neuroendocrine markers (**Fig. 2c**).

Response of CDXs to cisplatin and etoposide

Mice bearing passage-4 CDX3, CDX2 and CDX4 were treated with cisplatin and etoposide. CDX3 exhibited the greatest response to therapy, where no treated tumor reached four times initial tumor volume over the experimental time course and median maximum tumor regression was 95% (range 72–96%), which is significantly greater than CDX2 or CDX4 ($P < 0.0001$; **Fig. 3a–c** and **Supplementary Fig. 3**). CDX2 exhibited an intermediate response to cisplatin and etoposide, with a significant increase, compared to vehicle-treated group, in time to four times initial tumor volume ($P < 0.0001$) and median maximal tumor regression of 51% (range 0–67%; $P < 0.0001$). CDX4 did not respond to therapy. The response of CDXs to therapy closely mirrored overall survival of the corresponding patients (9.7, 3.5 and 0.9 months for patients 3, 2 and 4, respectively).

Doubling time analysis of tumors that regrew after regression (or that did not regress) revealed no significant difference in tumor growth of treated as compared to control tumors for any CDX (**Fig. 3d**). These data imply that degree of tumor regression is consistent with a resistant subpopulation of preexisting cells, the proportion of which dictates therapy response.

Genomic analysis of CDXs

Next-generation sequencing (NGS) of matching left and right flank tumors (CDX1 and CDX3) and single tumors (CDX2 and CDX4) confirmed that genomic profiles of CDXs maintained previously described characteristics of SCLC[13,14]. Copy number aberration (CNA) analysis showed clear patient-specific patterns of gains and losses, with CDX1 and CDX3 showing prominent CNA losses and gains in contrast to CDX2 and CDX4, which were characterized by CNA losses but far fewer gains (**Fig. 4a**). Left and right flank tumors from CDX1 and CDX3 were broadly similar to each other but with some differences that may reflect tumor evolution[23,24] either before or after CTC implantation. For example, CDX1L but not CDX1R exhibited loss of chromosome 2p, harboring *MYCN*, as well as additional copies of *BCL2* and *SOX2*. We detected an additional copy of *BCL2* in CDX3R but not CDX3L (**Fig. 4a** and **Supplementary Table 1**). For all six tumors, there were deletions affecting 13q (containing *RB1*), 17p (containing *TP53*) and 10q (containing *PTEN*). CNA analysis of 13 individual genes frequently altered in SCLC[14,15] confirmed allelic loss of *RB1*, *TP53* and *PTEN* in all six tumors and loss of *RASSF1* and *FHIT* in all CDXs except CDX4 (**Fig. 4b** and **Supplementary Table 1**). However, characteristic SCLC-associated CNA increases of genes including *SOX2* (refs. 14,15) were detected only in CDX1 and CDX3 (**Fig. 4b** and **Supplementary Table 1**), a pattern also seen with extended analysis of 6,341 cancer-related genes (see Online Methods and below).

Sequence analysis of CDX1L, CDX1R and CDX2 revealed large numbers of single nucleotide variants (SNVs) and smaller numbers of insertions and deletions (indels) (**Supplementary Table 2**). We performed targeted Sanger sequencing on 18 of the SNV loci identified by NGS whole-genome sequencing (WGS), where in 16/18 cases we identified the identical SNV (**Supplementary Table 3**). The overall confirmation rate of 89% is consistent with previously reported false discovery rates[14]. Although the absence of patient germline DNA limited unambiguous identification of somatically acquired mutations in CDXs, a large number of identified SNVs were previously identified or predicted as deleterious, associated with disease or

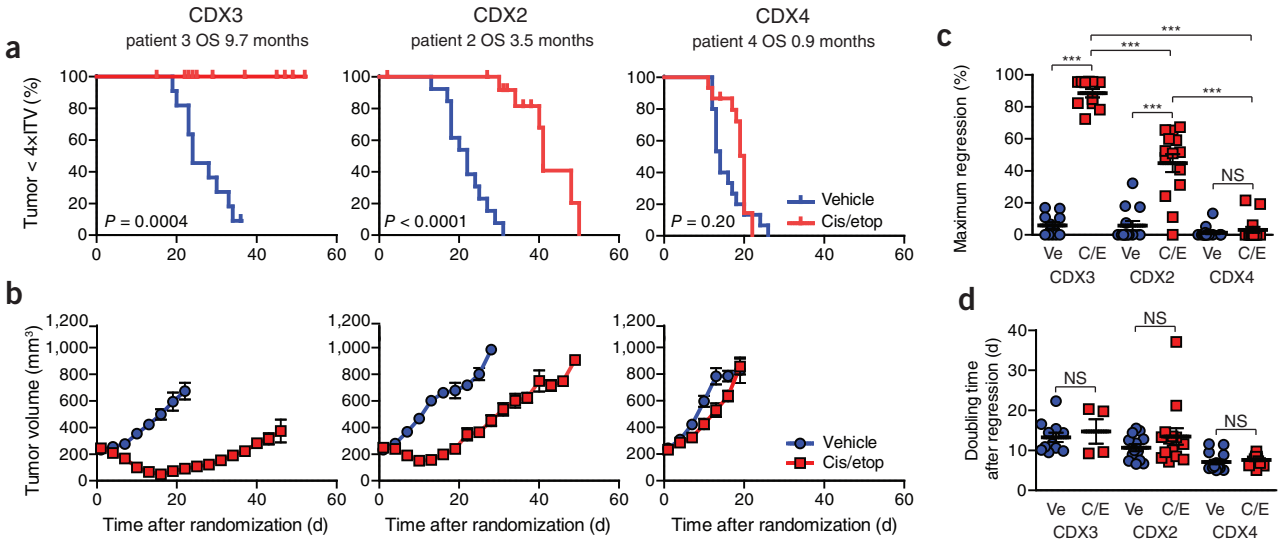


Figure 3 CDXs mirror patient response to therapy. Mice bearing passage-4 CDX3, CDX2 or CDX4 were treated with cisplatin and etoposide or vehicle control, and the tumor volume monitored. CDX3, $n = 11$ per group; CDX2 $n = 14$ in cisplatin-and-etoposide group and $n = 13$ in vehicle group; CDX4 $n = 15$ per group. (**a**) Kaplan-Meier survival curves comparing vehicle and cisplatin-and-etoposide–treated groups from randomization until the tumor reaches 4× initial tumor volume (4×ITV). *P* calculated by log-rank test. (**b**) Tumor volume of vehicle and cisplatin-and-etoposide–treated groups over time after randomization. Data represent mean ± s.e.m. (**c**,**d**) The maximum regression observed for each tumor relative to initial tumor volume (**c**) and the doubling time, calculated after growth had recommenced if regression was observed (**d**), for each CDX and treatment group (Ve, vehicle; C/E, cisplatin and etoposide). Line and error bars represent mean and s.e.m. NS, not significant; ***$P < 0.001$ according to unpaired two-tailed *t*-test. Patient overall survival is time from blood draw until death.

both (**Supplementary Table 2**). Furthermore, although CDX1L and CDX1R share a large proportion of genetic lesions, we detected >25% SNVs in only one of the tumors (**Supplementary Table 2**), indicating tumor heterogeneity and evolution[23,25]. As expected from the high mutation frequency in SCLC[25–27], we identified patient-specific mutations in both *RB1* and *TP53* (**Fig. 4c**,**d** and **Supplementary Table 4**). Short tandem repeat DNA fingerprint profiles of CDX samples (**Supplementary Table 5**) failed to match cell lines in the American

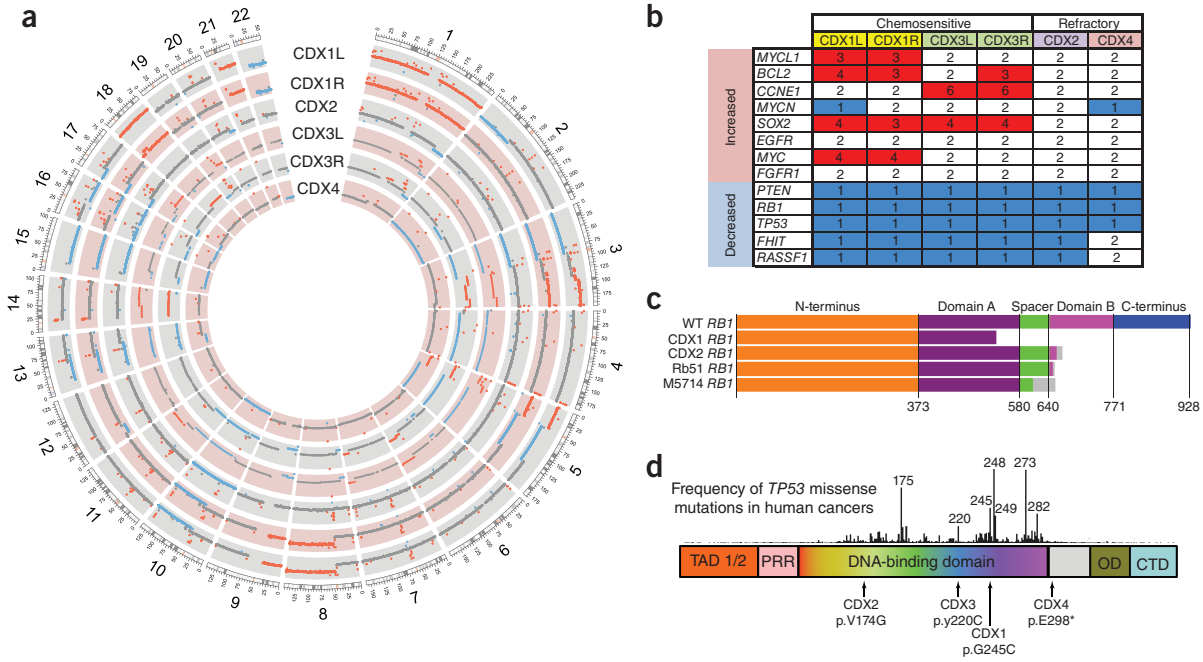


Figure 4 Genomic analysis of CDX. The left (L) and right (R) flank tumors from CDX1 and CDX3 and the single-flank CDX2 and CDX4 tumors were subjected to whole-genome and focused NGS as described in Online Methods. (**a**) Circos plots summarizing CNA data from all CDXs. Gains are shown in red, losses in blue and no change in gray. (**b**) CNA analysis of the genes showing frequent gains or losses in SCLC. (**c**) Position of predicted truncations in *RB1* from CDX1 and CDX2 (no *RB1* changes were detected in CDX3 or CDX4) alongside two reported *RB1* mutations[43,44]. (**d**) Position of *TP53* mutations from CDXs 1–4 alongside the frequency of reported *TP53* mutations[45]. TAD 1/2, transactivation domain 1 and 2; PRR, proline-rich region; OD, oligomerization domain; CTD, C terminal domain. p.E298* indicates a nonsense mutation. All indicated mutations were verified independently by a single round of NGS and Sanger sequencing in duplicate.

Type Culture Collection or our internal database, effectively ruling out the possibility of cell line contamination.

Comparison of patient CTCs and CDXs

To determine whether CDXs were derived from the same CTC pool enriched by CellSearch, we compared genomic profiles of CTCs isolated from the parallel enumeration of blood samples from patients 2 and 4 to their corresponding CDXs. We isolated CellSearch-enriched CTCs by DEPArray followed by whole-genome amplification (WGA) and WGS-based CNA analysis. We isolated six single CTCs, two pools of ten CTCs and ten white blood cells (WBCs) (germline samples) from patient 2 and two single CTCs, a pool of ten CTCs and ten WBCs from patient 4. Both principal component analysis (PCA) of genome-wide CNAs (**Fig. 5a**) and hierarchical clustering of copy number values for 6,341 cancer-related genes (**Fig. 5b**) of isolated CTCs strongly correlated with those of their corresponding CDXs and were distinct from those of unrelated CDXs and WBCs. CTC5 from patient 2 exhibited substantial CNA differences from the other patient 2 CTCs and CDX2, suggesting a degree of CTC heterogeneity. Indeed, CTC5 from patient 2 also exhibited an increased size and larger nucleus compared to the other five single CTCs from patient 2 (**Fig. 5c**).

We performed targeted Sanger sequencing on patient 2 CTCs at *TP53* and *RB1* loci (shown to be mutated in CDX2) (**Fig. 5c**). The *TP53* c.440T>G transversion was present in all CTC samples, with wild-type

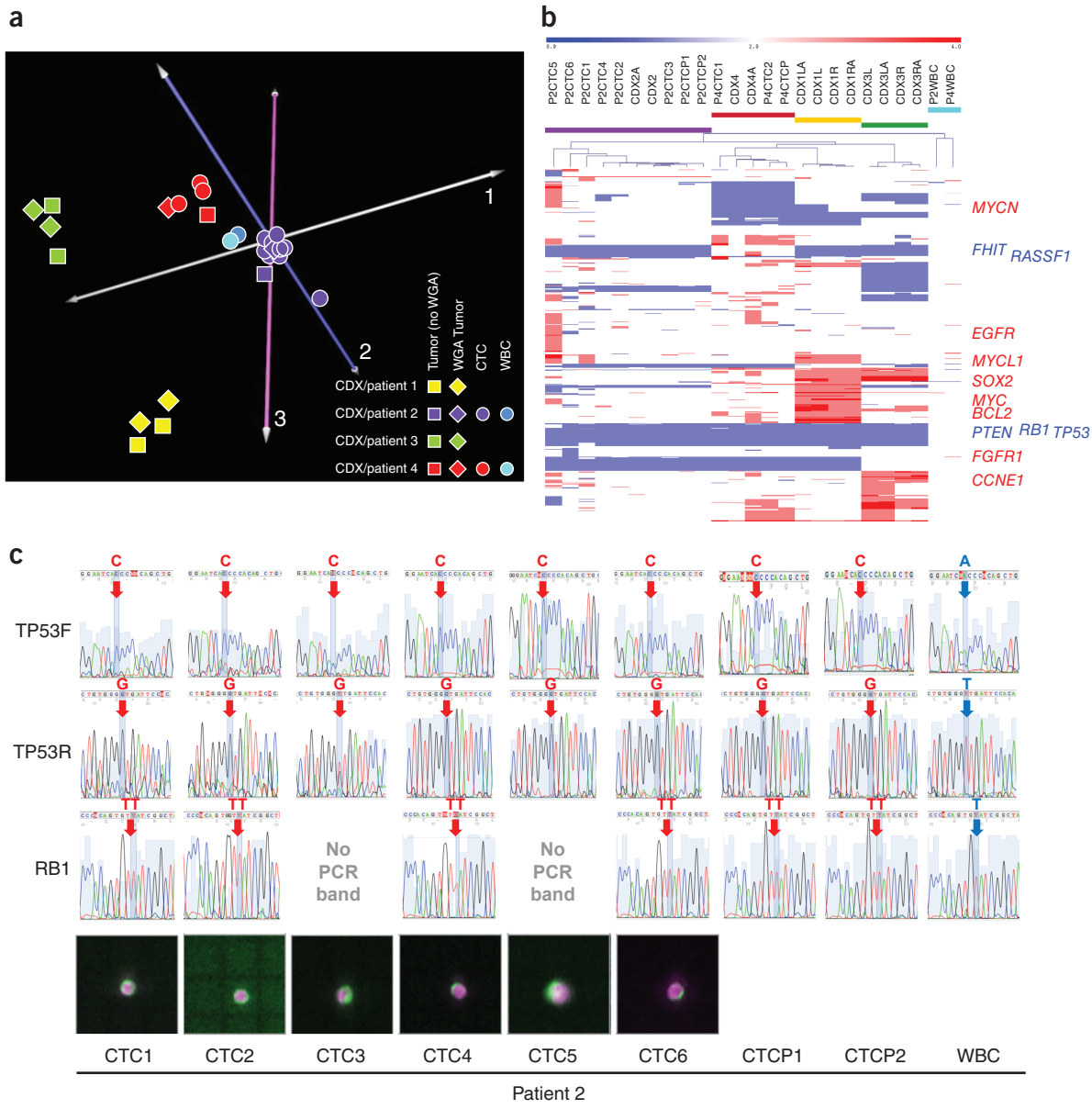


Figure 5 Molecular comparison of CDXs and patient CTCs. Single CTCs, pools of ten CTCs and pools of ten WBCs isolated from patients 2 and 4 were whole-genome–amplified along with 1 ng of DNA from CDXs 1–4. CNA analysis was carried out on the amplified material and unamplified tumor DNA. (**a**) PCA of genome-wide CNA data. (**b**) Hierarchical clustering of copy number values of 6,341 selected cancer-related genes. The positions of 13 genes showing frequent gains or losses in SCLC are indicated to the right of the heatmap. (**c**) Sanger sequencing of the *RB1* and *TP53* mutations (detected in CDX2) in six single CTCs, two pools of ten CTCs and a pool of ten WBCs isolated from patient 2, with images of a cytokeratin-stained single CTC from DEPArray. Red arrows indicate somatic mutations, and blue arrows indicate the corresponding unmutated regions. Sanger sequencing of all indicated samples was carried out in duplicate with representative traces presented.

sequence in corresponding WBC samples. Sanger sequencing revealed the presence of the *RB1* c.1963_1963insT in all CTC samples for which locus-specific PCR was obtained, but it was absent in the WBC sample. These data support the hypothesis that CellSearch-enriched CTCs are genetically highly related to tumorigenic SCLC CTCs.

DISCUSSION

We previously demonstrated that the number of CTCs detected by CellSearch (expressing EpCAM and cytokeratins) has independent prognostic significance in SCLC[19], suggesting their biological and functional importance. Prior to the current study, the viability and tumor-initiating capacity of SCLC CTCs was assumed but unknown. Here, we demonstrate for the first time, to our knowledge, that CTCs from patients with SCLC can form tumors in immunocompromised mice with preserved morphological and genetic characteristics. CDXs also faithfully recapitulate responses of donor patients to platinum and etoposide, the standard-of-care chemotherapy for SCLC, enabling clinically relevant studies of SCLC biology and a readily generated and sustainable patient-derived model to test targeted therapeutics. These data demonstrate formally that CTCs are tumorigenic and that the tumors they form recapitulate donor patients' tumor biology, which we believe confirms the assumed importance of CTCs in disease progression. Moreover, the combined WGS data from isolated CTCs, CDXs and WBCs confirm that CellSearch-enriched CTCs ($EpCAM^+cytokeratin^+$) are representative of, or closely related to, the tumor initiating cells present in the blood of patients with SCLC. These new findings open up the possibility of developing and routinely implementing personalized medicine strategies for patients with SCLC based on simple blood collection with subsequent and rapidly reported molecular analysis of CTCs. This is particularly relevant in a disease where repeat tumor biopsies are rarely obtained.

In our current study, only blood samples with >400 $EpCAM^+cytokeratin^+$ CTCs per 7.5 ml blood gave rise to CDXs. In ongoing investigations, we attempted engraftment of CTCs in mice from 19 chemotherapy-naive, extensive-stage patients with SCLC (including the six presented) with at least 5 months of follow-up to detect tumor formation (C.L.H., C.J.M. and C.D., unpublished data). Tumor engraftment occurred for 9/19 patients (including the four CDXs reported in the current study: 47% take rate). The $EpCAM^+cytokeratin^+$ CTC number was 160–7,687 per 7.5 ml blood (median 901, mean 1,974) for the patients whose CTCs generated CDXs and 0–2,048 per 7.5 ml blood (median 31, mean 274) in patients whose CTCs did not. Eight patients had a CTC number >400 per 7.5 ml blood with CDX generation from seven of these eight patients' samples. Another recent study reported tumor formation from CTCs directly explanted from patients with metastatic breast cancer[28]; three hundred and fifty patients were recruited and CTCs injected into femurs of 118 immunocompromised mice to generate CTC tumors from three patients. We speculate that the vast discrepancy in take rate compared to our study is due to significantly higher CTC burden in SCLC (2,915 ± 8,115 CTCs per 7.5 ml blood, mean ± s.d.[21]) compared to breast cancer (84 ± 885 CTCs per 7.5 ml blood, mean ± s.d.[29]). A previous study demonstrated that buffy coat preparations from patients with prostate or colorectal cancer formed tumors[30], though neither the presence of CTCs in the buffy coat nor the demonstration that the mouse tumors were of human prostate or colorectal origin were reported.

The presence of matched somatic *TP53* mutation in all CTCs examined in our study and the high degree of overall similarity in CNA patterns of CTCs and CDXs suggests that CTCs from a patient with extensive-stage SCLC are largely homogeneous. This is consistent with recent CTC analysis from patients with lung cancer that also showed a high degree of similarity amongst CTCs from the same patient[31]. Indeed, a study of disseminated cancer cells (DCCs) in metastatic breast and prostate cancer revealed relatively homogeneous genomes suggesting expansion of a dominant clone[32], whereas heterogeneity of CTCs and DCCs is primarily seen in early-stage breast cancer DCCs[32] and colorectal CTCs[33,34].

The current study builds upon and extends previous CTC molecular analysis[33,35–37] by providing a means of functionally testing molecular findings via selective interventions in corresponding CDXs. The rapid progression of SCLC prevents 'avatar trials'[38] using CDXs, but our study has implications for treatment of inherent and acquired drug-resistant disease. Although previous studies found no significant differences in overall genomic architecture between resected (likely to be chemosensitive) and autopsy (likely to be chemoresistant) SCLC cases[14], this does not exclude the existence of a predictive genomic signature for inherent resistance to conventional chemotherapy. In the present study, we observed gene copy number losses in chemorefractory and chemosensitive samples, but although CNA gains were frequent in chemosensitive CDXs and CTCs, they were far rarer in chemorefractory samples. The chemorefractory CNA ratio seen here was seldom observed in previous SCLC CNA studies[14]. One possible explanation is that CNA gains (seen in the majority of patients with SCLC[14]) are responsible for conferring initial chemosensitivity, which is also observed in most patients with SCLC. As our biobank of SCLC CTCs isolated from patients with known clinical outcomes grows, the hypothesis that lack of CNA gain is associated with inherent chemotherapy resistance can be tested. Another notable difference between CDXs generated from chemosensitive and chemorefractory patients is the faster growth rate of CDXs from the chemorefractory patients (**Fig. 1c** and **Supplementary Fig. 1b**). This may reflect the more aggressive disease in the chemorefractory patients who had shorter overall survival. However, this hypothesis needs to be tested in a larger cohort.

Detection of circulating tumor DNA in in the plasma of patients with cancer shows great potential as a 'liquid biopsy' and has been applied to SCLC[39]. Serial monitoring of circulating tumor DNA with altered prevalence of mutations in patients with emerging resistance to targeted therapies[40,41] could be used for future treatment decision making. We consider the molecular analysis of CTCs as a complementary approach[42]. Although more technically challenging, CTC analysis offers advantages, including detection of co-expressed genetic defects within tumor cells, which is of likely importance in understanding drug resistance mechanisms, and comparison of single CTCs with CDXs to model drug-imposed selection and therapy responses.

Our CDX models complement previously reported PDX models, which recapitulated patient responses to chemotherapy[11,12]. The main advantage of our CDX approach is the potential to examine mechanisms that underpin the acquired drug resistance commonly observed in SCLC. A patient's blood sample acquired before and after drug-resistant relapse can now be used to generate CDX models for comparison. In summary, these unique CDX models, generated from sequentially available, minimally invasive clinical samples now provide an unprecedented opportunity to study SCLC biology from diagnosis through treatment to progression. CDX models will also facilitate the search for new druggable targets in SCLC and enable routine *in vivo* testing of targeted therapies for a disease with clear unmet medical need.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Next-generation sequencing data have been deposited in the NCBI Sequence Read Archive with BioSample accession codes SAMN02803803, SAMN02803804, SAMN02803805, SAMN02803806, SAMN02803807 and SAMN02803808.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We are indebted to the patients who agreed to donate their blood samples for this study. We thank R. Marais, N. Jones and D. Ogilvie for their constructive comments on the manuscript. We thank M. Dawson, M. Lancashire, S. Bramley, J. Halstead and J. Castle, who enumerated CTCs using CellSearch. We thank A. Jardine for administrative support and M. Greaves, our laboratory manager. This research was supported by Cancer Research UK via core funding to the Cancer Research UK Manchester Institute (C5759/A12328), the Manchester Experimental Cancer Medicine Centre (C1467/A15578), the Manchester Cancer Research Centre (A12197) and their Translational Research Award for 2012. Funding to support this work was also provided via the European Union CHEMORES FP6 (contract number LSHG-CT-2007-037665). R.L.M. and L.C. were supported by education grants from Cancer Research UK and AstraZeneca.

AUTHOR CONTRIBUTIONS

C.L.H., P.K. and B.B. performed *in vivo* studies, F.T., R.P., K.L.S. and D.N. conducted histopathological examinations, D.G.R., D.J.B., S.D.P., A.G., J.A., M.G.K., M.A., L.C. and S.F. conducted the genomic analyses, Y.L., C.T., C.J. Miller and G.B. performed the bioinformatic analysis, K.M. oversaw CTC enumeration by CellSearch, R.L.M., L.C., L.P. and F.B. recruited and consented patients and collected blood samples, C.J. Morrow, C.J. Miller, G.B., F.B. and C.D. conceived and directed the study, interpreted the data and wrote the manuscript. All authors discussed the results and commented on the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Einhorn, L.H., Fee, W.H., Farber, M.O., Livingston, R.B. & Gottlieb, J.A. Improved chemotherapy for small-cell undifferentiated lung cancer. *J. Am. Med. Assoc.* **235**, 1225–1229 (1976).
2. Evans, W.K. *et al.* VP-16 and cisplatin as first-line therapy for small-cell lung cancer. *J. Clin. Oncol.* **3**, 1471–1477 (1985).
3. Sierocki, J.S. *et al.* cis-Dichlorodiammineplatinum(II) and VP-16–213: an active induction regimen for small cell carcinoma of the lung. *Cancer Treat. Rep.* **63**, 1593–1597 (1979).
4. Gazdar, A.F. *et al.* Establishment of continuous, clonable cultures of small-cell carcinoma of lung which have amine precursor uptake and decarboxylation cell properties. *Cancer Res.* **40**, 3502–3507 (1980).
5. Oboshi, S., Tsugawa, S., Seido, T., Shimosato, Y. & Koide, T. A new floating cell line derived from human pulmonary carcinoma of oat cell type. *Gann* **62**, 505–514 (1971).
6. Pleasance, E.D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).
7. Joshi, M., Ayoola, A. & Belani, C.P. Small-cell lung cancer: an update on targeted therapies. *Adv. Exp. Med. Biol.* **779**, 385–404 (2013).
8. William, W.N. Jr. & Glisson, B.S. Novel strategies for the treatment of small-cell lung carcinoma. *Nat. Rev. Clin. Oncol.* **8**, 611–619 (2011).
9. Meuwissen, R. *et al.* Induction of small cell lung cancer by somatic inactivation of both Trp53 and Rb1 in a conditional mouse model. *Cancer Cell* **4**, 181–189 (2003).
10. Kwon, M.C. & Berns, A. Mouse models for lung cancer. *Mol. Oncol.* **7**, 165–177 (2013).
11. Daniel, V.C. *et al.* A primary xenograft model of small-cell lung cancer reveals irreversible changes in gene expression imposed by culture *in vitro*. *Cancer Res.* **69**, 3364–3373 (2009).
12. Poupon, M.F. *et al.* Response of small-cell lung cancer xenografts to chemotherapy: multidrug resistance and direct clinical correlates. *J. Natl. Cancer Inst.* **85**, 2023–2029 (1993).
13. Davenport, R.D. Diagnostic value of crush artifact in cytologic specimens. Occurrence in small cell carcinoma of the lung. *Acta Cytol.* **34**, 502–504 (1990).
14. Peifer, M. *et al.* Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nat. Genet.* **44**, 1104–1110 (2012).
15. Rudin, C.M. *et al.* Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. *Nat. Genet.* **44**, 1111–1116 (2012).
16. Cristofanilli, M. *et al.* Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N. Engl. J. Med.* **351**, 781–791 (2004).
17. de Bono, J.S. *et al.* Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. *Clin. Cancer Res.* **14**, 6302–6309 (2008).
18. Hayes, D.F. *et al.* Circulating tumor cells at each follow-up time point during therapy of metastatic breast cancer patients predict progression-free and overall survival. *Clin. Cancer Res.* **12**, 4218–4224 (2006).
19. Hou, J.M. *et al.* Clinical significance and molecular characteristics of circulating tumor cells and circulating tumor microemboli in patients with small-cell lung cancer. *J. Clin. Oncol.* **30**, 525–532 (2012).
20. Krebs, M.G. *et al.* Evaluation and prognostic significance of circulating tumor cells in patients with non-small-cell lung cancer. *J. Clin. Oncol.* **29**, 1556–1563 (2011).
21. Hou, J.M. *et al.* Evaluation of circulating tumor cells and serological cell death biomarkers in small cell lung cancer patients undergoing chemotherapy. *Am. J. Pathol.* **175**, 808–816 (2009).
22. Evans, W.K. *et al.* VP-16 alone and in combination with cisplatin in previously treated patients with small cell lung cancer. *Cancer* **53**, 1461–1466 (1984).
23. Swanton, C. Intratumor heterogeneity: evolution through space and time. *Cancer Res.* **72**, 4875–4882 (2012).
24. Martinez, P. *et al.* Parallel evolution of tumour subclones mimics diversity between tumours. *J. Pathol.* **230**, 356–364 (2013).
25. Arriola, E. *et al.* Genetic changes in small cell lung carcinoma. *Clin. Transl. Oncol.* **10**, 189–197 (2008).
26. Mori, N. *et al.* Variable mutations of the *RB* gene in small-cell lung carcinoma. *Oncogene* **5**, 1713–1717 (1990).
27. Wistuba, I.I. & Gazdar, A.F. & Minna, J.D. Molecular genetics of small cell lung carcinoma. *Semin. Oncol.* **28**, 3–13 (2001).
28. Baccelli, I. *et al.* Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. *Nat. Biotechnol.* **31**, 539–544 (2013).
29. Allard, W.J. *et al.* Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clin. Cancer Res.* **10**, 6897–6904 (2004).
30. Pretlow, T.G. *et al.* Prostate cancer and other xenografts from cells in peripheral blood of patients. *Cancer Res.* **60**, 4033–4036 (2000).
31. Ni, X. *et al.* Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. *Proc. Natl. Acad. Sci. USA* **110**, 21083–21088 (2013).
32. Klein, C.A. Selection and adaptation during metastatic cancer progression. *Nature* **501**, 365–372 (2013).
33. Gasch, C. *et al.* Heterogeneity of epidermal growth factor receptor status and mutations of KRAS/PIK3CA in circulating tumor cells of patients with colorectal cancer. *Clin. Chem.* **59**, 252–260 (2013).
34. Fabbri, F. *et al.* Detection and recovery of circulating colon cancer cells using a dielectrophoresis-based device: KRAS mutation status in pure CTCs. *Cancer Lett.* **335**, 225–231 (2013).
35. Heitzer, E. *et al.* Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing. *Cancer Res.* **73**, 2965–2975 (2013).
36. Klein, C.A. *et al.* Genetic heterogeneity of single disseminated tumour cells in minimal residual cancer. *Lancet* **360**, 683–689 (2002).
37. Klein, C.A. *et al.* Combined transcriptome and genome analysis of single micrometastatic cells. *Nat. Biotechnol.* **20**, 387–392 (2002).
38. Morelli, M.P. *et al.* Prioritizing phase I treatment options through preclinical testing on personalized tumorgraft. *J. Clin. Oncol.* **30**, e45–e48 (2012).
39. Board, R.E. *et al.* Isolation and extraction of circulating tumor DNA from patients with small cell lung cancer. *Ann. NY Acad. Sci.* **1137**, 98–107 (2008).
40 Bettegowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci .Transl. Med.* **6**, 224ra224 (2014).
41. Dawson, S.J. *et al.* Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N. Engl. J. Med.* **368**, 1199–1209 (2013).
42. Krebs, M.G. *et al.* Molecular analysis of circulating tumour cells—biology and biomarkers. *Nat. Rev. Clin. Oncol.* **11**, 129–144 (2014).
43. Lohmann, D.R. *et al.* Constitutional RB1-gene mutations in patients with isolated unilateral retinoblastoma. *Am. J. Hum. Genet.* **61**, 282–294 (1997).
44. Szijan, I., Lohmann, D.R., Parma, D.L., Brandt, B. & Horsthemke, B. Identification of RB1 germline mutations in Argentinian families with sporadic bilateral retinoblastoma. *J. Med. Genet.* **32**, 475–479 (1995).
45. Joerger, A.C. & Fersht, A.R. The tumor suppressor p53: from structures to drug discovery. *Cold Spring Harb. Perspect. Biol.* **2**, a000919 (2010).

ONLINE METHODS

Patient selection and blood collection. From August 2012 to February 2013, 55 patients were recruited to our broader program of SCLC biomarker research. Patients had histologically or cytologically confirmed chemotherapy-naive SCLC and were referred to a tertiary cancer center, The Christie Hospital NHS Trust. The study was prospectively approved by the NHS NorthWest 9 Research Ethical Committee. Clinical and demographic data were collected. During this period, we initiated our CDX study, and 11 patients provided additional informed consent that specified their samples could be used for *in vivo* studies and genetic analysis in accordance with UK regulatory requirements. The 11 patients approached were selected for the CDX study as their clinic appointments coincided with the capacity within the *in vivo* research team for blood processing and enriched CTC implantation in mice. Seven of the 11 patients had the required clinical features of extensive (metastatic stage) disease and were chemotherapy naive. One of these patients was excluded from the study because the recipient mouse showed signs of ill health (confirmed to be non–cancer related), was culled 62 days following CTC implantation and was therefore uninterpretable with respect to CDX formation.

Blood was drawn at CDX study entry before administration of chemotherapy and immediately transferred to the laboratory for processing. Blood (10 ml) was drawn into CellSave tubes (Janssen Diagnostics) for CTC enumeration using the CellSearch platform[19]. CTCs thus defined expressed EpCAM and cytokeratins (cytokeratins, 8, 18 and 19), were >4 μm in diameter and had an intact DAPI-stained nucleus. A paired blood sample (10 ml) was drawn into EDTA vacutainers (Becton Dickinson). Patients' subsequent response to treatment was evaluated by computed tomography (CT) imaging performed before and following 4 cycles of chemotherapy, or earlier if clinically indicated. Patients who had a radiological response to chemotherapy that was sustained for greater than 3 months following completion of therapy were classified as chemotherapy sensitive. Patients with no evidence of response to therapy or progression within 3 months following completion of therapy were classified as chemotherapy refractory as previously described[22].

CTC enrichment before implantation into mice. An EDTA blood sample from a patient with SCLC was mixed with 500 μl RosetteSep Human Circulating Epithelial Tumor Cell Cocktail (Stem Cell Technology) and incubated for 20 min at room temperature with constant mixing. Blood was diluted with 10 ml 9:1 Hank's Balanced Saline Solution (HBSS) (Life Technologies): HITES medium (RPMI 1640 (Life Technologies), 5 μg ml^{-1} insulin, 10 μg ml^{-1} transferrin, 10 nM β-estradiol, 30 nM sodium selenite, 10 nM hydrocortisone (Sigma)), layered over 15 ml Ficoll-Plaque Plus (GE Heathcare) and centrifuged at 1,200 × *g* for 20 min. Cells at the medium-Ficoll boundary were collected, diluted with 30 ml 9:1 HBSS:HITES and centrifuged at 250 × *g* for 5 min. The cell pellet was resuspended in 100 μl ice cold HITES and mixed with 100 μl Matrigel (BD Biosciences) and kept on ice.

Growth of CTC tumors in immunocompromised mice. All procedures were carried out in accordance with Home Office Regulations (UK) and the UK Coordinating Committee on Cancer Research guidelines and by approved protocols (Home Office Project license no. 40-3306 and Cancer Research UK Manchester Institute Animal Welfare and Ethical Review Advisory Board). 100–200 μl of CTCs/HITES/Matrigel was injected subcutaneously into one or both flanks of 8-16 week old female NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG) mice (Jackson Laboratories). Mice were housed in individually vented caging systems in a 12-h light/12-h dark environment and maintained at uniform temperature and humidity. Mice were monitored twice weekly for signs of tumor growth, and once a palpable tumor was present this was measured twice a week by calipers and tumor volume calculated as tumor length × tumor width^2/2. When the total tumor burden reached 1,000 mm^3 or there were demonstrable signs of ill health, the animal was killed, tumor fragments were passaged into NSG mice and the remainder of the tumor was harvested for IHC analysis or DNA or RNA extraction. The internal organs were also harvested for further analysis. No statistical method was used to predetermine sample size as no previous data were available.

Cisplatin and etoposide treatment. Thirty female NSG mice were implanted with passage 4 CDX2, CDX3 or CDX4 with the expectation that 10 tumors would not grow successfully, leaving at least 10 animals per treatment group (drugs vs. vehicle). When tumors reached 200–250 mm^3, they were randomized by sequential assignment to cisplatin and etoposide or vehicle treatment groups. Animals were treated by intraperitoneal injection with 5 mg kg^{-1} cisplatin (Sigma) dissolved in 0.9% saline solution on day 1 and 8 mg kg^{-1} etoposide (Sigma) dissolved in 12.5:1 0.9% saline solution:0.1% citric acid in 1-methyl-2-pyrrolidinone on days 1, 2 and 3, or corresponding vehicle only. Tumor volume was monitored blinded to treatment group every three days until tumor reached four times initial tumor volume (4xITV) or until animal health deteriorated (censored in survival analysis). Survival analysis was performed in Graphpad Prism with comparison of survival curves by log-rank (Mantle Cox) test. Doubling time was calculated by nonlinear curve fitting of an exponential growth equation (Graphpad Prism). If the tumor exhibited regression, the doubling time was calculated from the point tumor growth recommenced. For comparison of maximum regression and tumor doubling time, normal distribution was assessed by D'Agostino and Pearson omnibus normality test and groups compared by unpaired two-tailed *t*-test.

Human- and mouse-specific qPCR. Tissue was collected from autopsied animals and gDNA isolated using the Ambion RecoverAll kit (Life Technologies). Quantitative PCR was then performed using Bioline SensiFAST qPCR reagents (Bioline) and 10 ng of gDNA from each sample with murine- and human-specific primers targeting the prostaglandin E receptor 2 (*Ptger2/PTGER2*) and phosphoserine aminotransferase 1 (*Psat1/PSAT*) genes as described previously[46,47]. Dissociation curve analysis was used to distinguish between human- and murine-specific amplification products. Normal murine lung gDNA and HNV gDNA were included in all experiments as controls.

Immunohistochemistry. IHC was performed on formalin-fixed, paraffin-embedded 4-μm tumor and normal tissue sections using antibodies to cytokeratins (pan-cytokeratin antibody to cytokeratins 1–8, 10, 13–16 and 19, mouse AE1/AE3, M3515, 1:60, Dako), CD56 (mouse, 1B6, NCL-CD56-1B6, 1:100, Novocastra), chromogranin A (mouse, LK2H10 + PHE5, MP-010-CM1, 1:600, Menapath), synaptophysin (mouse, 27G12, NCL-L-SYNAP-299, 1:200, Novocastra), cleaved PARP (mouse, Asp214, 51-9000017 1 : 100, BD Pharmingen) and Ki67 (mouse, MIB-1, M7240, 1:600, Dako). A human-specific anti-mitochondria antibody (rabbit, 1113-1, 1:500, Abcam) was used to detect micrometastases in mouse tissues. Antibody incubations and detection were carried out at room temperature on a Menarini IntelliPATH FLX (A. Menarini Diagnostics) using Menarini's reagent buffer and detection kits unless otherwise noted. Antigen retrieval was performed in a pressure cooker using access super retrieval fluid (MP-606-PG1) for all antisera except AE1/AE3, which was 10-min incubation with protease (MP-960-K15) on IntelliPATH, and anti-mitochondria, which was performed in citrate buffer pH 6 and microwaved for 15 min at 98 °C. Isotype controls used were rabbit immunoglobulin fraction and mouse IgG1 from Dako. Digital images of whole tissue sections were acquired using a Leica SCN400 histology scanner (Leica Microsystems). Ki67 and cPARP positive index were evaluated using Definiens Developer XD version 2.0.4 and the Tissue Studio Portal version 3.51 (Definiens AG). Regions of interest (ROIs) within the tissue sections were first identified using Definiens Tissue Studio via machine learning technology across pathological samples and tissue control, so that the full range of contrast was defined. Within these ROIs, nuclei were detected and classified as positive or negative based on IHC staining thresholds.

Tumor DNA extraction and WGS. CTC tumors were disaggregated using a sterile scalpel and gDNA isolated using the QiaAmp DNA Mini kit (Qiagen, Hilden, Germany). DNA libraries were generated from 50 ng gDNA in the NEBNext Ultra DNA Library kit (NEB) and sequenced on an Illumina HiSeq2500 instrument using the TruSeq PE Cluster Kit V3 and TruSeq SBSv3 chemistry.

CTC isolation and WGA. CTCs and WBCs (pre-stained with antibody to CD45, pan-CK and DAPI) were aspirated from the CellSearch cartridge used for the CTC enumeration, and single and groups of cells were isolated using the DEPArray system (Silicon Biosystems) as per manufacturer's instructions. WGA of single CTCs, pools of 10 CTCs or WBCs and genomic DNA (1 ng input) was performed using the Ampli1 WGA kit (Silicon Biosystems) according to manufacturer's instructions[48].

NGS of WGA samples. To remove the Ampli1 amplification primer, all WGA products (250ng) were digested with MSE1 (New England Biolabs) following

supplier's instructions. Digested samples were quantified using Qubit (Life Technologies) and sonicated with a Bioruptor UCD-200 (Diagenode) for 10 cycles (T1 = 30 s, T2 = 30 s) to produce fragments of about 300–350 bp that were checked by a Bioanalyzer (2100 Bioanalyzer, Agilent Life Sciences and Chemical Analysis). DNA libraries were prepared using the NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs) according to manufacturer's instructions. Final library PCR products were quantified using KAPPA Library Quantification Kits for Illumina (KAPABiosystems), Bioanalyzer and the Quant-iT assay using Qubit Quantitation Platform according to the manufacturer's instructions. Sequencing was carried out on an Illumina MiSeq System with paired-end 150-bp runs, and the resultant reads were base called, filtered by quality metrics and aligned to the human reference sequence as recommended by the manufacturer.

GeneRead of CDX3 and 4 tumor and CTCs. Genomic DNA from CDX3, CDX4 and WGA product from patient 4 WBCs were analyzed using the GeneRead DNAseq Human Lung Cancer panel (Qiagen) as described in the manufacturer's protocol. Briefly, 80 ng of DNA was PCR amplified using targeted multiplexed amplicons, and the resulting material cloned into a NEBNext Ultra DNA Library. The libraries were then run on an Illumina MiSeq and analyzed using the Qiagen NGS portal (http://ngsdataanalysis.sabiosciences.com/NGS/).

Sanger sequencing (confirmation of NGS). Target amplicons (**Supplementary Table 3**) of tumor-associated genes were amplified by PCR from 10 ng of genomic DNA from the original xenograft tumor samples and from HNV blood. Each amplicon was purified with a PCR cleanup kit (Qiagen) and subjected to Sanger sequencing on an ABI 3130 Genetic Analyzer using the same primers used for PCR. ABI sequencing files were analyzed using 4Peaks software (http://nucleobytes.com/index.php/4peaks/) and publically available web based BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and alignment tools (http://www.ebi.ac.uk/Tools/msa/clustalo/).

WGS analysis of tumors CDX1 and CDX2. Illumina HiSeq data were aligned to the human genome (version hg19) using SMALT (http://smalt.sourceforge.net/ with default strategies: http://www.sanger.ac.uk/resources/software/smalt/). To identify potential contaminating reads of murine origin, the same data was aligned to the mouse reference genome (version mm9). Reads aligning to both human and mouse genomes were discarded. In excess of 1.45 billion reads mapped uniquely to the human genome for CDX1L and CDX1R samples and >1.28 billion reads for the CDX2 samples. Aligned paired-end reads were used to identify SNVs and short indels for each sample using GATK[49]. Duplicate read removal, realignment around known indels and base and variant quality score recalibration[50] were performed as pre- and post-processing. Variant calling was performed using unifiedGenotyper with default settings. The putative SNVs and indels identified by GATK were annotated using ANNOVAR[51].

CNA analysis of WGA products. Illumina MiSeq whole-genome data from nineteen WGA samples (6 single cell CTCs, two 10-cell CTC pools and one WBC 10-cell pool from CDX2, 2 single-cell CTCs, one 10-cell CTC pools and one WBC 10-cell pool from CDX4, plus WGA products generated from 1 ng CDX1L, CDX1R, CDX2, CDX3L, CDX3R and CDR4 and three WGS samples (WGS for CDX3L, CDX3R and CDX4) were aligned to the human genome using SMALT (number of mapped reads reported in **Supplementary Table 6**). FREEC[52] was used to identify copy number variations (window size: 50 kb; step size 10 kb for WGS samples CDX1L, CDX1R and CDX2; the adaptive window size was used for all other samples) of both the genomic (HiSeq) and the WGA (MiSeq) data. Mappability data for HG19 with an edit distance of 1 were downloaded from the FREEC web site (http://bioinfo-out.curie.fr/projects/freec/). For each sample, an estimated copy number was assigned to every cytoband of the human genome (version hg19) by calculating the weighted mean of the overlapping copy number estimates (as computed by FREEC) that map to the given cytoband and passed to Circos[53] (**Fig. 4a**). FREEC predicted copy number data was averaged across cytobands, as before, and imported into MeV to generate the PCA data (median centering mode with recommended MeV algorithm; **Fig. 5a**). These copy numbers were also mapped to genome coordinates using the Bioconductor package annmap to provide ENSEMBL version 70 annotation[54] and clustered in MeV using Pearson correlation average linkage (**Fig. 5b**).

Cancer-related genes. The geneRIF database was downloaded from the NCBI web site (http://www.ncbi.nlm.nih.gov/gene/about-generif/) on 3 February 2014. A human gene was regarded as a cancer-related gene if its RIF text contains at least one of the 10 key words or word stems (carcinogen, cancer, carcinoma, tumor, leukemia, tumour, oncogen, leukaemia, oncolog, malignan). A list of 6,682 cancer-related protein-coding genes were compiled by mapping ENSEMBL gene ID and gene symbol in the geneRIF database to ENSEMBL (v70). Only autosomal chromosomes were used in CNA analysis, so 268 genes on the X, Y or mitochondrial chromosome were removed from the list.

Verification of the single-cell CNA approach. As part our evaluation of the WGA/CNA approach, we examined matched single and pooled WBCs and CTCs. Six WBC samples (two pools of 10 WBCs; four single WBCs) and six CTC samples (two pools of 10 CTCs; four single CTCs) were subjected to WGA and NGS (number of mapped reads reported in **Supplementary Table 6**), with resultant Illumina MiSeq data analyzed for CNA at the cytoband and cancer-related gene level. From this analysis, we detected the expected clear separation of CTC and WBC samples, and within each group the single-cell and 10-cell samples gave comparable results (**Supplementary Fig. 4**). Based on this evaluation, we also identified a small numbers of potentially unreliable loci (0.8% for cytobands and 1.1% for cancer genes) with reported loss or gain in at least three WBC samples. These loci (**Supplementary Table 7**) were subsequently removed for CNA analysis of the CDX samples, which reduced the number of cancer-related genes to 6,341.

Single tandem repeat DNA fingerprinting. Purified DNA from CDXs was processed using the Powerplex 21 kit (Promega) according to manufacturer's instructions. STR profiles were compared to the ATCC cell line database and to our internal database of all cell lines in use at the CRUK Manchester Institute.

46. Alcoser, S.Y. *et al.* Real-time PCR-based assay to quantify the relative amount of human and mouse tissue present in tumor xenografts. *BMC Biotechnol.* **11**, 124 (2011).
47. Thierry, A.R. *et al.* Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. *Nucleic Acids Res.* **38**, 6159–6175 (2010).
48. Peeters, D.J. *et al.* Semiautomated isolation and molecular characterisation of single or highly purified tumour cells from CellSearch enriched blood samples using dielectrophoretic cell sorting. *Br. J. Cancer* **108**, 1358–1367 (2013).
49. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
50. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
51. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
52. Boeva, V. *et al.* Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization. *Bioinformatics* **27**, 268–269 (2011).
53. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
54. Yates, T., Okoniewski, M.J. & Miller, C.J. X:Map: annotation and visualization of genome structure for Affymetrix exon array analysis. *Nucleic Acids Res.* **36**, D780–D786 (2008).