Diplomarbeit

Identification of driver mutations in plasma DNA from lung cancer patients

eingereicht von

Sebastian Bachmann

zur Erlangung des akademischen Grades

Doktor(in) der gesamten Heilkunde

(Dr. med. univ.)

an der

Medizinischen Universität Graz

ausgeführt am

Institut für Humangenetik

unter der Anleitung von

Assoz. Prof. Mag. Dr.rer.nat. Ellen Heitzer

Graz, am 08.05.2017

Eidesstattliche Erklärung

Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe, andere als die angegebenen Quellen nicht verwendet habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am 08.05.2017 *Sebastian Bachmann, eh*

Danksagungen

Zunächst möchte ich mich bei Martina und Julie für die geduldige, immer freundliche, motivierende Einführung und Unterstützung bei der Laborarbeit und den verschiedenen Arbeitsmethoden bedanken. Gerade ich, als quasi blutiger Anfänger im Labor, wusste dies sehr zu schätzen.

Besonderer Dank geht an Ellen, die mir stets mit Rat, neuen Ideen und schließlich der Korrektur meiner Arbeit zur Seite stand. Die oft spontanen, immer konstruktiven aber vor allem netten und entspannten Besprechungen werden mir in guter Erinnerung bleiben.

Vielen Dank auch an alle Freunde, die mir motivierend und aufbauend zur Seite gestanden sind, wenn meine Arbeit ins Stocken geriet.

Schlussendlich geht ein besonderes Dankeschön an meine Eltern für ihre stetige Unterstützung, die Hilfe bei der Korrektur und für ihre Geduld. Danke auch an meinen Bruder für seine moralische Ermutigung und das Interesse an meiner Arbeit.

Table of content

Abbreviations

°C	Degree Celsius
µl	microliter
BEAMing	Beads, emulsions, amplification and magnetics
bp	Base pairs
BWA	Burrows Wheeler Alignment
cfDNA	Cell-free DNA
COSMIC	Catalogue of Somatic Mutations in Cancer
CRC	Colorectal cancer
CTCs	Circulating tumor cells
ctDNA	Circulating tumor DNA
DST	Digital sequencing technology
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor
FFPE	Formalin fixed paraffin embedded
gDNA	Genomic DNA
MAF	Mutant allele frequency
mFAST-SeqS	Modified Fast Aneuploidy Screening Test-Sequencing System
ml	milliliter
NBF	Neutral buffered formalin
ng	nanogram
NGS	Next-generation sequencing
NSCLC	Non-small cell lung cancer

PCR	Polymerase chain reaction
pg	picogram
RT	Room temperature
SCNAs	Somatic copy number alterations
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variants
TKIs	Tyrosine kinase inhibitors

Zusammenfassung

Lungenkrebs ist nach wie vor eine der tödlichsten Krebserkrankungen und aktuell befinden sich mehrere zielgerichtete Therapien in Entwicklung bzw. Erprobung um die Überlebensrate von Patienten zu verbessern. Das Wissen um die genetischen Veränderungen, welche das Fortschreiten der Tumorerkrankung verursachen oder begünstigen, wird über die gesamte Erkrankungsdauer dringend benötigt um rechtzeitig mit zielgerichteten Therapien auf angreifbare Mutationen reagieren zu können.

Nach wie vor basiert die Detektion von relevanten Mutationen hauptsächlich auf Tumorbiopsien deren invasiver Charakter eine hohe Belastung für die Patienten darstellt. Ein neuer Zugang zur Mutationsdetektion ist die sogenannte „flüssige Biopsie“. Hierbei handelt es sich u.a. um die Analyse von zirkulierender Tumor-DNA (ctDNA) aus dem Blut. Durch den minimal-invasiven Charakter der Blutabnahme könnte diese Methode eine schnelle und verträglichere Variante zu herkömmlichen Gewebsbiopsien ermöglichen.

Das Hauptaugenmerk der vorliegenden Diplomarbeit war es, verschiedene Fragestellungen bezüglich einer möglichst robusten Mutationsdetektion von Tumor-DNA aus dem Blutplasma zu behandeln. Besonderer Fokus lag dabei auf den präanalytischen Methoden, wie z.B. der Auswahl von Blutröhrchen und dem Vergleich verschiedener Methoden zur Extraktion von Plasma DNA. Die untersuchten Parameter waren die Ausbeute, Größenverteilung und der Tumorgehalt der Plasma-DNA um somit Standards für weitere Arbeiten zu etablieren. Außerdem wurde ein Gen-Panel von QIAGEN mit einer Krebszelllinie getestet um seine Auflösungsgrenze zur Mutationsdetektion festzulegen. Im Anschluss wurde das Panel an sechs Lungenkrebsproben getestet.

Zu diesem Zweck wurde Blut von zwei gesunden Kontrollpersonen und zwei metastasierten Kolonkarzinompatienten gesammelt und die frei-zirkulierende DNA (cfDNA) mittels eines manuellen, vollautomatischen und eines Vakuum-basierten Verfahrens extrahiert. Es konnten keine signifikanten Unterschiede zwischen den Blutröhrchen gefunden werden. Bezüglich der DNA Ausbeute war die Vakuum-basierte Extraktion den anderen Verfahren überlegen.

Die Auflösungsgrenze des Gen-Panels von QIAGEN wurde mittels Verdünnungsreihe der Krebszelllinie HCT116 festgelegt. Dabei wurde gezeigt, dass eine Mutationsdetektion unter einer mutierten Allelfrequenz (MAF) von 5% nicht möglich war. Nach der Sequenzierung und manuellen Variantenauswahl der sechs Lungenkrebsproben wurden 51 Varianten identifiziert, welche allerdings nicht mit einer unabhängigen Methode bestätigt wurden.

Zusammenfassend konnte gezeigt werden, dass die drei getesteten Blutröhrchen gleich gut in der Konservierung der frei-zirkulierenden DNA abschnitten. Die Vakuum-basierte Extraktion erwies sich hinsichtlich der DNA Ausbeute als das beste Verfahren und bietet sich daher als bevorzugte Methode zur Extraktion an, vor allem wenn man den geringen Anteil der ctDNA im Plasma in Betracht zieht.

Die etablierte Auflösungsgrenze des Gen-Panels von QIAGEN von circa 5% MAF hat sich als nicht ausreichend erwiesen um Mutationen aus dem Plasma sensitiv genug nachzuweisen wenn man bedenkt, dass der Anteil von ctDNA nur 0.01% betragen kann. Dies bestätigte sich nach der Austestung des Gen-Panels mit den Lungenkrebsproben, wobei eine hohe Anzahl an Varianten auffiel und bestimmte Mutationen in mehreren Proben zu sehen waren. Dies könnte für eine hohe falsch-positiv Rate oder Artefakte sprechen.

Abstract

Lung cancer remains one of the deadliest tumor entities and currently various targeted therapies are under investigation to improve overall and progression-free survival. Knowledge of the genetic alterations affecting tumor progression in individual patients is urgently needed throughout the course of the disease to readily respond if actionable mutations for targeted therapies emerge. To date mutation detection mainly relies on tumor biopsies which impose a burden to patients. A novel approach for mutation detection in cancer patients is termed liquid biopsy. Among other things, it refers to the analysis of circulating tumor DNA (ctDNA) from blood and could offer a fast and minimal-invasive alternative to tissue biopsies.

The focus of this thesis was to address several issues which affect robustness of mutation detection in plasma DNA. Especially preanalytical considerations were investigated such as the selection of blood tubes and different extraction methods of cell-free DNA (cfDNA) from plasma. Yield, size distribution and tumor content of the plasma DNA were compared to establish standards for further research. Furthermore, a targeted gene panel was tested using a cancer cell line in order to assess the resolution limit for mutation detection and subsequently several lung cancer samples were sequenced using this panel.

To this end, blood of two healthy controls and two metastasized colon carcinoma patients was drawn into three different blood tubes and cfDNA was extracted from plasma using three different protocols including a manual, a vacuum-based and a fully-automated version. No significant differences were found between the different blood tubes. However with respect to DNA yield, the vacuum extraction method outperformed the two other methods.

The resolution limit for the QIAGEN GeneRead Lung Cancer Panel was established using a dilution series with the cancer cell line HCT116 and additionally the panel was tested on six lung cancer samples. It was shown that the detection of mutations below a mutant allele frequency (MAF) of about 5% was not possible with the tested panel. A total of 51 variants were identified after sequencing and manual variant prioritization of the six lung cancer samples. However, the variants were not confirmed with independent methods.

In conclusion, the three blood tubes performed equally well in preserving cfDNA if processed in a timely manner. The vacuum extraction proved to be superior to the other methods concerning DNA yield and this suggests its preferential use for further extractions taking the low abundance of ctDNA into account.

The established resolution limit of about 5% MAF for the QIAGEN GeneRead Lung Cancer Panel proved not to be sufficient for sensitive mutation detection in plasma DNA where ctDNA fraction can be as low as 0.01%. This was confirmed when testing it on the lung cancer samples which showed a high number of variants and certain mutations that were seen across the sample range, indicating a high number of false-positives or artifacts.

1 Introduction

1.1 Lung cancer in the era of molecular medicine

Lung cancer remains the leading cause of cancer mortality in Europe in men and is expected to become the neoplasm with the highest mortality rate also in women.(1) Furthermore most patients have advanced disease stages when they consult a physician due to clinical symptoms caused by a lung malignancy.(2) In the past there has only been modest improvement in the median survival of patients and therefore the need for new therapeutics, which act in novel pathways, arose.(3) Nowadays a lot of progress is being made in uncovering the molecular pathways promoting malignancy in lung cancer, especially in non-small cell lung cancer (NSCLC). This subtype of lung cancer accounts for about 85% of all lung cancers and the adenocarcinoma being its most common histological subtype.(4) This has led to the introduction of therapies such as tyrosine kinase inhibitors (TKIs) or monoclonal antibodies that are aimed at specific targets which maintain cancerogenesis through frequently mutated genes in lung adenocarcinoma and a variety of other cancers. Several studies showed that targeted therapies are superior to chemotherapy if actionable driver mutations of these frequently mutated genes like *EGFR* (encoding epidermal growth factor receptor) or *ALK* (encoding anaplastic lymphoma receptor tyrosine kinase) are present.(5) Another benefit of these therapies is that they mainly target tumor cells in contrast to systemic chemotherapies which eradicate all quickly proliferating cells.

1.1.1Genetic aspects of tumorigenesis

The formation and progression of every cancer underlies mutational events of a single cell that provide the cell with an evolutionary advantage compared to other cells. These mutational events allow the cell to proliferate autonomously and subsequently invade other tissues and metastasize.(6) Mutations allowing fast growth to cells and thereby providing positive selection to those cells are frequently involved in tumorigenesis and referred to as driver mutations.(6) The main hallmarks of driver mutations to promote

cancerogenesis enable cancer cells to resist apoptosis, to proliferate autonomously, to invade other tissues and thereby metastasize and to promote angiogenesis to allow fast growth. Genetic alterations which maintain tumorigenesis are often mutually exclusive, which means that just one mutation drives cancer proliferation. Many other mutations that promote the oncogenic phenotype to a much lesser degree are found alongside the driver mutation and are referred to as "passenger mutations". By mainly proliferating through the driver mutation, the cancer cell becomes oncogene-addicted because it relies on the constant signal of the driver oncogene. This is the reason why driver mutations make promising biomarkers for selecting targeted therapies. Many driver mutations affect genes that encode different receptor tyrosine kinases, i.e. high-affinity cell surface receptors for growth factors. These receptors play a crucial role in normal cell regulation and those genes are frequently mutated in a wide variety of cancers. Already back in 2000, after the completion of the Human Genome Project, the sequencing of the human genome showed that it contains 90 tyrosine kinase genes of which 58 are receptor type proteins.(7) Renowned examples for such genes are e.g. *EGFR*, *RET*, *ALK*, *KRAS* and if mutated they represent actionable driver oncogenes for a variety of tyrosine kinase inhibitors (TKIs) such as erlotinib and crizotinib.(8)

1.1.2 Driver mutations in lung cancer

In recent years numerous driver mutations for lung cancer, mainly for the subtype of NSCLC, have been elucidated and many new ones are under investigation.(8) A well-established example are mutations in the gene *EGFR* that encodes the epidermal growth factor receptor, a tyrosine-kinase receptor which is part of an important activating pathway. A study by J. Guillermo Paez et al. revealed that patients who harbored mutations in *EGFR* showed clinical response to the TKI gefitinib.(9) The detection of aberrations in several other genes like *ALK* and *ROS1* followed and patients with changes in the genes mentioned above profited from a therapy with the TKI crizotinib.(10, 11) According to a study by the College of American Pathologists it is now recommended to test for mutations in *EGFR* and *ALK* to guide treatment decisions.(12) In a nationwide French study 17,664 patients with NSCLC were screened for mutations in *EGFR*, *ERBB2*, *KRAS*, *BRAF*, *PIK3CA* and *ALK* rearrangements during a 1-year period. In about 50% of patients a genetic alteration

was found which subsequently affected first-line treatment for another 51% in this patient subset. In this study an improvement in overall survival (16.5 vs 11.8 months) was observed in those patients harnessing a genetic alteration compared to those without any.(13) Between 2009 and 2012 the Lung Cancer Mutation Consortium analyzed 10 genes in 773 patients with metastatic lung adenocarcinomas for targetable driver mutations using multiplex genotyping. In 466 of 733 patients a driver mutation was found with *KRAS* being most frequently mutated (25% of the 733 tumors), followed by sensitizing *EGFR* mutations (17%) and *ALK* rearrangements (8%).(14) A recent genetic profiling of 230 lung adenocarcinoma samples performed by The Cancer Genome Atlas Research Network showed an increase in the proportion of cases with an activation of the receptor tyrosine kinase (RTK)/RAS/RAF pathway from 62% to 76%.(15) This pathway, also called mitogen-activated protein kinase (MAPK) pathway, regulates proliferation and differentiation of cells and includes many proteins encoded by genes that are frequently involved in tumorigenesis when mutated. Renowned examples of such genes are *EGFR*, *ALK*, *RET* and *ROS1* and the aforementioned study added amplifications in *MET* and *ERBB2* and mutations in *NF1* and *RIT1* as potential driver events in lung adenocarcinoma and proposes implementing MET and ERBB2/HER2 inhibitors in treatment regimens.(15) Furthermore this study showed that tumors that otherwise lack such driver events and are therefore considered to be oncogene-negative, harness significantly more mutations in genes acting as tumor suppressors like *TP53*, *KEAP1* and *NF1*.(15)

1.1.3 Tissue biopsies and their limitations

The fast pace at which new potential therapeutic targets are being elucidated and the development of tailored drugs brings up the question of how to acquire the genetic information that is needed to guide those emerging or existing therapies. A well-established method for somatic mutation testing is the tissue biopsy and subsequent microdissection of tumor cells to get the desired tumor DNA. The classical workflow encompasses histological assessment, hematoxylin and eosin staining to confirm diagnosis and finding an area with a high proportion of tumor to get a high number of tumor cells for microdissection and subsequent sequencing. The DNA is then extracted from formalin fixed paraffin embedded (FFPE) material.(16) However, several important issues come along

with tissue biopsies such as availability of sufficient tissue, storage conditions or fixation processes which can have an impact on DNA fragmentation and degradation. The presence of material like necrotic tissue and hemoglobin can also have an impact on the efficiency of the polymerase chain reaction (PCR).(16) Currently the acquisition of specimens to test for mutations relies on invasive procedures such as open biopsies, fine-needle aspiration or surgical resections.(12) Like with any invasive procedure, there are several risks inherent with tissue biopsies like infections or surgical complications. Furthermore, they impose a burden to patients, increase healthcare costs and are thereby a limiting factor for the monitoring of driver mutations. Additionally, conventional biopsies only depicture the genetic makeup of a specific area and probably do not represent overall tumor heterogeneity and mutational status of metastases. Another problem is that tumor cells are under constant selection pressure through treatment regimens and often acquire resistance to targeted and systemic therapy such as the well-described secondary mutation T790M in *EGFR* which inhibits response to TKIs.(17) Interestingly, in the same study by Lecia V. Sequist et al., it was shown that this acquired mutation disappeared in a patient in a subsequent biopsy taken after 10 months of discontinuation with TKI-therapy.(17)

Thus it became apparent that serial monitoring of the mutational status of tumors is necessary to assess tumor heterogeneity in the course of time in order to be able to readily respond to changes that could impact treatment decisions.(18) Alternatives to the established biopsies are needed, preferably offering a cost-efficient, fast and non-invasive way to get hold of the desired genetic information. This has led to a rapidly increasing interest in the so-called liquid biopsy which seems to present such a new approach. Liquid biopsy refers to the acquisition of circulating tumor cells (CTCs) and cell-free circulating tumor DNA (ctDNA) from the blood which act as potential surrogates for the entire tumor genome.(19)

1.2 Cell-free DNA as a tool for liquid biopsies

In our study we focused on the extraction and analysis of cell-free DNA (cfDNA). Some advantages compared to CTC based liquid biopsies include the fact that there is no need for special equipment and that cell-free circulating tumor DNA (ctDNA), which is a fraction of the cell-free DNA, is more abundant than CTCs in the circulation.(19, 20) CfDNA was detected in the blood of healthy individuals as early as 1948.(21) A long time passed until Leon et al. demonstrated increased cfDNA levels in cancer patients.(22) The detection of mutations in genes like *KRAS* in the plasma or serum of pancreatic carcinoma patients and the suggestion to use these findings as novel biomarkers for diagnosis, prognosis and treatment response followed.(23) The most successful early application for cfDNA detection was prenatal screening for aneuploidy after it was discovered by Lo and colleagues that fetal DNA is circulating in the mother's bloodstream and can be detected in plasma samples as low as 10µl.(24) A recent study by Bianchi et al in which 2052 pregnant women were enrolled showed significantly lower false positive rates for trisomy 21 and 18 screening with cfDNA testing compared to standard screening (0.5% vs. 4.2%).(25)

1.2.1 Properties of cfDNA

The mechanisms by which cf- and ctDNA is released into the circulation are still not fully understood and the main hypotheses state that it is derived from apoptotic as well as necrotic cells but might also be actively released from tumor cells.(26, 27) CfDNA concentrations observed in various studies vary considerably and cancer patients generally showed higher concentrations with a range from 0 to >1000ng/ml compared to healthy individuals with cfDNA levels ranging from 0 to 100ng/ml.(28) Recently it has been shown that the majority of cfDNA is derived from apoptotic processes which is reflected in the size distribution around 160bp and multiples thereof, which corresponds to the size of nucleosomes into which DNA is degraded during apoptosis.(29)

With respect to clearance in pregnant women, it was shown that 2 hours postpartum there was no fetal DNA detectable in the plasma anymore in most women which is indicative of the short half-life of cfDNA.(30) A more recent study on the clearance of circulating fetal

DNA after delivery indicates that it occurs in two phases, a rapid phase with a half-life of about an hour and a slower phase with the mean half-life being around 13 hours. The same study also showed that renal excretion was involved in the clearance of circulating fetal DNA but that it was not the main route for it.(31) So despite numerous studies, the understanding of clearance and release of cfDNA remains limited.

1.2.2 CtDNA in tumor monitoring

As aforementioned ctDNA represents a varying fraction of cfDNA in cancer patients, the concentration of it probably depending upon factors like the extent of tumor load and metastases. As summarized in the review by Molina-Vila et al. the fraction of ctDNA in cfDNA can range from below 0.1% up to 10% and more. This depends on factors such as tumor burden, response to therapy, vascularization of the tumor and apoptotic rate and it is probably derived from lysed apoptotic and necrotic cells as well as released actively or through digestion of circulating tumor cells by macrophages.(32)

In a 2004 study by Gautschi et al. it was shown that increased plasma DNA concentrations in NSCLC patients significantly correlated with lower survival rates and advanced tumor stage and that tumor progression following chemotherapy was also associated with higher plasma DNA concentrations.(33) A more recent study by Bettegowda et al. compared patients with localized disease to patients with metastases for the presence of ctDNA in a variety of cancers and it was shown that ctDNA was detectable in higher percentages in the group of cancer patients with metastases (49-79% vs. 86%-100%).(34) In the same study a strong correlation between tumor stage and detectable ctDNA levels was found, which continually increased from stage I to stage IV disease (47, 55, 69, and 82%).(34) In addition to a higher cellular turnover due to larger tumor sizes, a lower activity of DNase compared to healthy individuals could be another reason for higher cfDNA levels in cancer patients.(35)

Despite recent improvements in the genetic analysis of ctDNA, many hurdles have to be overcome, e.g. low fractions and absolute amounts of ctDNA, high degradation or contamination with genomic DNA (gDNA) by lysed blood cells. Moreover, ctDNA analysis is

complicated by the heterogeneity of multiple subclonal populations that only share some somatic mutations among all cells.(36)

1.2.3 Approaches for ctDNA detection in lung adenocarcinoma

From a clinical perspective lung adenocarcinoma is a prime example for ctDNA mutation screening with novel 3rd generation TKIs that target the frequent secondary *EGFR* T790M mutation and have recently been proven of being superior to standard chemotherapy.(37)

There are various molecular assays available to detect mutations in lung adenocarcinoma. Ideally they should be cost efficient, offer high sensitivity and cover all clinically important targets.(16) For example methods for detecting *EGFR* mutations can roughly be separated into "screening" and "targeted" approaches with the first enabling detection of all mutations including novel variants and the latter only detecting specific mutations.(38) Two examples for targeted assays are the therascreen® EGFR Plasma Kit by QIAGEN and the cobas® EGFR Mutation Test by Roche. Both assays showed a good sensitivity for detecting specific mutations like *EGFR*-sensitizing mutations or the secondary T790M *EGFR* mutation and are based on modified real-time PCR techniques.(16, 32) The AURA trials evaluated the clinical activity of osimertinib in patients who were tested for the *EGFR* T790M resistance mutation using the cobas® EGFR Mutation Test. As part of these clinical trials *EGFR* mutation testing of tissue was additionally performed by using ctDNA and four different platforms for *EGFR* mutation detection in plasma DNA were compared.(39) Two non-digital detection methods, namely the cobas® EGFR Mutation Test and the therascreen® EGFR amplification refractory mutation system (ARMS) assay and two digital techniques, the Droplet Digital PCR (ddPCR™) by Bio-Rad and beads, emulsions, amplification and magnetics (BEAM)ing dPCR by Sysmex Inostics were compared. Sensitivity, specificity and concordance were assessed for the common T790M and L858R mutations and exon 19 deletions in *EGFR* with tissue tests as reference. All platforms showed a high sensitivity (78%-100%) and specificity (93%-100%) for L858R mutation detection. However, the digital techniques were superior in detecting T790M mutations.(39) In a further assessment with 72 plasma samples, BEAMing dPCR was compared to cobas® EGFR Mutation test and results showed high sensitivity (81% vs. 73%) and good specificity (58% vs. 67%) for T790M mutation testing. The "reduced" specificity

compared to tissue samples could be explained with tumor heterogeneity being favorably depictured by ctDNA and highlights an advantage of ctDNA analysis.(39) The same study showed that clinical response rate to osimertinib was almost identical between plasma and tissue (59% vs. 61%) and thereby further suggests plasma ctDNA as an alternative to tissue biopsy testing.(39) The recent AURA 3 trial showed significantly better progression-free survival (PFS) in patients with confirmed T790M-positive tumor and plasma who were treated with osimertinib compared to patients treated with platinum-pemetrexed (8.2 vs 4.2 months).(37) Both tumor and plasma samples were tested with the cobas EGFR Mutation Test with plasma testing showing a relatively low sensitivity of 51% and the authors recommended additional biopsy testing in patients with a T790M-negative plasma test result.(37) Another plasma genotyping study using the BEAMing method achieved a sensitivity of 70% for the same mutation compared to tumor testing and PFS and objective response rate (ORR) in T790M-positive samples were similar for plasma and tumor (ORR 63% vs 62% and PFS both 9.7 months).(40)

These aforementioned modified PCR techniques are commercially available and comparatively inexpensive, time-efficient and already used in a clinical setting. The disadvantage of these methods is, however, that they can only detect a few mutations within a specific gene.(32) Next-generation sequencing (NGS) can address these limitations and is already well established for tissue-based testing.(32) In order to apply NGS for ctDNA detection, deep sequencing for a limited number of genes is used to improve sensitivity. Studies using this approach showed that detection is possible for MAF as low as 0.2% in cfDNA and accurate mutation detection in genes such as *ALK*, *ROS1*, *EGFR*, *MET* and *PIK3CA* was achieved with sensitivities ranging from 58% to 77%.(41, 42) A new NGS-based approach called cancer personalized profiling by deep sequencing (CAPP-Seq) that was introduced by Newman et al. showed 100% sensitivity for ctDNA detection in stage II-IV NSCLC and 50% for patients with stage I disease with a MAF down to about 0.02%.(43) Recently, Paweletz et al. used a NGS panel targeting 11 driver oncogenes with a novel methodology optimizing on-target reads and minimizing artifacts.(42) In this study mutations were detected down to a MAF of 0.4% in DNA dilutions and a sensitivity of 77% was achieved when tested for 62 known driver mutations in 48 patients.(42)

1.3 Preanalytical considerations

In order to establish the liquid biopsy in clinical practice a standardization of preanalytical procedures is needed. Benchmarks for the use of blood collection tubes, the use serum versus plasma and for the processing of blood samples to maximize tumor DNA yield are urgently needed.(44) As shown by Thierry AR et al., plasma is superior to serum when it comes to cfDNA in oncology due to the fact that in serum samples DNA is released from blood cells during the clotting process. This results in a contamination with gDNA of blood cells.(45) Concerning blood sampling, an important issue to address is to minimize gDNA contamination by lysed blood cells due to delays in processing, suboptimal storage conditions or agitation during transport. Thierry et al. demonstrated that cfDNA concentrations in blood samples stored up to 4 hours at room temperature (RT) or stored at a temperature of 4°C remained stable but noted a significant increase after 6 hours of storage at RT.(46) Ideally blood samples should be processed, i.e. the centrifugation, isolation and freezing of the blood plasma, within 3 hours.(47) In a clinical setting it might often be difficult to promptly process blood samples and this circumstance led to the development of various blood tubes containing reagents to prevent cellular lysis and thereby limit gDNA contamination from blood cells. Several studies have been conducted to compare different blood tubes for cfDNA collection such as the widely used EDTA tubes, the Cell-free DNA BCT™ (Streck) or the PAXgene Blood DNA tubes (QIAGEN).(48-51) The main conclusions in those studies were that up to 6 hours all tubes performed equally well in preserving ctDNA. Nevertheless, BCT tubes from Streck were superior in preventing cell lysis and cellular release of gDNA when shipped or stored at RT for several days.(49-51) A major disadvantage of the BCT from Streck is its relatively high cost compared to standard EDTA tubes which may be bypassed by adding a neutral buffered solution containing formalin to stabilize cell membranes into EDTA tubes.(29)

Regarding preanalytical considerations starting from the blood collection, plasma preparation and storage, the final step, i.e. the extraction of cfDNA from the plasma, remains. There are several kits for cfDNA extraction available which have been compared to each other with respect to DNA yield and other parameters.(52-54) A substantial number of those extraction kits function by selectively binding the DNA to a silica-based membrane

with the kits by QIAGEN being prominent examples. In one of those studies four commercially available kits were tested for recovery of cfDNA using spike-in of reference templates and the QIAamp® DNA Blood Mini Kit and QIAamp® Circulating Nucleic Acid Kit (CNA) gave the highest recovery rates.(54) Also in another study the CNA kit outperformed two other extraction kits regarding cfDNA yield, the NucleoSpin Plasma XS kit by Macherey-Nagel and the FitAmp plasma/serum DNA isolation kit by Epigentek.(52)

1.4 Aim of this thesis

In the era of personalized medicine, the liquid biopsy stands out as one of the most promising tools to monitor the genetic makeup of a broad range of tumors. It seems to offer an option to readily respond to subtle changes in treatment response caused by e.g. upcoming molecular resistances to the current therapy. More and more private companies even promote the liquid biopsy as an option for prescreening supposedly healthy individuals for hidden malignant transformations at a very early stage thereby excessively increasing the chances of curative treatment. At this early stage there are still many issues in particular with respect to preanalytical and analytical considerations concerning standardization and best practice for isolating, processing and analyzing cell-free plasma DNA. The aim of this thesis was to address such issues. The first objective of this thesis was to compare different blood tubes and extraction methods for cfDNA yield, size distribution and tumor content. The second objective was to test the QIAGEN GeneRead Lung Cancer Panel, consisting of 45 genes associated with lung cancer, to establish a resolution limit using the cancer line HCT116 and to test the panel using several plasma DNA samples from lung cancer patients.

1.4.1 Comparison of extraction methods and blood tubes

The following parameters should be assessed for comparison: yield, size distribution and tumor content in the plasma DNA. By comparing the aforementioned parameters our goal was to distinguish differences between the different blood tubes and different purification methods.

1.4.2 GeneRead Panel: Dilution with cancer cell line and test on lung cancer samples

In the second part a commercially available kit by QIAGEN for amplicon-based enrichment of lung cancer associated genes should be tested. Therefore, a dilution series of the cancer cell line HCT116 as well as several lung cancer plasma samples were enriched, using the GeneRead™ DNAseq Targeted Panels V2 and sequenced on an Illumina MiSeq. The acquired data was then analyzed using an in-house analysis algorithm and Annovar software and compared to mutations that are listed in COSMIC.

2 Materials and methods

2.1 Patient samples

For tube comparisons blood was taken from four individuals. Two of them (one male and one female donor) were presumably healthy. In addition, two patients with metastasized colorectal cancer (CRC) were recruited at the Division of Oncology at the Medical University of Graz.

Moreover, for testing the GeneRead panel a set of 6 plasma DNA samples from patients with lung cancer, which were available from an ongoing study were used.

2.2 Blood collection

Blood samples from all donors were drawn into three different blood tubes, i.e. a conventional Vacuette K3EDTA blood tube to which neutral buffered formalin (NBF) was added immediately after the blood draw, the PAXgene blood DNA blood tube by QIAGEN and the Cell-Free DNA BCT by Streck. The NBF was added to the Vacuette K3EDTA blood tube to stabilize nucleated blood cells and thereby prevent the contamination with genomic DNA. The PAXgene blood DNA tube and the BCT by Streck contain a proprietary mixture of reagents to prevent coagulation and ensure stabilization of nucleated blood cells. Samples were gently inverted and stored at room temperature (RT) after the blood draw.

2.3 Sample Processing

The tubes were centrifuged at 200 x g for 10 min with a subsequent centrifugation step at 1600 x g for 10 min and then the supernatant (plasma without any cells) was collected and transferred into new tubes. After another centrifugation at 1600 x g for 10 min the plasma was transferred to sterile 1.5ml Eppendorf tubes and stored at -80°C for further processing.

2.4 DNA isolation from plasma

Three different approaches were used to isolate the cfDNA from plasma samples. The principal workflow is depicted in **Figure 1** and consist of the four steps lyse, bind, wash and elute. Lysis is necessary due to the fact that cfDNA is usually bound to proteins or enveloped in vesicles and is performed by adding proteinase K and Buffer AL/ACL depending on the purification method. Under the presence of elevated temperature the inactivation of DN/RNases takes place and the release of cfDNA is carried out. In the vacuum protocol the Buffer ACB is additionally added to enable optimal binding of the cfDNA to the silica membrane of the QIAamp Mini columns. During a brief vacuum or centrifugation step the cfDNA is then absorbed to the silica membrane of the Mini columns. Subsequently two washing steps with the Buffers AW1/ACW1 and AW2/ACW2 are performed to further purify cfDNA and remove contaminants. Finally cfDNA is eluted in nuclease-free water. All three approaches, i.e. fully-automated purification with the QIAcube, manual purification using the spin protocol and vacuum based purification with the QIAvac 24 Plus vacuum manifold were applied to all samples. For all extractions 1ml of plasma was used from a total of 36 samples.

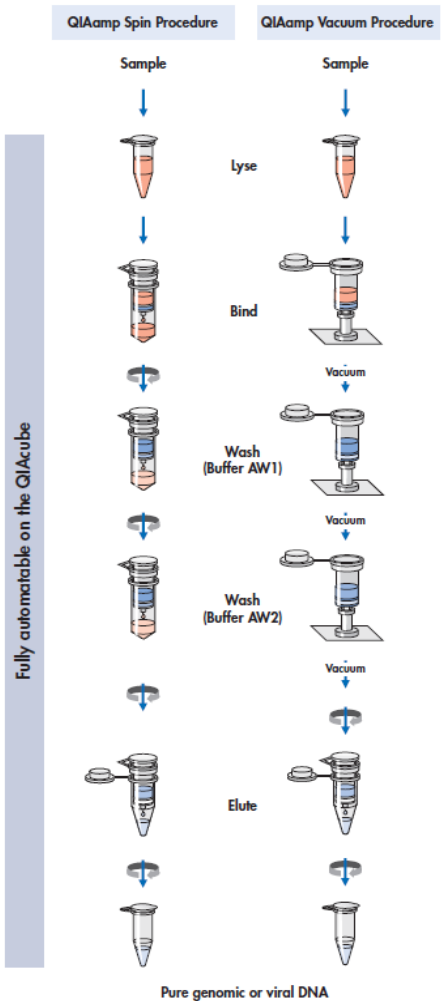

***Figure 1**. cfDNA purification on QIAamp Mini spin columns (QIAamp DNA Mini and Blood Mini Handbook 06/2012)*

2.4.1 DNA purification from blood (Spin protocol)

We used a slightly modified protocol from the ”QIAamp DNA Mini Kit and QIAamp DNA Blood Mini Kit Handbook”. For the manual extraction of 1ml plasma, two aliquots with 50µl proteinase K and 500µl plasma in 2ml microcentrifuge tubes were prepared. We proceeded

with adjusted quantities of Buffer AL and ethanol accordingly to the protocol. The mixture was applied to the QIAamp Mini spin column and the subsequent washing steps with Buffer AW1 and AW2 followed. The last step, i.e. the elution, was performed with 45µl of nuclease-free water and after incubation and centrifugation the samples were stored at -20°C.

2.4.2 Automated DNA purification on the QIAcube

The aforementioned spin protocol consists of the steps lyse, bind, wash and elute which can also be fully automated on the QIAcube. The same reagents as in the QIAamp DNA Mini Kit are used, with the exception that 3 aliquots of 333µl of plasma were used for extraction and pooled after the completion of the QIAcube program.

2.4.3 DNA purification with the QIAvac 24 Plus (Vacuum protocol)

The QIAvac 24 Plus works through vacuum processing of samples. It allows fast processing by applying a vacuum to the QIAamp Mini spin columns through which the samples and wash solutions are then drawn. Before the purification, the buffers were prepared according to the instructions. The QIAamp® Circulating Nucleic Acid (CNA) protocol for the processing of 1ml of plasma was used and accordingly 5.3ml of Buffer ACL and 33.8µl of carrier RNA in Buffer AVE were used for 6 samples. We performed the processing two times for a total of 12 samples being processed and only deviated from the protocol by performing the elution with 90µl of nuclease-free water instead of using the Buffer AVE.

2.5 Analysis of cfDNA

After sample processing DNA concentration and integrity (size distribution) of the cfDNA were assessed in order to compare the different blood tubes and the different DNA purification methods. The DNA concentration was measured through fluorescence with Qubit™. The evaluation of DNA fragment size distribution was performed using the Agilent Bioanalyzer. It allows to assess DNA integrity, i.e. the ratio of longer segments to shorter

segments, and to check for contamination with high molecular weight DNA, indicating the presence of gDNA by e.g. white blood cells.

2.5.1 DNA yield quantification with Qubit

The purified cfDNA was quantified using the Qubit dsDNA HS Assay Kit, which employs a fluorescent dye which binds to the DNA and emits a signal that is being measured. The samples were prepared for measuring by adding 5µl of the sample to 195µl of the working solution which consists of the Qubit dsDNA HS reagent and the Qubit dsDNA HS buffer. The prepared samples were then vortexed and incubated at room temperature for 2 min. Before the sample quantification two standards were measured to run a new calibration.

2.5.2 Sizing and analysis of DNA fragments with the Agilent High Sensitivity DNA Kit

We used the microfluidics-based platform Agilent High Sensitivity DNA Kit to assess the size distribution of the cfDNA samples and thereby draw conclusions about DNA integrity. CfDNA fragment sizes of healthy individuals normally peak at 166bp due to enzymatic breakdown of nucleosomes during e.g. apoptosis. A peak around 320bp is often seen in tumor patients and a contamination with high molecular weight DNA is indicative for DNA from blood cells. We used a normalized amount of 800pg of sample DNA for the Bioanalyzer Chip. Therefore, the volume containing 1600pg was calculated for each sample and concentrated to 2µl in a Speed Vac Eppendorf. After following the Agilent High Sensitivity DNA Assay Protocol for preparing the Bioanalyzer chip, 1µl of the sample was loaded onto the chip. We only deviated from the protocol by vortexing the chip at 2000 instead of 2400rpm.

2.5.3 Evaluation of tumor content using mFAST-SeqS method

In order to estimate overall tumor content of the two colon carcinoma patients we used the untargeted Modified Fast Aneuploidy Screening Test-Sequencing System (mFAST-SeqS) method, which was previously established at the Institute of Human Genetics.(55) This method is a fast and cost-effective way of prescreening plasma DNA samples for the presence of sufficient tumor content for further analyses by sequencing.(55) By

amplification of uniquely mappable *LINE1*-sequences across the genome, somatic copy number alterations (SCNAs) are detected on chromosome-arm level. In the process chromosome-arm specific and genome wide z-scores, i.e. the comparison of read count deviations per chromosome which are summed up and compared to a non-cancer population, are established. In short, the z-score displays the number of standard deviations from a population mean. The genome-wide z-score is highly correlated to mutant allele frequencies and therefore reflects tumor content in the plasma sample.(55) The mFAST-SeqS protocol consists of two PCR reactions (i.e. a target-specific PCR in order to amplify uniquely mappable *LINE1*-sequences followed by an index PCR) that are each followed by a PCR clean up using AMPure XP beads. The master mix for one PCR reaction was composed of nuclease free water, Phusion HF Buffer (5x), 0.5 µM forward primer, 0.5µM reverse primer, 10mM dNTPs and 2U Phusion Hot Start II Polymerase. After mixing aforementioned contents the plasma-DNA sample was added and everything was mixed again and followed by a brief centrifugation. PCR was performed under following cycling conditions: denaturation at 98°C for 2 min followed by 8 cycles of 98°C for 10s, 57°C for 2 min and 72°C for 2 min. PCR products were purified accordingly to the protocol by adding 70µl of AMPure XP beads to each sample, followed by incubation for 5 min at room temperature and then 5 min on a magnet. Afterwards the supernatant was removed and the bead pellets were washed twice with 200µl of 70% ethanol each time. In the next step the pellets were resuspended by adding 12µl of 1xTE buffer and were again incubated for 2 min at room temperature and 2 min on the magnet. Finally, 10µl of the supernatant from each sample were transferred to a new tube in preparation for the subsequent PCR. The master mix for the second PCR remained the same, except for the primers of which the reverse primer acted as an index primer. Cycling conditions for the second PCR were 2 min at 98°C followed by a total of 18 cycles of 98°C for 10s, 65°C for 15s and 72°C for 15s. We then proceeded with the second PCR clean up as described above. Before sequencing the L1 amplicon libraries on the Illumina MiSeq, the amplification products were quantified using the Agilent High Sensitivity DNA Kit.

2.6 GeneRead^{TM} DNAseq Targeted Panels V2: Lung Cancer Panel

The GeneRead Targeted Panels use overlapping primer sets for the coding (exonic) regions of disease-focused genes that are separated into different pools to optimize target coverage and specificity. A specific primer design algorithm aims at minimizing nonspecific amplification and at maximizing coverage. After targeted enrichment library construction follows to enable next-generation sequencing and subsequent data analysis of the samples as depicted in **Figure 2**.

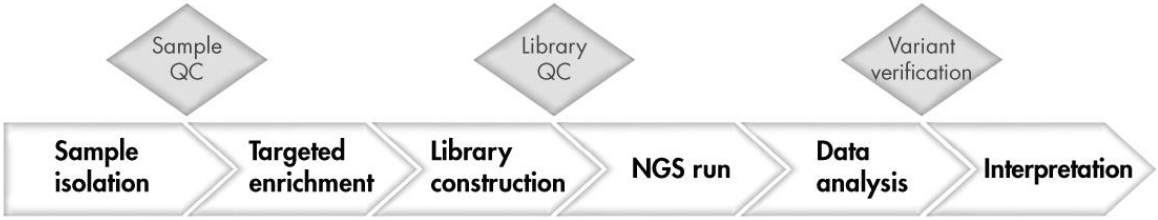

Figure 2. *Workflow of the GeneRead DNASeq Targeted Panels V2 (GeneRead DNAseq Targeted Panels V2 Handbook 06/2015)*

In this study the Human Lung Cancer Panel (NGHS-005X) was used. It encompasses 45 clinically relevant genes for lung cancer (listed in the Appendix) which were selected from using whole genome and exome sequencing studies from databases such as the Cancer Genome Atlas and COSMIC (Catalogue of Somatic Mutations in Cancer) and it claims to cover more than 95% of the coding (exonic) regions. These specific regions of interest, i.e. genes or regions which are associated with cancerogenesis, underwent a targeted enrichment using multiplex PCR. After completion of the PCR reaction the reactions were pooled and purified using AMPure XP beads. Subsequent sequencing was performed on an Illumina MiSeq NGS platform. Data analysis was performed using an in-house analysis algorithm and Annovar software.

2.6.1 Dilution with cancer cell line HCT116

The GeneRead Human Lung Cancer Panel was tested using the human colon cancer cell line HCT116. In order to assess the analytical sensitivity, a dilution series by spiking Promega control DNA with HCT116 to achieve dilutions of 12.5%, 6.25%, 3.125% and 1.56% was performed. To this end the DNA concentrations of the human control DNA (Promega) and

HCT116 DNA were measured with the Qubit dsDNA assay (Thermo Fisher) and set to 25ng/µl. The HCT116 DNA was then serially diluted from 50% down to 1.56%.

2.6.2 Enrichment of target regions by PCR

As recommended by the manufacturer a total of 40ng DNA (16µl of a 2.5ng/µl dilution) were used for the PCR. The 12.5%, 6.25%, 3.125% and 1.56% HCT116 dilutions were therefore diluted with NF-H2O to produce a 2.5ng/µl concentration for further processing.

Plasma DNA samples from lung cancer patients were previously extracted using the QIAamp® CNA Kit. Due to low absolute plasma DNA concentrations (L15 (59.6pg/µl; 86µl total volume), L18 (492pg/µl; 22µl), L27 (54.8pg/µl; 75µl), L38 (312.5pg/µl; 32µl), L48 (123pg/µl; 64µl) and L53 (101pg/µl; 69µl) the samples, were concentrated to 16µl using the SpeedVac system. Then 4µl of each DNA sample were aliquoted into each PCR reaction of the 4-pool panel and PCR reactions were set according to the manufacturer´s recommendation. **(Table 1)**

Table 1. *Preparation of the master mix*

Component	**Per *n* samples (µl)**
GeneRead DNAseq Panel PCR Buffer (5x)	4.4 x *n*
Primer mix pool (2x)	11 x *n*
GeneRead HotStar Taq DNA Polymerase (6 U/µl)	1.5 x *n*
Dnase-free water	0.7 x *n*
Final volume	**17.6 x *n***

16µl of master mix were mixed with each DNA sample and placed in the thermocycler to be amplified under reaction parameters depicted in **Table *2***. As the number of primer pairs per pool was between 400-1200, a total of 18 PCR cycles were performed in the second step of the cycling conditions. For the lung cancer plasma DNA samples 22 cycles were performed.

Table 2. *PCR program; 18 cycles were used for HCT116 samples and 22 cycles for lung cancer plasma DNA samples;*

Cycle	Temperature	Time
1	95°C	15 min
18/22	95°C	15 s
	60°C	4 min
1	72°C	10 min
1	4°C	∞

2.6.3 Pooling and purification

After completion of the cycling program, all four PCR reactions from each sample were combined in a new tube and mixed thoroughly. 40µl were then transferred to a new tube and 36µl (0.9 x volume) of AMPure XP beads were added. Samples were incubated at room temperature on a magnetic rack as described in the protocol. As soon as the solution was clear 70µl of the supernatant were transferred into a new tube. Beads, which contained unwantedly large DNA fragments were discarded. Subsequently 64µl (1.6 x the original volume of the PCR product) AMPure XP beads were added to the supernatant and incubated for 5 min at room temperature and 5 min on a magnetic rack. The supernatant was discarded and the beads containing the PCR products were washed with 80% ethanol and dried on the magnetic rack. The PCR products were finally eluted in 28µl NF-H2O. The tubes were put on a magnetic rack and after 2 min 25µl supernatant were transferred to a new tube. For the lung cancer plasma DNA samples, the whole PCR product was used for further processing. This means that 74µl PCR product were transferred to a new tube and 66.6µl (0.9 x volume) of AMPure XP beads were added. 129µl of the supernatant were put into a new tube and the beads were discarded. In the second clean-up 118.4µl (1.6 x the original volume of the PCR product) AMPure XP beads were added to the supernatant for all lung cancer samples. Finally, the DNA concentration was determined using the Bioanalyzer High Sensitivity kit.

2.6.4 Library Preparation using GeneRead Library Prep Kits for Illumina

The first step in the library construction consisted of the End Repair of the PCR products. 2.5µl End-Repair Buffer and 2.0µl End-Repair Enzyme Mix were combined with 20.5µl PCR-

enriched DNA from the pooling and purification step and incubated in a thermocycler for 30 min at 25°C, followed by 20 min at 75°C. In order to improve adapter ligation an A-addition followed. To this end 25µl of the end-repaired DNA were mixed with 3µl A-addition Buffer and 3µl Klenow Fragment (3'→5' exo-) and incubated for 30 min at 37°C, followed by 10 min at 75°C. Afterwards adapter ligation was performed by mixing 31µl of the DNA from the previous steps with 45µl Ligation Buffer, 1µl Adaptor, 4µl T4 DNA Ligase and 9µl DNase-free water and incubating the reaction at 25°C for 10 min in a thermocycler. In order to remove unbound adapters, the library was purified with AMPure XP beads as described in the manufacturer's protocol. Adapter-ligated fragments were then amplified by PCR. 25µl HiFi PCR Master Mix, 1.5µl Primer Mix, 6.5µl RNase-free water and 17µl of the library DNA were mixed and amplified under following cycling conditions. **(Table 3)** Finally a further cleanup with AMPure XP beads for the amplified library was conducted and the library was stored at -20°C prior to quantification with Bioanalyzer High Sensitivity kit.

***Table 3**. Cycling conditions for the amplification of the purified library*

Cycle	Temperature	Time
1	98°C	2 min
4	98°C	20 s
	60°C	30 s
	72°C	30 s
1	72°C	1 min
1	4°C	∞

2.6.5 Variant prioritization after sequencing

Sequencing was performed on an Illumina MiSeq NGS platform. Alignment and variant calling was done using the Burrows Wheeler Alignment (BWA) and Annovar open source algorithms followed by manual variant prioritization. The raw data was filtered for variants in exonic regions of the 45 genes that were enriched with the Human Lung Cancer Panel. Exclusion criteria were synonymous variants and variants with low coverage, low quality or strand bias. Additionally, frequent polymorphisms (>1%) that are listed in the esp6500 and 1000g databases were excluded. Single nucleotide polymorphisms (SNP) that are listed in the snp137 database and are known to be non-pathogenic were also eliminated.

3 Results

3.1 Comparison of different blood collection systems and plasma DNA extraction protocols

To test the different blood collection systems and library extraction protocols we used two healthy controls and two metastasized CRC patients. One parameter to assess the different tubes and extraction kits was plasma DNA yield after quantification with the Qubit dsDNA HS Assay Kit. **Table 4** summarized concentrations for the different tubes and extraction methods used.

Table 4. *Comparison of plasma DNA concentrations between tubes and extraction methods*

Sample	Tube	DNA Mini Kit (Spin)	QiaCube	CNA Kit (Vakuum)
		ng/ml plasma	ng/ml plasma	ng/ml plasma
Male control	Pax	3,1	7,2	12,3
	Streck	3,8	4,1	25,9
	EDTA	3,0	4,0	19,2
Female control	Pax	2,3	4,6	14,1
	Streck	3,2	5,8	23,9
	EDTA	3,6	6,8	17,8
Colon74	Pax	14,4	19,1	32,4
	Streck	7,5	20,6	41,4
	EDTA	10,8	17,9	37,4
Colon 75	Pax	670,5	1908,0	3888,0
	Streck	189,9	1836,0	5112,0
	EDTA	572,4	1053,0	4698,0

As expected plasma DNA concentrations and yields were significantly increased in CRC patients compared to healthy controls. There was a 10- to 1000-fold increase in plasma DNA concentrations when comparing the mean values of the tubes and extraction methods combined for healthy controls (mean: 9.2ng/ml; range: 2.3-25.9ng/ml) and cancer patients C74 (mean: 22.1ng/ml; range: 7.5-41.4ng/ml) and C75 (mean: 2199.3ng/ml; range: 189.9-5112ng/ml).

In contrast, no significant difference was observed between the different blood collection tubes. **(Figure 3)**

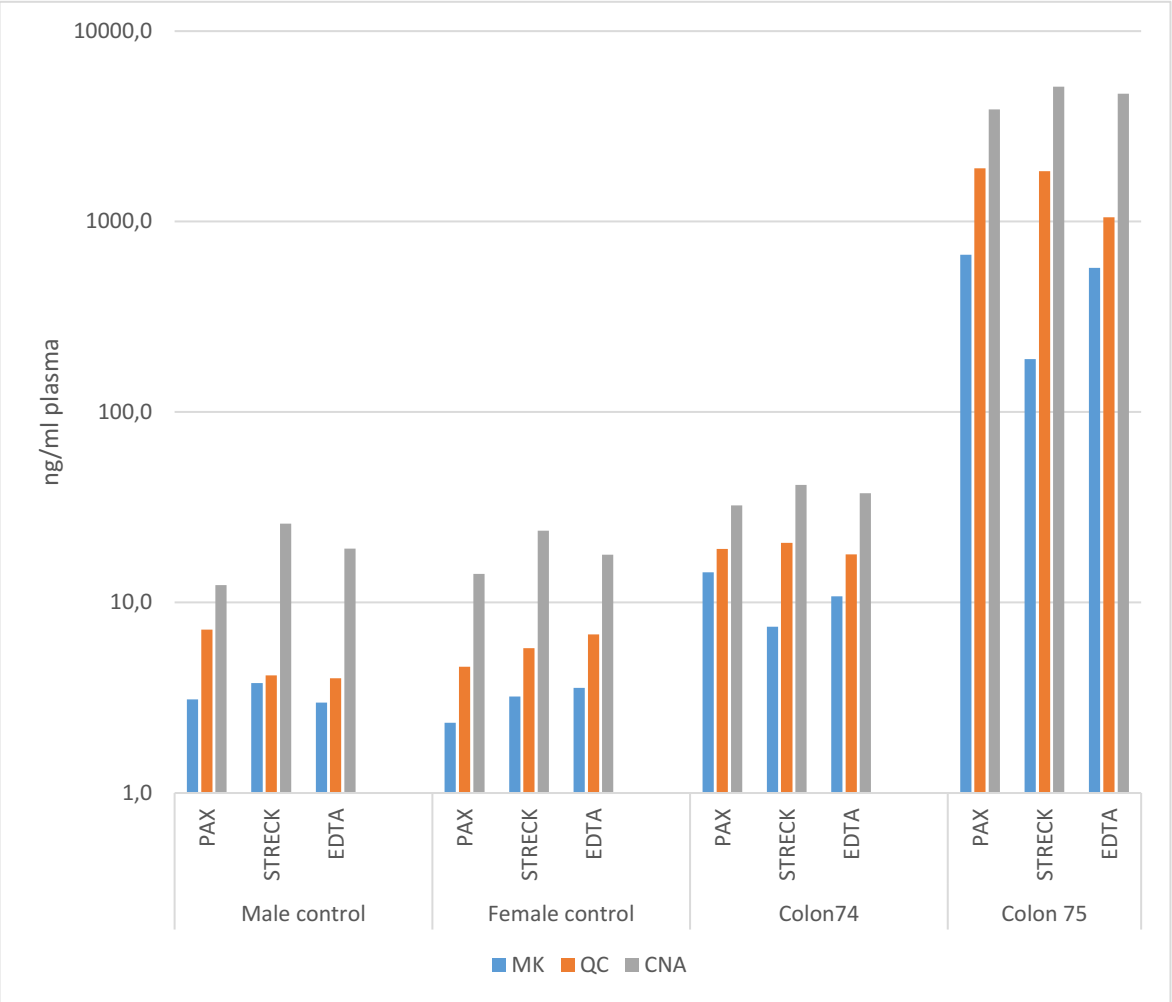

Figure 3. *cfDNA concentration differences between blood tubes (STRECK: Cell-free DNA BCT by Streck; PAX: PAXgene Blood DNA tube by QIAGEN; EDTA: Vacuette K3EDTA blood tube + 200µl 10% NBF) and different DNA purification methods (QC: automated on the QIAcube; MK: QIAamp DNA Mini Kit- Spin protocol; CNA: QIAamp Circulating Nucleic Acid Kit -Vacuum protocol). Results are displayed in a logarithmic scale due to the huge differences between healthy controls and cancer patients.*

However, with respect to DNA extraction kits significant differences in yield and concentration were observed. While the automated extraction using the DNA Mini Kit yielded higher amounts of plasma DNA than the manual protocol, the CNA Kit (Vacuum protocol) outperformed the other two protocols. **(Figure 4)** Due to the substantial differences in concentrations, samples C74 and C75 were plotted separately. **(Figure 5)**

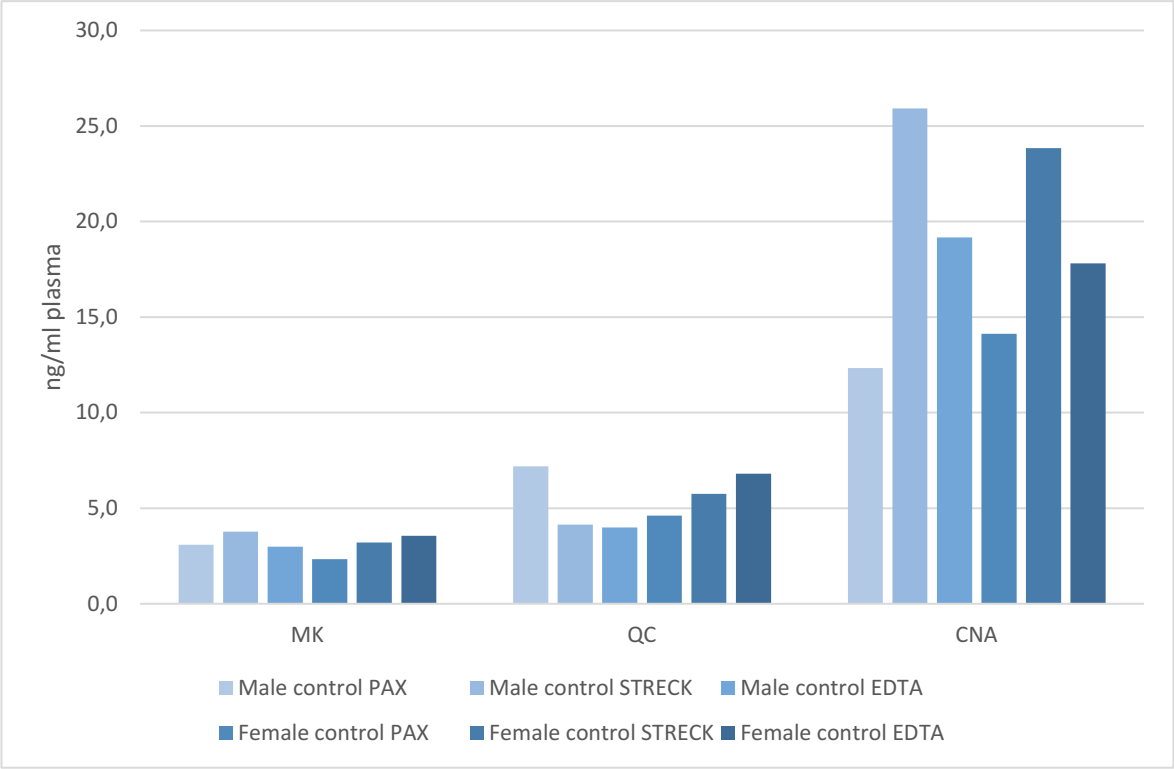

Figure 4. *cfDNA concentration differences between blood tubes (STRECK: Cell-free DNA BCT by Streck; PAX: PAXgene Blood DNA tube by QIAGEN; EDTA: Vacuette K3EDTA blood tube + 200µl 10% NBF) and different DNA purification methods (QC: automated on the QIAcube; MK: QIAamp DNA Mini Kit- Spin protocol; CNA: QIAamp Circulating Nucleic Acid Kit -Vacuum protocol) for healthy controls. CfDNA yield was highest with the CNA Kit.*

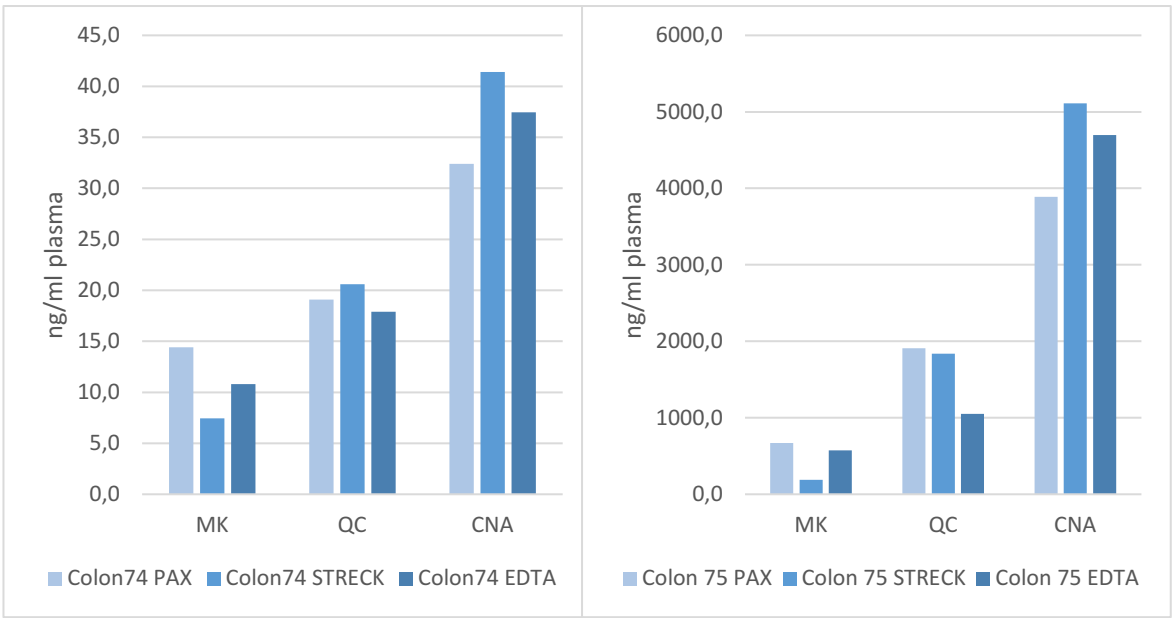

Figure 5*. Comparison of extraction methods (QC: automated on the QIAcube; MK: QIAamp DNA Mini Kit- Spin protocol; CNA: QIAamp Circulating Nucleic Acid Kit -Vacuum protocol) and blood tubes (STRECK: Cell-free DNA BCT by Streck; PAX: PAXgene Blood DNA tube by QIAGEN; EDTA: Vacuette K3EDTA blood tube + 200µl 10% NBF) for samples C74 and C75.*

In **Table 5** the mean plasma DNA concentrations and the respective fold-changes for the different extraction methods are displayed. There was a mean 2-fold increase in DNA yield of the Spin protocol compared to the automated purification on the QIAcube and a further up to 3.5-fold increase to the CNA Kit. **(Table 5)**

Table 5. *Differences in DNA yield using the different extraction methods*

Sample	DNA Mini Kit (Spin)	QiaCube	CNA Kit (Vakuum)	QC vs. Mini	CNA vs. QC	CNA vs. Mini
	ng/ml plasma	ng/ml plasma	ng/ml plasma	x	x	x
Healthy controls	3,2	5,4	18,9	1,7	3,5	6,0
Colon74	10,9	19,2	37,1	1,8	1,9	3,4
Colon 75	477,6	1599,0	4566,0	3,3	2,9	9,6

3.2 Evaluation of DNA integrity with the Agilent High Sensitivity DNA Kit

In order to assess the DNA integrity of the plasma DNA and to check for contamination with high molecular weight DNA, all samples were analyzed on an Agilent Bioanalyzer.

Consistent with the apoptotic origin of cfDNA, all samples showed an enrichment of DNA fragments around 160bp. Higher peaks in this range were observed with cancer patient samples. Healthy controls mainly showed a monophasic size distribution as depicted in **Figure 6**. Additional peaks around 250-400bp which are associated with dinucleosomes were mainly seen with cancer patients with an example shown in **Figure 7**. Interestingly, this biphasic size distribution was more distinct in patient sample C75 which also had a considerably higher plasma DNA concentration. For instance, the samples of C74 and C75 which were collected in a BCT by Streck and manually extracted with the Spin protocol both showed a biphasic size distribution. The maximum second peak of C74 was at 297bp with a concentration of 10.4pg/µl in this fragment size whereas the maximum second peak of the C75 sample being at 313bp had a concentration of 54pg/µl. **(Figure 8)(Figure 9)** Contamination with gDNA, observable in the form of a third peak with fragment lengths of >1000bp, was only seen in few samples and relatively low in extent. **(Figure 7)(Figure 8)**

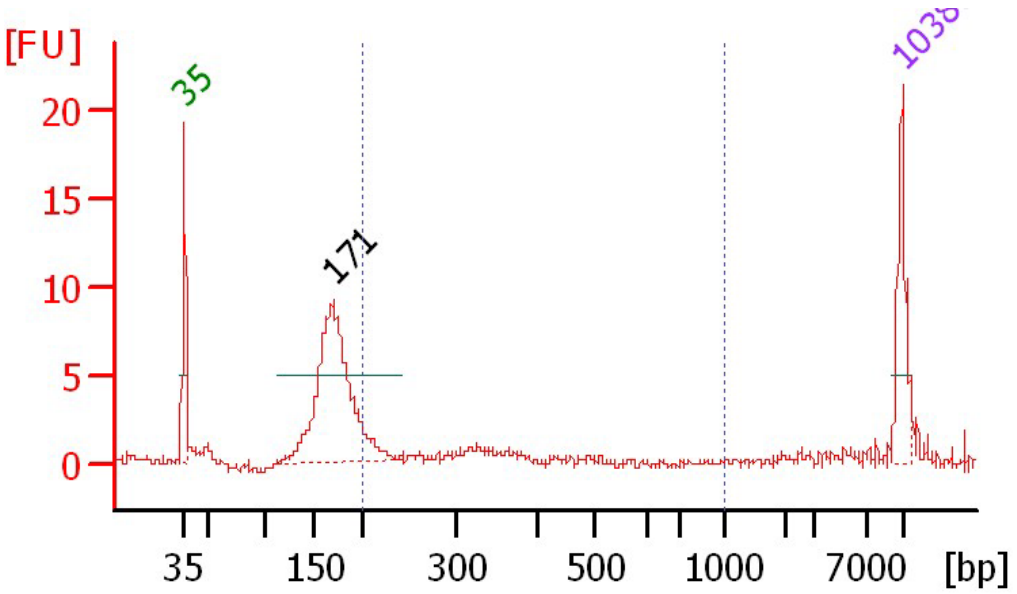

Figure 6. *DNA size distribution for the healthy male sample (PAXgene Blood DNA tube by QIAGEN). Plasma DNA was extracted with the Spin protocol. Specified DNA fragments are used as Lower and Upper Markers (left, 35bp and right, 10380bp). The peaks in-between those borders show the length of the cfDNA fragments.*

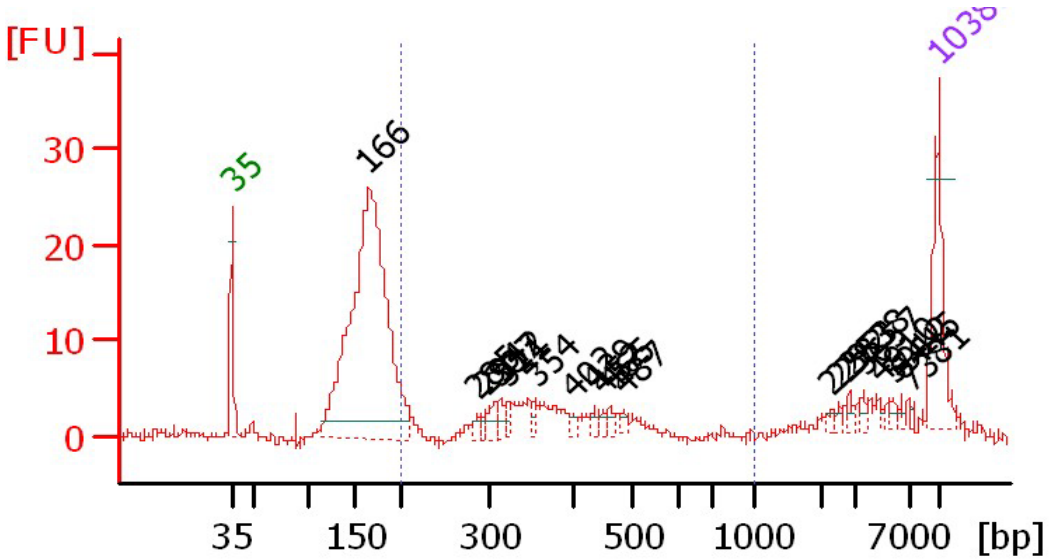

Figure 7. *DNA size distribution for the C74 sample (PAXgene Blood DNA tube by QIAGEN). Plasma DNA was extracted with the Spin protocol. Specified DNA fragments are used as Lower and Upper Markers (left, 35bp and right, 10380bp).*

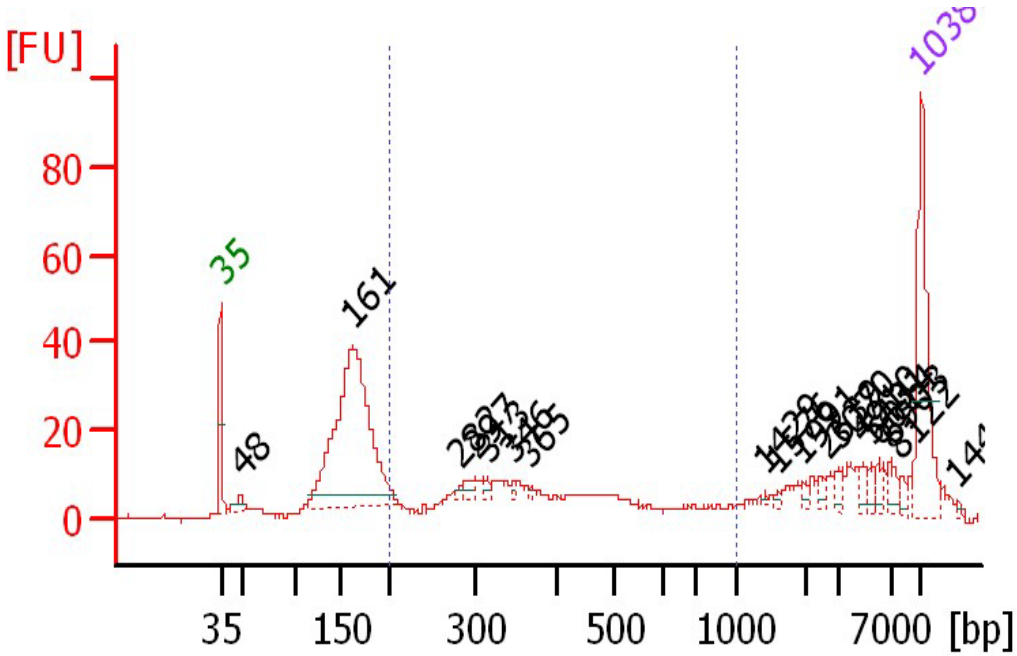

Figure 8. *DNA size distribution for the C74 sample (Cell-free DNA BCT by Streck). Plasma DNA was extracted with the Spin protocol. Specified DNA fragments are used as Lower and Upper Markers (left, 35bp and right, 10380bp).*

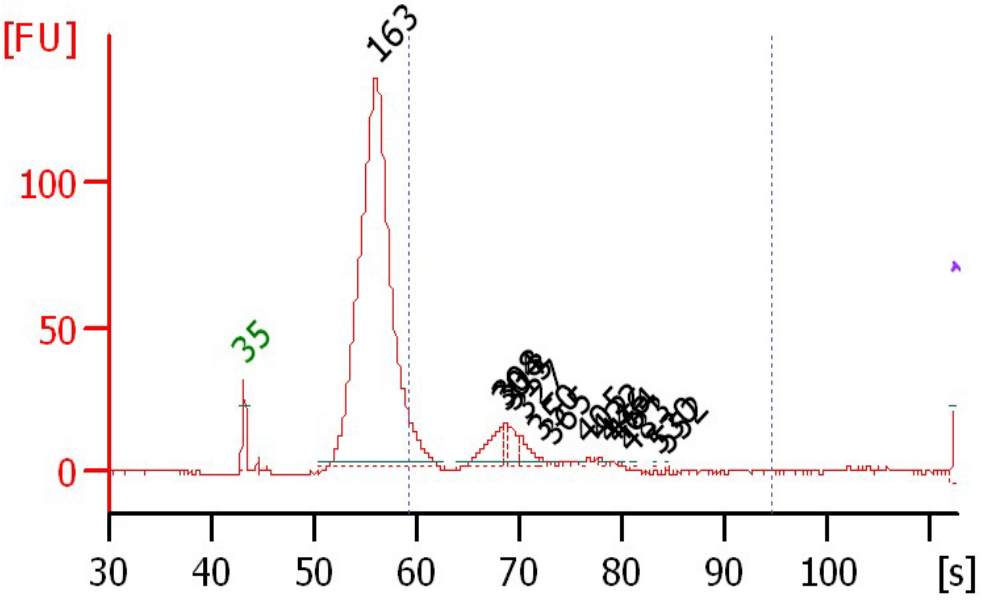

Figure 9. *DNA size distribution for the C75 sample (Cell-free DNA BCT by Streck). Plasma DNA was extracted with the Spin protocol. Specified DNA fragments are used as Lower and Upper Markers (left, 35bp and right, 10380bp).*

3.3 Evaluation of tumor content in samples C74 and C75

In order to estimate the overall tumor content in the metastasized CRC patients and to compare it between the different blood tubes and extraction methods, we used the mFAST-SeqS method. Using this method chromosome-arm specific and genomewide z-score can be calculated based on read count analysis of *LINE1* sequences. The genomewide z-score can be used as an overall measure of aneuploidy and is highly correlated to mutant allele frequencies of ctDNA.

A heat map was used to display chromosome arm specific z-scores in which blue bars are indicating loss of chromosomal material (z-scores below -5) and red bars show overrepresentations of chromosomal material (z-scores above 5). The overall z-score and cfDNA concentration for each sample are shown above each column. **(Figure 10)**

The results displayed in the heat map are highly consistent throughout the different blood tubes and extraction methods. As expected both the chromosome arm specific z-scores and the genome wide z-scores were much higher in patient C75 correlating with the higher amounts of ctDNA present in this patient sample.

Figure 10. *Heat map for samples C74 and C75 to show differences in tumor content using chromosome arm specific z-scores. Blue bars indicate z-scores below -5 and red bars show z-scores above 5. Additionally, genomwide z-scores and plasma DNA concentration (ng/ml plasma) are shown above the heat map. (STRECK: Cell-free DNA BCT by Streck; PAX: PAXgene Blood DNA tube by QIAGEN; EDTA: Vacuette K3EDTA blood tube + 200µl 10% NBF; QC: automated on the QIAcube; MK: QIAamp DNA Mini Kit- Spin protocol; CNA: QIAamp Circulating Nucleic Acid Kit -Vacuum protocol)*

3.4 Analysis of HCT116 and several dilutions with Promega control DNA

In order to determine a resolution limit for the GeneRead Targeted Panel V2 we sequenced several dilutions of HCT116 DNA (12%, 6%, 3%, 1.5%) in the background of a Promega control DNA. To assess the resolution limit of the kit we focused on genes with known HCT116 mutations that are listed in the COSMIC database and that were enriched with the GeneRead lung cancer panel.

Table 6. *HCT116 samples with a selection of heterozygous mutations and percentages of mutated fractions.*

Mutated gene	*ATM*	*CDKN2A*	*EPHA5*	*FGFR2*	*KMT2D*	*KRAS*	*Expected*
HCT116	47%	58%	55%	33%	48%	51%	50.00%
12% HCT116	4%	4%	5%	4%	4%	3%	6.00%
6% HCT116	1%	3%	2%	2%	3%	2%	3.00%
3% HCT116	1%	1%	1%	1%	2%	1%	1.50%
1.5% HCT116	2%	1%	1%	1%	0%	1%	0.75%

Mutated gene	*LRP1B*	*MET*	*NOTCH1*	*NOTCH1*	*PIK3CA*	*RARB*	*Expected*
HCT116	55%	48%	49%	49%	50%	50%	50.00%
12% HCT116	2%	4%	6%	5%	3%	4%	6.00%
6% HCT116	2%	3%	4%	0%	2%	2%	3.00%
3% HCT116	6%	0%	2%	0%	1%	1%	1.50%
1.5% HCT116	0%	2%	0%	0%	1%	1%	0.75%

A total of 12 mutations were identified in undiluted HCT116 DNA with MAF ranging from 33-58%. In the 12% and 6% dilutions, all mutations could be identified with MAF ranging from 2-6% which corresponds to the expected values. However, lower HCT116 dilutions did not yield the expected values. **(Table 6)(Table 7)**

In contrast, MAF did not show any differences between the 3% and 1.5% dilution indicating that a reliable detection of mutations below MAF of 5% was not possible with the GeneRead panel. **(Table 6)(Table 7)**

Table 7. *HCT116 samples with a selection of heterozygous mutations and percentages of mutated fractions.*

Samples		HCT116		12% HCT116		6% HCT116		3% HCT116		1.5% HCT116	
Genes GR	Mutation GenRead	% mutated	depth	% mutated	depth	% mutated	depth	% mutated	depth	% mutated	depth
ATM	c.C3380T:p.A1127V	47%	561	4%	567	1%	408	1%	339	2%	357
CDKN2A	c.97delG:p.E33fs	58%	895	4%	276	3%	379	1%	357	1%	653
EPHA5	c.2098delA:p.R700fs	55%	4156	5%	4312	2%	4482	1%	3184	1%	4009
FGFR2	c.C1409T:p.P470L	33%	4051	4%	2533	2%	2771	1%	1756	1%	2774
KMT2D	c.15G>Ap.K5K	48%	948	4%	854	3%	994	2%	571	0%	983
KRAS	c.G38A:p.G13D	51%	4215	3%	4739	2%	5631	1%	4308	1%	6118
LRP1B	c.C10622T:p.A3541V	55%	227	2%	248	2%	235	6%	138	0%	176
MET	c.710delT:p.V237fs	48%	334	4%	220	3%	160	0%	186	2%	196
NOTCH1	c.G3583A:p.G1195R	49%	556	6%	167	4%	295	2%	341	0%	404
NOTCH1	c.344delG:p.G115fs	49%	126	5%	117	0%	33	0%	37	0%	38
PIK3CA	c.A3140G:p.H1047R	50%	2231	3%	2329	2%	2381	1%	1970	1%	2367
RARB	c.C871T:p.H291Y	50%	3460	4%	4041	2%	3449	1%	2823	1%	3554

3.5 Lung cancer samples

To test the applicability of the GeneRead panel for plasma DNA samples, we used a total of six plasma DNA samples from lung cancer patients, which were available from an ongoing study. To this end samples with high and low mFAST-SeqS z-scores, i.e. high and low tumor content were selected. After enrichment and sequencing of these samples, the average coverage was calculated and variant prioritizing was done to identify relevant mutations. To attain a certain degree of confidence for particular base calls it is necessary to achieve a sufficient level of NGS coverage (read depth). Sequencing coverage refers to the number of reads that align to a reference base sequence. Higher levels of coverage mean that more sequence reads align to the bases covered and thereby the degree of confidence is increased. Various factors can influence the NGS coverage such as the size of the reference genome or the size of the target region. For reliable SNV (single nucleotide variants) calling a coverage for 30-50x is necessary to distinguish between sequencing errors and true heterozygous SNVs. However, for the detection of low frequency variants - as it is often the case in tumors or plasma DNA - read depths of up to 10.000 x are needed to detect variants down to 0.01%.

Table 8. *Average coverage of the lung cancer samples*

Sample	L15	L18	L27	L38_2	L48	L53
Average coverage	1530.0	440.0	834.4	437.4	240.3	395.0
% region > 100x	87.0	69.8	88.6	79.0	61.8	71.8

Sequencing of the lung cancer samples L15, L18, L27, L38, L48 and L53 showed average coverages as depicted in **Table 8**. The highest coverage was achieved in sample L15 with 1530x and the lowest coverage with 240.3x was seen in sample L48. The percentage of regions with coverages above 100x ranged from 61.8%-88.6%.

After variant prioritizing a total of 51 variants were identified for the samples L15, L18, L27, L38, L48 and L53. Due to repetitive mutations across the samples and the high number of detected variants we refrained from validating the variants with an independent method.

4 Discussion

Liquid biopsy is on the doorstep of implementation in a routine clinical setting. Standardization of preanalytical procedures is slowly becoming reality and more and more targeted assays for ctDNA detection are being commercially available and sensitive enough compared to tissue-based mutation testing.(16, 39, 46, 48, 52, 54) However, there is still room for optimization and evaluation of blood collection, extraction and analysis of ctDNA. In the present thesis three commonly used blood tubes for cfDNA collection and three different methods for plasma DNA extraction were evaluated. Finally, a NGS gene panel covering 45 lung cancer associated genes was evaluated for clinical use.

In the first part of the thesis cfDNA yield, size distribution and tumor content were compared between three different blood tubes (Vacuette K3EDTA blood tube with added neutral buffered formalin, PAXgene blood DNA tube by QIAGEN, Cell-Free DNA BCT by Streck) and three different methods of binding the cfDNA to the silica membrane based QIAamp Mini spin columns were used.

The extraction by centrifugation with the QIAamp DNA Mini Kit was either performed manually or fully automated on the QIAcube whereas the QIAamp CNA Kit used a vacuum manifold for extraction. Several studies showed the superiority of the CNA Kit by QIAGEN compared to other cfDNA extraction kits.(52-54, 56) By comparing the QIAamp DNA Kit with the QIAamp CNA Kit, we wanted to distinguish, if there are substantial differences between those kits provided by the same company which are operating on the same principle.

With respect to the plasma DNA extraction there was a 2-fold increase of cfDNA concentrations on average between the manual extraction and the QIAcube and a further up to 3.5-fold increase with the vacuum extraction. The QIAcube uses the same principle of binding the DNA to the membrane by centrifugation as in the manually performed protocol. However, the QIAcube offers a more standardized method due to the automation of the steps which might have resulted in the higher DNA yields observed. The reason for the much higher yield via the vacuum extraction might be due to omitting the aliquotation step to two (Spin protocol) or three (QIAcube) different columns, as used in the QIAamp

DNA Mini Kit. In the CNA Kit protocol the whole sample is processed through one column which probably results in the higher yield due to less DNA being caught in the membrane. Another reason could be the faster suction through the membrane by the deployed vacuum as opposed to centrifugation. Beside the advantage of more efficient plasma DNA extraction, the vacuum protocol allows processing up to 5ml of plasma sample compared to 1ml in the other protocols.

As expected, the estimated tumor content using the mFAST-SeqS method showed no significant differences between the extraction methods. Not surprisingly, patient C75 who had much higher cfDNA concentrations than patient C74 also had higher chromosome arm specific and genome wide z-scores. Concerning discrepancies between the different tubes, the Streck tubes showed the highest tumor fractions for the manual and the vacuum protocol in patient C75. However, this was not the case with patient C74 though.

With respect to DNA integrity no significant differences were observed for the different blood tubes. As previously described, blood samples should be processed within 4 hours following blood draw to maintain cfDNA concentration and fragmentation.(46) When blood samples are processed timely, there seems to be no significant contamination with gDNA from white blood cells although studies have shown that BCT tubes by Streck are superior in preserving cfDNA when stored or shipped at room temperature for extended times compared to PAXgene or EDTA tubes.(48, 50, 51)

In our study BCT, PAXgene and EDTA+NBF tubes performed equally well in preserving cfDNA concentrations and impeding contamination with gDNA by white blood cells. We only observed minimal contamination with high-molecular weight DNA in some samples with fragment lengths of >400bp but it is difficult to distinguish whether this is due to contamination with gDNA from white blood cells, derived from other sources like necrotic cells or simply caused by suboptimal handling or performance of the Agilent Bioanalyzer.

In the two plasma DNA samples from colorectal cancer patients (C74, C75) a biphasic DNA size distribution could be observed. In an early study, examining the origins of cfDNA in cancer patients, it was found that most DNA fragments were in the range of 180bp and sometimes multiples of 180bp, corresponding to mono- and dinucleosomal DNA and are probably mainly derived from apoptotic processes.(26) In a more recent study the biphasic

size distribution was previously described in a subset of colorectal carcinoma patients and correlated with higher plasma DNA concentrations.(29) This observation is in-line with the small sample size of cancer patients in the present thesis, where the biphasic distribution was present in both patients but more pronounced in C75 which also had a higher plasma DNA concentration. There might be several reasons for the occurrence of biphasic size distribution in cancer patients, such as extensive cell destruction, shorter degradation times of tumor DNA in the bloodstream or saturation of DNA degradation mechanisms.(29)

Summing up the results, no significant differences could be observed between the three different blood collection tubes used in this study concerning size distribution, tumor content and plasma DNA concentration. The same is true for the different extraction methods with the exception of DNA yield. We could show that the QIAamp CNA Kit outperformed the two other methods, which was not surprising, as this kit was specifically designed for cfDNA extraction in contrast to the QIAamp DNA Mini Kit of which the original use is the extraction of total (genomic, mitochondrial and viral) DNA.

In the second part of the present thesis we evaluated the potential clinical use of the QIAGEN GeneRead Lung Cancer Panel. To this end we first assessed the detection limit using cell line DNA. Furthermore, clinical plasma samples from six lung cancer patients were tested with this gene panel. The tested GeneRead Lung Cancer Panel uses multiplex PCR-based targeted enrichment with a specific primer design to amplify and enrich 45 lung cancer-associated genes. The gene panel was developed for the processing of FFPE tissue samples as source material and an input of 40ng DNA per sample for the 4-pool panel is advised. The dilution series of HCT116 (12%, 6%, 3%, 1.5%) showed a resolution limit about 5% MAF. Considering the wide variation of ctDNA fraction, in which can be as low as 0.01%, this resolution limit is not sufficient for sensitive profiling for mutations using plasma DNA.

After the establishment of the resolution limit, the sequencing of six lung cancer samples followed also using the GeneRead targeted panel. Despite the low input DNA which ranged from about 5ng-11ng instead of the recommended 40ng, it was possible to prepare libraries with the recommended protocol and most of the targeted regions were covered after sequencing. In total 51 mutations were detected in the six lung cancer samples after manual variant prioritization. Considering this high number of variants and the fact that the same mutations were frequently seen across different samples, it is highly probable that

most of the variants were in fact artifacts. It also has to be taken into account that the established resolution limit for this panel of about 5% MAF is not sensitive enough for the reliable detection of low frequency mutations. Especially in lung cancer somatic mutations are present at low prevalence in cfDNA due to e.g. contamination with gDNA of germline origin and this therefore increases false-positive results.(57)

The average coverage of sequenced lung cancer samples in this study ranged from 240-1530x and the percentage of regions covered >100x ranged from 62%-87%. In a recent study by Couraud et al. it was not possible to detect mutations in some tumor samples, even with a high coverage of >30,000x while in other samples with relatively low coverages of 400-800x, some mutations could be detected.(41) Thus, the authors hypothesized that variables in tumor biology probably are the main factors impacting mutation detection in cfDNA although high sensitivity is crucial for the detection of low MAF.(41)

A study from 2015 evaluated the concordance in mutation detection between cfDNA and genomic tumor DNA (tDNA) from biopsies, using next generation digital sequencing technology (DST) on a targeted panel with 54 genes frequently involved in cancerogenesis.(58) Despite achieving high sensitivity with the cfDNA DST method compared to tissue-based analysis for *KRAS* and *BRAF* mutations and an overall concordance rate of 85.9% between cf- and tDNA, the authors also considered tumor biology, e.g. if the tumor does not shed significant DNA amounts into the bloodstream, as limiting factor for mutation detection.(58)

In conclusion, clinical applications for ctDNA analysis that have robust prognostic and therapeutic value seem feasible and some of them are already clinically implemented, for example the testing for *EGFR* sensitizing or resistance mutations. However, for the analysis of larger gene panels we currently lack analytical sensitivity. Recently a number of kits including the molecular barcoding technology, which offer error suppression, became commercially available and might help to overcome this issue. Additionally, important parameters such as the exact origin, metabolization and size distribution of ctDNA remain insufficiently understood and offer a vast field for further research.

5 Literature

1. Malvezzi M, Bertuccio P, Rosso T, Rota M, Levi F, La Vecchia C, et al. European cancer mortality predictions for the year 2015: does lung cancer have the highest death rate in EU women? Annals of oncology : official journal of the European Society for Medical Oncology / ESMO. 2015;26(4):779-86.
2. Feinstein AR, Wells CK. A clinical-severity staging system for patients with lung cancer. Medicine. 1990;69(1):1-33.
3. Breathnach OS, Freidlin B, Conley B, Green MR, Johnson DH, Gandara DR, et al. Twenty-two years of phase III trials for patients with advanced non-small-cell lung cancer: sobering results. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2001;19(6):1734-42.
4. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA: a cancer journal for clinicians. 2015;65(1):5-29.
5. Sholl LM. Biomarkers in lung adenocarcinoma: a decade of progress. Archives of pathology & laboratory medicine. 2015;139(4):469-80.
6. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009;458(7239):719-24.
7. Robinson DR, Wu YM, Lin SF. The protein tyrosine kinase family of the human genome. Oncogene. 2000;19(49):5548-57.
8. Gerber DE, Gandhi L, Costa DB. Management and future directions in non-small cell lung cancer with known activating mutations. American Society of Clinical Oncology educational book / ASCO American Society of Clinical Oncology Meeting. 2014:e353-65.
9. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science (New York, NY). 2004;304(5676):1497-500.
10. Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. The New England journal of medicine. 2010;363(18):1693-703.
11. Bergethon K, Shaw AT, Ou SH, Katayama R, Lovly CM, McDonald NT, et al. ROS1 rearrangements define a unique molecular class of lung cancers. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2012;30(8):863-70.
12. Lindeman NI, Cagle PT, Beasley MB, Chitale DA, Dacic S, Giaccone G, et al. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. The Journal of molecular diagnostics : JMD. 2013;15(4):415-53.
13. Barlesi F, Mazieres J, Merlio JP, Debieuvre D, Mosser J, Lena H, et al. Routine molecular profiling of patients with advanced non-small-cell lung cancer: results of a 1-year nationwide programme of the French Cooperative Thoracic Intergroup (IFCT). Lancet (London, England). 2016;387(10026):1415-26.
14. Kris MG, Johnson BE, Berry LD, Kwiatkowski DJ, Iafrate AJ, Wistuba, II, et al. Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. Jama. 2014;311(19):1998-2006.
15. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511(7511):543-50.

16. Khoo C, Rogers TM, Fellowes A, Bell A, Fox S. Molecular methods for somatic mutation testing in lung adenocarcinoma: EGFR and beyond. Translational lung cancer research. 2015;4(2):126-41.
17. Sequist LV, Waltman BA, Dias-Santagata D, Digumarthy S, Turke AB, Fidias P, et al. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. Science translational medicine. 2011;3(75):75ra26.
18. Heitzer E, Ulz P, Geigl JB. Circulating tumor DNA as a liquid biopsy for cancer. Clinical chemistry. 2015;61(1):112-23.
19. Heitzer E, Auer M, Ulz P, Geigl JB, Speicher MR. Circulating tumor cells and DNA as liquid biopsies. Genome medicine. 2013;5(8):73.
20. Pantel K, Brakenhoff RH, Brandt B. Detection, clinical relevance and specific biological properties of disseminating tumour cells. Nature reviews Cancer. 2008;8(5):329-40.
21. Mandel P. Les acides nucleiques du plasma sanguin chez l'homme. CR Acad Sci Paris. 1948;142:241-3.
22. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the serum of cancer patients and the effect of therapy. Cancer research. 1977;37(3):646-50.
23. Sorenson GD, Pribish DM, Valone FH, Memoli VA, Bzik DJ, Yao SL. Soluble normal and mutated DNA sequences from single-copy genes in human blood. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 1994;3(1):67-71.
24. Lo YM, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, et al. Presence of fetal DNA in maternal plasma and serum. Lancet (London, England). 1997;350(9076):485-7.
25. Bianchi DW, Rava RP, Sehnert AJ. DNA sequencing versus standard prenatal aneuploidy screening. The New England journal of medicine. 2014;371(6):578.
26. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer research. 2001;61(4):1659-65.
27. Stroun M, Maurice P, Vasioukhin V, Lyautey J, Lederrey C, Lefort F, et al. The origin and mechanism of circulating DNA. Annals of the New York Academy of Sciences. 2000;906:161-8.
28. Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. Nature reviews Cancer. 2011;11(6):426-37.
29. Heitzer E, Auer M, Hoffmann EM, Pichler M, Gasch C, Ulz P, et al. Establishment of tumor-specific copy number alterations from plasma DNA of patients with cancer. International journal of cancer Journal international du cancer. 2013;133(2):346-56.
30. Lo YM, Zhang J, Leung TN, Lau TK, Chang AM, Hjelm NM. Rapid clearance of fetal DNA from maternal plasma. American journal of human genetics. 1999;64(1):218-24.
31. Yu SC, Lee SW, Jiang P, Leung TY, Chan KC, Chiu RW, et al. High-resolution profiling of fetal DNA clearance from maternal plasma by massively parallel sequencing. Clinical chemistry. 2013;59(8):1228-37.
32. Molina-Vila MA, Mayo-de-Las-Casas C, Gimenez-Capitan A, Jordana-Ariza N, Garzon M, Balada A, et al. Liquid Biopsy in Non-Small Cell Lung Cancer. Frontiers in medicine. 2016;3:69.

33. Gautschi O, Bigosch C, Huegli B, Jermann M, Marx A, Chasse E, et al. Circulating deoxyribonucleic Acid as prognostic marker in non-small-cell lung cancer patients undergoing chemotherapy. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2004;22(20):4157-64.
34. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Science translational medicine. 2014;6(224):224ra24.
35. Tamkovich SN, Cherepanova AV, Kolesnikova EV, Rykova EY, Pyshnyi DV, Vlassov VV, et al. Circulating DNA and DNase activity in human blood. Annals of the New York Academy of Sciences. 2006;1075:191-6.
36. Volik S, Alcaide M, Morin RD, Collins C. Cell-free DNA (cfDNA): Clinical Significance and Utility in Cancer Shaped By Emerging Technologies. Molecular cancer research : MCR. 2016;14(10):898-908.
37. Mok TS, Wu YL, Ahn MJ, Garassino MC, Kim HR, Ramalingam SS, et al. Osimertinib or Platinum-Pemetrexed in EGFR T790M-Positive Lung Cancer. The New England journal of medicine. 2017;376(7):629-40.
38. Ellison G, Zhu G, Moulis A, Dearden S, Speake G, McCormack R. EGFR mutation testing in lung cancer: a review of available methods and their use for analysis of tumour tissue and cytology samples. Journal of clinical pathology. 2013;66(2):79-89.
39. Thress KS, Brant R, Carr TH, Dearden S, Jenkins S, Brown H, et al. EGFR mutation detection in ctDNA from NSCLC patient plasma: A cross-platform comparison of leading technologies to support the clinical development of AZD9291. Lung cancer (Amsterdam, Netherlands). 2015;90(3):509-15.
40. Oxnard GR, Thress KS, Alden RS, Lawrance R, Paweletz CP, Cantarini M, et al. Association Between Plasma Genotyping and Outcomes of Treatment With Osimertinib (AZD9291) in Advanced Non-Small-Cell Lung Cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2016;34(28):3375-82.
41. Couraud S, Vaca-Paniagua F, Villar S, Oliver J, Schuster T, Blanche H, et al. Noninvasive diagnosis of actionable mutations by deep sequencing of circulating free DNA in lung cancer from never-smokers: a proof-of-concept study from BioCAST/IFCT-1002. Clinical cancer research : an official journal of the American Association for Cancer Research. 2014;20(17):4613-24.
42. Paweletz CP, Sacher AG, Raymond CK, Alden RS, O'Connell A, Mach SL, et al. Bias-Corrected Targeted Next-Generation Sequencing for Rapid, Multiplexed Detection of Actionable Alterations in Cell-Free DNA from Advanced Lung Cancer Patients. Clinical cancer research : an official journal of the American Association for Cancer Research. 2016;22(4):915-22.
43. Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nature medicine. 2014;20(5):548-54.
44. Esposito A, Bardelli A, Criscitiello C, Colombo N, Gelao L, Fumagalli L, et al. Monitoring tumor-derived cell-free DNA in patients with solid tumors: clinical perspectives and research opportunities. Cancer treatment reviews. 2014;40(5):648-55.
45. Thierry AR, Mouliere F, Gongora C, Ollier J, Robert B, Ychou M, et al. Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. Nucleic acids research. 2010;38(18):6159-75.

46. El Messaoudi S, Rolet F, Mouliere F, Thierry AR. Circulating cell free DNA: Preanalytical considerations. Clinica chimica acta; international journal of clinical chemistry. 2013;424:222-30.
47. Tatsumi N, Miwa S, Lewis SM. Specimen collection, storage, and transmission to the laboratory for hematological tests. International journal of hematology. 2002;75(3):261-8.
48. Norton SE, Lechner JM, Williams T, Fernando MR. A stabilizing reagent prevents cell-free DNA contamination by cellular DNA in plasma during blood sample storage and shipping as determined by digital PCR. Clinical biochemistry. 2013;46(15):1561-5.
49. 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. Journal of clinical laboratory analysis. 2013;27(4):305-11.
50. Toro PV, Erlanger B, Beaver JA, Cochran RL, VanDenBerg DA, Yakim E, et al. Comparison of cell stabilizing blood collection tubes for circulating plasma tumor DNA. Clinical biochemistry. 2015;48(15):993-8.
51. Kang Q, Henry NL, Paoletti C, Jiang H, Vats P, Chinnaiyan AM, et al. Comparative analysis of circulating tumor DNA stability In K3EDTA, Streck, and CellSave blood collection tubes. Clinical biochemistry. 2016;49(18):1354-60.
52. Devonshire AS, Whale AS, Gutteridge A, Jones G, Cowen S, Foy CA, et al. Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification. Analytical and bioanalytical chemistry. 2014;406(26):6499-512.
53. Mauger F, Dulary C, Daviaud C, Deleuze JF, Tost J. Comprehensive evaluation of methods to isolate, quantify, and characterize circulating cell-free DNA from small volumes of plasma. Analytical and bioanalytical chemistry. 2015;407(22):6873-8.
54. Page K, Guttery DS, Zahra N, Primrose L, Elshaw SR, Pringle JH, et al. Influence of plasma processing on recovery and analysis of circulating nucleic acids. PloS one. 2013;8(10):e77963.
55. Belic J, Koch M, Ulz P, Auer M, Gerhalter T, Mohan S, et al. Rapid Identification of Plasma DNA Samples with Increased ctDNA Levels by a Modified FAST-SeqS Approach. Clinical chemistry. 2015;61(6):838-49.
56. Perez-Barrios C, Nieto-Alcolado I, Torrente M, Jimenez-Sanchez C, Calvo V, Gutierrez-Sanz L, et al. Comparison of methods for circulating cell-free DNA isolation using blood from cancer patients: impact on biomarker testing. Translational lung cancer research. 2016;5(6):665-72.
57. Oxnard GR, Paweletz CP, Kuang Y, Mach SL, O'Connell A, Messineo MM, et al. Noninvasive detection of response and resistance in EGFR-mutant lung cancer using quantitative next-generation genotyping of cell-free plasma DNA. Clinical cancer research : an official journal of the American Association for Cancer Research. 2014;20(6):1698-705.
58. Kim ST, Lee WS, Lanman RB, Mortimer S, Zill OA, Kim KM, et al. Prospective blinded study of somatic mutation detection in cell-free DNA utilizing a targeted 54-gene next generation sequencing panel in metastatic solid tumor patients. Oncotarget. 2015;6(37):40360-9.

6 Appendix

Gene	Description[1]	Drugs[1]
AKT1	AKT1 is one of 3 closely related serine/threonine-protein kinases (AKT1, AKT2 and AKT3) called the AKT kinase, which regulate many processes including metabolism, proliferation, cell survival, growth and angiogenesis. This is mediated through serine and/or threonine phosphorylation of a range of downstream substrates.	**Carboplatin , Cisplatin** (DNA synthesis inhibitors), **Everolimus** (mTOR inhibitor)
ALK	This gene encodes a receptor tyrosine kinase, which belongs to the insulin receptor superfamily. This gene has been found to be rearranged, mutated, or amplified in a series of tumours including anaplastic large cell lymphomas, neuroblastoma, and non-small cell lung cancer.	**Crizotinib** (C-MET/ALK inhibitor), **Ceritinib** , **Bevacizumab** (VEGF antagonist)
APC	This gene encodes a tumor suppressor protein that acts as an antagonist of the Wnt signaling pathway. It is also involved in other processes including cell migration and adhesion, transcriptional activation, and apoptosis. Defects in this gene cause familial adenomatous polyposis (FAP), an autosomal dominant pre-malignant disease that usually progresses to malignancy.	
ATM	The protein encoded by this gene belongs to the PI3/PI4-kinase family. This protein is an important cell cycle checkpoint kinase that phosphorylates; thus, it functions as a regulator of a wide variety of downstream proteins, including tumor suppressor proteins p53 and BRCA1, checkpoint kinase CHK2, checkpoint proteins RAD17 and RAD9, and DNA repair protein NBS1. This protein and the closely related kinase ATR are thought to be master controllers of cell cycle checkpoint signaling pathways that are required for cell response to DNA damage and for genome stability.	**Olaparib** (a PARP inhibitor, inhibiting poly ADP ribose polymerase), **Carboplatin**
BAI3	This p53-target gene encodes a brain-specific angiogenesis inhibitor, a seven-span transmembrane protein, and is thought to be a member of the secretin receptor family. Brain-specific angiogenesis proteins BAI2 and BAI3 are similar to BAI1 in structure, have similar tissue specificities, and may also play a role in angiogenesis.	
BAP1	The encoded enzyme by this gene binds to the breast cancer type 1 susceptibility protein (BRCA1) via the RING finger domain of the latter and acts as a tumor suppressor. In addition, the enzyme may be involved in regulation of transcription, regulation of cell cycle and growth, response to DNA damage and chromatin dynamics. Germline mutations in this gene may be associated with tumor predisposition syndrome (TPDS), which involves increased risk of cancers including malignant mesothelioma, uveal melanoma and cutaneous melanoma.	
BRAF	This gene encodes a protein that plays a role in regulating the MAP kinase/ERKs signaling pathway, which affects cell division, differentiation, and secretion. Mutations in this gene have also been associated with various cancers, including non-Hodgkin lymphoma, colorectal cancer, malignant melanoma, thyroid carcinoma, non-small cell lung carcinoma, and adenocarcinoma of the lung.	**Sorafenib** (Raf kinases and tyrosine kinases inhibitor), **Dabrafenib** (BRAF kinase inhibitor)
CDKN2A	This gene is frequently mutated or deleted in a wide variety of tumors, and is known to be an important tumor suppressor gene.Capable of inducing cell cycle arrest in G1 and G2 phases. Acts as a tumor suppressor. Binds to MDM2 and blocks its nucleocytoplasmic shuttling by sequestering it in the nucleolus. This inhibits the oncogenic action of MDM2 by blocking MDM2-induced degradation of p53 and enhancing p53-dependent transactivation and apoptosis.	
EGFR	The protein encoded by this gene is a transmembrane glycoprotein that is a member of the protein kinase superfamily. EGFR is a cell surface protein that acts as an receptor for members of the epidermal growth factor family. Binding of the protein to a ligand induces receptor dimerization and tyrosine autophosphorylation and leads to cell proliferation. Mutations in this gene are associated with lung cancer.	**Gefitinb, Erlotinib** and **Cetuximab** (EGFR inhibitors)
EPHA5	This gene belongs to the ephrin receptor subfamily of the protein-tyrosine kinase family. EPH and EPH-related receptors have been implicated in mediating developmental events, particularly in the nervous system. Diseases associated with this gene are acrodermatitis chronica athrophicans and large cell lung carcinoma.	**Paclitaxel** (Tubulin and Bcl2 inhibitor)
ERBB2	This gene encodes a member of the epidermal growth factor (EGF) receptor family of receptor tyrosine kinases. This gene has been reported in numerous cancers, including breast and ovarian tumors.	**Lapatinib, Pertuzumab, Trastuzumab, Gefitinib** (EGFR/HER2 Inhibitors)
ERBB4	This gene is a member of the Tyr protein kinase family and the epidermal growth factor receptor subfamily.The protein binds to and is activated by neuregulins and other factors and induces a variety of cellular responses including mitogenesis and differentiation.Mutations in this gene have been associated with cancer.	**Gefitinib, Afatinib** (EGFR Inhibitors)
FBXW7	Mutations in this gene are detected in ovarian and breast cancer cell lines, implicating the gene's potential role in the pathogenesis of human cancers.	
FGFR1	The protein encoded by this gene is a member of the fibroblast growth factor receptor (FGFR) family Tyrosine-protein kinase that acts as cell-surface receptor for fibroblast growth factors and plays an essential role in the regulation of embryonic development, cell proliferation, differentiation and migration. Chromosomal aberrations involving this gene are associated with stem cell myeloproliferative disorder and stem cell leukemia lymphoma syndrome.	**Lenvatinib, Nintedanib** (VEGFR inhibitors)

FGFR2	The protein encoded by this gene is a member of the fibroblast growth factor receptor family. FGFR family members differ from one another in their ligand affinities and tissue distribution. An important paralog of this gene is FGFR1.	**Palifermin** (keratinocyte growth factor (KGF) receptor binder), **Lenvatinib, Nintedanib** (VEGFR inhibitors)
GRM8	G-protein coupled receptor for glutamate. Glutamatergic neurotransmission is involved in most aspects of normal brain function and can be perturbed in many neuropathologic conditions. Diseases associated with GRM8 include autism spectrum disorder and schizophrenia.	
KDR	Vascular endothelial growth factor (VEGF) is a major growth factor for endothelial cells. This gene encodes one of the two receptors of the VEGF. This receptor, known as kinase insert domain receptor, is a type III receptor tyrosine kinase. It functions as the main mediator of VEGF-induced endothelial proliferation, survival, migration, tubular morphogenesis and sprouting.	**Sunitinib, Sorafenib** (multi-targeted receptor tyrosine kinase (RTK) inhibitor)
KEAP1	KEAP1 (Kelch Like ECH Associated Protein 1) is a Protein Coding gene. Diseases associated with KEAP1 include Lung Papillary Adenocarcinoma.	**Dimethyl fumarate** (Transcription factor binding)
KIT	This gene encodes a Tyrosine-protein kinase that acts as cell-surface receptor for the cytokine KITLG/SCF and plays an essential role in the regulation of cell survival and proliferation, hematopoiesis, stem cell maintenance, gametogenesis, mast cell development, migration and function, and in melanogenesis. Mutations in this gene are associated with gastrointestinal stromal tumors, mast cell disease, acute myelogenous lukemia, and piebaldism.	**Sunitinib** (VEGFR/PDGFRβ/ KIT/ FLT3/RET/CSF-1R inhibitor), **Dasatinib** (SRC/BCR-ABL tyrosine kinase inhibitor)
KMT2D	The encoded protein is part of a large protein complex called ASCOM, which has been shown to be a transcriptional regulator of the beta-globin and estrogen receptor genes.	
KRAS	This gene, a Kirsten ras oncogene homolog from the mammalian ras gene family, encodes a protein that is a member of the small GTPase superfamily. A single amino acid substitution is responsible for an activating mutation. The transforming protein that results is implicated in various malignancies, including lung adenocarcinoma, mucinous adenoma, ductal carcinoma of the pancreas and colorectal carcinoma.	**Cetuximab, Panitumumab** (EGFR inhibitors), **Everolimus** (mTOR inhibitor)
LRP1B	LRP1B belongs to the low density lipoprotein (LDL) receptor gene family. These receptors play a wide variety of roles in normal cell function and development due to their interactions with multiple ligands.	
MDM2	The encoded protein by this gene can promote tumor formation by targeting tumor suppressor proteins, such as p53, for proteasomal degradation. This gene is itself transcriptionally-regulated by p53. Overexpression or amplification of this locus is detected in a variety of different cancers.	
MET	The proto-oncogene MET product is the hepatocyte growth factor receptor and encodes tyrosine-kinase activity.Various mutations in the MET gene are associated with papillary renal carcinoma.	**Crizotinib** (C-MET/ALK inhibitor), **Cetuximab** (EGFR Inhibitor)
MLH1	This gene was identified as a locus frequently mutated in hereditary nonpolyposis colon cancer (HNPCC). It is a human homolog of the E. coli DNA mismatch repair gene mutL, consistent with the characteristic alterations in microsatellite sequences (RER+phenotype) found in HNPCC.	
MUC16	This gene encodes a protein that is thought to provide a protective, lubricating barrier against particles and infectious agents at mucosal surfaces. Diseases associated with MUC16 include e.g. ovarian cancer.	**Oregovomab** (Monoclonal Antibody against CA125)
MYC	The protein encoded by this gene is a multifunctional, nuclear phosphoprotein that plays a role in cell cycle progression, apoptosis and cellular transformation. It functions as a transcription factor that regulates transcription of specific target genes. Mutations, overexpression, rearrangement and translocation of this gene have been associated with a variety of hematopoietic tumors, leukemias and lymphomas, including Burkitt lymphoma.	
NF1	This gene product appears to function as a negative regulator of the ras signal transduction pathway. Mutations in this gene have been linked to neurofibromatosis type 1, juvenile myelomonocytic leukemia and Watson syndrome.	
NFE2L2	The encoded transcription factor by this gene regulates genes which contain antioxidant response elements (ARE) in their promoters; many of these genes encode proteins involved in response to injury and inflammation which includes the production of free radicals.	**Dimethyl fumarate** (Transcirption factor binding; up-regulates a pathway that is activated in response to oxidative stress)
NOTCH1	Members of this Type I transmembrane protein family share structural characteristics including an extracellular domain consisting of multiple epidermal growth factor-like (EGF) repeats, and an intracellular domain consisting of multiple different domain types.	

PDGFRA	This gene encodes a cell surface tyrosine kinase receptor for members of the platelet-derived growth factor family. These growth factors are mitogens for cells of mesenchymal origin. Studies suggest that this gene plays a role in organ development, wound healing, and tumor progression.	**Sunitinib** (VEGFR, PDGFRbeta and KIT inhibitor), **Pazopanib** (VEFGR, c-Kit, and PDGFR inhibitor)
PIK3CA	This gene has been found to be oncogenic and has been implicated in cervical cancers.	**Everolimus** (mTOR inhibitor)
PIK3CG	Phosphoinositide 3-kinases (PI3Ks) phosphorylate inositol lipids and are involved in the immune response.The protein encoded by this gene is a class I catalytic subunit of PI3K.This gene is located in a commonly deleted segment of chromosome 7 previously identified in myeloid leukemias.	
PKHD1	Mutations in this gene cause autosomal recessive polycystic kidney disease, also known as polycystic kidney and hepatic disease-1.	
PTEN	This gene was identified as a tumor suppressor that is mutated in a large number of cancers at high frequency. Acts as a dual-specificity protein phosphatase, dephosphorylating tyrosine-, serine- and threonine-phosphorylated proteins. Antagonizes the PI3K-AKT/PKB signaling pathway by dephosphorylating phosphoinositides and thereby modulating cell cycle progression and cell survival.	**Cetuximab, Erlotinib** (EGFR Inhibitor), **Everolimus** (mTOR inhibitor)
RARB	This gene encodes the retinoic acid receptor beta, a member of the thyroid-steroid hormone receptor superfamily of nuclear transcriptional regulators. It binds retinoic acid, the biologically active form of vitamin A which mediates cellular signalling in embryonic morphogenesis, cell growth and differentiation. It is thought that this protein limits growth of many cell types by regulating gene expression. The gene was first identified in a hepatocellular carcinoma where it flanks a hepatitis B virus integration site.	
RB1	The protein encoded by this gene is a negative regulator of the cell cycle and was the first tumor suppressor gene found. The encoded protein also stabilizes constitutive heterochromatin to maintain the overall chromatin structure. Defects in this gene are a cause of childhood cancer retinoblastoma (RB), bladder cancer, and osteogenic sarcoma.	
RET	This gene encodes one of the receptor tyrosine kinases, which are cell-surface molecules that transduce signals for cell growth and differentiation. It plays a crucial role in neural crest development and it can undergo oncogenic activation in vivo and in vitro by cytogenetic rearrangement. Mutations in this gene are associated with the disorders multiple endocrine neoplasia (types IIA, IIB), Hirschsprung disease and medullary thyroid carcinoma.	**Regorafenib** (VEGFR/PDGFR/FGFR/mutant kit/RET/Raf-1 inhibitor)
ROS1	This proto-oncogene, highly-expressed in a variety of tumor cell lines, belongs to the subfamily of tyrosine kinase insulin receptor genes. May activate several downstream signaling pathways related to cell differentiation, proliferation, growth and survival including the PI3 kinase-mTOR signaling pathway.	**Crizotinib** (ALK (anaplastic lymphoma kinase) and ROS1 (c-ros oncogene 1) inhibitor), **Ceritinib** (2nd generation ALK inhibitor used after tumours develop drug resistance to Crizotinib)
RUNX1T1	This gene encodes a member of the myeloid translocation gene family which interact with DNA-bound transcription factors and recruit a range of corepressors to facilitate transcriptional repression. The t(8;21)(q22;q22) translocation is one of the most frequent karyotypic abnormalities in acute myeloid leukemia.	
SMAD4	Mutations or deletions in this gene have been shown to result in pancreatic cancer, juvenile polyposis syndrome, and hereditary hemorrhagic telangiectasia syndrome.	**Capecitabine, Gemcitabine** (DNA synthesis inhibitors)
SMARCA4	The protein encoded by this gene is a member of the SWI/SNF family of proteins which are thought to regulate transcription of certain genes by altering the chromatin structure around those genes.	
SOX2	This intronless gene encodes a member of the SRY-related HMG-box (SOX) family of transcription factors involved in the regulation of embryonic development and in the determination of cell fate.	
STK11	This gene, which encodes a member of the serine/threonine kinase family, regulates cell polarity and functions as a tumor suppressor.	
TP53	This gene encodes a tumor suppressor protein containing transcriptional activation, DNA binding, and oligomerization domains. The encoded protein responds to diverse cellular stresses to regulate expression of target genes, thereby inducing cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism. Mutations in this gene are associated with a variety of human cancers, including hereditary cancers such as Li-Fraumeni syndrome.	**Cisplatin, Fluorouracil** (DNA synthesis inhibitors), **Docetaxel** (Microtubulin disassembly inhibitor)

[1] Modified from http://www.genecards.org/