Doctoral theses at NTNU, 2023:327

Anine Larsen Ottestad

Challenges in implementing circulating tumor DNA analyses in lung cancer research

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Medicine and Health Sciences
Department of Clinical and Molecular Medicine

NTNU
Norwegian University of
Science and Technology

Anine Larsen Ottestad

Challenges in implementing circulating tumor DNA analyses in lung cancer research

Thesis for the Degree of Philosophiae Doctor

Trondheim, October 2023

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Clinical and Molecular Medicine

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Medicine and Health Sciences
Department of Clinical and Molecular Medicine

© Anine Larsen Ottestad

ISBN 978-82-326-7354-4 (printed ver.)
ISBN 978-82-326-7353-7 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2023:327

Printed by NTNU Grafisk senter

Name of candidate: Anine Larsen Ottestad

Department: Department of Clinical and Molecular Medicine

Faculty: Faculty of Medicine and Health Sciences

Main supervisor: Bjørn Henning Grønberg

Co-supervisor: Hong Yan Dai

Funding: The Liaison Committee for Education, Research, and Innovation in Central Norway

This thesis is found to be worthy of public defense for the degree of Philosophiae Doctor in Medicine and Health Sciences at the Norwegian University of Science and Technology

The public defense takes place in auditorium MTA in the Fred Kavli building on October 13th, 2023

Utfordringer ved å innføre sirkulerende tumor DNA-analyser i lungekreftforskning

Lungekreft er den ledende årsaken til kreftrelaterte dødsfall over hele verden, inkludert i Norge. Dette prosjektet handlet om ikke-småcellet lungekreft (NSCLC) som er den vanligste typen lungekreft. Til tross for at det har vært store fremskritt innen behandling er prognosen for NSCLC fortsatt svært varierende. Noen pasienter opplever god effekt av behandling, andre en kortvarig eller dårlig effekt, og mange opplever at kreftsykdommen kommer tilbake. Det er behov for bedre verktøy for å klassifisere pasienter med NSCLC.

Analyse av sirkulerende tumor DNA (ctDNA) som finnes i blodet til kreftpasienter kan muligens være et slik verktøy og i dette prosjektet undersøkte vi ulike aspekter ctDNA som en biomarkør. Det er nemlig flere utfordringer med ctDNA-analyser på grunn av de lave nivåene i blodet, særlig hos pasienter med tidlig stadium kreft. Det var derfor et sentralt mål med prosjektet å utvikle en tilnærming for ctDNA-analyse som var sensitiv, spesifikk og samtidig gjennomførbar i rutinediagnostikk.

Vi testet en "tumor-informert" tilnærming til ctDNA-analyse, som innebar analyse av pasientens tumor-DNA og deretter tilpassing av ctDNA-analysen til den enkelte pasient. Ved å legge til informasjon om kvaliteten på tumor DNA, ble data enklere å tolke og mer pålitelig. Resultatene viste videre at påvisning av ctDNA ved bruk av denne metoden var en negativ prognostisk faktor i vår pasientkohort. Med andre ord, de som hadde ctDNA i blodet levde kortere og kreften kom raskere tilbake. Vi så likevel at fordelen med slike analyser var størst hos pasienter der vi klarte å påvise ctDNA, og de tilhørte mindretallet.

Videre undersøkte vi sammenhengen mellom ctDNA-mengde i blodet og den metabolske aktiviteten i tumor, altså hvor mye sukker som tas opp i tumor for å lage energi, enkelt forklart. Resultatene tydet på at det kan være en sammenheng mellom disse to karakteristikkene, men det kreves videre forskning for å bekrefte disse resultatene og forklare de underliggende mekanismene.

Resultatene fra dette prosjektet bidrar i en større innsatts for å forbedre persontilpassede strategier for behandling av NSCLC.

Abstract

Lung cancer is the leading cause of cancer-related deaths worldwide, including Norway. This study focused on non-small cell lung cancer (NSCLC), the most common subtype, and aimed to enhance the classification of NSCLC patients using circulating tumor DNA (ctDNA) as a biomarker. Despite advancements in treatment, the prognosis for NSCLC remains highly variable, with some patients experiencing positive responses, while others face relapse or no response at all.

The thesis investigated the potential of ctDNA, which carries tumor-specific mutations, to improve prognostic accuracy and promote more individualized and thereby improved cancer management. The theoretical utility is high, but ctDNA analyses pose challenges due to its low levels in the blood, particularly in early-stage patients. Therefore, a key objective of this project was to develop an approach for ctDNA analysis that could be implemented in routine diagnostic laboratories.

To address these challenges, a 'tumor-informed' ctDNA analysis approach was tested and validated. This involved analyzing the patient's tumor DNA and customizing the analysis to individual patients. By incorporating quality assessment of the tumor DNA, the analysis could be tailored to improve data quality and reliability. The results demonstrated that ctDNA detection using this customized approach served as a negative prognostic factor in our NSCLC patient cohort. However, it was found that the benefit of ctDNA detection was limited to a minority of patients who had detectable ctDNA.

Furthermore, the study explored the relationship between ctDNA levels and tumor metabolic activity. Preliminary findings indicated a correlation between these two characteristics, suggesting the potential involvement of tumor metabolism in ctDNA release. However, further research is necessary to confirm these results and elucidate the underlying mechanisms.

The findings of this project contribute to the ongoing efforts to improve personalized treatment strategies and outcomes for NSCLC patients.

Acknowledgements

This work has been carried out at the Department of Clinical and Molecular Medicine at the Norwegian University of Science and Technology (NTNU). The PhD project, including my position, was funded by The Liaison Committee for Education, Research, and Innovation in Central Norway. Additionally, I received funding from the Cancer Foundation at St. Olavs Hospital, Trondheim University Hospital.

I would like to thank my main supervisor Bjørn Henning Grønberg for initiating this project, for everything you have taught me and the support along the way. Likewise, I would like to thank my co-supervisor Hong Yan Dai for sharing your knowledge and challenging me during this project. I am grateful for the advice and feedback from the both of you, the perspectives you have shown me, and I appreciate the chance to work in an interdisciplinary environment.

Thanks to Sissel and Elisabeth for your invaluable contributions, lively discussions, and social companionship in the office, and to Tarje for valuable feedback and support in statistics even though I insisted on learning R. My grateful thanks to all the other co-authors for your contributions.

I am grateful to the patients who participate in the biobank and for everyone involved in the practical organization of our research activity. Thanks to the Pathology department at St. Olavs Hospital and the Genome Core Facility (GCF) at NTNU for letting me use the laboratory facilities and to GCF providing excellent DNA sequencing services. I am also grateful to Oddgeir and his colleagues at HUNT Cloud for providing invaluable services in data storage and computing, and for prompt and positive solutions to every issue.

Lastly, my deepest gratitude to Sindre for continuous encouragement and support and to Kristian for making life more fun.

Anine Larsen Ottestad

Trondheim, June 2023

List of papers

Paper I

Ottestad AL, Wahl SGF, Grønberg BH, Skorpen F, Dai HY. **The relevance of tumor mutation profiling in interpretation of NGS data from cell-free DNA in non-small cell lung cancer patients.** Exp Mol Pathol. 2019;:104347. doi: 10.1016/j.yexmp.2019.104347

Paper II

Ottestad AL, Emdal EF, Grønberg BH, Halvorsen TO, Dai HY. **Fragmentation assessment of FFPE DNA helps in evaluating NGS library complexity and interpretation of NGS results.** Exp Mol Pathol. 2022;126:104771. doi: 10.1016/j.yexmp.2022.104771

Paper III

Ottestad AL, Dai HY, Halvorsen TO, Emdal EF, Wahl SGF, Grønberg BH. **Associations between tumor mutations in cfDNA and survival in non-small cell lung cancer.** Cancer Treat Res Commun. 2021;29:100471. doi: 10.1016/j.ctarc.2021.100471

Paper IV

Ottestad AL, Johansen H, Halvorsen TO, Dai HY, Wahl SGF, Emdal EF, Grønberg BH. **Associations between detectable circulating tumor DNA and tumor glucose uptake measured by ^{18}F-FDG PET/CT in early-stage non-small cell lung cancer.** BMC Cancer – under review.

Abbreviations

^{18}F-FDG	^{18}F-fluorodeoxyglucose
ADC	Adenocarcinoma
AKT	AKT Serine/Threonine Kinase
ALK	Anaplastic lymphoma kinase
BRAF	Neurotrophic Receptor Tyrosine Kinase
CDKN2A	Tumor Protein P53
cfDNA	Circulating cell-free DNA
CHIP	Clonal hematopoiesis of indeterminant potential
CT	Computed Tomography
CTC	Circulating tumor cells
ctDNA	Circulating tumor DNA
ddPCR	digital droplet polymerase chain reaction
EANM	The European Association of Nuclear Medicine
EGFR	*Epidermal growth factor receptor*
ES	Extensive stage
FFPE	Formalin-fixed paraffin-embedded
hGE	Human genome equivalent
IHC	Immunihistochemistry
KRAS	Kirsten Rat Sarcoma Viral Oncogene Homolog
LOD	Limit of detection
LS	Limited stage
MAF	Mutant allele frequency
MTV	Metabolic tumor volume
NGS	Next-generation sequencing
NSCLC	Non-small cell lung cancer
NTRK	Neurotrophic Receptor Tyrosine Kinase
OS	Overall survival
PCR	Polymerase chain reaction
PD-1	Programmed cell death protein 1
PD-L1	Programmed death-ligand 1
PET	Positron emission tomography
PFS	Progression-free survival

PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
PS	Eastern Cooperative Oncology Group performance status
PTEN	Phosphatase And Tensin Homolog
qPCR	quantitative polymerase chain reaction
RB1	RB Transcriptional Corepressor 1
RET	Ret Proto-Oncogene
ROS1	ROS Proto-Oncogene 1, Receptor Tyrosine Kinase
SCC	Squamous cell carcinoma
SCLC	Small-cell lung cancer
STK11	Serine/Threonine Kinase 11
SUV	Standardized uptake value
TKI	Tyrosine kinase inhibitor
TLG	Total lesion glycolysis
TMB	Tumor mutation burden
TP53	Tumor Protein P53
UMI	Unique molecular identifier

Table of Contents

Utfordringer ved å innføre sirkulerende tumor DNA-analyser i lungekreftforskning ..... **V**

Abstract ..... **VI**

Acknowledgements ..... **VII**

List of papers ..... **VIII**

Abbreviations ..... **IX**

1 Introduction **1**

2 Background **2**

2.1 Epidemiology 2

2.2 Survival 3

2.3 Subtypes of lung cancer 4

2.4 Molecular biology of NSCLC and implications for treatment 5

2.5 Diagnostic procedures 9

2.6 Treatment modalities and prognosis 12

2.7 Current challenges in the classification of patients with NSCLC 14

2.8 Candidate classification systems in NSCLC 15

2.9 Circulating tumor DNA (ctDNA) 17

3 The rationale for the project **21**

4 Aims and research questions **21**

4.1 Research questions paper I 22

4.2 Research questions paper II 22

4.3 Research questions paper III 22

4.4 Research questions paper IV 22

5 Material and methods **23**

5.1 Patient selection 23

5.2 Ethics 24

5.3 Statistical considerations 24

5.4 Data collection 24

5.5 Tumor specimens 25
5.6 ^{18}F-FDG PET/CT 25
5.7 Plasma samples 26
5.8 Next-generation sequencing (NGS) 26
5.9 Funding 28
6 Summary of results 29
6.1 Paper I 29
6.2 Paper II 31
6.3 Paper III 32
6.4 Paper IV 34
7 Discussion 36
7.1 Tumor-informed ctDNA analysis 36
7.2 Pre-treatment ctDNA detection as a prognostic factor in NSCLC 40
7.3 Methodological considerations 45
7.4 Challenges in the implementation of ctDNA analysis 47
8 Conclusion 51
9 Future perspectives on the use of ctDNA in lung cancer research 52
10 References 54
Paper I-IV 63

1 Introduction

Survival for non-small cell lung cancer (NSCLC) has improved considerably in the last decades [1,2]. One reason is an improved understanding of the biology of NSCLC leading to development of targeted therapies, in addition to better methods for staging, less invasive surgical techniques, stereotactic and conformal radiotherapy, and immunotherapy. The patient's health, the stage of disease and histopathological and molecular classification of the tumor are the main factors taken considered when selecting therapy for individual patients. However, there are large variations in treatment outcomes and need for better therapy. It is evident that we need better tools for the classification of NSCLC, which will enable more individualized and, thereby, better treatment of patients and facilitate the development of new, effective therapies.

Analysis of circulating tumor DNA (ctDNA) released from tumor cells into the circulation and identified from a blood sample has emerged as a promising method for such classification [3]. It has been hypothesized that ctDNA analyses can be used for, e.g., early detection of cancer, molecular tumor classification, prediction of treatment response and prognosis, response evaluation, and early detection of disease progression. However, the current clinical utility is limited, and several challenges must be overcome before ctDNA can be implemented in routine clinical practice.

This Ph.D. thesis explores the implementation of ctDNA analyses in translational research of NSCLC, from selecting the optimal analytical methods to investigating the clinical relevance. Three specific aims were formulated for the project:

- Find an approach for ctDNA detection that is sensitive, specific, and applicable for all NSCLC patients, while simple enough to be implemented in routine diagnostics.
- Investigate whether ctDNA detection provides prognostic information in patients with early-stage NSCLC.
- Investigate whether ctDNA analysis provides more prognostic information than established diagnostic procedures.

2 Background

2.1 Epidemiology

2.1.1 Etiology

Tobacco smoking is the main risk factor for lung cancer, and about 80% of patients are current or former smokers [4]. Environmental carcinogens such as radon and asbestos are believed to be the second most important risk factors in Norway, and it is estimated that radon annually causes 370 incidents, ~12% of new cases, and 300 deaths annually [5]. Other known common risk factors include exposure to asbestos, diesel exhaust, nickel, and chrome [6]. In addition, risk factors such as air pollution and high arsenic levels in drinking water are important risk factors globally [7].

A proportion of patients develop lung cancer with no apparent cause, indicating that there are other unknown factors. It is known that age is a risk factor for cancer in general, yet only a small percentage of the older population develops lung cancer [8]. It is likely that genetics also play a role [9], and it is estimated that hereditary factors cause 5-10% of all cancers, although the specific mechanisms are largely unknown [9].

2.1.2 Incidence

Lung cancer is the second most frequent cancer, with 2.2 million new cases globally in 2020 [10]. The incidence rates are highest in Western countries and Asia and lowest in Africa and Central America. It is the second most frequent cancer in Norway, with 1694 new cases among women and 1772 cases among men in 2022. Lung cancer is a rare disease in individuals below 50 years, reflected by the median age at diagnosis being 73 years [11].

Due to the strong association with tobacco smoking, the trends in lung cancer incidence follow the smoking trends with about 30 years delay [11]. After 70 years of an increasing incidence rate among Norwegian men, the rate stabilized around the 1990s and has decreased since 2010 (Figure 1). The incidence rate among women has been increasing until today since smoking became popular among women later than men, but it may have reached its maximum now. The Norwegian Cancer Registry has estimated that the number of new cases annually will increase until 2030, mainly due to an aging and growing population [12].

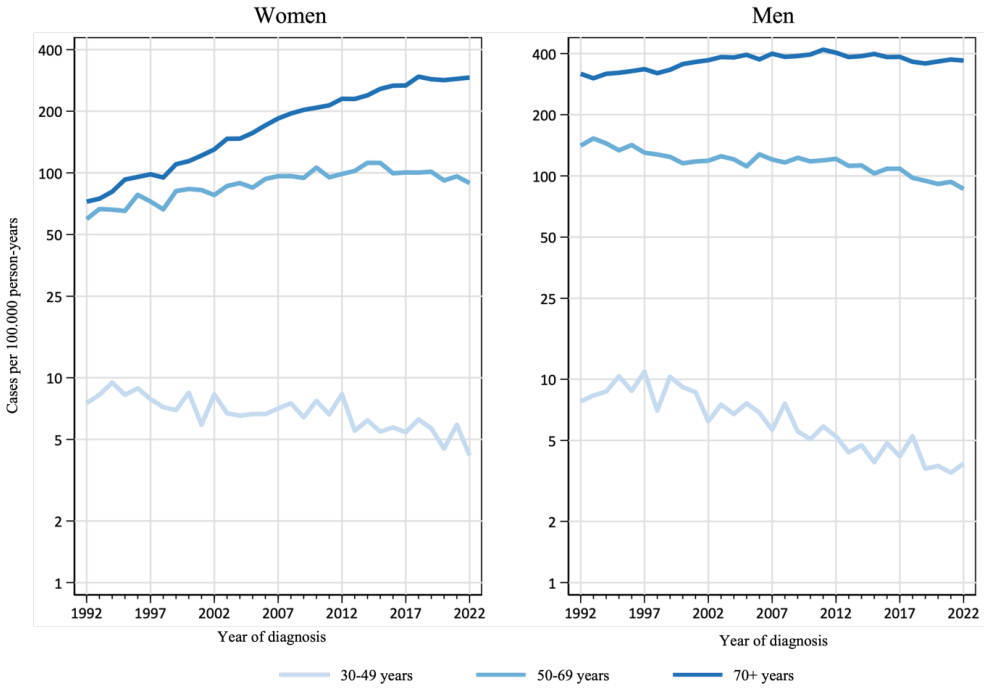


Figure 1: Lung cancer incidents per 100.000 person-years among Norwegian women (left) and men (right). Modified with permission from [11].

2.2 Survival

The survival rate of lung cancer is low compared to other cancer types, and it remains the most common cause of cancer death with 1.8 million deaths worldwide in 2020 [10]. Similarly, lung cancer is ranked highest in Norway and caused 2202 deaths in 2022 [11].

Patients with a single small tumor have the highest 5-year relative survival rate at ~70% compared to ~8% for advanced disease patients. Almost 45% of patients have metastatic disease at the time of diagnosis, which contributes to the poor survival of the lung cancer population as a whole [11]. There have, however, been encouraging improvements in survival since the start of this millennium. The overall 5-year survival rate has increased from 9% to 30% (Figure 2). Similarly, the 1- and 2-year survival rates have increased significantly in this period since more patients are offered potentially curative therapy, but also because patients with advanced disease are offered more effective therapies. Consequently, the lung cancer prevalence in Norway has almost doubled in the past decade from around 6000 patients living with the diagnosis to 11,500 [11].

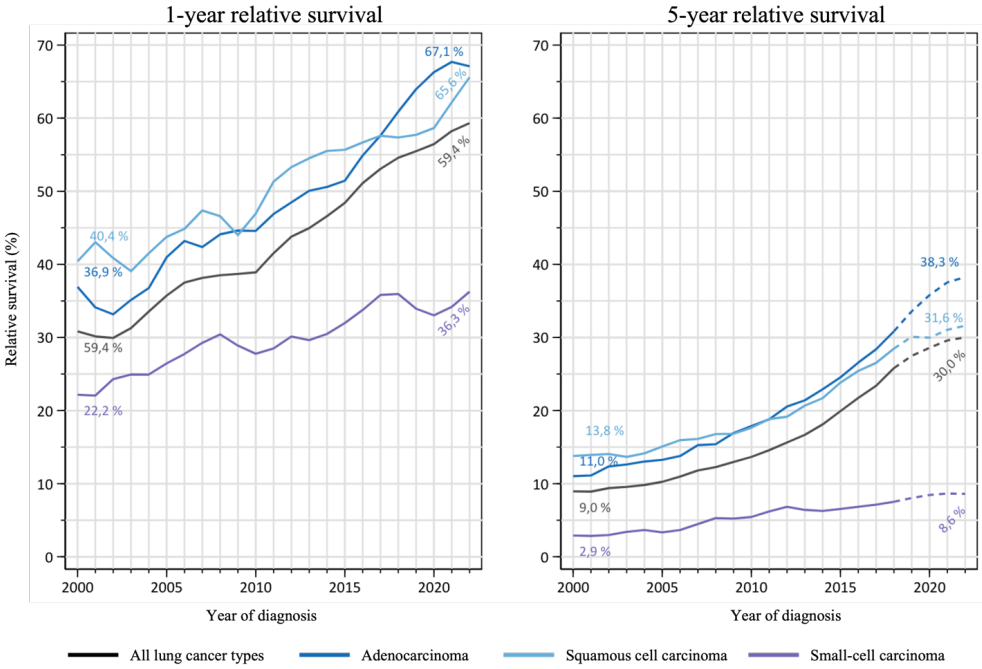


Figure 2: 1- and 5-year relative lung cancer survival for all patients and according to subtype. Modified with permission from [11].

2.3 Subtypes of lung cancer

Lung cancer is a collective term for malignant tumors that develop in the tissues of the lung, typically epithelial cells lining the airways, and are thereby defined as lung carcinomas. Two main subtypes, small-cell lung cancer (SCLC) and NSCLC, were defined in the 1960s based on different development and treatment outcomes [13]. The research in this thesis focused on NSCLC, and little attention is paid to SCLC in the following chapters.

2.3.1 Non-small cell lung cancer (NSCLC)

About 85% of lung cancer cases are classified as NSCLC [14]. Compared to SCLC, these tumors are generally less aggressive, and with more heterogeneity concerning disease development and response to therapy. Historically, the most common subtype was squamous cell carcinoma (SCC), likely due to its strong association with smoking, particularly with the early types of cigarettes [15]. SCC arises from squamous epithelial cells and is often located centrally in the lungs, close to the large airways (bronchi). The proportion of adenocarcinoma (ADC) cases has increased rapidly in the last decades and

ADCs now account for 60% of cases, while the proportion of SCC has decreased to 25% [16]. The reason is probably changes in chemical content of tobacco and the introduction of filtered cigarettes leading to deeper inhalation of smoke, and this subtype often occurs peripherally in the lungs [17]. ADC is also most common subtype among never-smokers [18]. Various types of epithelial cells give rise to ADC, and the best characterized precursor cells are alveolar type II cells, which produce pulmonary surfactant vital to maintain surface tension and prevent alveoli from collapsing [19]. Some tumors contain a combination of ADC and SCC morphology and are classified as adenosquamous carcinoma. Large cell carcinoma is another common entity and there is a wide range of rare subtypes. In some cases, tumors do not resemble any defined subtype and is classified as NSCLC-not otherwise specified [20].

2.3.2 Small-cell lung cancer

About 15% of cases are classified as SCLC, characterized by rapid tumor growth, early formation of metastases (particularly to the brain), and rapid and profound response to chemotherapy, which is usually followed by treatment resistance. The overall poor prognosis and has remained more or less unchanged the last 20 years [21]. SCLC is almost exclusively seen in smokers, at least in Western countries, and similar to SCC, the primary tumor often arises in the bronchi [22]. The precursor cells are believed to be neuroendocrine epithelial cells because the tumors often express neuroendocrine markers, and SCLC is one of the most common causes of paraneoplastic syndromes [23]. The tumor cells are small to medium in size, have a high nucleus-to-cytoplasm ratio, grow fast, and metastasize early [24]. Some tumors contain a mixture of NSCLC and SCLC tumor cells but are classified and treated as SCLC since this component is usually the most aggressive and clinically dominant [25].

2.4 Molecular biology of NSCLC and implications for treatment

Cancer cells are characterized by hallmark capabilities that distinguish them from their normal counterparts, such the ability to sustain proliferation, avoid detection by the immune system, invade surrounding tissue, and metastasize [26]. These hallmarks are believed to be acquired through inflammation, epigenetic programming, and genomic instability. Tobacco smoke causes inflammation, e.g., via the release of reactive oxygen species, which trigger a production of growth factors and cytokines, and chronic inflammation by consistent exposure lead to an imbalance and consistent production of these factors [27,28]. This creates a tumor-promoting milieu by, for example, providing proliferative signals, inducing angiogenesis, and recruitment and selection of tumor-promoting immune cells. Additionally, both inflammatory mediators and tobacco

carcinogens cause epigenetic changes and direct changes in DNA [29,30]. Studies have demonstrated that individual CpG sites are hypomethylated and hypermethylated in smoking-related ADC compared to non-smokers [31]. Gene expression is repressed by DNA hypermethylated promoter regions which may be carcinogenic if it occurs in tumor suppressor genes. These changes are reversible, and a meta-analysis showed that the methylation pattern mostly returns to a normal level within five years of smoking cessation [32]. Hypomethylation, on the other hand, causes genomic instability and aneuploidy, for example, by loss or gain of chromosome arms (reviewed by [33]). Furthermore, tobacco smoking is associated with a higher tumor mutation burden (TMB, i.e., mutations/megabase) and predominance of cytosine to adenine (C>A) transversion. In contrast, cytosine to thymine (C>T) transitions are predominant in non-smokers.

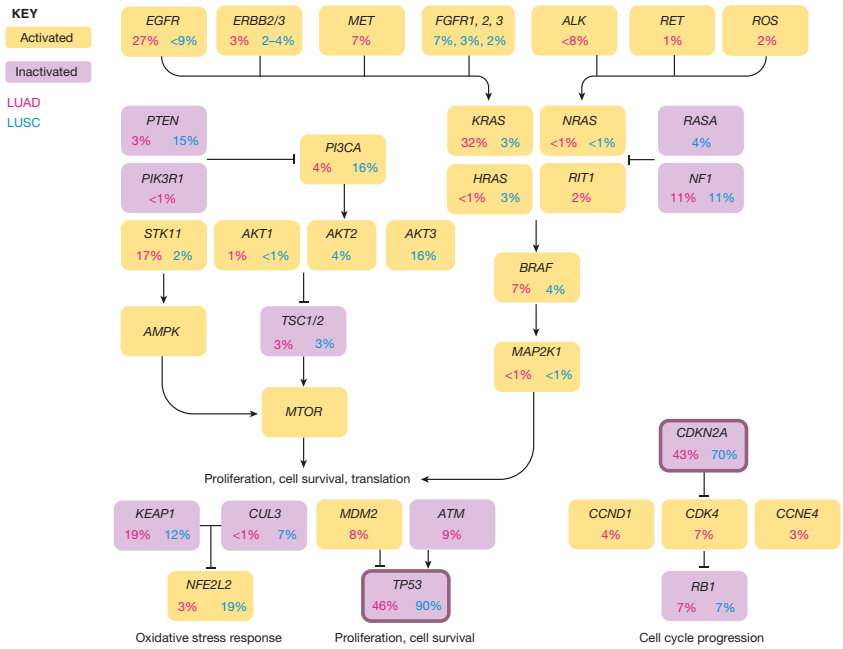


Figure 3: Alterations in targetable oncogenic pathways in NSCLC, with percentages showing the prevalence in ADC (LUAD) and SCC (LUSC). Reused with permission from Herbst *et al*, 2018 [1].

Direct changes in the DNA can functionally alter oncogenes and activate growth-promoting signaling pathways (Figure 3). The first targetable driver gene to be identified in lung cancer was the epidermal growth factor receptor (*EGFR*), a transmembrane tyrosine kinase. Upon ligand binding, the receptor dimerizes and activates the RAS/RAF/MEK/MAPK pathway leading to cell proliferation and survival. EGFR is over-

expressed in many NSCLCs, which lead to treatment with monoclonal antibodies and, later intracellular tyrosine kinase inhibitors (TKIs), which demonstrated an effect in a minority of patients [34]. In 2004, activating *EGFR* mutations were discovered and suspected to be associated with TKI response, confirmed by the IPASS study in 2009 [35–37]. Targetable mutations in *EGFR* are found in 15-20% of ADC patients in Western countries and 45-50% of the Asian population and are more prevalent in female non-smokers (Figure 4). The most frequent mutations are deletions in exon 19 and the single nucleotide variant L858R in exon 21. Tumors develop resistance upon EGFR TKI treatment by acquiring an exon 20 insertion or the T790M in *EGFR,* via mechanisms activating the same pathway but bypassing EGFR signaling, or by histological transformation to SCLC [38]. Third-generation TKI is effective in patients with a T790M mutation but lead to other resistance mutations, such as C979S, demonstrating a central issue of targeted treatment [39].

A second important discovery was the gene fusion involving the transmembrane tyrosine kinase anaplastic lymphoma kinase (*ALK*) in 2007 and the subsequent identification of an effective TKI [40]. *ALK* fusions are found in ~4% of ADC patients, primarily female light- or never-smokers [41]. In the following years, more driver gene alterations were identified, such as *RET* [42], *ROS1* [43], and *NTRK* fusions [44], succeeded by the development of effective TKIs [45–47], as well as an effective treatment for ADC with a V600E mutation in *BRAF* [48]. Alterations in these driver genes are most often mutually exclusive, each event is rare and occurs in less than 5% of ADC patients (Figure 4) and most often in non-smokers [49].

The second most frequent driver gene in ADC is *KRAS*, a known driver gene in several cancer types, including lung cancer, for decades, but targeted treatment development proved difficult [50]. However, a *KRAS* G12C inhibitor was recently proven effective [51]. *KRAS* mutations are more frequent in the Western population than in Asia (~25% vs. ~10%) and in male smokers. About 80% of mutations occurring in codon 12 (Figure 4). Among Norwegian ADC patients, *KRAS* is mutated in about 38% and 17% had G12C mutation, qualifying them for targeted treatment [52]. The mutated protein constitutively activates the downstream signaling pathway, effectively bypassing EGFR signaling, and mutations in these two driver genes rarely co-exist.

About 25-50% of ADC patients harbor no targetable driver gene alteration [53], but may harbor alterations in other driver genes. Examples are over-expression of the transcription factor MYC or mutated genes in the PI3K/AKT/mTOR pathway, such as *PIK3CA*, *PTEN*, and *AKT*, which lead to inhibition of apoptosis [54].

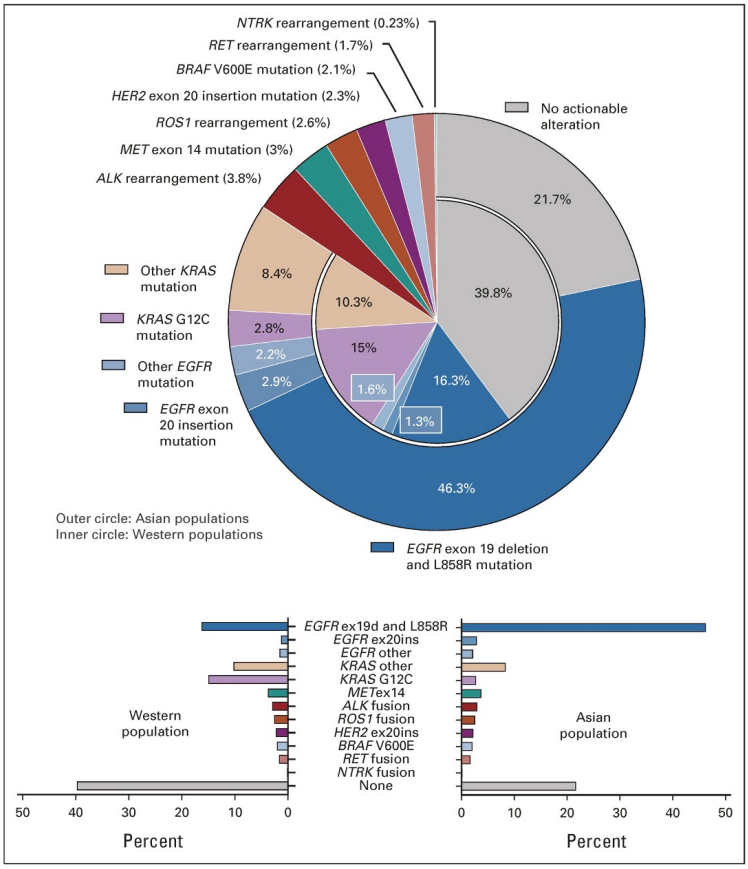


Figure 4: The most frequent actionable alterations in ADC patients in the Western and Asian populations. Reused with permission from Tan and Tan, 2021 [53].

The inactivation of tumor suppressor genes leads to the repression of growth-inhibiting signaling pathways (Figure 3). For example, the protein p53 is activated by DNA damage and prevents further accumulation of genetic damage by arresting the cell in the G1 phase (growth) and subsequently induce DNA repair or apoptosis [55]. The gene encoding p53, *TP53*, is mutated in 40-80% of NSCLC and is also frequently lost by deletion of the chromosomal region 17p13 causing a homozygous loss-of-function [56,57]. Other frequently mutated tumor suppressors are *CDK2NA* and *RB1* (the latter more prevalent in SCLC than NSCLC), which control the cell cycle progression from G1 to S phase (DNA synthesis), and *STK11* which controls cell polarity, motility, and differentiation. *STK11* often correlates with *KRAS* mutations and, thus, is often found in male smokers with ADC. Loss of heterozygosity by chromosome arm 3p, which contains several tumor suppressor genes, is also found in most NSCLCs [58]. Currently no targeted treatments are available for alterations in tumor suppressor genes, possibly because it is more challenging to restore loss-of-function than to block gain-of-function.

Another hallmark of NSCLC that has implications for treatment is the ability to escape detection by the immune system, which is especially relevant in lung cancer because the high TMB increases the likelihood of producing neoantigens recognized T-cells. Many NSCLC tumors have increased expression of programmed cell death ligand-1 (PD-L1), which binds to the programmed death receptor 1 (PD-1) of T-cells and inhibits the cytotoxic activity [59].

2.5 Diagnostic procedures

Typical symptoms of lung cancer are coughing, hemoptysis, dyspnea, chest pain, fatigue, and weight loss [60]. If lung cancer is suspected, patients undergo a set of diagnostic procedures to confirm the cancer diagnosis, assess the disease stage, histological subtype, molecular profile, and the patient's tolerability to treatment.

2.5.1 Cancer staging

Lung cancer is staged according to the TNM system, i.e., the tumor size (T, stage 1-4), lymph node involvement (N, stage 0-3), and presence of distant metastases (M, stage 0-1)[61]. According to the T, N and M descriptors, an overall disease stage is defined from I to IV (Table 1). The TNM staging system is continuously revised and updated due to changes in diagnostic technology, available therapies, and treatment policies, and the TNM system is designed to be a prognostic factor to separate the four stages in a Kaplan-Meier analysis [61]. The 9th version is expected to be introduced in 2024.

SCLC has traditionally been staged as "limited stage" (LS) or "extensive stage" (ES) since this is the only separation that is used to select therapy. LS is defined as having confined disease to one hemithorax and regional lymph nodes, whereas all other patients have ES. Various definitions of LS have been employed, and TNM staging is also recommended for SCLC to understand better the clinical impact of disease extent [62].

Table 1: TNM version 8. Source Goldstraw *et al,* 2015 [61].

	N0	**N1**	**N2**	**N3**
T1	IA	IIB	IIIA	IIIB
T2a	IB	IIB	IIIA	IIIB
T2b	IIA	IIB	IIIA	IIIC
T3	IIB	IIIA	IIIB	IIIC
T4	IIIA	IIIA	IIIB	IIIC
M1a	IVA	IVA	IVA	IVA
M1b	IVA	IVA	IVA	IVA
M1c	IVB	IVB	IVB	IVB

2.5.2 ^{18}F-fluorodeoxyglucose positron emission tomography/computed tomography (^{18}F-FDG PET/CT)

All suspected lung cancer patients are examined by chest computed tomography (CT), and those considered for potentially curative treatment based on this initial staging are then referred for an ^{18}F-fluorodeoxyglucose positron emission tomography/CT (^{18}F-FDG PET/CT) scan since this is the most sensitive and specific imaging technique for assessing malignant lesions [63]. The patient receives an injection with radioactively labeled glucose (^{18}F-fluorodeoxyglucose) which emit radiation that is detected, transformed into a visual signal, and then transposed onto the CT images.

Most tumor tissues have a high glycose uptake due to the Warburg effect of malignant cells, which is the process of switching to anaerobic glycolysis even in the presence of oxygen [64]. The level of glucose uptake can be quantified as the standard uptake value (SUV, g/mL), and used to identify malignant lesions [65]. It is, however, important to remember that some organs (e.g., kidneys and brain) and benign conditions (e.g., inflammation) have a constant high glucose uptake or high perfusion, making it challenging to detect of metastases in these organs by ^{18}F-FDG PET/CT. Patients with a high risk of brain metastases, e.g., stage III patients, should therefore undergo magnetic resonance imaging of the brain.

^{18}F-FDG PET/CT is also used to define target volumes for definitive radiotherapy, ensuring that all lesions are included while at the same time limiting the radiotherapy fields to ^{18}F-FDG PET/CT-positive lesions instead of, e.g., routinely including all regional lymph node stations [63]. This approach limits normal tissue irradiation, reducing radiotoxicity to a minimum.

2.5.3 Histological classification

Histological classification of the tumor requires a biopsy or a cytological specimen, which are obtained via bronchoscopy, surgery, or transthoracic sampling. Tissue is fixated with formalin which creates crosslinks between the proteins and keeps the morphology is intact [66]. The fixed tissue is then embedded in paraffin to enable cutting thin sections (3 μM) of the formalin-fixed paraffin-embedded (FFPE) tissue block and allows stable long-time storage. The tissue sections are treated with a hematoxylin and eosin stain that make cellular structures visible to the pathologist who evaluates the specimen by microscopy. Additional slides are often stained using immunohistochemistry (IHC) to visualize proteins which are commonly expressed by different subgroups of lung cancer.

2.5.4 Molecular diagnostics

Molecular diagnostics support the histological classification and identify molecular characteristics that are targetable by specific therapies and might be prognostic. The expression of the protein PD-L1 is routinely assessed using IHC since this is essential for deciding whether patients who are ineligible for targeted therapy should be offered mono- or combination immunotherapy. Additionally, tumor DNA extracted from advanced NSCLC is analyzed by next-generation sequencing (NGS) to identify actionable targets. Advanced ADCs are additionally tested for translocations of *ALK* and *ROS1* using IHC and confirmed using fluorescent *in situ* hybridization (FISH). Available targeted therapies at public hospitals outside clinical trials in Norway are EGFR, ALK, ROS1, NTRK, RET, and BRAF inhibitors.

There are no established clinical implications of molecular testing for SCLC.

2.5.5 Patient health and preferences

Measuring lung function is essential to ensure sufficient lung capacity after surgery or radiotherapy. Adequate bone marrow, kidney, and liver function is required to tolerate most systemic therapies, especially cytotoxic chemotherapy. The patient's overall performance status (PS) is usually assessed according to the WHO/Eastern Cooperative Oncology Group Performance Status [67]. Patients with normal activity have a PS of 0, those with close to normal daily activities have a PS of 1, while those with a score of 2 are unable to work, yet in bed <50% of the time. Patients with a score of 3 are in bed >50% of the time, and patients with a score of 4 are completely bedbound.

There are several methods to measure the number and types of co-existing conditions and diseases. Numerous studies show that comorbidity is associated with shorter survival time and more treatment toxicity, but the clinical application when selecting treatment for individual patients remains to be established [68].

Finally, the patient's preferences should always be taken into consideration. Compliance and satisfaction depend on sufficient information about the disease, treatment options, and potential benefits and toxicity.

2.6 Treatment modalities and prognosis

2.6.1 Potentially curative treatment of NSCLC

About 40% of Norwegian NSCLC patients receive potentially curative treatment with surgery or radiotherapy [11]. As shown in Figure 5, operable patients with stage I disease are treated with surgery alone, while patients with stage II-III disease are offered adjuvant chemotherapy after surgery [69], and patients with a targetable *EGFR* mutation receive an EGFR TKI [70]. Overall, 75% of operated patients are alive after five years, ranging from 50% of stage III patients to 90% of stage I patients [11].

Medically inoperable patients with stage I-II tumors without nodal involvement can be offered stereotactic body radiotherapy, in which several beams of energy are focused on the tumor and deliver high-intensity radiation [71]. Patients with larger tumors or lymph node involvement are offered conventional high-dose radiotherapy. Medically inoperable stage III patients are treated with concurrent chemotherapy and radiotherapy followed by adjuvant immunotherapy for up to one year [72]. Regardless of treatment modality, the 5-year relative survival is 68%, 50% and 24% for stages I, II, and III, respectively [11]. Notably, the population of stage III patients encompasses both curative and palliative treatment.

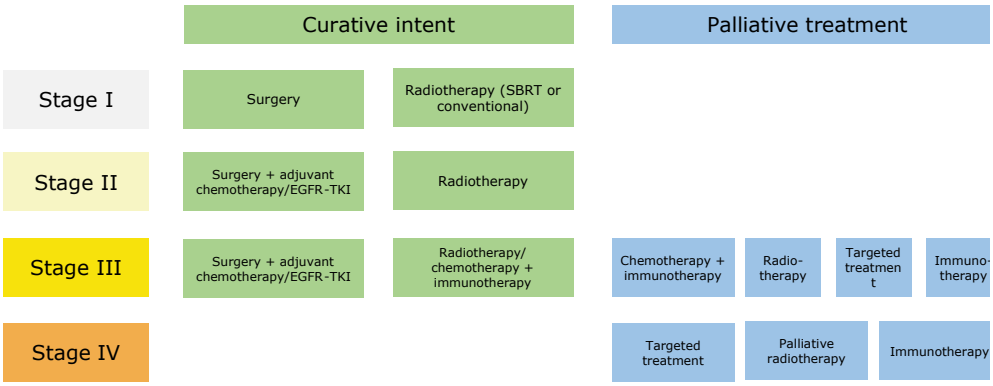


Figure 5: Curative and palliative treatment of NSCLC, according to the Norwegian treatment guidelines for lung cancer [60]. SBRT: stereotactic body radiotherapy.

2.6.2 Palliative treatment of NSCLC

The goal for patients who cannot be offered curative treatment is to prolong life, achieve symptom control and maintain quality of life. The main anti-cancer treatments are chemotherapy, targeted therapy, immunotherapy, and palliative radiotherapy (Figure 5).

The first systemic palliative therapy for NSCLC was cytotoxic chemotherapy. The survival benefit was, however, limited, and severe toxicity was common. Typically, 1/3 of patients respond, 1/3 have stable disease, and 1/3 progress during standard, first-line platinum-doublet chemotherapy [73,74]. The first major advancement in systemic therapy for NSCLC was the identification of TKI response in patients with a targetable *EGFR* mutation [37]. Third-generation EGFR TKIs lead to response rates of 80%, median response durations of 17 months, and significantly prolong survival compared with cytotoxic chemotherapy [75]. Targeted therapies provide higher response rates, prolonged survival compared to chemotherapy and are usually less toxic [76]. However, virtually all patients experience disease progression due to treatment resistance, and unfortunately, most patients with driver alterations appear to benefit less from immunotherapy than patients without [77].

Checkpoint inhibitor immunotherapy has become the backbone of systemic therapy for NSCLC without driver alterations. NSCLCs, regardless of histologic subtype, with high PD-L1 expression are offered mono-immunotherapy [78], and patients with low or no PD-L1 expression are offered combined chemotherapy and immunotherapy [79].

Palliative radiotherapy for NSCLC includes irradiation of progressing, systemic treatment-resistant, thoracic tumors compressing central airways and large vessels; bone metastases causing pain or at risk of fracturing; and ulcerous skin metastases [80]. Surgery is an alternative for patients with an immediate need of symptom control or with radioresistant tumors, usually limited to patients without general progression or eligible for subsequent highly effective systemic therapy.

There are examples of patients with advanced disease who achieve long-term disease control and are alive five years after treatment with immunotherapy [81]. These patients may have been cured even though the treatment intention initially was palliative. However, the 5-year relative survival for patients with stage IV disease is 8% [11].

2.6.3 Treatment of SCLC

In contrast to NSCLC, surgery plays little role in SCLC treatment since most patients have too widespread disease at diagnosis [82]. The main treatment for SCLC is platinum doublet chemotherapy [83]. Patients with LS receive concurrent radiotherapy to all lesions, whereas ES patients receive chemotherapy and immunotherapy [84]. Prophylactic cranial irradiation reduces the risk of brain metastases and prolongs survival in LS SCLC [85]. Patients who experience relapse may be offered more chemotherapy [86]. About 70-90%

respond to primary treatment and approximately 1/3 of LS patients are cured by chemoradiotherapy, but most experience disease relapse [87]. The 5-year overall survival rate is 25-30% for LS patients [88], whereas 3-year survival has increased from 5-6% to 15-18% for ES patients with the recent addition of immunotherapy [89].

2.6.4 Prognostic and predictive factors used in treatment selection

Prognostic factors are associated with the disease outcome regardless of treatment modality. PS, sex, disease stage, and appetite loss are the strongest prognostic factors in NSCLC [90–92]. Good PS, female sex, lower disease stage, and no appetite loss are associated with better prognosis.

A predictive marker provides information about the outcomes of specific interventions. The best examples in NSCLC are molecular alterations for which there are specific targeted treatments available. PD-L1 expression is another predictive factor, although not perfect, associated with response to immunotherapy for NSCLC, except patients with a targetable alteration [79].

2.7 Current challenges in the classification of patients with NSCLC

Although there has been considerable improvement in lung cancer survival, there are considerable variations in treatment response, time to progression, survival, and treatment toxicity among patients who receive potentially curative treatment. Current challenges include:

1. Accurate mapping of the disease extent – who should be offered adjuvant treatment?

The inability to accurately assess the disease extent is likely a reason why many patients experience relapse despite surgical removal or irradiation of all known lesions – or radiological complete therapeutic response, or metastases in an initially believed unaffected organ. For example, the improved ability to detect the extent of micro-metastases in the primary setting or as a residual disease after initial therapy would enable us to avoid unnecessary treatment and potentially toxic interventions and avoid adjuvant therapies in patients who are already cured.

2. Predict disease development – whom should we treat and whom should we monitor?

Some tumors grow slowly and do not influence the patient's health much. These patients may benefit from surveillance rather than receiving potentially unnecessary or toxic treatment. The problem is that we cannot predict growth rate or metastatic potential. We

know that prophylactic cranial irradiation reduces the risk of brain metastases in SCLC. However, not all patients are at risk of brain metastases, and irradiation can cause significant cognitive dysfunction [93].

3. Predicting and evaluating treatment effect

Identifying characteristic(s) that distinguish patients with poor response from the good responders would help us selecting the most effective treatment for every patient. Furthermore, many patients experience temporary or no response to treatment. If severe treatment toxicity occurs, it would be of great help to know whether the treatment is effective. Today, treating patients for relatively long intervals is common before response evaluation by imaging is performed. In the worst case, patients' conditions might deteriorate due to treatment toxicity or disease progression, prohibiting switching to other, effective therapy.

4. Evaluation of treatment tolerability – who are at risk for experiencing treatment toxicity?

It is unclear why some patients experience severe toxicity from the same treatment that others tolerate. Identifying at-risk patients would further help select the most optimal treatment for the patient.

5. Prognosis – predictions of the future

Current prognostic and predictive factors cannot accurately estimate disease trajectories, such as disease-free survival after treatment response and total survival time. Improved prognostication would help clinicians decide the frequency of follow-up visits and would likely help patients and their families.

Following the identification of *who*, it is also relevant to investigate *how* or *why* tumor development and treatment response differ between patients in order to support the classification and explore new strategies for cancer management.

2.8 Candidate classification systems in NSCLC

The development of effective *EGFR* TKIs began the era of biomarker-driven treatment of NSCLC [37]. One challenge with patient classification based on tissue markers is the limited available tumor tissue. Only a minority of lung cancer patients undergo surgery, and tissue samples are frequently obtained through biopsy and, in many cases, cytology. Thus, blood biomarkers emerged as an attractive alternative or a complementary source

of biomarkers for patient classification. Blood sampling is easy, virtually risk-free and opens up new possibilities, such as repeated sampling.

The term "liquid biopsy" was introduced in 2010 to describe the collection and analysis of circulating tumor cells (CTC) [94]. CTCs originate in the tumor and theoretically provide insight into the genomic, proteomic, and transcriptomic landscape. Moreover, detecting CTCs in the blood was shown to be a negative prognostic factor in itself. A considerable challenge for CTC analysis is the low count in blood, especially for early-stage cancer, and the definition of a CTC-elevated sample is ≥ 5 CTCs/7.5 mL whole blood. Furthermore, the blood must be processed soon after sampling to avoid cell lysis of the few CTCs that were collected [95]. Now, the liquid biopsy concept includes many different circulating molecules, and several of these are under investigation for a role in the management of NSCLC (Figure 6). Examples include different types of cell-free circulating nucleic acids, tumor-educated platelets, and extracellular vesicles. Additionally, saliva, urine, and pleural fluid are sources of these biomarkers [96].

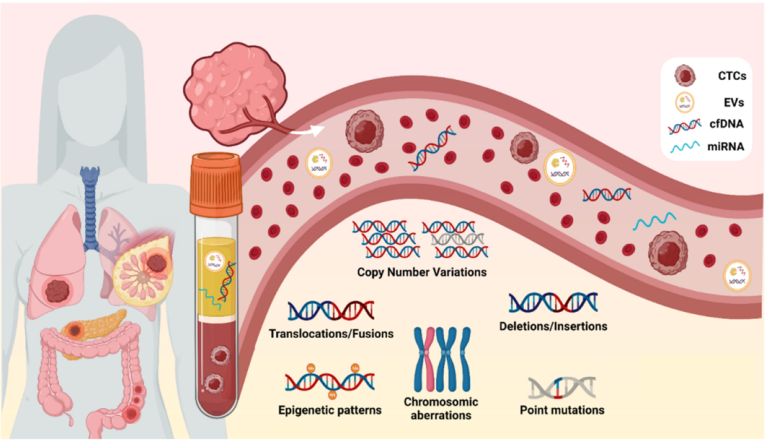


Figure 6: Examples of liquid biopsy components which are investigated for a role in NSCLC classification and management. Modified with permission from Palacín-Aliana *et al.* 2021 [97].

In addition to liquid biopsy, other candidate classification systems are under investigation, such as characterizing the tumor microenvironment and the microbiome, especially focusing on predicting response to immunotherapy [98]. New technology enables multimodal analysis of DNA, RNA, and protein expression at a much higher resolution than before, even in morphologically intact tissue, which may identify new tumor traits associated with specific patient outcomes [99]. This is a rapidly evolving field, especially with the improvement and availability of artificially intelligent software for image analysis.

2.9 Circulating tumor DNA (ctDNA)

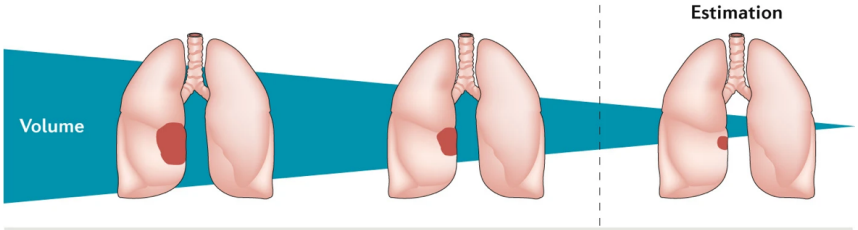
2.9.1 Discovery

In the 1960-70s, Stroun and Anker performed experiments indicating that eukaryotic cells actively secreted DNA to the medium. The suggestion that cfDNA existed was so controversial that the researchers were accused of not performing "real science" and had trouble receiving funding (reviewed in [100]). However, more evidence of cfDNA came to light, and its relevance in oncology was introduced when Leon *et al.* in 1977 showed that the cfDNA level was higher in the blood of cancer patients than in healthy individuals [101]. In the following decades, Stroun, Anker, and their collaborators demonstrated that cancer patients had tumor-derived cfDNA in the blood and that it harbored tumor-specific mutations, thus defining the cfDNA fraction circulating *tumor* DNA – ctDNA [100].

2.9.2 Characteristics of ctDNA

ctDNA is believed to originate from dying cells, based on the most abundant fragment size of 143 bp [102]. This length equals DNA wrapped around a nucleosome, thus protected from DNase digestion during apoptosis. Fragmented ctDNA is released into the circulation, where it exists in several forms with a short half-life of 0.5-2 hr. Contrary to how it is depicted in illustrations, ctDNA is probably rarely circulating in the form of naked DNA since it would be degraded by blood DNases and trigger the immune system. ctDNA is likely protected by protein complexes or associated with various types of vesicles [100]. Most ctDNA extraction protocols disrupt vesicles and protein complexes, and consequently, all cfDNA is extracted together, and information on its circulating form is lost.

A major challenge in the analysis of ctDNA is that it comprises a very small fraction of the total cfDNA. The ctDNA fraction is commonly measured by the mutant allele frequency (MAF, sometimes termed variant allele frequency, VAF), which is the percentage of haploid DNA genomes with a given mutation. Abbosh *et al.* observed that a tumor of 10 cm^3, equivalent to stage T1c, corresponded to a 0.1% MAF (Figure 7) [103]. The authors estimated that a tumor of 1 cm^3 would correspond to a 0.008% MAF. Consequently, 12,500 unique cfDNA genomes (haploid genome equivalents, hGEs) must be analyzed to detect one mutated ctDNA hGE. This is a challenge even with ultrasensitive detection methods since 4 mL of plasma from a 10 mL blood draw contain, on average, 6000 hGEs [104].



MAF	1.4% (0.62–3.1%)	0.1% (0.06–0.18%)	0.008% (0.002–0.03%)
Nodule diameter	5.8 cm	2.6 cm	1.2 cm
Nodule volume	100 cm^3	10 cm^3	1 cm^3
T stage	T3	T1c	T1b

Figure 7: Illustrates the correlation between tumor volume and expected MAF of a ctDNA mutation and how this corresponds to the T-stage according to TNM version 8. Reused with permission from Abbosh *et al.* 2018 [103].

Even a sample with enough hGEs poses a challenge because the true ctDNA mutation could be difficult to distinguish from technical noise and non-tumor-relevant variants. Examples are mutations that arise in hematopoietic stem cell differentiation (clonal hematopoiesis of indeterminant potential, CHIP) [105], spontaneous changes in DNA post blood draw such, or artifactual mutations that arise during DNA preparation for analysis.

2.9.3 Analysis of ctDNA

There are two main approaches for identifying ctDNA; tumor-informed and tumor-naïve, distinguished by whether analysis of matching tumor DNA is included (Figure 8). Tumor-informed approaches typically involve analysis of a wide range of genes in tumor DNA by NGS, and the mutation profile is then exploited in one of two ways. The first option is to analyze a pre-defined set of genes in cfDNA (commonly by NGS) and then use the tumor mutation list to identify ctDNA mutations (Figure 8A). This approach allows parallel analysis of tumor DNA and cfDNA and equal, standardized treatment of all patient samples. Alternatively, the tumor mutation profile is used to select regions cfDNA analysis (Figure 8B), either by designing customized NGS panels [106,107] or using single-gene detection methods such as droplet digital polymerase chain reaction (ddPCR)[108]. This approach has a longer turnaround time but enables higher coverage (i.e., deep sequencing) in a limited region and, thus, highly sensitive analysis.

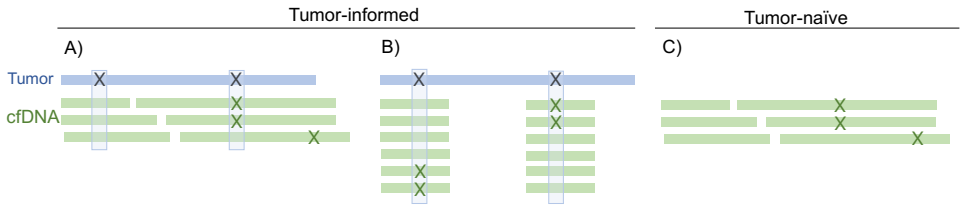


Figure 8: (A) A tumor-informed approach, in which cfDNA is analyzed by a broad NGS panel and data interpretation is guided by the matching tumor mutation profile. **(B)** An alternative tumor-informed approach, in which the tumor mutation profile is used to select specific regions in cfDNA for analysis, i.e., a customized approach. This approach may be more sensitive due to the lower number of analyzed genomic positions, allowing "deeper" NGS. **(C)** A tumor-naïve approach, in which ctDNA is analyzed without information from the matching tumor and requires sophisticated software for error correction and filtering non-tumor relevant mutations.

Tumor-naïve cfDNA analysis requires sophisticated bioinformatic analysis for distinguishing biological variants from technical errors and true tumor mutations from CHIP or other non-tumor related mutations (Figure 8C). Optionally, matching white blood cells can be analyzed by equally deep sequencing to exclude CHIP mutations. In 2014, the CAncer Personalized Profiling by deep sequencing (CAPP-Seq) method was described by Newman *et al.* which was optimized for NGS library preparation of cfDNA and targeted ~125 kb in 139 genes, theoretically covering 96% of patients with ADC or SCC [109]. The sensitivity of CAPP-seq for ctDNA detection in NSCLC patients (n=13) was 85%, and 93% in a subsequent study of 40 stage I-III NSCLC patients [110]. Another tumor-naïve method was described by Phallen *et al.* in 2017, called Targeted Error Correction sequencing (TEC-Seq), which involved the removal of duplicate sequences and filtering sequencing artifacts and CHIP mutations [111], and showed ctDNA detection in 45-83% of stage I-IV NSCLC patients (n=71).

These studies, in combination with the results from the TRACERx study detecting ctDNA in 48% of operable NSCLC patients, reported encouraging results that ctDNA detection was feasible in early-stage NSCLC patients [106]. In addition, they demonstrated the need to apply highly sensitive detection methods, and ctDNA still did not seem to be detectable in all patients.

2.9.4 Clinical and research applications

The optimal analysis approach and analytical sensitivity requirement depends on the application setting. The sensitivity is often measured as the lowest detectable MAF and termed limit of detection (LOD), and applications in early-stage cancer usually require higher sensitivity (i.e., lower LOD) than in the setting of advanced disease (Figure 9). Currently, ctDNA analysis is approved to guide targeted therapy in advanced stage NSCLC. Since the MAF generally increases with tumor volume, the likelihood of ctDNA detection is

higher in samples from patients with advanced disease, and most translational studies on ctDNA focused on advanced cancer patients due to the lack of sufficiently sensitive technology. These patients also comprise the majority.

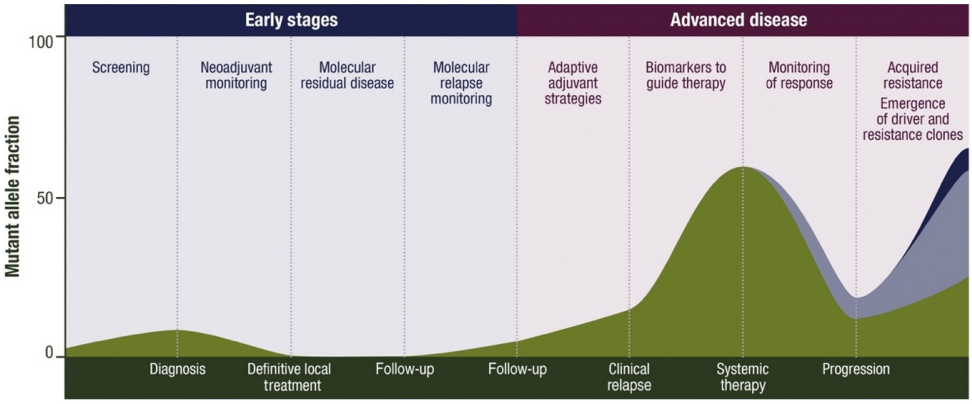


Figure 9: Illustration of different cfDNA analysis applications throughout the disease development. The Y-axis estimates the MAF in the different settings and illustrates the varying requirement in assay LOD. Reused with permission from Pascual *et al.*, 2022 [112].

3 The rationale for the project

It is reasonable to believe that there may be a difference in tumor behavior between ctDNA "shedders" and "non-shedders", and that the amount of ctDNA in plasma may be a surrogate marker for tumor volume, turn-over rate, or aggressiveness [106]. In this setting, pre-treatment ctDNA detection is a potential prognostic marker that could contribute to an improved classification of NSCLC patients. Most studies on ctDNA have included patients with advanced NSCLC, and it is not clear whether ctDNA analysis provides prognostic information in early-stage NSCLC patients who receive potentially curative treatment. Furthermore, we do not know whether such analyses provide information independent of established prognostic factors in these patients.

The technical challenges of ctDNA analysis are more prominent in the early-stage setting than advanced stage due to the limitations as mentioned earlier in both the absolute and relative quantity of ctDNA in a blood sample. The best strategies for ctDNA analysis and identification that are sensitive, specific, applicable for patients with various mutations, and additionally feasible to implement in routine diagnostics remain to be defined.

4 Aims and research questions

To investigate the potential of ctDNA as a biomarker for the classification of NSCLC, we aimed to:

- Find an approach for ctDNA detection that is sensitive, specific, and applicable for all NSCLC patients while simple enough to be implemented in routine diagnostics.
- Investigate whether ctDNA detection provides prognostic information in patients with early-stage NSCLC.
- Investigate whether ctDNA analysis provides more prognostic information than established diagnostic procedures.

4.1 Research questions paper I

Theoretically, increasing the NGS gene panel size increases the likelihood of detecting ctDNA mutations in most patients.

- Is it feasible to analyze ctDNA using a commercially available large NGS panel and bioinformatic analysis software?
- Is genotyping of matching tumor DNA and leucocyte DNA necessary to interpret cfDNA NGS data?

4.2 Research questions paper II

Tissue genotyping relies on DNA extracted from routinely obtained FFPE tissue, varying in quality and quantity.

- Does quality assessment of individual FFPE DNA samples provide helpful information in NGS data interpretation?

4.3 Research questions paper III

- Is it feasible to apply customized NGS panels for ctDNA detection?
- Is ctDNA detection associated with progression-free survival (PFS) and overall survival (OS) in NSCLC patients?

4.4 Research questions paper IV

Both ctDNA detection and assessment of tumor metabolism by ^{18}F-FDG PET/CT seem to provide information on tumor aggressiveness.

- Is ctDNA quantity/detectability associated with tumor metabolic activity in early-stage NSCLC patients?
- Do ctDNA analysis and routinely performed ^{18}F-FDG PET/CT provide independent prognostic information for these patients?

5 Material and methods

5.1 Patient selection

The patient selection criteria for the current project were having confirmed NSCLC with an available tumor tissue sample and a plasma sample drawn before treatment. This project overlapped in time with another PhD project in which we screened ADC tumor specimens for a mutation in *KRAS* (Figure 10). *KRAS* negative patients were included in the current project, and a handful of patients with non-ADC originally registered as ADCs. Patients underwent diagnostic workups, received treatment, and were followed according to national Norwegian guidelines and local routines.

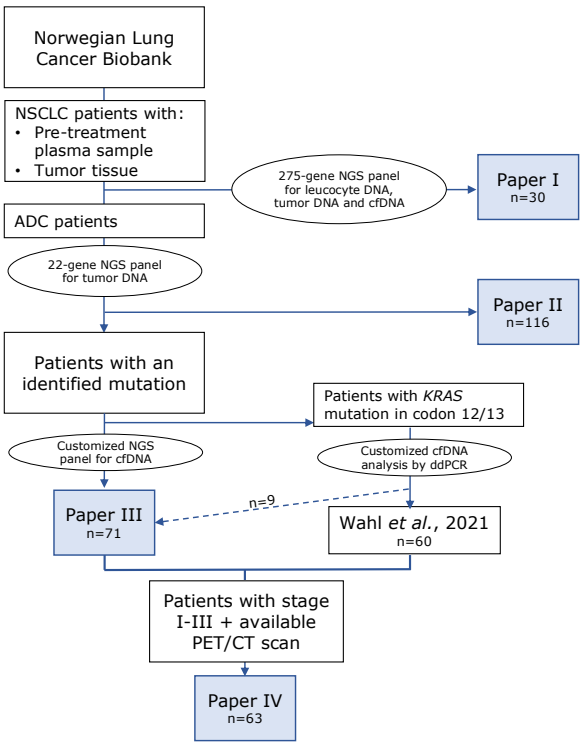


Figure 10: The patient selection for the different cohorts. There was an overlap of nine patients between the study of Wahl *et al.* [113] and Paper III since these samples were used in method validation.

5.2 Ethics

Biological material and clinical data were obtained from the Norwegian Lung Cancer Biobank, a regional biobank organized through Biobank 1. Per 01. January 2019, the biobank included 852 patients with all subtypes and stages of lung cancer.

The biobank was approved by the Central Norway Regional Committee for Medical and Health Research Ethics (approval ID: 2015/356), the Ministry of Health and Care Services, and the Norwegian Data Protection Authority. The research was conducted in accordance with the Helsinki Declaration. Patients above 18 years old with suspected lung cancer referred to St. Olavs Hospital, Trondheim University Hospital are invited to sign a broad consent for participation in the regional biobank. Participation involves an extra blood draw at inclusion and a blood draw at each follow up visit. Any results obtained are solely for research purposes and do not influence the patient's treatment. The consent also allows the use of diagnostic tissue samples and grants access to the patient's medical records.

5.3 Statistical considerations

In this exploratory project, we analyzed a convenient number of samples, limited in number by what was reasonable to analyze given access to lab, resources, and funding. Thus, no further sample size was calculated. However, the size is within the range of similar studies and believed to be sufficient for establishing a pipeline for later, more extensive ctDNA research projects.

PFS was defined as the time from lung cancer diagnosis until progression or death of any cause. OS was defined as the time from diagnosis until death of any cause. The median follow-up times for PFS and OS were estimated using the reverse Kaplan–Meier method, and the median PFS and OS was estimated using the Kaplan-Meier method. Survival analyses were performed using Cox proportional hazard models. The multivariable model was adjusted for sex, age, PS, and disease stage. The threshold for statistical significance was set at 0.05. All statistical analyses were performed using R Studio [114].

5.4 Data collection

Comprehensive clinical data were collected from patients' hospital medical records and validated by trained healthcare personnel. The disease stage was classified according to TNM version 8 [61], and disease progression was defined according to the RECIST criteria [115]. The histopathological classification was evaluated by a trained lung cancer pathologist (Sissel Gyrid Freim Wahl).

5.5 Tumor specimens

Diagnostic tumor tissue and histological sections utilized in this project were stored and handled by the Department of Pathology at St. Olavs Hospital. The material comprised a combination of resected tumors, biopsies obtained through different techniques, and cytological specimens, all retrieved by routine diagnostic procedures. Per routine at the department, tumor tissue was treated with a 4% phosphate-buffered formaldehyde solution before the preparation of FFPE blocks. The pathologist assessed all specimens before inclusion and instructed which FFPE block and frequently which tissue region of the block should be selected for DNA extraction to maximize the tumor cell-ratio. Tissue sections of 10 µm were made using a microtome followed by DNA extraction.

5.6 ^{18}F-FDG PET/CT

^{18}F-FDG PET/CT was not available at our hospital until autumn 2013. Thus, patients underwent ^{18}F-FDG PET/CT at three different hospitals: Haukeland University Hospital, Bergen (n=3), Oslo University Hospital, Oslo (n=9), and St. Olavs Hospital (n=51). All hospitals used scanners from Siemens Healthcare (Erlangen, Germany). The European Association of Nuclear Medicine (EANM) granted an EANM Research Ltd. ^{18}F-FDG PET/CT accreditation in September 2015 for the ^{18}FDG PET/CT scanner at St. Olavs Hospital. The accreditation status for the other centers between 2011-13 is unknown.

Image reconstruction was performed with iterative reconstruction, point-spread-function (PSF), decay-, attenuation-, and scatter-correction. Time-of-flight was used when available. Different matrix sizes were applied at different sites. All examinations were done following the EANM procedure guidelines for tumor imaging version 2.0 [116]. Patients fasted at least four hours (median 14h) before 4 MBq ^{18}F-FDG/kg were administered. Blood glucose level was 4.4-9.5 mmol/L (median 5.6 mmol/L), and the interval between ^{18}F-FDG administration and the start of the acquisition was 51-159 minutes (median 60 minutes). A low-dose CT for attenuation correction and anatomical localization was done in the same session.

Datasets were transferred from the hospital's picture archiving and communication systems and reprocessed using standard clinical software (AW Server 3.2 Ext. 3.0, General Electric Company) by a nuclear medicine physician (Håkon Johansen). The physician was blinded for the ctDNA data but not the previous ^{18}F-FDG PET/CT reports. A 3D isocontour model with a threshold of standardized uptake value (SUV) of 2.5 was used when computing metabolic tumor volume (MTV) and tumor lesion glycolysis (TLG) (=product of MTV and SUVmean). MTV and TLG were calculated manually in separate sessions for each lesion when the ^{18}F-FDG uptake conflated. The highest value of SUVmax in any lesion was

used for each patient. For both MTV and TLG, the sum of all lesions was used for statistical analyses. The raw PET data were unavailable and thus, the original AC-PET reconstructions were used to assess MTV and TLG.

5.7 Plasma samples

Plasma was prepared from a 10 mL blood sample within two hours after the blood draw. Samples included before 2016 were centrifuged once at 2500x g for 10 min. After 2016, samples were centrifuged twice, 1500x g for 15 min and at 10,000x g for 10 min. Plasma was transferred to cryotubes and stored at -80 °C until DNA extraction using QIAamp Circulating Nucleic Acid Kit (Qiagen, Hilden, Germany).

5.8 Next-generation sequencing (NGS)

The specific process of DNA preparation into an NGS library varies between protocols, but the goal is to create short nucleotide fragments (<400 bp) with a genetic sequence of interest flanked by synthetic nucleotide sequences. In targeted NGS, the NGS gene panel defines which genes and regions are selected, and there are two main approaches for this selection. Amplicon-based targeted NGS utilizes polymerase chain reaction (PCR) with target-specific primers, while hybrid-capture-based NGS utilizes target-specific probes which hybridize to DNA and enable extraction of the probe-DNA complex.

The synthetic sequences genomic sequence, known as adapters, include a unique sample index, sequences that are used during the preparation and sequencing, and optionally a unique molecular index (UMI). UMIs are short (~12 bp) random nucleotide sequences added to each original DNA fragment and used in the data analysis to identify duplicate reads with the same UMI [117]. This project involved targeted amplicon NGS using the QIAseq Human Targeted DNA Panels (Qiagen, Hilden, Germany), including UMIs (Figure 11).

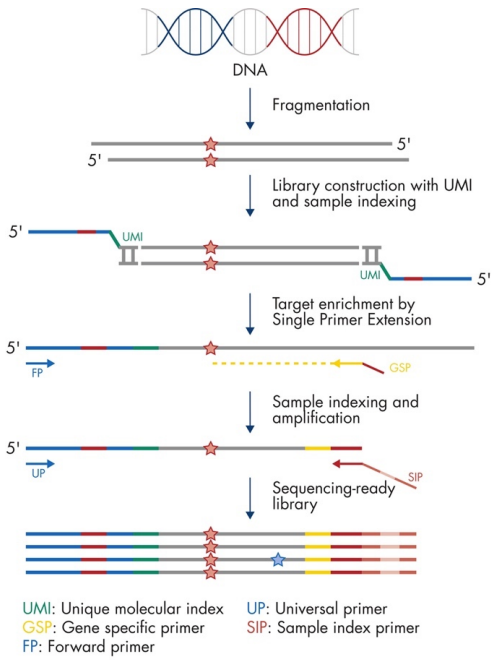


Figure 11: Illustration of the QIAseq Targeted DNA Panels workflow utilized in this project. Source [118].

In Illumina sequencing, which was applied in this project, the NGS libraries prepared from different samples are added to a "flowcell", a flat surface with a lawn of oligonucleotides that bind the NGS library fragments via their adapter. Each fragment then gives rise to millions of identical fragments by PCR to create a cluster that produces a strong enough signal during the sequencing. The nucleotide sequence of each fragment (cluster) is obtained by synthesizing the complementary strands and registering which nucleotide is incorporated based on the fluorescence emitted upon incorporation [119].

The nucleotide sequence generated from the clusters are called "reads" and are sorted into separate files based on the sample index in the adapter sequence. These FASTQ files are the basis for variant detection, and this project utilized the software CLC Genomics Workbench (Qiagen, Hilden, Germany) for this purpose. In brief, all reads with the same UMI (i.e., UMI families) were collapsed into consensus UMI reads since any discrepancies in a UMI family are believed to be technical artifacts. The UMI reads were then aligned to the human reference genome (version 37). The number of raw reads that align to a specific genomic position is known as the "read coverage" of that position. In contrast, the number of UMI families that cover a position is known as the "UMI coverage" or "unique coverage" and reflects the number of hGEs analyzed. Further analysis was restricted to the target

regions defined by the NGS panel. A list of discrepant variants was subjected to quality filtering that finally gave a list of "called" variants.

5.9 Funding

This Ph.D. project was funded by The Liaison Committee for Education, Research, and Innovation in Central Norway (2018/42794) and the Cancer Foundation at St. Olavs Hospital. The funding bodies had no role in the project beyond providing financial support.

6 Summary of results

6.1 Paper I

In total, 30 patients were included, of which most (97 %) had stage I-III disease, 50% had ADC, 43% had SCC and 6% had other NSCLC diagnoses. Tumor DNA, leucocyte DNA, and cfDNA were sequenced using the QIAseq Targeted Human Comprehensive Cancer Panel targeting ~800 kb in 275 genes commonly mutated in cancer. The mean unique coverage within the target region was 743× (225×-1643×) in tumor DNA, 207× (63×-443×) in cfDNA, and 493× (209×-1061×) in leucocyte DNA. Variants were reported by the bioinformatic software in all samples.

Since the main goal was to investigate how many patients had common mutations likely related to their cancer in tumor DNA and cfDNA, a key aspect was to distinguish those true tumor mutations from the myriad reported variants. We considered tumor DNA the gold standard and classified a variant as a "tumor mutation" if it was not found in the matching leucocyte DNA and was confirmed or predicted to be pathogenic. This resulted in 0-12 detected mutations per tumor sample. At least one mutation was detected in 97 % of patients (29/30) compared to a 63-78% detection rate in studies using smaller NGS panels (<60 genes) [120,121].

The tumor mutations were compared to the variants reported in the matching cfDNA samples (Figure 12). In two cases, 5/5 and 9/10 tumor mutations were identified in the matching cfDNA with mean MAFs 8.6% and 5.4%, respectively. In 27/29 patients, no tumor DNA mutations were identified in cfDNA, and, notably, no lung cancer-specific pathogenic mutations were identified in cfDNA alone. We then manually inspected the cfDNA UMI reads and identified at least one matching tumor mutation in 13 patients with a mean MAF of 1.9%. All 13 patients had additional tumor DNA mutations not identified in the matching cfDNA.

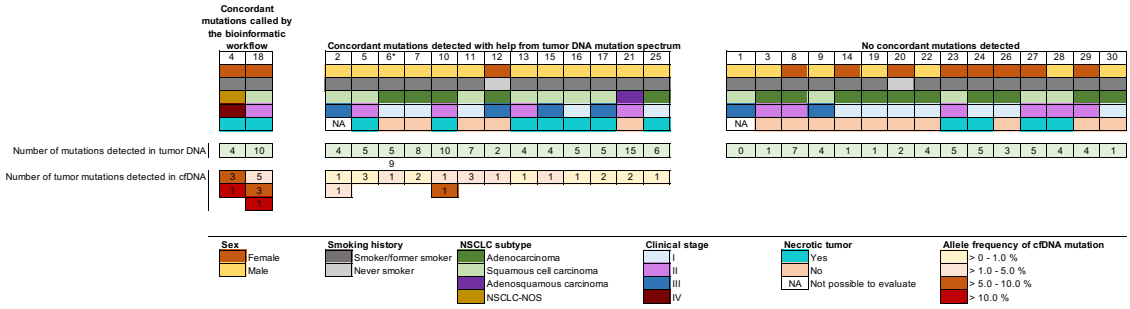


Figure 12: Summary of patient characteristics and concordance between tumor DNA and matched cfDNA sequencing. ctDNA mutations were readily detected in two patients (left) and in additional 13 patients by tumor-guided analysis. No tumor mutations were detected in ctDNA from the remaining 15 patients. Reused with permission from Ottestad *et al*, 2019 [122].

This study demonstrated that applying a commercially available comprehensive NGS panel and bioinformatic workflow was feasible to detect cancer-relevant mutations in tumor DNA and cfDNA from NSCLC patients. This was relevant because it showed that mutation analysis was available without specialized expertise in NGS panel design or bioinformatics. Further, it showed that using a large NGS panel sequencing was advantageous to ensure the detection of at least one tumor DNA mutation but disadvantageous for cfDNA analysis since there is a trade-off between panel size, cost per sample, and analytic sensitivity. Our conclusion from this study was that the most effective is to use a large NGS panel to ensure the detection of tumor DNA mutations, and then apply a focused NGS panel in an analysis of the matching cfDNA to maximize the analytic sensitivity.

6.2 Paper II

The DNA quality was evaluated in 116 DNA samples extracted from diagnostic FFPE tumor tissue from NSCLC patients using a qPCR assay. The number of hGEs with an intact 300 bp region in the gene *FCGR3b* was estimated relative to a sample of high-quality DNA extracted from leucocytes. As expected, most samples (72%) were over ten times more fragmented than leucocyte DNA, meaning that 100 ng of leucocyte DNA and 100 ng of 10-fold fragmented FFPE-derived DNA contain 100 ng and 10 ng "amplifiable" DNA, respectively, which may serve as a template for NGS library preparation.

There was a stronger relationship between amplifiable input DNA amount and NGS library quality than between ng DNA input and NSG library quality. In general, a higher amount of amplifiable DNA gave higher unique coverage (Figure 13) and fewer duplicates per UMI family. Such high-quality libraries are considered more complex and give higher analytic sensitivity since more unique genomes are sequenced.

High-quality DNA samples were expected to give high-quality NGS libraries, and a simple solution is to use more input DNA to increase the amplifiable amount. Since the available patient material largely dicates the DNA amount and quality, we rather aimed to demonstrate the advantage of assessing the DNA quality. The qPCR assay applied in this study did not add considerable cost, time, or sample loss, considering the value gained from the assay. Since there was a relationship between amplifiable DNA and unique coverage, it is possible to omit the quality control if UMIs are used. However, many diagnostic NGS library preparation protocols do not incorporate UMIs, and two samples made from 100 ng and 10 ng amplifiable DNA may seem equally complex. In this setting, input DNA quality control is even more important.

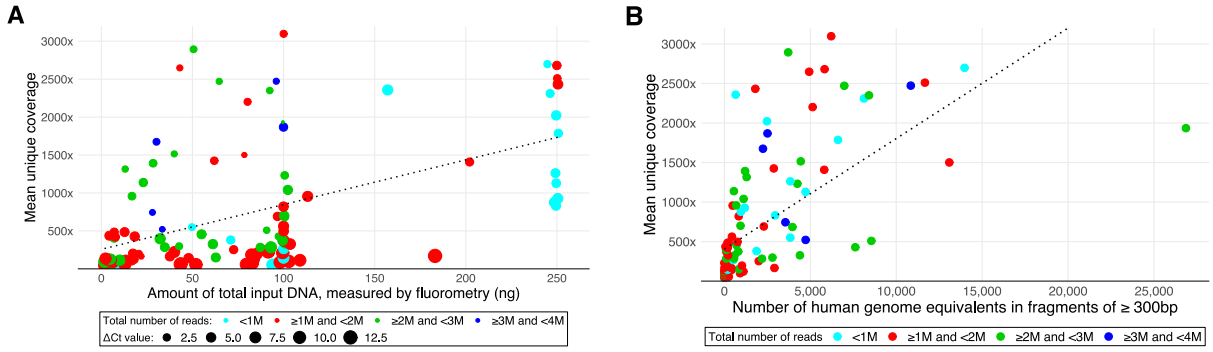


Figure 13: Mean unique coverage for each of the 116 libraries versus **(A)** the number of input DNA measured by fluorometry and **(B)** the amount of amplifiable hGE templates measured by qPCR. All reads passing the quality filters were analyzed, and the color shows the number of reads analyzed in each sample. In figure A, the point size increases with increasing fragmentation degree of the input DNA. Reused with permission from Ottestad *et al*, 2022 [123].

6.3 Paper III

The first aim of this study was to test a customized, tumor-informed NGS approach for sequencing only the region(s) in cfDNA where a mutation was identified in the matching tumor DNA. The approach was tested by constructing four DNA solutions from FFPE-derived DNA and leucocyte DNA with a MAF of *KRAS* G12C mutation between 0.0016 % and 1.15 %. The mutation was detected in the solutions with MAFs 1.15 %, 0.17 %, and 0.016 % (Figure 14A). The results were confirmed by ddPCR, which additionally detected the mutation in the solution with the lowest MAF. Similarly, concordance was observed between the two approaches in 8/9 patient cfDNA samples. In one patient, the mutation was detected by NGS only with a MAF of 1.7% (Figure 14B).

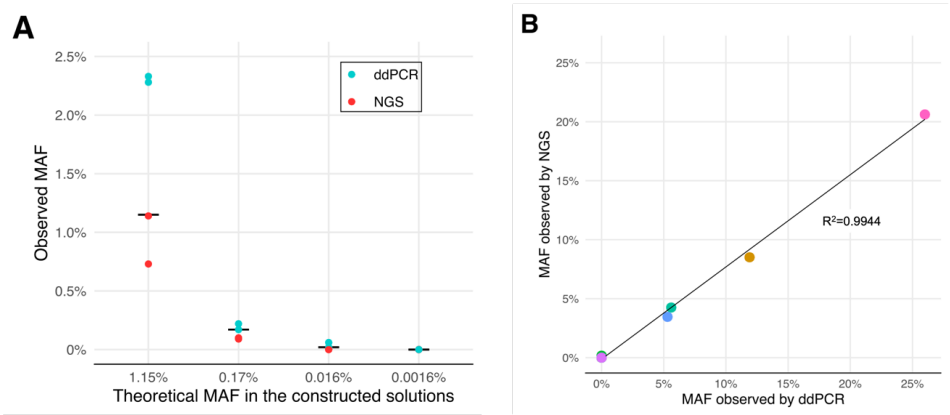


Figure 14: **(A)** Four DNA solutions were constructed to mimic cfDNA with *KRAS* G12C mutation with MAF 1.15%, 0.17%, 0.016%, and 0.0016%. Each solution was analyzed in duplicates. The horizontal line represents the theoretical MAF. The mutation was detected by ddPCR only in the solution with MAF 0.0016%. **(B)** MAF of *KRAS* codon 12 mutations in nine patient cfDNA samples analyzed by both NGS and ddPCR. The mutation was detected in five samples and undetected by both technologies in four samples. Note that the R^2 was incorrect in the published version of the article. cfDNA: circulating cell-free DNA, ddPCR: droplet digital polymerase chain reaction, NGS: next-generation sequencing, MAF: mutant allele frequency. Reused with permission from Ottestad *et al*, 2021 [124].

The customized NGS approach was then applied to cfDNA from 71 of 107 NSCLC patients, from which 1-3 mutations were identified in the matching tumor DNA by QIAseq Human Actionable Solid Tumor Panel (targeting 22 genes). Among the 71 patients, the median age was 68 years (range 48-86), 51% were female, and 88% were current or former smokers. A total of 91% had ADC, 4% had SCC, 1% had adenosquamous carcinoma, and 5% had NSCLC not otherwise specified. A total of 33% had stage I disease, 8% stage II, 30% stage III, and 29% stage IV.

One mutation was targeted in 61 cfDNA samples and 2-3 mutations in ten samples. The median unique coverage in the cfDNA samples was 973× (266×-5955×). Tumor mutation(s) were detected in cfDNA from 29 patients (41%) with median MAF 1.8% (0.05-65.7%). ctDNA was in 4% of stage I patients, 29 % of stage II, 75% of stage III and 55% of stage IV. All tumor mutations were detected in 8/10 patients with >1 targeted mutation.

Next, we investigated if ctDNA detectability was associated with disease development or survival. The median follow-up time for PFS was 88.7 months (95 % CI: 45.2-105.9), and the median PFS was significantly shorter for patients with detectable ctDNA compared to those with undetectable ctDNA (9.6 months vs. 41.3 months, HR: 2.9, 95% CI: 1.6-5.2, p<0.001, see Figure 15A). In the multivariable analysis, higher PS and disease stage, but not detectable ctDNA, were significant negative prognostic factors for PFS. The median follow-up time for OS was 65.6 months (95% CI 45.2-106.0). The median OS was significantly shorter for patients with detectable ctDNA than those without (13.6 months vs. 115.0 months, HR: 4.0, 95% CI: 2.1-7.6, p<0.001, see Figure 15B). The multivariable analysis showed that detectable ctDNA, stage IV, and PS of 2 were significant, negative prognostic factors for OS.

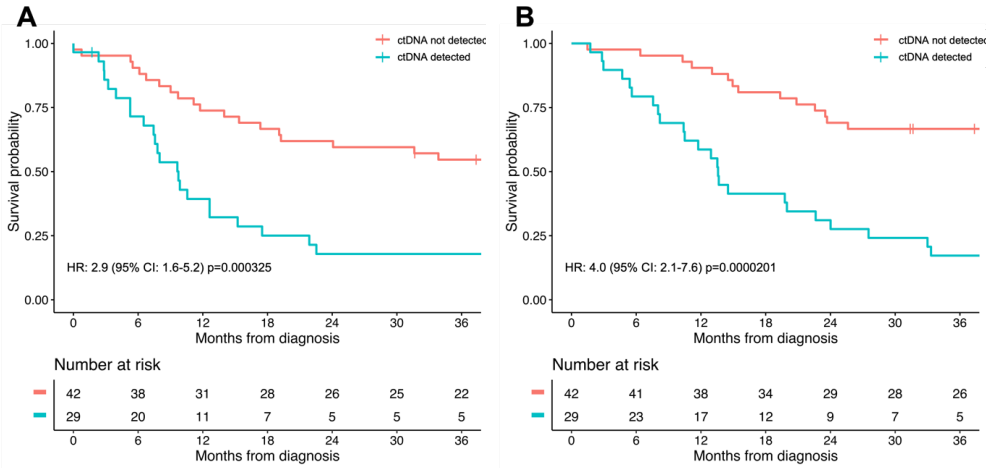


Figure 15: Kaplan-Meier plot for **(A)** progression-free survival and **(B)** overall survival. Note that the color legend was incorrect (opposite) in the published version of the article. CI: confidence interval, ctDNA: circulating tumor DNA, HR: hazard ratio. Reused with permission from Ottestad *et al*, 2021 [124].

6.4 Paper IV

This study included patients with stage I-III NSCLC who had available ^{18}F-FDG PET/CT scans and results from cfDNA analysis from paper III or Wahl *et al.* [113]. A total of 63 patients diagnosed between 2009 and 2018 met these inclusion criteria. The median age was 70, 60% were female and 94% were current or former smokers. Most patients (90%) had ADC, 3% had SCC, 5% had NSCLC not otherwise specified, and 2% had large cell neuroendocrine carcinoma. Twenty-eight patients (44%) had stage I disease, 12 (19%) stage II, and 23 (37%) stage III. ctDNA was detected in 19 patients (30%), of which most (74%) had stage III.

Patients with detectable ctDNA had significantly higher tumor glucose uptake, measured as MTV, TLG, and SUVmax ($p<0.05$). There was a correlation between the ctDNA quantity (highest MAF) and MTV (Spearman's $\rho=0.53$, $p=0.021$) and TLG (Spearman's $\rho=0.53$, $p=0.021$), but not SUVmax (Spearman's $\rho=0.34$, $p=0.15$), see Figure 16. MTV (OR 1.03, 95% CI: 1.01-1.05, $p=0.019$) and TLG (OR 1.00, 95% CI: 1.00-1.01, $p=0.038$), but not SUVmax (OR: 1.07, 95% CI: 0.98-1.20, $p=0.19$), were significantly associated with ctDNA detection independently of disease stage and histology.

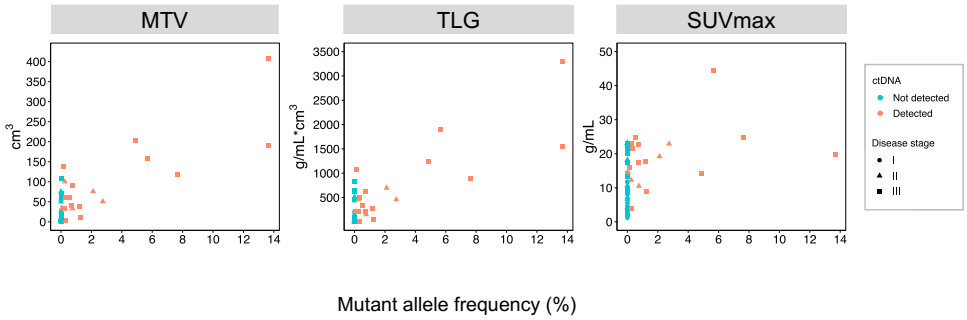


Figure 16: MTV, TLG, SUVmax, and the ctDNA quantity, measured as the highest variant allele frequency (i.e., MAF). TLG: total lesion glycolysis, MTV: metabolic tumor volume. SUVmax: maximum standardized uptake value

The median follow-up time for PFS was 57.0 months (95% CI: 50.7-65.6). ctDNA detection and higher MTV, TLG, and SUVmax were significantly associated with worse PFS in univariable analyses, while none remained independently associated with PFS in multivariable analyses. The median follow-up time for OS was 57.0 months (95% CI: 50.7-64.0). Multivariable analyses indicated that ctDNA detection was associated with worse OS independently of MTV (HR: 2.70, 95% CI: 1.07-6.82, $p=0.035$) and TLG (HR: 2.63, 95% CI: 1.06-6.51, $p=0.036$), but not SUVmax (HR: 2.30, 95% CI: 0.977-5.42, $p=0.056$).

Lastly, the results indicated a combined prognostic value of ctDNA analysis and ^{18}F-FDG PET/CT. Patients with detectable ctDNA and high (above median value) MTV, TLG or SUVmax had shorter PFS and OS (Figure 17) than those without detectable ctDNA, although the differences were not statistically significant.

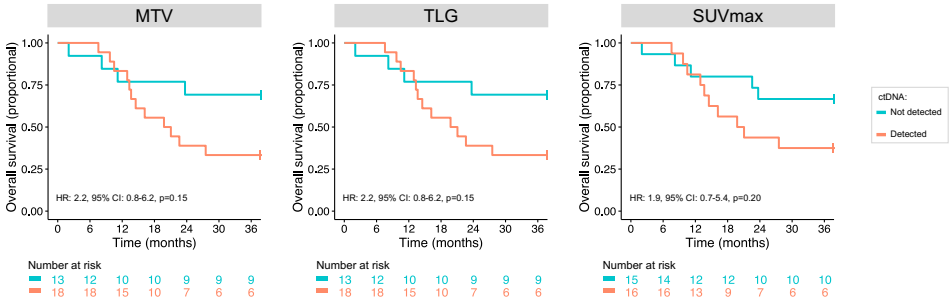


Figure 17: Kaplan-Meier plots showing OS for patients with MTV, TLG and SUVmax above the median value and split on ctDNA status.

7 Discussion

This project explored practical and methodological aspects of implementing ctDNA as a biomarker in translational lung cancer research. In paper I, we experienced the practical and analytical benefit of applying tumor-informed ctDNA analysis, and the approach was further optimized to a customized assay in paper III. Based on the results in paper II, we argue that DNA quality assessment of FFPE-derived DNA optimizes the analytical quality of individual samples. These methodological experiences affect our ongoing and planned research on patient material. Furthermore, paper III indicated that ctDNA detection was a negative prognostic factor, though most patients considered for curative treatment had undetectable ctDNA levels, questioning the cost-effectiveness of this analysis since established prognostic factors remain the most important. Lastly, in paper IV, we explored the relationship between ctDNA detection and tumor metabolic activity, and we believe these preliminary results encourage further research into this association and the biological mechanism of ctDNA release.

7.1 Tumor-informed ctDNA analysis

Tumor-informed interpretation of non-customized ctDNA sequencing data has been applied in other studies in addition to ours [125–127], and we report several benefits of this approach. First, the ctDNA detection rate in paper I increased from 10% to 52%, demonstrating increased detection sensitivity and that a tumor-informed approach can enable the utilization of data initially considered unreliable or uninformative. Tumor DNA genotyping enabled the exclusion of patients with no identifiable tumor mutation, which by a tumor-naïve approach would have resulted in false negative or potentially false positive results in cfDNA analysis. Chabon *et al.* showed that ctDNA was detectable in 10/17 initially negative cfDNA samples using tumor-naïve CAPP-seq when the number of targeted mutations was increased [125]. This underscores the difficulty of designing one optimal panel for all samples and supports the use of tumor-informed analysis.

Secondly, the approach was time-saving in variant interpretation due to the high number of variants generated per sample. These properties are particularly beneficial for data obtained from large panel sequencing or NGS data from low quality or quantity cfDNA. Suboptimal sample material is a frequent issue in clinical research, and quality improvement is often outside the researchers' control, particularly in prospective studies.

In contrast to broad panel NGS, customized NGS panels generate mostly clinically relevant data and potentially less data in total, thus requiring less time and computational resources in analysis. Furthermore, we showed in paper III that using a customized NGS approach achieved lower LOD at a lower cost per sample than when large a NGS panel

was used (mean unique coverage 1523× vs. 207×, at ~¼ of the cost). Customized NGS has been successfully applied in several other studies [106,126,128,129]. Similar to our study, McDonald *et al.* used ddPCR to validate the MAF reported by the customized NGS assay [130]. In contrast to our panels which targeted a median 1 mutation, the median number of targets in other studies ranged from 18 to 200 mutations or 10^2-10^4 different loci, likely contributing to the higher ctDNA detection rates reported than our study, particularly in stage I-II patients (48-66% vs. 10% of our stage I/II patients) [106,128]. Several studies reported the benefit of targeting many mutations [106,126,129,130], and one study reported that targeting more than 50 mutations improved the sensitivity of samples with low cfDNA input, i.e., less than 20 ng [128]. The median cfDNA input in our study was 21 ng. Although increasing the number of mutations increases the likelihood of detecting at least one mutation in ctDNA, it does not change the likelihood of detecting one specific mutation, such as a targetable alteration. If anything, this likelihood is reduced at the expense of a broader NGS panel. Compared with the tumor-informed approach in paper I, the turnaround time is longer because tumor DNA and cfDNA are sequenced sequentially due to the need for individual NGS panel design. Which tumor-informed approach is most appropriate depends largely on the quantity of cfDNA, required LOD (often dependent on the clinical setting), maximum cost per sample, and available resources for data analysis.

Other studies have applied tumor-naïve ctDNA analysis, commonly attempting to develop a tool for early cancer detection or detecting minimal residual disease [109–111,125,131]. Studies that applied CAPP-seq or TEC-seq successfully detected low MAF mutations in ctDNA confirmed in the matching tumor DNA, in addition to mutations detected only in ctDNA [111,132]. When TEC-seq was applied, 18% of patients had ctDNA mutations that could not be confirmed in the patients' tumor DNA, and the results indicated that tumor heterogeneity could be one cause of discordant mutations [111]. ctDNA was detected in 45-83% of NSCLC patients, depending on the disease stage, which is similar to what is reported in tumor-informed studies [128]. In contrast, another study detected ctDNA mutations in 93% of stage I-III NSCLC patients, a higher rate than in any other study, and did not include matching tumor DNA for confirmation [110]. A recent study using the CAPP-seq panel demonstrated that most ctDNA mutations in NSCLC patients were patient-specific CHIP mutations [125], and the authors further developed a machine learning model using tumor-naïve CAPP-seq in combination with analysis of cfDNA fragment length and mutation profile to identify true ctDNA mutations. This approach showed similar sensitivity to tumor-informed ctDNA analyses with ctDNA detection rate of ~30-90% in stage I-IV NSCLC. Another recent study combined tumor-naïve cfDNA sequencing with protein analysis for early cancer detection and identification of the organ

or origin [131]. This "cancerSEEK" method correctly classified ~60% of stage I-III NSCLC patients.

Since these approaches are independent of tumor DNA analysis, the turnaround time is shorter, and the analyses apply to all NSCLC patients in almost any setting since the only requirement is a blood draw. However, the risk of false positive results seems high without additional supporting information. Although CHIP mutations may be removed by equally deep sequencing of white blood cells, sophisticated bioinformatic software is still needed to remove technical artifacts. A study comparing tumor-informed and tumor-naïve approaches in the same cfDNA samples demonstrated inferior sensitivity of tumor-naïve approaches in samples with low ctDNA fraction, which usually was the case in samples from early-stage patients [133].

It is well established that FFPE-derived DNA is fragmented in varying degree and contain unpredictable amounts of amplifiable DNA [134,135]. Previous studies have applied various methods, including qPCR used in our study, to assess the quality of FFPE-derived DNA destined for NGS (reviewed by [136]). The main result from this study was the positive association between the amount of amplifiable input DNA and the NGS quality, measured by the mean unique coverage. In contrast, one study concluded that high read coverage in NGS could be achieved regardless of whether input DNA was determined based on methods that estimate the amplifiable DNA amount or measure DNA in nanograms [137]. However, we believe that read coverage is not a suitable NGS quality metric since two samples can have comparable mean read coverage while considerably different mean *unique* coverage and, thereby, different LOD in variant detection, as demonstrated in our study. A recent study confirmed our results, and few other studies have described the association between DNA input and unique coverage, possibly because the use of UMI is a relatively new invention, and many protocols still do not utilize it [138].

NGS library protocols more frequently recommend evaluating the amount of amplifiable DNA, possibly since NGS is more often used in routine clinical practice. Large companies offering NGS panels, such as Illumina and Qiagen, offer accompanying qPCR quality assessment assays and recommend excluding low-quality samples from NGS. Rather than excluding samples, we argue that quality assessment enables individual sample adjustment. McNulty *et al.* used unfragmented DNA to systematically demonstrate that adding more input DNA generates higher mean unique coverage [138]. Our study confirms this and also demonstrates that the amount of *amplifiable* DNA is more important than the amount of DNA in nanograms when adjusting the input amount from FFPE tissue. As previously mentioned, tumor samples are often small, and increasing the amount of input DNA is not always possible. Nevertheless, small samples might still provide sufficient hGEs for detecting mutations with confidence if the DNA quality or tumor cell ratio is high.

However, a negative result when analyzing small samples is more difficult to interpret, and collecting a re-biopsy is optimal but not always feasible.

The risk of false negative and false positive results may be reduced by such quality assessment through a more accurate estimation of the individual sample LOD. The results further suggest that it may be sufficient to estimate the LOD by either applying quality assessment of input DNA or incorporating UMIs. Several NGS panels used in diagnostic laboratories do not include UMIs, probably because it requires additional, time-consuming data processing. Routinely implementing a qPCR assay might be faster and cheaper than establishing a different NGS panel and analysis pipeline. That said, UMIs also enable error correction, which is an advantage in detecting low MAF variants, or identifying artifacts in poor-quality samples where such variants may arise above the MAF threshold. The combination of quality assessment of crude DNA and incorporation of UMIs enables individual sample adjustment of input DNA, technical error correction, and accurate estimation of sample LOD.

Some conditions of FFPE tissue preparation and DNA extractions may be optimized to reduce the extent of DNA fragmentation [139]. It is essential to limit the time until fixation and incubation time as much as possible since the DNA damage occurs in a time-dependent manner in these steps. Large specimens should be cut into smaller pieces before fixation, and the incubation time should be adjusted according to the tissue size since small biopsies are more vulnerable to over-fixation. Optimal fixation temperature, formalin concentration, and including a heating step during DNA extraction also positively affect the DNA quality. Other factors, such as increased perioperative ischemic time and storage time, also decrease the DNA quality but may be more challenging to control. An alternative to FFPE tissue is to obtain fresh frozen tissue since it is not exposed to formalin. However, this material is not optimal for detailed histopathological assessment or IHC. Thus, FFPE biopsies still have to be collected and currently, it may be practically challenging to acquire an additional biopsy for freezing. Additionally, the valuable combination of the pathologist's assessment and mutation profile will be lost if the second biopsy is obtained from another tumor region. Furthermore, storing fresh frozen tissue requires considerably more space, electricity, and safety measures due to the severe consequences of a power outage.

ctDNA is also fragmented, and the bulk of cfDNA is shorter than most FFPE-derived DNA samples [102]. A notable difference is that cfDNA fragmentation occurs *in vivo* and is difficult to prevent. Individual quality assessment of cfDNA samples is not equally relevant because the quality is relatively even among samples. Further, the input amount of cfDNA for NGS library preparation would likely be maximized in most cases regardless of the quality due to limited cfDNA available. Lastly, most cfDNA NGS protocols include UMIs. Importantly, there has been much development in NGS library preparation

chemistry which has improved the preparation from short fragments from FFPE-derived DNA and cfDNA.

7.2 Pre-treatment ctDNA detection as a prognostic factor in NSCLC

Patients with detectable ctDNA in paper III (41%) had significantly shorter OS than patients without detectable ctDNA, which is in line with the findings in other studies on early-stage NSCLC [113,128,140] and advanced NSCLC [141] (Table 2). A recent publication from the TRACERx study, including 197 operable NSCLC patients, showed that ctDNA detection was a negative prognostic factor for both PFS and OS in ADC but not in non-ADC patients [106]. In ADC patients, ctDNA detection was associated with clinically occult lymph node disease and the development of extrathoracic metastasis. Similarly, another large study (n=330) on early-stage NSCLC found that ctDNA detection was a negative prognostic factor in ADCs only [121]. The authors suggest that ctDNA release reflects micrometastatic disease in ADC and may be a more passive result of necrosis in SCC. Other studies have confirmed that the ctDNA detection rate is higher in SCC than ADC, which could be explained by SCC generally having a higher number of mutations, a higher rate of clonal mutations, or because the mutation profile and, thus, the molecular biology differ [106,142]. Abbosh *et al.* did not find a difference in tumor purity or clonality between ADC shedders and non-shedders [128].

Two studies on early-stage NSCLC found no significant association between pre-treatment ctDNA detection and OS, but there were substantial methodological differences. The first one was the abovementioned study which detected ctDNA in 95% of stage I-III patients using a tumor-naïve approach [99]. It is reasonable to assume that the detection rate reflects patient selection and methodology differences, which might influence the overall results. The second study investigated the ctDNA *level*, rather than ctDNA detection as a categorical factor, and found no significant association with OS [125]. It is possible that the difference in tumor behavior between non-shedders and shedders represents a categorical shift in biology, and that there is another mechanism which explains the different ctDNA level in the blood.

ctDNA detection alone was associated with shorter PFS in our study. A study on advanced NSCLC patients did not find any significant association with PFS [143], in contrast to studies on early-stage NSCLC [113,125,127,128,140], including two prospective observational studies. These studies analyzed the largest cohorts and were designed to study minimal residual disease and, thus, involved systematic evaluation of treatment relapse. The PFS calculation may be influenced by differences in CT intervals,

definition of relapse, cancer treatment, and since the association was not independent of other patient characteristics in our cohort, possibly by differences in the patient selection.

It is complicated to compare studies investigating associations between ctDNA detection and treatment outcomes since patient cohorts are not directly comparable, for example, concerning disease stage distribution and histological subgroups. Although these characteristics are commonly included in multivariable analyses, these analyses only answer whether ctDNA detection is a prognostic factor *independently* of the characteristics. Subgroup analyses are needed to explore the role of ctDNA analysis within specific patient groups, as was indicated for ADC [127,128], and such studies require rather large cohorts. Another issue is the variation in the ctDNA detection approach (tumor-informed vs. tumor-naïve, number of mutations included), assay LOD and individual sample LOD, which is not readily comparable, and many studies do not clearly describe these metrics. It does not matter if an assay can detect one mutated genome in a million if only 1000 genomes are sequenced. Chabon *et al.* showed that patients with undetectable ctDNA had higher sample LOD than those with detectable ctDNA, likely not unique to this study [125].

Table 2: Comparison of ctDNA detection rate, assay LOD and survival analysis in studies comparable to Paper II.

Author, year	Disease stage	ctDNA approach	No. of patients	Histology	ctDNA detected (%)	Assay LOD + lowest reported MAF	PFS (multivariable Cox)	OS (multivariable Cox)
Chaudhuri, 2017 [110]	I-III	Tumor-naïve, CAPP-seq	40	48% ADC	I-III: 93%	Median MAF 0.62%	Not analyzed	ctDNA concentration (continuous): HR 1.00, 95 % CI (1.00-1.01), p=0.08
Zulato, 2020 [143]	III-IV	Tumor-informed, ddPCR *KRAS*	58	ADC, *KRAS* mutated	III-IV: 48%	LOD: 0.2% Lowest MAF:0.46%	ctDNA detected: HR 0.9, 95 % CI (0.5-1.6), p=0.73	ctDNA detected: HR 2.1, 95 % CI (0.9-4.7), p=0.075
Peng, 2020 [140]	I-IV (operated)	Tumor informed, NGS 127 genes	77	52% ADC	I-II: 33% III-IV: 75%	Not stated	ctDNA detected: HR 3.41, 95 % CI (1.36–8.51), p=**0.009**	ctDNA detected: HR 3.78, 95 % CI (1.54-9.30), p=**0.004**
Chabon, 2020 [125]	I-III	Tumor-informed, CAPP-seq (255 genes)	85	75% ADC	I: 42% II: 67% III: 88%	Lowest MAF: 0.00029%	ctDNA MAF, log10 (continuous): HR 2.15 95% CI (1.17-3.96), p=**0.014**	ctDNA MAF, log10 (continuous): HR 9.91 95% CI (0.87-4.19), p=0.106
Wahl, 2021 [113]	I-IV	Tumor informed, ddPCR *KRAS*	60	ADC, *KRAS* mutated	I-II: 14% III-IV: 72%	LOD: 0-0.69 copies/mL. Lowest reported: 1.1 copies/mL	ctDNA detected: HR 2.76, 95% CI (1.064–7.162), p = **0.037**	ctDNA detected: HR 3.609, 95% CI 1.261–10.328, p = **0.017**

Ottestad, 2021 [124]	I-IV	Tumor-informed, customized NGS targeting 1-3 mutations	71	90% ADC	I-II: 10% III-IV: 65%	LOD: 0.016% Lowest MAF: 0.05%	ctDNA detected: HR 1.58, 95 % CI (0.75-3.32), p=0.23	ctDNA detected: HR 2.49, 95 % CI (1.08-5.74), p=**0.033**
Jee, 2022 [141]	IV	Tumor-informed and tumor-naïve. NGS 22/129 genes	457	86% ADC	IV: 65%	Lowest MAF: <0.1%	Not analyzed	ctDNA detected: HR 1.90, 95 % CI (1.47, 2.46). p-value not stated, but likely <0.05
Xia, 2022 [127]	I-III	Tumor-informed, NGS 769 genes	280	ADC	I-III: 12%	Not stated	ctDNA detected: HR 6.0, 95% CI, 3.4–10.6, p **<0.001**	Not analyzed
			50	Non-ADC	I-III: 75%		ctDNA detected: HR 2.4, 95% CI, 0.5–10.7, p=0.241	Not analyzed
Abbosh, 2023 [128]	I-III (operated)	Tumor informed, customized NGS targeting 70-201 mutations	93	ADC	I-III: 42%	LOD: 0.01% with 90% sensitivity with >20 ng cfDNA	ctDNA <u>not</u> detected: HR 0.13, 95% CI (0.04-0.39), p<**0.001**	ctDNA <u>not</u> detected: HR 0.15, 95% CI (0.05-0.48), p=**0.001**
			85	Non-ADC	I-III: 92%		ctDNA <u>not</u> detected: HR 0.81, 95% CI (0.15-4.45), p=0.805	ctDNA <u>not</u> detected: HR 0.78, 95% CI (0.19-3.12), p=0.723

In agreement with our results, others have shown a correlation between ctDNA level and MTV/TLG in patients with early-stage [110,125] and advanced stage NSCLC [141,144,145]. One study additionally confirmed our finding that MTV predicted ctDNA detection independently of the disease stage in stage I-III NSCLC (n=85), indicating that MTV provides more information than only tumor volume [125]. ctDNA level did not correlate with SUVmax in our and another study [144], while other studies reported a correlation [141,145]. Studies that correlated the total cfDNA level and ^{18}F-FDG PET/CT-derived parameters reported conflicting results, possibly since estimating the total cfDNA level is not optimal for estimating the ctDNA level [140–143].

There are important variations between these studies, with differences in sample size (n=40-128), distribution of disease stage and histological subtypes, whether patients were treatment-naïve or not, type of treatment administered, number of included ^{18}F-FDG PET/CT-derived parameters, ctDNA detection approach, and assay, and whether the association between ^{18}F-FDG PET/CT and ctDNA analysis was the primary aim of the study. One of the most comprehensive studies investigated the association between ctDNA level and MTV, TLG, SUVmax, CT volume, and somatic mutations in 69 patients with advanced NSCLC [145]. All three ^{18}F-FDG PET/CT derived parameters, CT volume, and mutations in *TP53* and *EGFR* were associated with the ctDNA level. Notably, the CT volume was comparable to the MTV, and was associated with ctDNA detection independently of *TP53* and *EGFR* mutations but not *KRAS* mutations, further supporting that there may be several tumor characteristics that explain ctDNA release. The cohort included advanced stage patients who were actively or previously treated, which might not be optimal. ctDNA was detected in 95% of patients, as expected for metastatic cancer. However, the results are not necessarily representative for early-stage patients, which usually comprise a larger proportion of non-shedders.

Chabon *et al.* reported that ctDNA detection was a negative prognostic factor for OS independently of MTV, supporting our results [125]. We also found that ctDNA was prognostic independently of TLG but not SUVmax. Further, we explored the combined prognostic value of ctDNA and ^{18}F-FDG PET/CT, as shown in previous studies on metastatic cancer, including NSCLC [141,149,150]. The results indicated that ctDNA might discriminate between patients with high MTV/TLG/SUVmax, thus enabling further risk stratification, though the results were not statistically significant. However, in support of these results, Jee *et al.* found that ctDNA detection was prognostic for patients with high and low MTV, although this study only included patients with stage IV NSCLC [141].

Elevated glucose uptake may indicate an elevated metabolic activity or a shift towards anaerobic metabolism, either due to the Warburg effect or hypoxic conditions. In contrast, a novel study measured glucose uptake and citric acid cycle flux in mouse solid tumors, including lung tumors [151]. Although the study confirmed higher glucose uptake

in tumor tissue than healthy tissue, the energy production in the tumors was slower, and ATP was mostly generated oxidatively. The Warburg effect assumes that mitochondria are non-functional. Results from kidney and pancreatic tumors indicated that tumor cells downregulated energy-demanding tissue-specific processes to sustain proliferation despite slower ATP production. Without complementary metabolic measurements to ^{18}F-FDG PET/CT, we cannot explain the metabolic state of the tumors in our study, and it is challenging to explain a potential relationship between glucose uptake and ctDNA release.

7.3 Methodological considerations

Results were not validated by others or in a different laboratory. On the other hand, the preparation of tissue material and tumor DNA extraction was performed according to the routines at the Pathology department at St. Olavs Hospital, which is validated directly and indirectly by either an internal control system or by participating in occasional external quality assurance programs. ctDNA analyses can be challenging to validate externally due to the limited sample material. However, a selection of NGS libraries from both tumor DNA and cfDNA was re-sequenced, tumor DNA was sequenced using different NGS panels, and cfDNA samples were validated using ddPCR. These internal validations showed consistent results. There is a lack of standardized pre-analytical factors, such as protocols for plasma preparation and cfDNA extraction. Although updated recommendations were recently published, many studies include stored samples that may not have been optimally processed [112]. Thus, the abovementioned limitations are not unique to our project, and there is no reason to believe that the biological samples included in this project differ considerably from those analyzed in other clinical laboratories or research biobanks. External validation of ctDNA assays is not common. However, recent studies performed applied assays approved by the Food and Drug Administration (FDA), although these mainly apply to advanced stage patients [141].

The low number of mutations targeted by the customized NGS panels was a limitation compared to similar studies. Matching tumor DNA from these patients was initially sequenced as part of another project in our group which screened patients for *KRAS* mutations. Consequently, few mutations were identified per patient, and about 1/3 of the included patients had no identifiable mutation.

Although all stages and the most common NSCLC subtypes were represented in our cohorts, the findings regarding ctDNA as a prognostic factor may not represent the whole NSCLC population. A patient cohort from a biobank does not necessarily represent the general patient population since immediate study inclusion is not a priority for those who need immediate treatment or otherwise are in very poor condition. Our studies required a blood draw before treatment, which may have caused a selection bias. Further,

our cohorts included a higher percentage of ADC than the Norwegian NSCLC population, and the cohort in paper IV consisted of a high proportion of *KRAS* mutated ADC (56% compared to ~38% in the Norwegian ADC population [52]). Lastly, there was a selection of patients by excluding those without a tumor mutation identifiable by the applied NGS panel, which targeted genes with potentially actionable alterations.

ctDNA was detected in few of the patients with stage I and II disease and none of the stage I patients included in paper IV. Based on previous studies, the detection rates in these stages were lower than expected and was likely attributable to the lower number of mutations targeted in our study. ctDNA detection as a prognostic factor independent of disease stage based on this project alone should therefore be interpreted with caution.

The ^{18}F-FDG PET/CT scans were performed at different sites using different protocols, though most scans were performed at St. Olavs Hospital using the same scanner and protocol. The values of the ^{18}F-FDG PET/CT-derived parameters are, in principle, dependent on the characteristics of the PET/CT camera, reconstruction parameters, matrix size, and point-spread function, which could cause a batch effect. Although raw data was unavailable from the other hospitals and different parameters may have influenced the calculated values, it has likely not affected the overall result.

Patients included in this project were diagnosed between 2007 and 2018, and the cancer treatment has changed in this period. Additionally, we cannot exclude that the routines for tissue sampling and preparation, disease staging, and follow-up have changed. However, all patients were re-staged according to the TNM version 8, although some patients may have been treated according to version 7.

All processes from sample inclusion and evaluation, laboratory experiments, clinical data collection, data analysis, and statistical analyses were performed by individuals in our research group specifically for our projects. An experienced lung pathologist assessed the tumor tissue to confirm the morphologic classification, evaluate the tumor cell content and ensure DNA extraction from a tumor-rich region. The same nuclear physician evaluated all the ^{18}F-FDG PET/CT scans. Comprehensive clinical data was collected directly from patients' medical records by trained healthcare/research personnel, and updated survival data was provided for each study. An accurate date of death is reported to the National Population Register (Folkeregisteret) in Norway. The median follow-up time for PFS and OS was long in study III (89 months and 66 months, respectively) and in study IV (57 months for both PFS and OS). Most NSCLC relapses occur within 24-30 months after primary therapy.

Our analyses were performed in the facilities of the Pathology department at St. Olavs hospital, and the department's routines and protocols were followed in procedures that overlap with their work. This close collaboration ensures that all procedures are, in theory, possible to include in routine diagnostic work.

7.4 Challenges in the implementation of ctDNA analysis

During this Ph.D. project, it has become clear that several important elements must be in place to implement ctDNA analyses in routine clinical practice and to facilitate further ctDNA research. It is of utmost importance that biopsies are as large as possible and optimally treated to minimize the DNA damage. Blood should be drawn before any treatment commences, and the samples need to be processed within two hours after it is drawn. Hospitals without a diagnostic laboratory should probably draw blood in tubes with reagents that prevent cell lysis for transport. Tumor DNA should be sequenced using a larger NGS panel than those currently applied in routine clinical diagnostics for actionable target identification. Standard operation procedures for sample processing, cfDNA extraction, and library preparation must be developed and validated to confirm a low risk of contamination between samples. The diagnostic laboratory should have a designated clean area for cfDNA handling separate from work involving tumor DNA. General recommendations for the implementation of ctDNA analyses were recently published by the European Society for Medical Oncology [112].

If the customized NGS approach described in paper III would be implemented, we imagine that the same gene-specific primers which are included in the tumor NGS panel should be available in separate tubes. Once tumor mutation(s) are identified, the customized cfDNA NGS panel is ready to be assembled by selecting the appropriate primers. Importantly, the NGS library protocol must include UMIs or another approach for error correction and identifying unique hGEs. When results are reported to the clinicians, the sample LOD (unique coverage at the position(s) of interest) should be included, especially for samples with no detected mutation. In case the ctDNA analysis is repeated later in the treatment course, any change in MAF should be interpreted with caution since NGS is not optimal for accurate quantification, and non-tumor related factors could affect the MAF.

A tumor-informed approach prolongs the time from sampling to results are reported to clinicians. However, the data analysis and interpretation are relatively fast due to the limited number of reads per sample, and the interpretation is limited to 1-3 genomic positions. Importantly, each sample generates almost exclusively clinically relevant data. Studies indicate a clinical role of ctDNA analysis in other cancer types, such as colon cancer [152]. The customized NGS approach is tumor agnostic, an advantage in the laboratory logistics since there are likely few eligible patients with the same diagnosis per month. Due to technical limitations in NGS and the short fragment size of ctDNA, the approach is most suitable for cancer types where ctDNA can be identified based on alterations including few nucleotides. Other alterations, such as copy number variations, large chromosomal aberrations, and gene fusions, are still more challenging to identify.

This ctDNA analysis approach is currently unsuitable in situations where a decision must be made fast, the patient is unfit for a biopsy, or tumor DNA is unavailable for NGS. Arguably, it is possible to apply an NGS panel targeting the most characteristic lung cancer alterations, such as hotspots in *EGFR* and *KRAS*. A cfDNA mutation in these genes would likely represent a true positive, especially if it can be supported by other clinical findings or patient characteristics, such as the patient's sex or smoking history. However, samples with no detected mutations will be much less non-informative, and most SCCs and 25-50% of ADCs harbor no alteration typically found in lung cancer. Also, considering that many patients have undetectable plasma ctDNA levels, most samples would be non-informative by this approach. Again, the optimal approach depends on the clinical setting.

Therefore, it would be advantageous if we at least could predict which patients are most likely to have detectable ctDNA, for example, by assessing the tumor metabolic activity by ^{18}F-FDG PET/CT. High MTV and TLG, which both incorporate the tumor volume, predicted ctDNA detection in ours and other studies. However, both estimation of SUVmax (used to calculate MTV) and volume definition may depend on biological factors and external factors such as the ^{18}F-FDG PET/CT protocol and the interpreting physician, and one should be careful with comparing exact values obtained at different sites. Other ^{18}F-FDG PET/CT-derived parameters may be more appropriate. For example, MTV was calculated as the combined volume of all tumors and would not have discriminated between a patient with a large primary tumor and lymph node metastasis and a patient with a small primary tumor and many lymph node metastases. Although this pattern is reflected, to a certain extent, by the TNM classification, ctDNA detectability might more accurately be predicted by combining the level of metabolic tumor activity, volume, and the metastatic pattern. One emerging approach is radiomics, which aims to describe image patterns more accurately or reproducibly than people. In general, the ctDNA detection rate increases with higher tumor volume, higher disease stage (potentially explained by increasing volume) and the rate is higher among SCCs than ADCs. Although these characteristics may help us predict whether a patient likely has detectable ctDNA, it is not given that the prognostic value of ctDNA analysis is higher in these patients.

Furthermore, detecting ctDNA in all patients is the ultimate goal in early cancer detection and tumor genotyping without access to tumor tissue. In contrast, this may not be an appropriate goal for using ctDNA as a prognostic or predictive biomarker. If "negative" results can be supported by other clinical findings or liquid biopsy analyses, samples with and without detectable ctDNA may provide equally valid information. Importantly, non-informative samples due to undetectable mutations must be interpreted in the context of the assay LOD and the sample LOD. Such results could be more useful if a clinically relevant ctDNA fraction is defined. For example, let us say that a fraction of ≥0.1% ctDNA has the most prognostic value, and identification of one mutated hGE is

sufficient for classifying a sample as positive. In that case, ≥1,000 hGE must be analyzed to detect a mutation at 0.1% MAF. Thus, samples with no detected mutation and LOD <1000 hGEs are non-informative, while those with no detected mutation and LOD >>1000 hGEs are more likely true negative. One complicating factor is that the total cfDNA level can rise in the blood due to non-tumor-related processes and falsely lead to the impression of a low sample LOD. Rolfo *et al.* recently suggested an alternative approach, in which an algorithm was used to estimate the ctDNA fraction based on several characteristics, such as aneuploidy [153]. The authors showed that the negative predictive value was lowest in samples with a ctDNA fraction < 1%.

Overall, patients with detectable ctDNA seem to have around a 2-fold risk of dying at any given time compared to patients without detectable ctDNA. This hazard ratio is informative on a group level but cannot be used to estimate the prognosis for individual patients accurately. Although 50 % of patients with detectable ctDNA died within 13.6 months in the cohort of paper III, this survival time was likely also affected by other cohort characteristics. Nevertheless, a positive sample may indicate that the patient should be followed more closely after primary therapy since other studies have supported that ctDNA detection is associated with shorter PFS. Furthermore, we generally observed that the ctDNA detection rate and MAF increased with increasing disease stage. Therefore, reevaluating or complementing the disease stage may be appropriate if a ctDNA mutation is detected at a suspiciously high MAF considering the assigned stage. ctDNA analyses provide complementary prognostic information, but other established factors are more important in clinical practice. It is especially difficult to see that the TNM staging system will become redundant in the foreseeable future.

One might imagine that the prognostic value of ctDNA detection in *EGFR*-mutated patients depends on whether EGFR TKI is administered. Theoretically, ctDNA might be easier to detect in patients with homogenous *EGFR*-mutated tumors, which might be associated with a more uniform, global response to TKIs. However, studies show that ctDNA detection is a negative prognostic factor both in patients who receive TKIs and those who do not [154,155], indicating that detectable ctDNA reflects more aggressive tumor biology rather than tumor clone purity, as suggested by Abbosh *et al.* [128]. Again, we need to understand better why some tumors shed more ctDNA into the bloodstream than others to fully utilize ctDNA analyses in clinical practice and research.

In our study, the median number of hGEs analyzed per sample was 973 (range 266-5955), resulting in a median LOD at 0.1% MAF. Even if the issue with DNA loss in NGS preparation was eliminated, the median input cfDNA amount was 6482, i.e., an absolute median LOD of 0.02%. The obtained amount of cfDNA dictated the available input amount, and our samples are likely representative. Ultimately, the amount of ctDNA obtained from a blood sample limits ctDNA detection. Paradoxically, ultrasensitive assays

that promise detection of mutations at <0.01% MAF require an input amount of hGEs that might mainly be achievable in samples from patients with advanced disease, which rarely require such low assay LOD for detection.

8 Conclusion

This project demonstrated the benefit of performing tumor-informed ctDNA interpretation and that ctDNA analyses were more sensitive and cheaper when customized NGS panels were applied. Quality assessment of FFPE DNA provided useful information on LOD in NGS data interpretation. It may contribute to fewer false positive and false negative results, especially in research on archival patient material.

In our cohort of NSCLC patients, ctDNA detection was a negative prognostic factor for OS and PFS, though the latter was not independent of established prognostic factors. The benefit of ctDNA analysis was limited to patients with detectable ctDNA, which was only a minority of those considered for potentially curative treatment. The low cost-benefit and the requirement for laboratory infrastructure and logistics do not support clinical implementation of pre-treatment ctDNA analysis based on this project, especially since it is unclear what interventions should be done based on the results. The investigation of a combined prognostic value of ctDNA detection and assessment of tumor metabolic activity by ^{18}F-FDG PET/CT was not conclusive in this project but encouraged further investigation since a such a benefit was not excluded.

9 Future perspectives on the use of ctDNA in lung cancer research

Over the past decade, liquid biopsy by ctDNA analysis has transformed from a dream into an available analysis for the identification of targetable alterations in advanced stage lung cancer patients. Technical challenges must now be overcome to expand the utility to early-stage patients and other clinical settings. Better NGS protocols, e.g., with less DNA loss and exploitation of both DNA strands, are continuously developed. The NGS technologies will likely improve and become less error-prone, we have likely only seen the beginning of PCR-free sequencing technologies, and yet unthought-of ultrasensitive technologies may emerge. Although current technical challenges may be solved, we must acknowledge the absolute limitation in the amount of ctDNA per mL of plasma obtained from the patients.

Exploiting other ctDNA characteristics, such as dynamic changes in quantity over time, or inherent characteristics, such as fragment length and epigenetic pattern, may improve the utility of ctDNA analyses. Furthermore, performing complementary analysis of another liquid biopsy component might improve the sensitivity and specificity. However, the challenge remains to develop feasible methods to implement in the routine laboratory. One aspect to remember is the right to equal access to health care. Thus, we should avoid developing analyses that require specialized equipment and personnel which may in practice mainly be available for patients living close to big hospitals or with enough personal funds.

Most ctDNA studies to date have been retrospective or observational studies. There are more than 100 ongoing clinical trials investigating different roles of ctDNA in lung cancer management, including several prospective interventional trials wherein decisions are made based on ctDNA analysis. For example, a Canadian trial (NCT04966663) in which early-stage NSCLC patients receive adjuvant chemo-immunotherapy if ctDNA is detectable after surgery. Such trials may give more definitive answers on the clinical utility of ctDNA and might also shed more light on the biological mechanism behind ctDNA release. ctDNA studies seem to mainly be focused on early cancer detection, e.g., by combining multianalyte tests with low-dose CT screening, assessing dynamic changes in ctDNA quantity to evaluate treatment response and for detecting minimal residual disease, and longitudinal monitoring to detect relapse. A new potential application of longitudinal ctDNA monitoring is to help decide when treatment de-escalation may be appropriate, which is becoming relevant in the adjuvant setting and for patients receiving consolidating immunotherapy.

In parallel with clinical trials, research must be aimed at understanding more of the biological mechanisms involved in ctDNA release since this may help us understand which

patients and clinical settings are most likely to benefit from ctDNA analysis. Additionally, it may provide scientific support for novel clinical trials. It is interesting to look further into the difference in prognostic value of ctDNA detection between ADC and SCC patients since this difference indicates several release mechanisms and suggests that pre-treatment ctDNA detection is not a global prognostic factor.

Our research group will further investigate the role of tumor behavior and histology since our focus is shifting towards patients with SCLC. We will continue to analyze ctDNA using tumor-informed customized NGS panels, though sequencing tumor DNA with a comprehensive NGS panel to increase the mutation detection rate and the number of mutations included per customized panel. Additionally, we will test an approach to quantify the amount of mutated ctDNA per mL plasma accurately, since we aim to study the dynamics of ctDNA quantity before, during, and after treatments in patients enrolled in clinical trials conducted by our group. In combination with ctDNA, we will analyze circulating miRNA and other tumor characteristics, such as copy number variations in DNA, global and spatial tissue gene expression, and use artificial intelligence to analyze morphological patterns in the tumor tissue.

10 References

1. Herbst, R.S.; Morgensztern, D.; Boshoff, C. The Biology and Management of Non-Small Cell Lung Cancer. *Nature* **2018**, *553*, 446–454, doi:10.1038/nature25183.
2. Thai, A.A.; Solomon, B.J.; Sequist, L.V.; Gainor, J.F.; Heist, R.S. Lung Cancer. *The Lancet* **2021**, *398*, 535–554, doi:10.1016/S0140-6736(21)00312-3.
3. Wan, J.C.M.; Massie, C.; Garcia-Corbacho, J.; Mouliere, F.; Brenton, J.D.; Caldas, C.; Pacey, S.; Baird, R.; Rosenfeld, N. Liquid Biopsies Come of Age: Towards Implementation of Circulating Tumour DNA. *Nat. Rev. Cancer* **2017**, *17*, 223–238, doi:10.1038/nrc.2017.7.
4. Pfeifer, G.P.; Denissenko, M.F.; Olivier, M.; Tretyakova, N.; Hecht, S.S.; Hainaut, P. Tobacco Smoke Carcinogens, DNA Damage and P53 Mutations in Smoking-Associated Cancers. *Oncogene* **2002**, *21*, 7435–7451, doi:10.1038/sj.onc.1205803.
5. Hassfjell, C.S.; Grimsrud, T.K.; Standring, W.J.F.; Tretli, S. Lungekreftforekomst knyttet til radoneksponering i norske boliger. *Tidsskr. Den Nor. Legeforening* **2017**, doi:10.4045/tidsskr.16.0127.
6. Malhotra, J.; Malvezzi, M.; Negri, E.; La Vecchia, C.; Boffetta, P. Risk Factors for Lung Cancer Worldwide. *Eur. Respir. J.* **2016**, *48*, 889–902, doi:10.1183/13993003.00359-2016.
7. Schabath, M.B.; Cote, M.L. Cancer Progress and Priorities: Lung Cancer. *Cancer Epidemiol. Biomarkers Prev.* **2019**, *28*, 1563–1579, doi:10.1158/1055-9965.EPI-19-0221.
8. Shay, J.W. Role of Telomeres and Telomerase in Aging and Cancer. *Cancer Discov.* **2016**, *6*, 584–593, doi:10.1158/2159-8290.CD-16-0062.
9. Teixeira Loiola De Alencar, V.; Nirvana Formiga, M.; Cordeiro De Lima, V.C. Inherited Lung Cancer: A Review. *ecancermedicalscience* **2020**, *14*, doi:10.3332/ecancer.2020.1008.
10. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **2021**, *71*, 209–249, doi:10.3322/caac.21660.
11. Cancer Registry of Norway *[Annual Report 2022 With Results and Improvement Actions From the National Quality Register For Lung Cancer]*; Oslo, 2023;
12. Kreftregisteret *[Annual Report 2021 With Results and Improvement Actions From the National Quality Register For Lung Cancer]*; Kreftregisteret: Oslo, 2022;
13. Haddadin, S.; Perry, M.C. History of Small-Cell Lung Cancer. *Clin. Lung Cancer* **2011**, *12*, 87–93, doi:10.1016/j.cllc.2011.03.002.
14. Molina, J.R.; Yang, P.; Cassivi, S.D.; Schild, S.E.; Adjei, A.A. Non-Small Cell Lung Cancer: Epidemiology, Risk Factors, Treatment, and Survivorship. *Mayo Clin. Proc.* **2008**, *83*, 584–594, doi:10.4065/83.5.584.
15. Witschi, H. A Short History of Lung Cancer. *Toxicol Sci* **2001**, *64*, 4–6.
16. Lortet-Tieulent, J.; Soerjomataram, I.; Ferlay, J.; Rutherford, M.; Weiderpass, E.; Bray, F. International Trends in Lung Cancer Incidence by Histological Subtype: Adenocarcinoma Stabilizing in Men but Still Increasing in Women. *Lung Cancer* **2014**, *84*, 13–22, doi:10.1016/j.lungcan.2014.01.009.
17. Stellman, S.D.; Muscat, J.E.; Hoffmann, D.; Wynder, E.L. Impact of Filter Cigarette Smoking on Lung Cancer Histology. *Prev. Med.* **1997**, *26*, 451–456, doi:10.1006/pmed.1997.0212.
18. Muscat, J.E.; Wynder, E.L. Lung Cancer Pathology in Smokers, Ex-Smokers and Never Smokers. *Cancer Lett.* **1995**, *88*, 1–5, doi:10.1016/0304-3835(94)03608-L.
19. Sainz de Aja, J.; Dost, A.F.M.; Kim, C.F. Alveolar Progenitor Cells and the Origin of Lung Cancer. *J. Intern. Med.* **2021**, *289*, 629–635, doi:10.1111/joim.13201.
20. Nicholson, A.G.; Tsao, M.S.; Beasley, M.B.; Borczuk, A.C.; Brambilla, E.; Cooper, W.A.; Dacic, S.; Jain, D.; Kerr, K.M.; Lantuejoul, S.; et al. The 2021 WHO Classification of Lung Tumors: Impact of Advances Since 2015. *J. Thorac. Oncol.* **2022**, *17*, 362–387, doi:10.1016/j.jtho.2021.11.003.
21. Byers, L.A.; Rudin, C.M. Small Cell Lung Cancer: Where Do We Go from Here? *Cancer* **2015**, *121*, 664–672, doi:10.1002/cncr.29098.
22. Khuder, S.A. Effect of Cigarette Smoking on Major Histological Types of Lung Cancer: A Meta-Analysis. *Lung Cancer* **2001**, *31*, 139–148, doi:10.1016/S0169-5002(00)00181-1.
23. Soomro, Z.; Youssef, M.; Yust-Katz, S.; Jalali, A.; Patel, A.J.; Mandel, J. Paraneoplastic Syndromes in Small Cell Lung Cancer. *J. Thorac. Dis.* **2020**, *12*, 6253–6263, doi:10.21037/jtd.2020.03.88.
24. Nicholson, S.A.; Beasley, M.B.; Brambilla, E.; Hasleton, P.S.; Colby, T.V.; Sheppard, M.N.; Falk, R.; Travis, W.D. Small Cell Lung Carcinoma (SCLC): A Clinicopathologic Study of 100 Cases With Surgical Specimens. *Am. J. Surg. Pathol.* **2002**, *26*, 1184–1197, doi:10.1097/00000478-200209000-00009.

25. Qin, J.; Lu, H. Combined Small-Cell Lung Carcinoma. *OncoTargets Ther.* **2018**, *Volume 11*, 3505–3511, doi:10.2147/OTT.S159057.
26. Hanahan, D.; Weinberg, R.A. Hallmarks of Cancer: The Next Generation. *Cell* **2011**, *144*, 646–674, doi:10.1016/j.cell.2011.02.013.
27. Churg, A.; Dai, J.; Tai, H.; Xie, C.; Wright, J.L. Tumor Necrosis Factor-α Is Central to Acute Cigarette Smoke–Induced Inflammation and Connective Tissue Breakdown. *Am. J. Respir. Crit. Care Med.* **2002**, *166*, 849–854, doi:10.1164/rccm.200202-097OC.
28. Valavanidis, A.; Vlachogianni, T.; Fiotakis, K. Tobacco Smoke: Involvement of Reactive Oxygen Species and Stable Free Radicals in Mechanisms of Oxidative Damage, Carcinogenesis and Synergistic Effects with Other Respirable Particles. *Int. J. Environ. Res. Public. Health* **2009**, *6*, 445–462, doi:10.3390/ijerph6020445.
29. Leanderson, P.; Tagesson, C. Cigarette Smoke-Induced DNA Damage in Cultured Human Lung Cells: Role of Hydroxyl Radicals and Endonuclease Activation. *Chem. Biol. Interact.* **1992**, *81*, 197–208, doi:10.1016/0009-2797(92)90034-I.
30. Kawanishi, S.; Ohnishi, S.; Ma, N.; Hiraku, Y.; Murata, M. Crosstalk between DNA Damage and Inflammation in the Multiple Steps of Carcinogenesis. *Int. J. Mol. Sci.* **2017**, *18*, 1808, doi:10.3390/ijms18081808.
31. Alexandrov, L.B.; Ju, Y.S.; Haase, K.; Van Loo, P.; Martincorena, I.; Nik-Zainal, S.; Totoki, Y.; Fujimoto, A.; Nakagawa, H.; Shibata, T.; et al. Mutational Signatures Associated with Tobacco Smoking in Human Cancer. *Science* **2016**, *354*, 618–622, doi:10.1126/science.aag0299.
32. Joehanes, R.; Just, A.C.; Marioni, R.E.; Pilling, L.C.; Reynolds, L.M.; Mandaviya, P.R.; Guan, W.; Xu, T.; Elks, C.E.; Aslibekyan, S.; et al. Epigenetic Signatures of Cigarette Smoking. *Circ. Cardiovasc. Genet.* **2016**, *9*, 436–447, doi:10.1161/CIRCGENETICS.116.001506.
33. Yan, X.; Liu, S.M.; Liu, C. Clinical Applications of Aneuploidies in Evolution of NSCLC Patients: Current Status and Application Prospect. *OncoTargets Ther.* **2022**, *Volume 15*, 1355–1368, doi:10.2147/OTT.S380016.
34. Kris, M.G.; Natale, R.B.; Herbst, R.S.; Lynch, Jr, T.J.; Prager, D.; Belani, C.P.; Schiller, J.H.; Kelly, K.; Spiridonidis, H.; Sandler, A.; et al. Efficacy of Gefitinib, an Inhibitor of the Epidermal Growth Factor Receptor Tyrosine Kinase, in Symptomatic Patients With Non–Small Cell Lung Cancer: A Randomized Trial. *JAMA* **2003**, *290*, 2149, doi:10.1001/jama.290.16.2149.
35. Lynch, T.J.; Bell, D.W.; Sordella, R.; Gurubhagavatula, S.; Okimoto, R.A.; Brannigan, B.W.; Harris, P.L.; Haserlat, S.M.; Supko, J.G.; Haluska, F.G.; et al. Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib. *N. Engl. J. Med.* **2004**, *350*, 2129–2139, doi:10.1056/NEJMoa040938.
36. Paez, J.G.; Jänne, P.A.; Lee, J.C.; Tracy, S.; Greulich, H.; Gabriel, S.; Herman, P.; Kaye, F.J.; Lindeman, N.; Boggon, T.J.; et al. *EGFR* Mutations in Lung Cancer: Correlation with Clinical Response to Gefitinib Therapy. *Science* **2004**, *304*, 1497–1500, doi:10.1126/science.1099314.
37. Mok, T.S.; Wu, Y.-L.; Thongprasert, S.; Yang, C.-H.; Chu, D.-T.; Saijo, N.; Sunpaweravong, P.; Han, B.; Margono, B.; Ichinose, Y.; et al. Gefitinib or Carboplatin–Paclitaxel in Pulmonary Adenocarcinoma. *N. Engl. J. Med.* **2009**, *361*, 947–957, doi:10.1056/NEJMoa0810699.
38. Camidge, D.R.; Pao, W.; Sequist, L.V. Acquired Resistance to TKIs in Solid Tumours: Learning from Lung Cancer. *Nat. Rev. Clin. Oncol.* **2014**, *11*, 473–481, doi:10.1038/nrclinonc.2014.104.
39. Thress, K.S.; Paweletz, C.P.; Felip, E.; Cho, B.C.; Stetson, D.; Dougherty, B.; Lai, Z.; Markovets, A.; Vivancos, A.; Kuang, Y.; et al. Acquired EGFR C797S Mutation Mediates Resistance to AZD9291 in Non–Small Cell Lung Cancer Harboring EGFR T790M. *Nat. Med.* **2015**, *21*, 560–562, doi:10.1038/nm.3854.
40. Soda, M.; Choi, Y.L.; Enomoto, M.; Takada, S.; Yamashita, Y.; Ishikawa, S.; Fujiwara, S.; Watanabe, H.; Kurashina, K.; Hatanaka, H.; et al. Identification of the Transforming EML4–ALK Fusion Gene in Non-Small-Cell Lung Cancer. *Nature* **2007**, *448*, 561–566, doi:10.1038/nature05945.
41. Koivunen, J.P.; Mermel, C.; Zejnullahu, K.; Murphy, C.; Lifshits, E.; Holmes, A.J.; Choi, H.G.; Kim, J.; Chiang, D.; Thomas, R.; et al. *EML4-ALK* Fusion Gene and Efficacy of an ALK Kinase Inhibitor in Lung Cancer. *Clin. Cancer Res.* **2008**, *14*, 4275–4283, doi:10.1158/1078-0432.CCR-08-0168.
42. Ju, Y.S.; Lee, W.-C.; Shin, J.-Y.; Lee, S.; Bleazard, T.; Won, J.-K.; Kim, Y.T.; Kim, J.-I.; Kang, J.-H.; Seo, J.-S. A Transforming *KIF5B* and *RET* Gene Fusion in Lung Adenocarcinoma Revealed from Whole-Genome and Transcriptome Sequencing. *Genome Res.* **2012**, *22*, 436–445, doi:10.1101/gr.133645.111.
43. Rikova, K.; Guo, A.; Zeng, Q.; Possemato, A.; Yu, J.; Haack, H.; Nardone, J.; Lee, K.; Reeves, C.; Li, Y.; et al. Global Survey of Phosphotyrosine Signaling Identifies Oncogenic Kinases in Lung Cancer. *Cell* **2007**, *131*, 1190–1203, doi:10.1016/j.cell.2007.11.025.

44. Vaishnavi, A.; Capelletti, M.; Le, A.T.; Kako, S.; Butaney, M.; Ercan, D.; Mahale, S.; Davies, K.D.; Aisner, D.L.; Pilling, A.B.; et al. Oncogenic and Drug-Sensitive NTRK1 Rearrangements in Lung Cancer. *Nat. Med.* **2013**, *19*, 1469–1472, doi:10.1038/nm.3352.
45. Shaw, A.T.; Ou, S.-H.I.; Bang, Y.-J.; Camidge, D.R.; Solomon, B.J.; Salgia, R.; Riely, G.J.; Varella-Garcia, M.; Shapiro, G.I.; Costa, D.B.; et al. Crizotinib in *ROS1* -Rearranged Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **2014**, *371*, 1963–1971, doi:10.1056/NEJMoa1406766.
46. Lin, J.J.; Kummar, S.; Tan, D.S.-W.; Lassen, U.N.; Leyvraz, S.; Liu, Y.; Moreno, V.; Patel, J.D.; Rosen, L.S.; Solomon, B.M.; et al. Long-Term Efficacy and Safety of Larotrectinib in Patients with TRK Fusion-Positive Lung Cancer. *J. Clin. Oncol.* **2021**, *39*, 9109–9109, doi:10.1200/JCO.2021.39.15_suppl.9109.
47. Drilon, A.; Oxnard, G.R.; Tan, D.S.W.; Loong, H.H.F.; Johnson, M.; Gainor, J.; McCoach, C.E.; Gautschi, O.; Besse, B.; Cho, B.C.; et al. Efficacy of Selpercatinib in *RET* Fusion–Positive Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **2020**, *383*, 813–824, doi:10.1056/NEJMoa2005653.
48. Planchard, D.; Kim, T.M.; Mazieres, J.; Quoix, E.; Riely, G.; Barlesi, F.; Souquet, P.-J.; Smit, E.F.; Groen, H.J.M.; Kelly, R.J.; et al. Dabrafenib in Patients with BRAFV600E-Positive Advanced Non-Small-Cell Lung Cancer: A Single-Arm, Multicentre, Open-Label, Phase 2 Trial. *Lancet Oncol.* **2016**, *17*, 642–650, doi:10.1016/S1470-2045(16)00077-2.
49. Harada, G.; Yang, S.-R.; Cocco, E.; Drilon, A. Rare Molecular Subtypes of Lung Cancer. *Nat. Rev. Clin. Oncol.* **2023**, *20*, 229–249, doi:10.1038/s41571-023-00733-6.
50. Kwan, A.K.; Piazza, G.A.; Keeton, A.B.; Leite, C.A. The Path to the Clinic: A Comprehensive Review on Direct KRASG12C Inhibitors. *J. Exp. Clin. Cancer Res.* **2022**, *41*, 27, doi:10.1186/s13046-021-02225-w.
51. Skoulidis, F.; Li, B.T.; Dy, G.K.; Price, T.J.; Falchook, G.S.; Wolf, J.; Italiano, A.; Schuler, M.; Borghaei, H.; Barlesi, F.; et al. Sotorasib for Lung Cancers with *KRAS* p.G12C Mutation. *N. Engl. J. Med.* **2021**, *384*, 2371–2381, doi:10.1056/NEJMoa2103695.
52. Wahl, S.G.F.; Dai, H.Y.; Emdal, E.F.; Berg, T.; Halvorsen, T.O.; Ottestad, A.L.; Lund-Iversen, M.; Brustugun, O.T.; Førde, D.; Paulsen, E.-E.; et al. The Prognostic Effect of KRAS Mutations in Non-Small Cell Lung Carcinoma Revisited: A Norwegian Multicentre Study. *Cancers* **2021**, *13*, 4294, doi:10.3390/cancers13174294.
53. Tan, A.C.; Tan, D.S.W. Targeted Therapies for Lung Cancer Patients With Oncogenic Driver Molecular Alterations. *J. Clin. Oncol.* **2022**, *40*, 611–625, doi:10.1200/JCO.21.01626.
54. Larsen, J.E.; Minna, J.D. Molecular Biology of Lung Cancer: Clinical Implications. *Clin. Chest Med.* **2011**, *32*, 703–740, doi:10.1016/j.ccm.2011.08.003.
55. Kastan, M.B.; Onyekwere, O.; Sidransky, D.; Vogelstein, B.; Craig, R.W. Participation of P53 Protein in the Cellular Response to DNA Damage. *Cancer Res.* **1991**, *51*, 6304–6311.
56. The Cancer Genome Atlas Research Network Comprehensive Genomic Characterization of Squamous Cell Lung Cancers. *Nature* **2012**, *489*, 519–525, doi:10.1038/nature11404.
57. The Cancer Genome Atlas Research Network Comprehensive Molecular Profiling of Lung Adenocarcinoma. *Nature* **2014**, *511*, 543–550, doi:10.1038/nature13385.
58. Iniesta, P.; González-Quevedo, R.; Morán, A.; García-Aranda, C.; Juan, C.D.; Sánchez-Pernaute, A.; Torres, A.; Díaz-Rubio, E.; Balibrea, J.L.; Benito, M. Relationship between 3p Deletions and Telomerase Activity in Non-Small-Cell Lung Cancer: Prognostic Implications. *Br. J. Cancer* **2004**, *90*, 1983–1988, doi:10.1038/sj.bjc.6601775.
59. Freeman, G.J.; Long, A.J.; Iwai, Y.; Bourque, K.; Chernova, T.; Nishimura, H.; Fitz, L.J.; Malenkovich, N.; Okazaki, T.; Byrne, M.C.; et al. Engagement of the Pd-1 Immunoinhibitory Receptor by a Novel B7 Family Member Leads to Negative Regulation of Lymphocyte Activation. *J. Exp. Med.* **2000**, *192*, 1027–1034, doi:10.1084/jem.192.7.1027.
60. Norwegian Directorate of Health [Treatment Guidelines for Lung Cancer, Mesothelioma and Thymoma] Available online: https://www.helsedirektoratet.no/retningslinjer/lungekreft-mesoteliom-og-thymom-handlingsprogram.
61. Goldstraw, P.; Chansky, K.; Crowley, J.; Rami-Porta, R.; Asamura, H.; Eberhardt, W.E.E.; Nicholson, A.G.; Groome, P.; Mitchell, A.; Bolejack, V.; et al. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J. Thorac. Oncol.* **2016**, *11*, 39–51, doi:10.1016/j.jtho.2015.09.009.
62. Vallières, E.; Shepherd, F.A.; Crowley, J.; Van Houtte, P.; Postmus, P.E.; Carney, D.; Chansky, K.; Shaikh, Z.; Goldstraw, P. The IASLC Lung Cancer Staging Project: Proposals Regarding the Relevance of TNM in the Pathologic Staging of Small Cell Lung Cancer in the Forthcoming (Seventh) Edition of the TNM Classification for Lung Cancer. *J. Thorac. Oncol.* **2009**, *4*, 1049–1059, doi:10.1097/JTO.0b013e3181b27799.
63. Ambrosini, V.; Nicolini, S.; Caroli, P.; Nanni, C.; Massaro, A.; Marzola, M.C.; Rubello, D.; Fanti, S. PET/CT Imaging in Different Types of Lung Cancer: An Overview. *Eur. J. Radiol.* **2012**, *81*, 988–1001, doi:10.1016/j.ejrad.2011.03.020.

64. Koppenol, W.H.; Bounds, P.L.; Dang, C.V. Otto Warburg's Contributions to Current Concepts of Cancer Metabolism. *Nat. Rev. Cancer* **2011**, *11*, 325–337, doi:10.1038/nrc3038.
65. Im, H.-J.; Bradshaw, T.; Solaiyappan, M.; Cho, S.Y. Current Methods to Define Metabolic Tumor Volume in Positron Emission Tomography: Which One Is Better? *Nucl. Med. Mol. Imaging* **2018**, *52*, 5–15, doi:10.1007/s13139-017-0493-6.
66. Metz, B.; Kersten, G.F.A.; Hoogerhout, P.; Brugghe, H.F.; Timmermans, H.A.M.; De Jong, A.; Meiring, H.; Ten Hove, J.; Hennink, W.E.; Crommelin, D.J.A.; et al. Identification of Formaldehyde-Induced Modifications in Proteins. *J. Biol. Chem.* **2004**, *279*, 6235–6243, doi:10.1074/jbc.M310752200.
67. Oken, M.M.; Creech, R.H.; Tormey, D.C.; Horton, J.; Davis, T.E.; McFadden, E.T.; Carbone, P.P. Toxicity and Response Criteria of the Eastern Cooperative Oncology Group. *Am. J. Clin. Oncol.* **1982**, *5*, 649–655.
68. Frank, M.S.; Bodtger, U. An Individualized Approach to Comorbidities in Lung Cancer. *J. Thorac. Oncol.* **2023**, *18*, 254–256, doi:10.1016/j.jtho.2022.12.007.
69. Winton, T.; Livingston, R.; Johnson, D.; Rigas, J.; Johnston, M.; Butts, C.; Cormier, Y.; Goss, G.; Inculet, R.; Vallieres, E.; et al. Vinorelbine plus Cisplatin vs. Observation in Resected Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **2005**, *352*, 2589–2597, doi:10.1056/NEJMoa043623.
70. Wu, Y.-L.; Tsuboi, M.; He, J.; John, T.; Grohe, C.; Majem, M.; Goldman, J.W.; Laktionov, K.; Kim, S.-W.; Kato, T.; et al. Osimertinib in Resected *EGFR* -Mutated Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **2020**, *383*, 1711–1723, doi:10.1056/NEJMoa2027071.
71. Ball, D.; Mai, G.T.; Vinod, S.; Babington, S.; Ruben, J.; Kron, T.; Chesson, B.; Herschtal, A.; Vanevski, M.; Rezo, A.; et al. Stereotactic Ablative Radiotherapy versus Standard Radiotherapy in Stage 1 Non-Small-Cell Lung Cancer (TROG 09.02 CHISEL): A Phase 3, Open-Label, Randomised Controlled Trial. *Lancet Oncol.* **2019**, *20*, 494–503, doi:10.1016/S1470-2045(18)30896-9.
72. Antonia, S.J.; Villegas, A.; Daniel, D.; Vicente, D.; Murakami, S.; Hui, R.; Yokoi, T.; Chiappori, A.; Lee, K.H.; de Wit, M.; et al. Durvalumab after Chemoradiotherapy in Stage III Non-Small-Cell Lung Cancer. *N Engl J Med* **2017**, *377*, 1919–1929, doi:10.1056/NEJMoa1709937.
73. Scagliotti, G.V.; Parikh, P.; Von Pawel, J.; Biesma, B.; Vansteenkiste, J.; Manegold, C.; Serwatowski, P.; Gatzemeier, U.; Digumarti, R.; Zukin, M.; et al. Phase III Study Comparing Cisplatin Plus Gemcitabine With Cisplatin Plus Pemetrexed in Chemotherapy-Naive Patients With Advanced-Stage Non–Small-Cell Lung Cancer. *J. Clin. Oncol.* **2008**, *26*, 3543–3551, doi:10.1200/JCO.2007.15.0375.
74. NSCLC Meta-Analyses Collaborative Group Chemotherapy in Addition to Supportive Care Improves Survival in Advanced Non–Small-Cell Lung Cancer: A Systematic Review and Meta-Analysis of Individual Patient Data From 16 Randomized Controlled Trials. *J. Clin. Oncol.* **2008**, *26*, 4617–4625, doi:10.1200/JCO.2008.17.7162.
75. Mok, T.S.; Wu, Y.-L.; Ahn, M.-J.; Garassino, M.C.; Kim, H.R.; Ramalingam, S.S.; Shepherd, F.A.; He, Y.; Akamatsu, H.; Theelen, W.S.M.E.; et al. Osimertinib or Platinum–Pemetrexed in *EGFR* T790M–Positive Lung Cancer. *N. Engl. J. Med.* **2017**, *376*, 629–640, doi:10.1056/NEJMoa1612674.
76. Liu, S.; Kurzrock, R. Toxicity of Targeted Therapy: Implications for Response and Impact of Genetic Polymorphisms. *Cancer Treat. Rev.* **2014**, *40*, 883–891, doi:10.1016/j.ctrv.2014.05.003.
77. Mazieres, J.; Drilon, A.; Lusque, A.; Mhanna, L.; Cortot, A.B.; Mezquita, L.; Thai, A.A.; Mascaux, C.; Couraud, S.; Veillon, R.; et al. Immune Checkpoint Inhibitors for Patients with Advanced Lung Cancer and Oncogenic Driver Alterations: Results from the IMMUNOTARGET Registry. *Ann. Oncol.* **2019**, *30*, 1321–1328, doi:10.1093/annonc/mdz167.
78. Reck, M.; Rodriguez-Abreu, D.; Robinson, A.G.; Hui, R.; Csoszi, T.; Fulop, A.; Gottfried, M.; Peled, N.; Tafreshi, A.; Cuffe, S.; et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. *N Engl J Med* **2016**, *375*, 1823–1833, doi:10.1056/NEJMoa1606774.
79. Paz-Ares, L.; Luft, A.; Vicente, D.; Tafreshi, A.; Gümüş, M.; Mazières, J.; Hermes, B.; Çay Şenler, F.; Csőszi, T.; Fülöp, A.; et al. Pembrolizumab plus Chemotherapy for Squamous Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **2018**, *379*, 2040–2051, doi:10.1056/NEJMoa1810865.
80. Fairchild, A.; Harris, K.; Barnes, E.; Wong, R.; Lutz, S.; Bezjak, A.; Cheung, P.; Chow, E. Palliative Thoracic Radiotherapy for Lung Cancer: A Systematic Review. *J. Clin. Oncol.* **2008**, *26*, 4001–4011, doi:10.1200/JCO.2007.15.3312.
81. Garon, E.B.; Hellmann, M.D.; Rizvi, N.A.; Carcereny, E.; Leighl, N.B.; Ahn, M.-J.; Eder, J.P.; Balmanoukian, A.S.; Aggarwal, C.; Horn, L.; et al. Five-Year Overall Survival for Patients With Advanced Non–Small-Cell Lung Cancer Treated With Pembrolizumab: Results From the Phase I KEYNOTE-001 Study. *J. Clin. Oncol.* **2019**, *37*, 2518–2527, doi:10.1200/JCO.19.00934.

82. Ignatius Ou, S.-H.; Zell, J.A. The Applicability of the Proposed IASLC Staging Revisions to Small Cell Lung Cancer (SCLC) with Comparison to the Current UICC 6th TNM Edition. *J. Thorac. Oncol.* **2009**, *4*, 300–310, doi:10.1097/JTO.0b013e318194a355.
83. Mascaux, C.; Paesmans, M.; Berghmans, T.; Branle, F.; Lafitte, J.J.; Lemaître, F.; Meert, A.P.; Vermylen, P.; Sculier, J.P. A Systematic Review of the Role of Etoposide and Cisplatin in the Chemotherapy of Small Cell Lung Cancer with Methodology Assessment and Meta-Analysis. *Lung Cancer* **2000**, *30*, 23–36, doi:10.1016/S0169-5002(00)00127-6.
84. Horn, L.; Mansfield, A.S.; Szczesna, A.; Havel, L.; Krzakowski, M.; Hochmair, M.J.; Huemer, F.; Losonczy, G.; Johnson, M.L.; Nishio, M.; et al. First-Line Atezolizumab plus Chemotherapy in Extensive-Stage Small-Cell Lung Cancer. *N Engl J Med* **2018**, *379*, 2220–2229, doi:10.1056/NEJMoa1809064.
85. Simon, G.R.; Turrisi, A. Management of Small Cell Lung Cancer. *Chest* **2007**, *132*, 324S-339S, doi:10.1378/chest.07-1385.
86. Rudin, C.M.; Ismaila, N.; Hann, C.L.; Malhotra, N.; Movsas, B.; Norris, K.; Pietanza, M.C.; Ramalingam, S.S.; Turrisi, A.T.; Giaccone, G. Treatment of Small-Cell Lung Cancer: American Society of Clinical Oncology Endorsement of the American College of Chest Physicians Guideline. *J. Clin. Oncol.* **2015**, *33*, 4106–4111, doi:10.1200/JCO.2015.63.7918.
87. Grønberg, B.H.; Halvorsen, T.O.; Fløtten, Ø.; Brustugun, O.T.; Brunsvig, P.F.; Aasebø, U.; Bremnes, R.M.; Tollåli, T.; Hornslien, K.; Aksnessæther, B.Y.; et al. Randomized Phase II Trial Comparing Twice Daily Hyperfractionated with Once Daily Hypofractionated Thoracic Radiotherapy in Limited Disease Small Cell Lung Cancer. *Acta Oncol. Stockh. Swed.* **2016**, *55*, 591–597, doi:10.3109/0284186X.2015.1092584.
88. Kubota, K.; Hida, T.; Ishikura, S.; Mizusawa, J.; Nishio, M.; Kawahara, M.; Yokoyama, A.; Imamura, F.; Takeda, K.; Negoro, S.; et al. Etoposide and Cisplatin versus Irinotecan and Cisplatin in Patients with Limited-Stage Small-Cell Lung Cancer Treated with Etoposide and Cisplatin plus Concurrent Accelerated Hyperfractionated Thoracic Radiotherapy (JCOG0202): A Randomised Phase 3 Study. *Lancet Oncol* **2014**, *15*, 106–113, doi:10.1016/S1470-2045(13)70511-4.
89. Goldman, J.W.; Dvorkin, M.; Chen, Y.; Reinmuth, N.; Hotta, K.; Trukhin, D.; Statsenko, G.; Hochmair, M.J.; Özgüroğlu, M.; Ji, J.H.; et al. Durvalumab, with or without Tremelimumab, plus Platinum-Etoposide versus Platinum-Etoposide Alone in First-Line Treatment of Extensive-Stage Small-Cell Lung Cancer (CASPIAN): Updated Results from a Randomised, Controlled, Open-Label, Phase 3 Trial. *Lancet Oncol.* **2021**, *22*, 51–65, doi:10.1016/S1470-2045(20)30539-8.
90. Paesmans, M. Prognostic and Predictive Factors for Lung Cancer. *Breathe* **2012**, *9*, 112–121, doi:10.1183/20734735.006911.
91. Stabellini, N.; Bruno, D.S.; Dmukauskas, M.; Barda, A.J.; Cao, L.; Shanahan, J.; Waite, K.; Montero, A.J.; Barnholtz-Sloan, J.S. Sex Differences in Lung Cancer Treatment and Outcomes at a Large Hybrid Academic-Community Practice. *JTO Clin. Res. Rep.* **2022**, *3*, 100307, doi:10.1016/j.jtocrr.2022.100307.
92. Dall'Olio, F.G.; Maggio, I.; Massucci, M.; Mollica, V.; Fragomeno, B.; Ardizzoni, A. ECOG Performance Status ≥2 as a Prognostic Factor in Patients with Advanced Non Small Cell Lung Cancer Treated with Immune Checkpoint Inhibitors—A Systematic Review and Meta-Analysis of Real World Data. *Lung Cancer* **2020**, *145*, 95–104, doi:10.1016/j.lungcan.2020.04.027.
93. Wolfson, A.H.; Bae, K.; Komaki, R.; Meyers, C.; Movsas, B.; Le Pechoux, C.; Werner-Wasik, M.; Videtic, G.M.M.; Garces, Y.I.; Choy, H. Primary Analysis of a Phase II Randomized Trial Radiation Therapy Oncology Group (RTOG) 0212: Impact of Different Total Doses and Schedules of Prophylactic Cranial Irradiation on Chronic Neurotoxicity and Quality of Life for Patients With Limited-Disease Small-Cell Lung Cancer. *Int. J. Radiat. Oncol.* **2011**, *81*, 77–84, doi:10.1016/j.ijrobp.2010.05.013.
94. Pantel, K.; Alix-Panabières, C. Circulating Tumour Cells in Cancer Patients: Challenges and Perspectives. *Trends Mol. Med.* **2010**, *16*, 398–406, doi:10.1016/j.molmed.2010.07.001.
95. Alix-Panabières, C.; Pantel, K. Liquid Biopsy: From Discovery to Clinical Application. *Cancer Discov.* **2021**, *11*, 858–873, doi:10.1158/2159-8290.CD-20-1311.
96. Tivey, A.; Church, M.; Rothwell, D.; Dive, C.; Cook, N. Circulating Tumour DNA — Looking beyond the Blood. *Nat. Rev. Clin. Oncol.* **2022**, *19*, 600–612, doi:10.1038/s41571-022-00660-y.
97. Palacín-Aliana, I.; García-Romero, N.; Asensi-Puig, A.; Carrión-Navarro, J.; González-Rumayor, V.; Ayuso-Sacido, Á. Clinical Utility of Liquid Biopsy-Based Actionable Mutations Detected via DdPCR. *Biomedicines* **2021**, *9*, 906, doi:10.3390/biomedicines9080906.
98. Schluter, J.; Peled, J.U.; Taylor, B.P.; Markey, K.A.; Smith, M.; Taur, Y.; Niehus, R.; Staffas, A.; Dai, A.; Fontana, E.; et al. The Gut Microbiota Is Associated with Immune Cell Dynamics in Humans. *Nature* **2020**, *588*, 303–307, doi:10.1038/s41586-020-2971-8.

99. Williams, C.G.; Lee, H.J.; Asatsuma, T.; Vento-Tormo, R.; Haque, A. An Introduction to Spatial Transcriptomics for Biomedical Research. *Genome Med.* **2022**, *14*, 68, doi:10.1186/s13073-022-01075-1.
100. Thierry, A.R.; El Messaoudi, S.; Gahan, P.B.; Anker, P.; Stroun, M. Origins, Structures, and Functions of Circulating DNA in Oncology. *Cancer Metastasis Rev.* **2016**, *35*, 347–376, doi:10.1007/s10555-016-9629-x.
101. Leon, S.A.; Shapiro, B.; Sklaroff, D.M.; Yaros, M.J. Free DNA in the Serum of Cancer Patients and the Effect of Therapy. *Cancer Res.* **1977**, *37*, 646–650.
102. Underhill, H.R.; Kitzman, J.O.; Hellwig, S.; Welker, N.C.; Daza, R.; Baker, D.N.; Gligorich, K.M.; Rostomily, R.C.; Bronner, M.P.; Shendure, J. Fragment Length of Circulating Tumor DNA. *PLOS Genet.* **2016**, *12*, e1006162, doi:10.1371/journal.pgen.1006162.
103. Abbosh, C.; Birkbak, N.J.; Swanton, C. Early Stage NSCLC — Challenges to Implementing CtDNA-Based Screening and MRD Detection. *Nat. Rev. Clin. Oncol.* **2018**, *15*, 577–586, doi:10.1038/s41571-018-0058-3.
104. Diamandis, E.P.; Fiala, C. Can Circulating Tumor DNA Be Used for Direct and Early Stage Cancer Detection? *F1000Research* **2017**, *6*, 2129, doi:10.12688/f1000research.13440.1.
105. Hu, Y.; Ulrich, B.C.; Supplee, J.; Kuang, Y.; Lizotte, P.H.; Feeney, N.B.; Guibert, N.M.; Awad, M.M.; Wong, K.-K.; Jänne, P.A.; et al. False-Positive Plasma Genotyping Due to Clonal Hematopoiesis. *Clin. Cancer Res.* **2018**, *24*, 4437–4443, doi:10.1158/1078-0432.CCR-18-0143.
106. Abbosh, C.; Birkbak, N.J.; Wilson, G.A.; Jamal-Hanjani, M.; Constantin, T.; Salari, R.; Le Quesne, J.; Moore, D.A.; Veeriah, S.; Rosenthal, R.; et al. Phylogenetic CtDNA Analysis Depicts Early-Stage Lung Cancer Evolution. *Nature* **2017**, *545*, 446–451, doi:10.1038/nature22364.
107. Tie, J.; Wang, Y.; Tomasetti, C.; Li, L.; Springer, S.; Kinde, I.; Silliman, N.; Tacey, M.; Wong, H.-L.; Christie, M.; et al. Circulating Tumor DNA Analysis Detects Minimal Residual Disease and Predicts Recurrence in Patients with Stage II Colon Cancer. *Sci. Transl. Med.* **2016**, *8*, 346ra92-346ra92, doi:10.1126/scitranslmed.aaf6219.
108. Bettegowda, C.; Sausen, M.; Leary, R.J.; Kinde, I.; Wang, Y.; Agrawal, N.; Bartlett, B.R.; Wang, H.; Luber, B.; Alani, R.M.; et al. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Sci. Transl. Med.* **2014**, *6*, 224ra24-224ra24, doi:10.1126/scitranslmed.3007094.
109. Newman, A.M.; Bratman, S.V.; To, J.; Wynne, J.F.; Eclov, N.C.W.; Modlin, L.A.; Liu, C.L.; Neal, J.W.; Wakelee, H.A.; Merritt, R.E.; et al. An Ultrasensitive Method for Quantitating Circulating Tumor DNA with Broad Patient Coverage. *Nat. Med.* **2014**, *20*, 548–554, doi:10.1038/nm.3519.
110. Chaudhuri, A.A.; Chabon, J.J.; Lovejoy, A.F.; Newman, A.M.; Stehr, H.; Azad, T.D.; Khodadoust, M.S.; Esfahani, M.S.; Liu, C.L.; Zhou, L.; et al. Early Detection of Molecular Residual Disease in Localized Lung Cancer by Circulating Tumor DNA Profiling. *Cancer Discov.* **2017**, *7*, 1394–1403, doi:10.1158/2159-8290.CD-17-0716.
111. Phallen, J.; Sausen, M.; Adleff, V.; Leal, A.; Hruban, C.; White, J.; Anagnostou, V.; Fiksel, J.; Cristiano, S.; Papp, E.; et al. Direct Detection of Early-Stage Cancers Using Circulating Tumor DNA. *Sci. Transl. Med.* **2017**, *9*, eaan2415, doi:10.1126/scitranslmed.aan2415.
112. Pascual, J.; Attard, G.; Bidard, F.-C.; Curigliano, G.; De Mattos-Arruda, L.; Diehn, M.; Italiano, A.; Lindberg, J.; Merker, J.D.; Montagut, C.; et al. ESMO Recommendations on the Use of Circulating Tumour DNA Assays for Patients with Cancer: A Report from the ESMO Precision Medicine Working Group. *Ann. Oncol.* **2022**, *33*, 750–768, doi:10.1016/j.annonc.2022.05.520.
113. Wahl, S.G.F.; Dai, H.Y.; Emdal, E.F.; Ottestad, A.L.; Dale, V.G.; Richardsen, E.; Halvorsen, T.O.; Grønberg, B.H. Prognostic Value of Absolute Quantification of Mutated KRAS in Circulating Tumour DNA in Lung Adenocarcinoma Patients Prior to Therapy. *J. Pathol. Clin. Res.* **2021**, cjp2.200, doi:10.1002/cjp2.200.
114. R: A Language and Environment for Statistical Computing.
115. Eisenhauer, E.A.; Therasse, P.; Bogaerts, J.; Schwartz, L.H.; Sargent, D.; Ford, R.; Dancey, J.; Arbuck, S.; Gwyther, S.; Mooney, M.; et al. New Response Evaluation Criteria in Solid Tumours: Revised RECIST Guideline (Version 1.1). *Eur J Cancer* **2009**, *45*, 228–247, doi:10.1016/j.ejca.2008.10.026.
116. Boellaard, R.; Delgado-Bolton, R.; Oyen, W.J.G.; Giammarile, F.; Tatsch, K.; Eschner, W.; Verzijlbergen, F.J.; Barrington, S.F.; Pike, L.C.; Weber, W.A.; et al. FDG PET/CT: EANM Procedure Guidelines for Tumour Imaging: Version 2.0. *Eur. J. Nucl. Med. Mol. Imaging* **2015**, *42*, 328–354, doi:10.1007/s00259-014-2961-x.
117. Kinde, I.; Wu, J.; Papadopoulos, N.; Kinzler, K.W.; Vogelstein, B. Detection and Quantification of Rare Mutations with Massively Parallel Sequencing. *Proc. Natl. Acad. Sci.* **2011**, *108*, 9530–9535, doi:10.1073/pnas.1105422108.
118. Qiagen QIAseq® Targeted DNA Panel Handbook 2017.

119. Goodwin, S.; McPherson, J.D.; McCombie, W.R. Coming of Age: Ten Years of next-Generation Sequencing Technologies. *Nat. Rev. Genet.* **2016**, *17*, 333–351, doi:10.1038/nrg.2016.49.
120. Guo, Q.; Wang, J.; Xiao, J.; Wang, L.; Hu, X.; Yu, W.; Song, G.; Lou, J.; Chen, J. Heterogeneous Mutation Pattern in Tumor Tissue and Circulating Tumor DNA Warrants Parallel NGS Panel Testing. *Mol. Cancer* **2018**, *17*, 131, doi:10.1186/s12943-018-0875-0.
121. Chen, K.; Zhang, J.; Guan, T.; Yang, F.; Lou, F.; Chen, W.; Zhao, M.; Zhang, J.; Chen, S.; Wang, J. Comparison of Plasma to Tissue DNA Mutations in Surgical Patients with Non–Small Cell Lung Cancer. *J. Thorac. Cardiovasc. Surg.* **2017**, *154*, 1123-1131.e2, doi:10.1016/j.jtcvs.2017.04.073.
122. Ottestad, A.L.; Wahl, S.G.F.; Grønberg, B.H.; Skorpen, F.; Dai, H.Y. The Relevance of Tumor Mutation Profiling in Interpretation of NGS Data from Cell-Free DNA in Non-Small Cell Lung Cancer Patients. *Exp. Mol. Pathol.* **2019**, 104347, doi:10.1016/j.yexmp.2019.104347.
123. Ottestad, A.L.; Emdal, E.F.; Grønberg, B.H.; Halvorsen, T.O.; Dai, H.Y. Fragmentation Assessment of FFPE DNA Helps in Evaluating NGS Library Complexity and Interpretation of NGS Results. *Exp. Mol. Pathol.* **2022**, *126*, 104771, doi:10.1016/j.yexmp.2022.104771.
124. Ottestad, A.L.; Dai, H.Y.; Halvorsen, T.O.; Emdal, E.F.; Wahl, S.G.F.; Grønberg, B.H. Associations between Tumor Mutations in CfDNA and Survival in Non-Small Cell Lung Cancer. *Cancer Treat. Res. Commun.* **2021**, *29*, 100471, doi:10.1016/j.ctarc.2021.100471.
125. Chabon, J.J.; Hamilton, E.G.; Kurtz, D.M.; Esfahani, M.S.; Moding, E.J.; Stehr, H.; Schroers-Martin, J.; Nabet, B.Y.; Chen, B.; Chaudhuri, A.A.; et al. Integrating Genomic Features for Non-Invasive Early Lung Cancer Detection. *Nature* **2020**, *580*, 245–251, doi:10.1038/s41586-020-2140-0.
126. Wan, J.C.M.; Heider, K.; Gale, D.; Murphy, S.; Fisher, E.; Mouliere, F.; Ruiz-Valdepenas, A.; Santonja, A.; Morris, J.; Chandrananda, D.; et al. CtDNA Monitoring Using Patient-Specific Sequencing and Integration of Variant Reads. *Sci. Transl. Med.* **2020**, *12*, eaaz8084, doi:10.1126/scitranslmed.aaz8084.
127. Xia, L.; Mei, J.; Kang, R.; Deng, S.; Chen, Y.; Yang, Y.; Feng, G.; Deng, Y.; Gan, F.; Lin, Y.; et al. Perioperative CtDNA-Based Molecular Residual Disease Detection for Non–Small Cell Lung Cancer: A Prospective Multicenter Cohort Study (LUNGCA-1). *Clin. Cancer Res.* **2022**, *28*, 3308–3317, doi:10.1158/1078-0432.CCR-21-3044.
128. Abbosh, C.; Frankell, A.M.; Harrison, T.; Kisistok, J.; Garnett, A.; Johnson, L.; Veeriah, S.; Moreau, M.; Chesh, A.; Chaunzwa, T.L.; et al. Tracking Early Lung Cancer Metastatic Dissemination in TRACERx Using CtDNA. *Nature* **2023**, doi:10.1038/s41586-023-05776-4.
129. Li, J.; Jiang, W.; Wei, J.; Zhang, J.; Cai, L.; Luo, M.; Wang, Z.; Sun, W.; Wang, S.; Wang, C.; et al. Patient Specific Circulating Tumor DNA Fingerprints to Monitor Treatment Response across Multiple Tumors. *J. Transl. Med.* **2020**, *18*, doi:10.1186/s12967-020-02449-y.
130. McDonald, B.R.; Contente-Cuomo, T.; Sammut, S.-J.; Odenheimer-Bergman, A.; Ernst, B.; Perdigones, N.; Chin, S.-F.; Farooq, M.; Mejia, R.; Cronin, P.A.; et al. Personalized Circulating Tumor DNA Analysis to Detect Residual Disease after Neoadjuvant Therapy in Breast Cancer. *Sci. Transl. Med.* **2019**, *11*, eaax7392, doi:10.1126/scitranslmed.aax7392.
131. Cohen, J.D.; Li, L.; Wang, Y.; Thoburn, C.; Afsari, B.; Danilova, L.; Douville, C.; Javed, A.A.; Wong, F.; Mattox, A.; et al. Detection and Localization of Surgically Resectable Cancers with a Multi-Analyte Blood Test. *Science* **2018**, *359*, 926–930, doi:10.1126/science.aar3247.
132. Newman, A.M.; Lovejoy, A.F.; Klass, D.M.; Kurtz, D.M.; Chabon, J.J.; Scherer, F.; Stehr, H.; Liu, C.L.; Bratman, S.V.; Say, C.; et al. Integrated Digital Error Suppression for Improved Detection of Circulating Tumor DNA. *Nat. Biotechnol.* **2016**, *34*, 547–555, doi:10.1038/nbt.3520.
133. Santonja, A.; Cooper, W.N.; Eldridge, M.D.; Edwards, P.A.W.; Morris, J.A.; Edwards, A.R.; Zhao, H.; Heider, K.; Couturier, D.; Vijayaraghavan, A.; et al. Comparison of Tumor-informed and Tumor-naïve Sequencing Assays for CtDNA Detection in Breast Cancer. *EMBO Mol. Med.* **2023**, e16505, doi:10.15252/emmm.202216505.
134. Ferrer, I.; Armstrong, J.; Capellari, S.; Parchi, P.; Arzberger, T.; Bell, J.; Budka, H.; Ströbel, T.; Giaccone, G.; Rossi, G.; et al. Effects of Formalin Fixation, Paraffin Embedding, and Time of Storage on DNA Preservation in Brain Tissue: A BrainNet Europe Study. *Brain Pathol.* **2007**, *17*, 297–303, doi:10.1111/j.1750-3639.2007.00073.x.
135. Greer, C.E.; Peterson, S.L.; Kiviat, N.B.; Manos, M.M. PCR Amplification from Paraffin-Embedded Tissues: *Effects of Fixative and Fixation Time*. *Am. J. Clin. Pathol.* **1991**, *95*, 117–124, doi:10.1093/ajcp/95.2.117.
136. Mathieson, W.; Thomas, G.A. Why Formalin-Fixed, Paraffin-Embedded Biospecimens Must Be Used in Genomic Medicine: An Evidence-Based Review and Conclusion. *J. Histochem. Cytochem.* **2020**, *68*, 543–552, doi:10.1369/0022155420945050.
137. Heydt, C.; Fassunke, J.; Künstlinger, H.; Ihle, M.A.; König, K.; Heukamp, L.C.; Schildhaus, H.-U.; Odenthal, M.; Büttner, R.; Merkelbach-Bruse, S. Comparison of Pre-Analytical FFPE Sample

Preparation Methods and Their Impact on Massively Parallel Sequencing in Routine Diagnostics. *PLoS ONE* **2014**, *9*, e104566, doi:10.1371/journal.pone.0104566.
138. McNulty, S.N.; Mann, P.R.; Robinson, J.A.; Duncavage, E.J.; Pfeifer, J.D. Impact of Reducing DNA Input on Next-Generation Sequencing Library Complexity and Variant Detection. *J. Mol. Diagn.* **2020**, *0*, doi:10.1016/j.jmoldx.2020.02.003.
139. Einaga, N.; Yoshida, A.; Noda, H.; Suemitsu, M.; Nakayama, Y.; Sakurada, A.; Kawaji, Y.; Yamaguchi, H.; Sasaki, Y.; Tokino, T.; et al. Assessment of the Quality of DNA from Various Formalin-Fixed Paraffin-Embedded (FFPE) Tissues and the Use of This DNA for next-Generation Sequencing (NGS) with No Artifactual Mutation. *PLOS ONE* **2017**, *12*, e0176280, doi:10.1371/journal.pone.0176280.
140. Peng, M.; Huang, Q.; Yin, W.; Tan, S.; Chen, C.; Liu, W.; Tang, J.; Wang, X.; Zhang, B.; Zou, M.; et al. Circulating Tumor DNA as a Prognostic Biomarker in Localized Non-Small Cell Lung Cancer. *Front. Oncol.* **2020**, *10*, doi:10.3389/fonc.2020.561598.
141. Jee, J.; Lebow, E.S.; Yeh, R.; Das, J.P.; Namakydoust, A.; Paik, P.K.; Chaft, J.E.; Jayakumaran, G.; Rose Brannon, A.; Benayed, R.; et al. Overall Survival with Circulating Tumor DNA-Guided Therapy in Advanced Non-Small-Cell Lung Cancer. *Nat. Med.* **2022**, *28*, 2353–2363, doi:10.1038/s41591-022-02047-z.
142. Zhang, B.; Niu, X.; Zhang, Q.; Wang, C.; Liu, B.; Yue, D.; Li, C.; Giaccone, G.; Li, S.; Gao, L.; et al. Circulating Tumor DNA Detection Is Correlated to Histologic Types in Patients with Early-Stage Non-Small-Cell Lung Cancer. *Lung Cancer* **2019**, *134*, 108–116, doi:10.1016/j.lungcan.2019.05.034.
143. Zulato, E.; Attili, I.; Pavan, A.; Nardo, G.; Del Bianco, P.; Boscolo Bragadin, A.; Verza, M.; Pasqualini, L.; Pasello, G.; Fassan, M.; et al. Early Assessment of KRAS Mutation in CfDNA Correlates with Risk of Progression and Death in Advanced Non-Small-Cell Lung Cancer. *Br. J. Cancer* **2020**, *123*, 81–91, doi:10.1038/s41416-020-0833-7.
144. Fiala, O.; Baxa, J.; Svaton, M.; Benesova, L.; Ptackova, R.; Halkova, T.; Minarik, M.; Hosek, P.; Buresova, M.; Finek, J.; et al. Combination of Circulating Tumour DNA and ^{18}F-FDG PET/CT for Precision Monitoring of Therapy Response in Patients With Advanced Non-Small Cell Lung Cancer: A Prospective Study. *Cancer Genomics - Proteomics* **2022**, *19*, 270–281, doi:10.21873/cgp.20319.
145. Lam, V.K.; Zhang, J.; Wu, C.C.; Tran, H.T.; Li, L.; Diao, L.; Wang, J.; Rinsurongkawong, W.; Raymond, V.M.; Lanman, R.B.; et al. Genotype-Specific Differences in Circulating Tumor DNA Levels in Advanced NSCLC. *J. Thorac. Oncol.* **2021**, *16*, 601–609, doi:10.1016/j.jtho.2020.12.011.
146. Morbelli, S.; Alama, A.; Ferrarazzo, G.; Coco, S.; Genova, C.; Rijavec, E.; Bongioanni, F.; Biello, F.; Dal Bello, M.G.; Barletta, G.; et al. Circulating Tumor DNA Reflects Tumor Metabolism Rather Than Tumor Burden in Chemotherapy-Naive Patients with Advanced Non–Small Cell Lung Cancer: ^{18}F-FDG PET/CT Study. *J. Nucl. Med.* **2017**, *58*, 1764–1769, doi:10.2967/jnumed.117.193201.
147. González de Aledo-Castillo, Jm.; Casanueva-Eliceiry, S.; Soler-Perromat, A.; Fuster, D.; Pastor, V.; Reguart, N.; Viñolas, N.; Reyes, R.; Vollmer, I.; Paredes, P.; et al. Cell-Free DNA Concentration and Fragment Size Fraction Correlate with FDG PET/CT-Derived Parameters in NSCLC Patients. *Eur. J. Nucl. Med. Mol. Imaging* **2021**, doi:10.1007/s00259-021-05306-2.
148. Nygaard, A.D.; Holdgaard, P.C.; Spindler, K.-L.G.; Pallisgaard, N.; Jakobsen, A. The Correlation between Cell-Free DNA and Tumour Burden Was Estimated by PET/CT in Patients with Advanced NSCLC. *Br. J. Cancer* **2014**, *110*, 363–368, doi:10.1038/bjc.2013.705.
149. Hyun, M.H.; Lee, E.S.; Eo, J.S.; Kim, S.; Kang, E.J.; Sung, J.S.; Choi, Y.J.; Park, K.H.; Shin, S.W.; Lee, S.Y.; et al. Clinical Implications of Circulating Cell-Free DNA Quantification and Metabolic Tumor Burden in Advanced Non-Small Cell Lung Cancer. *Lung Cancer* **2019**, *134*, 158–166, doi:10.1016/j.lungcan.2019.06.014.
150. Woff, E.; Kehagias, P.; Vandeputte, C.; Ameye, L.; Guiot, T.; Paesmans, M.; Hendlisz, A.; Flamen, P. Combining ^{18}F-FDG PET/CT–Based Metabolically Active Tumor Volume and Circulating Cell-Free DNA Significantly Improves Outcome Prediction in Chemorefractory Metastatic Colorectal Cancer. *J. Nucl. Med.* **2019**, *60*, 1366–1372, doi:10.2967/jnumed.118.222919.
151. Bartman, C.R.; Weilandt, D.R.; Shen, Y.; Lee, W.D.; Han, Y.; TeSlaa, T.; Jankowski, C.S.R.; Samarah, L.; Park, N.R.; da Silva-Diz, V.; et al. Slow TCA Flux and ATP Production in Primary Solid Tumours but Not Metastases. *Nature* **2023**, doi:10.1038/s41586-022-05661-6.
152. Tie, J.; Cohen, J.D.; Lahouel, K.; Lo, S.N.; Wang, Y.; Kosmider, S.; Wong, R.; Shapiro, J.; Lee, M.; Harris, S.; et al. Circulating Tumor DNA Analysis Guiding Adjuvant Therapy in Stage II Colon Cancer. *N. Engl. J. Med.* **2022**, *386*, 2261–2272, doi:10.1056/NEJMoa2200075.
153. Rolfo, C.D.; Madison, R.; Pasquina, L.W.; Brown, D.W.; Huang, Y.; Hughes, J.D.; Oxnard, G.R.; Husain, H. Utility of CtDNA Tumor Fraction to Inform Negative Liquid Biopsy (LBx) Results

and Need for Tissue Reflex in Advanced Non-Small Cell Lung Cancer (ANSCLC). *J. Clin. Oncol.* **2023**, *41*, 9076–9076, doi:10.1200/JCO.2023.41.16_suppl.9076.
154. Gray, J.E.; Okamoto, I.; Sriuranpong, V.; Vansteenkiste, J.; Imamura, F.; Lee, J.S.; Pang, Y.-K.; Cobo, M.; Kasahara, K.; Cheng, Y.; et al. Tissue and Plasma EGFR Mutation Analysis in the FLAURA Trial: Osimertinib versus Comparator EGFR Tyrosine Kinase Inhibitor as First-Line Treatment in Patients with EGFR-Mutated Advanced Non–Small Cell Lung Cancer. *Clin. Cancer Res.* **2019**, *25*, 6644–6652, doi:10.1158/1078-0432.CCR-19-1126.
155. Papadimitrakopoulou, V.A.; Han, J.; Ahn, M.; Ramalingam, S.S.; Delmonte, A.; Hsia, T.; Laskin, J.; Kim, S.; He, Y.; Tsai, C.; et al. Epidermal Growth Factor Receptor Mutation Analysis in Tissue and Plasma from the AURA3 Trial: Osimertinib versus Platinum-pemetrexed for T790M Mutation-positive Advanced Non–Small Cell Lung Cancer. *Cancer* **2020**, *126*, 373–380, doi:10.1002/cncr.32503.

Paper I-IV

Paper I

Experimental and Molecular Pathology 112 (2020) 104347

Contents lists available at ScienceDirect

Experimental and Molecular Pathology

journal homepage: www.elsevier.com/locate/yexmp

The relevance of tumor mutation profiling in interpretation of NGS data from cell-free DNA in non-small cell lung cancer patients

Anine Larsen Ottestad[a,b,*], Sissel G.F. Wahl[a,c], Bjørn Henning Grønberg[a,b], Frank Skorpen[a,b], Hong Yan Dai[a,c]

[a] Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
[b] Cancer Clinic, St. Olav's Hospital, Trondheim University Hospital, Trondheim, Norway
[c] Department of Pathology, Clinic of Laboratory Medicine, St. Olav's Hospital, Trondheim University Hospital, Trondheim, Norway

ARTICLE INFO

Keywords:
Non-small cell lung cancer (NSCLC)
Next-generation sequencing (NGS)
Circulating tumor DNA (ctDNA)

ABSTRACT

Studies have indicated that detection of circulating tumor DNA (ctDNA) prior to treatment is a negative prognostic marker in non-small cell lung cancer (NSCLC). ctDNA is currently identified by detection of tumor mutations. Commercial next-generation sequencing (NGS) assays for mutation analysis of ctDNA for routine practice usually include small gene panels and are not suitable for general mutation analysis. In this study, we investigated whether mutation analysis of cfDNA could be performed using a commercially available comprehensive NGS gene panel and bioinformatics workflow. Tumor DNA, plasma DNA and peripheral blood leukocyte DNA from 30 NSCLC patients were sequenced. In two patients (7%), tumor mutations in cfDNA were immediately called by the bioinformatic workflow. In 13 patients (43%), tumor mutations were not called, but were present in ctDNA and were identified based on the known tumor mutation profile. In the remaining 15 patients (50%), no concordant mutations were detected. In conclusion, we were able to identify tumor mutations in ctDNA from 57% of NSCLC patients using a comprehensive gene panel. We demonstrated that sequencing paired tumor DNA was helpful to interpret data and confirm ctDNA, and thus increased the ratio of patients with detectable ctDNA. This approach might be feasible for mutation analysis of ctDNA in routine diagnostic practice, especially in case of suboptimal plasma quality and quantity.

1. Introduction

Tumors release DNA, known as circulating tumor DNA, which makes up a small fraction of total cell-free DNA (cfDNA) in the blood (Stroun et al., 2001). Many studies have shown that tumor mutations can be detected in cfDNA from patients with non-small cell lung cancer (NSCLC) (Bettegowda et al., 2014; Phallen et al., 2017). It has also been shown that pre-treatment detection of tumor mutations in cfDNA is a negative prognostic factor. cfDNA is thereby emerging as an important biomarker.

A requirement for using cfDNA as a prognostic biomarker is to detect at least one tumor mutation in cfDNA prior to treatment. Because the mutation spectrum is diverse in NSCLC, it is often necessary to analyze many different genes to ensure mutation detection (Ono et al., 2019). Commercial cfDNA next-generation sequencing (NSG) assays for routine practice usually contain small gene panels with mainly targetable genes which are not suitable for this purpose.

Mutation analysis of cfDNA using a large gene panel is challenging. It generates large data sets and sometimes the data quality is poor. Suboptimal quality and quantity of plasma are the main factors that affect the quality of sequencing data due to insufficient amount and low quality of input DNA used for NGS library preparation. This is a frequent issue, especially in the routine practice.

It is promising that recent studies using large gene panels found concordant mutation in tumor DNA and cfDNA in 50–100% of early stage NSCLC patients (Phallen et al., 2017; Abbosh et al., 2017; Chaudhuri et al., 2016). In these studies, customized gene panels and own-developed bioinformatic workflows were used for mutation detection and interpretation of cfDNA sequencing data. These approaches may not be directly transferable to most routine diagnostic laboratories in which commercial assays are usually applied.

In this study we used a commercial comprehensive gene panel that included hot spots in 275 genes. We sequenced matched tumor DNA, peripheral blood leukocyte (PBL) DNA and cfDNA from 30 patients with NSCLC. The aim was to investigate whether tumor mutations could be detected in cfDNA using a large commercially available gene panel and

* Corresponding author at: The Cancer Clinic, St. Olav's Hospital, Trondheim University Hospital, P.O. Box 3250 Sluppen, 7006 Trondheim, Norway.
E-mail address: anine.ottestad@gmail.com (A.L. Ottestad).

https://doi.org/10.1016/j.yexmp.2019.104347
Received 7 August 2019; Received in revised form 11 November 2019; Accepted 19 November 2019
Available online 21 November 2019
0014-4800/ © 2019 Elsevier Inc. All rights reserved.

bioinformatic workflow.

2. Methods

2.1. Patient material and approvals

Tumor tissue and blood samples were retrieved from Biobank1, a local lung cancer biobank of tumor tissue, cytological specimens, blood samples and clinical data from > 900 patients with all stages and subtypes of lung cancer. The biobank is approved by the Norwegian Regional Committee for Medical and Health Research Ethics (REC) Central, the Norwegian Health Department, and the Norwegian Data Protection Authority. The REC central has approved the present study.

Patients with NSCLC and both tumor tissue and blood samples available for histological examination and for DNA extraction were included in this study. Blood samples were collected before treatment commenced. Tumor specimens were reviewed, classified and subtyped according to the 2015 World Health Organization classification of lung tumors by a lung cancer pathologist (SGFW) (Travis et al., 2015). Disease stage was assessed according to the 8th TNM Classification of Malignant Tumors for lung cancer (Brierley et al., 2017).

2.2. DNA extraction

Formalin-fixed (formaldehyde solution 4% phosphate buffered) paraffin-embedded (FFPE) tumor blocks were used for isolation of tumor DNA. Tissue sections of 10 μm were cut from areas with the highest proportion of tumor cells. Tumor cell content ranged from 5 to 50%. From seven tumors, two replicating DNA extracts were prepared by cutting the same tumor tissue twice. Both DNA extracts were sequenced. DNA was extracted using GeneRead DNA FFPE kit (Qiagen, Valencia, CA) or QIAamp DNA FFPE Tissue kit (Qiagen, Valencia, CA) according to the respective protocols. Treatment of DNA with uracil-DNA-glycosylase was performed either during extraction or after final elution to remove uracil. Spontaneous deamination of cytosine to uracil occurs over time, and this would lead to C > T and G > A sequence artifacts.

Plasma was prepared from 10 mL whole blood with EDTA or citrate anticoagulant immediately after sampling. Blood samples were either centrifuged once at 2500 ×*g* for 10 min, or first at 1500 ×*g* for 15 min and then at 10,000 ×*g* for 10 min. Plasma was transferred to cryotubes and stored at −80 °C. cfDNA was isolated using QIAamp Circulating Nucleic Acid kit (Qiagen, Valencia, CA) from 2.5–6 mL plasma and eluted with 30–50 μL supplied buffer. It was further concentrated by precipitation with 0.3 M sodium acetate, 0.05–1 μg/μL glycogen (Thermo Fisher Scientific, Waltham, MA) and 2.5× volume of 96% ethanol and finally resuspended in 10 μL molecular biology grade water.

DNA from PBLs was extracted from whole blood using QIAsymphony DSP DNA kit (Qiagen, Valencia, CA). DNA concentration was measured by Qubit® (Thermo Fisher Scientific, Waltham, MA), using dsDNA HS Assay Kit for cfDNA and dsDNA BR Assay Kit for tumor DNA and PBL DNA.

2.3. NGS library preparation

NGS libraries were prepared from 30 to 210 ng tumor DNA, 20 ng cfDNA and 20–40 ng PBL DNA using QIAseq Comprehensive Cancer Panel (Qiagen, Valencia, CA). The panel included 836,670 bases in 275 genes (Table S4). Briefly, DNA was fragmented, end-repaired and a 5′ adenine overhang was made. Synthetic fragments with a sequencing adapter, unique molecular index (UMI, a 12-nucleotide random sequence) and a 3′ thymine overhang were then ligated to all DNA fragments. An estimated size of 4^{12} different UMI sequence combinations ensured that all DNA fragments were tagged with a unique UMI.

In the first polymerase chain reaction (PCR) of six cycles, PCR products covering the genomic areas of interest were synthesized using gene-specific primers and primers complementary to the 5′ sequencing adapter. There were a total of 11,311 different gene-specific primers, which all had a common sequence, a primer site, at their 5′ end. The PCR products were then used as templates for the second PCR, in which complete NGS libraries were made using primers for the 5′ sequencing adapters and 3' sequencing adapters complementary to the common 5′ end of the templates.

Tumor DNA libraries and cfDNA libraries were pooled separately with 5 μL from each individual tumor DNA and cfDNA library, respectively. The pooled libraries were then purified using QIAquick PCR Purification kit (Qiagen, Valencia, CA). To acquire a library pool with even and proper fragment lengths, electrophoresis of both library pools was performed using DNA 300 Chip (Caliper Life Sciences, Hopkinton, MA) on LabChip XT (Caliper Life Sciences, Hopkinton, MA). Fragments with a length of 295 to 445 bp were extracted according to the procedure “Extract and Pause”. The two pools were then quantified using KAPA Library Quantification Kit – Illumina ABI Prism19 (KAPA Biosciences, Wilmington, MA). Libraries were sequenced on NextSeq 550 (Illumina, San Diego, CA) with 151 paired-end reading. NGS libraries of PBL DNA were prepared, but without the step of fragment purification, and sequenced separately in the same manner as tumor DNA and cfDNA.

2.4. Bioinformatic analysis

Biomedical Genomics Workbench version 5.0.1 (Qiagen, Valencia, CA) with the gene panel-specific plugin QIAseq Targeted Panel Analysis version 1.2 was applied for detection of variants. Briefly, adapters and UMIs were removed and the reads annotated with their UMI to allow mapping to the reference genome (hg19). Reads with same UMI (i.e PCR duplicates) were grouped into a “UMI family”. A consensus sequence was assembled from each UMI family with the variants that were present in ≥75% of the duplicates. The subsequent steps were performed with the consensus sequence. Sequences of two DNA fragments that were generated during adapter/UMI-ligation in the NGS library preparation, were removed. Inserts and deletions were then detected, followed by local realignment and removal of primer sequences. Variants were then called in the panel target area with the tool “Low frequency variant detection”.

For each patient, variants that were detected in ≥10 different UMI families in PBL DNA were removed from both tumor DNA and cfDNA. Tumor mutations were defined as follows; coding, non-synonymous, not located in a homopolymer, detected equally in read 1 and 2, and detected in ≥5 big UMI families (defined as families with ≥3 duplicates). cfDNA mutations were defined as follows; coding, non-synonymous, not located in a homopolymer, detected equally in read 1 and 2, and detected in ≥3 big UMI families.

The lists of mutations in tumor DNA passing these filters were uploaded to SNPnexus to evaluate pathogenicity (Dayem Ullah et al., 2018). A mutation was classified as pathogenic if it was registered in the catalogue of somatic mutations in cancer (COSMIC) (Tate et al., 2019) or predicted to be pathogenic by both SIFT (Kumar et al., 2009) and PolyPhen (Adzhubei et al., 2010). Only mutations classified as pathogenic were included for concordance analysis.

Mutation spectrum was compared between paired tumor DNA and cfDNA. Tumor mutation spectrum from two synchronous tumors were separately compared to the matched cfDNA in one patient (patient 6). The BAM files created after UMI family creation in the cfDNA samples were manually inspected for presence of matched tumor mutations that were not called by the bioinformatic workflow.

Cases with no concordant mutations were classified as undetected. Patients with concordant mutations were divided into two categories based on the method of detection. Cases with concordant mutations that were immediately called by the bioinformatic workflow in cfDNA were classified in the first category. Cases where concordant mutations

Table 1
Characteristic of the patients analyzed in the study.

Characteristics	Number of patients (%)
In total	30
Age, years:	
Mean: 70	
Range: 57–81	
Sex:	
Female	11 (37)
Male	19 (63)
Smoking history:	
Smoker/former smoker	28 (93)
Never smoker	2 (7)
NSCLC subtype:	
Adenocarcinoma	15 (50)
Squamous cell carcinoma	13 (43)
Adenosquamous carcinoma	1 (3)
NSCLC-not otherwise specified	1 (3)

NSCLC; non-small cell lung cancer

were identified in cfDNA with prior knowledge from tumor DNA were classified in the other category.

3. Results

Thirty-one patients diagnosed with NSCLC at St. Olav's Hospital, Trondheim University Hospital, Norway between 2009 and 2016 were included (Table S1). One patient was excluded from the analyses because of failed PBL DNA library sequencing. Patient characteristics of the remaining 30 patients are shown in Table 1. Histological diagnoses were adenocarcinoma (n = 15, 50%), squamous cell carcinoma (n = 13, 43%), adenosquamous carcinoma (n = 1, 3%) and NSCLC-not otherwise specified (NSCLC-NOS, n = 1, 3%). Twenty-two patients (77%) had stage I–II disease and seven patients (23%) had stage III–IV.

Sequencing coverage, i.e. the mean number of sequencing reads covering each position in the target area, was on average 2394× (range 1031×–4338×) across all tumor DNA samples, 2143× (range 1055×–3625×) across all cfDNA samples and 748× (range 515×–1630×) across all PBL DNA samples (Table S2).

The UMI family consensus sequence represents one original DNA fragment, and by extension, one haploid genome. Mean haploid genome coverage was on average 743× (range 225×–1643×) in all tumor DNA samples, 207× (range 69×–443×) in all cfDNA samples and 493× (range 209×–1061×) in all PBL DNA samples (Table S2). The median number of PCR duplicates in each UMI family was ≤3 in all samples.

3.1. Mutation analyses

At least one mutation was detected in tumor from 29/30 patients (97%), with a mean allele frequency (AF) of 24.4% (range 1.4–73.8%). Tumors that were sequenced in duplicates contained both concordant and discordant variants, but pathogenic mutations were concordant in all cases. Patient 6 was the only exception. Two synchronous tumors were sequenced from this patient and no mutations were concordant between the two tumors.

Mutations detected in tumor DNA were compared to those detected in the matching cfDNA (Fig. 1, Table S3). Tumor mutations were immediately called by the bioinformatic workflow in two patients. In one of these patients, all four tumor mutations were detected in cfDNA with a mean AF of 8.6% (range 5.5–12.2%) These mutations were detected in 17 different UMI families on average (range 13–24). In the other patient, nine out of ten tumor mutations were detected in cfDNA with a mean AF of 5.4% (range 2.6–12.2%). The mutations were detected in 8 UMI families on average (range 3–13). Variants were called by the bioinformatic workflow in the remaining 28 patients, but these were not confirmed in the matching tumor DNA.

Next, we manually investigated if tumor mutations were present in cfDNA but not called by the bioinformatic workflow. Through this approach, we identified tumor mutations in cfDNA from another 14 patients. The median AF of tumor mutations in cfDNA was 0.9% (range 0.26–15.2%). One example was an L858R mutation in *Epidermal Growth Factor Receptor* (*EGFR*) that was detected in tumor DNA from patient 12. It was detected in two UMI families in the matching cfDNA (AF 1.0%). One family was made from two reads, the other was a singleton UMI. In both cases, the mutation was only detected in read 1 because it was located in a non-overlapping area in the DNA fragment. Another example was a mutation in *Signal Transducer And Activator Of Transcription 3* (*STAT3*) that was detected in tumor DNA from patient 2. The same mutation was detected in four UMI families in the matching cfDNA (AF 2.2%). These UMI families were made from total 11 reads.

3.2. ctDNA and patient characteristics

In total, concordant mutations were detected in 15/29 patients (52%). Eight patients (53%) had squamous cell carcinoma, five (33%) had adenocarcinoma, one (6%) had NSCLC-NOS and one (6%) had adenosquamous carcinoma. Necrosis was observed in tumor from nine out of 14 patients (64%) that had evaluable tumor tissue and concordant mutations, and in four out of 14 patients (29%) with no concordant mutations. Concordant mutations were detected in 43% of patients with stage I-II disease and in 67% of patients with stage III–IV disease. In patient 6 with two synchronous tumors, a mutation from one tumor was identified in cfDNA by manual inspection. The mean haploid genome coverage in this patient was 116×.

4. Discussion

In this study of 30 NSCLC-patients, we sequenced matched tumor DNA and plasma cfDNA and compared mutation profiles. At least one concordant mutation was detected in 15/29 patients (52%) with mutations in tumor DNA. In 2/15 patients, tumor mutations were immediately called by the bioinformatic workflow in cfDNA. These cases were also different in that virtually all tumor mutations were detected in cfDNA. In 13/15 patients, concordant mutations were present in cfDNA, but were not called by the bioinformatic workflow. These mutations could be identified because they were detected in the matching tumor DNA. The remaining 14 patients (47%) had mutations in their tumor DNA, but these mutations were not detected in the matching cfDNA.

The result demonstrates the benefit of using a large gene panel for sequencing tumor DNA from NSCLC patients. In a previous study we used a 26-gene panel and detected mutations in only a subset of tumors from NSCLC patients *(unpublished data)*, and studies that used panels of < 60 genes detected mutations in 63–78% of patients with early stage NSCLC (Phallen et al., 2017; Guo et al., 2018; Chen et al., 2017; Chen et al., 2016; Guo et al., 2016). In this study, we detected at least one tumor tissue mutation in 97% of patients with the 275-gene panel. Similarly, a recent study detected mutations in 94% of NSCLC tumor samples with a 546-gene panel (Zhang et al., 2019).

We observed mutation concordance in tumor DNA and cfDNA in 52% of patients. This is consistent with previous research. Studies on stage I-III NSCLC reported concordance in 33–50% of patients (Phallen et al., 2017; Abbosh et al., 2017; Guo et al., 2016, 2018; Chen et al., 2016, 2017; Zhang et al., 2019), though, Chaudhuri et al. (Zhang et al., 2019) detected concordant mutations in 100% of patients with available tumor tissue using their Cancer Personalized Profiling by deep Sequencing (CAPP-Seq) assay. A possible explanation is that all mutations, not only known pathogenic or driver mutations, were included in the study.

Most patients with concordant mutations in our study had squamous cell carcinoma histology. The study by Abbosh et al. (2017) also found that non-adenocarcinoma histology was an independent

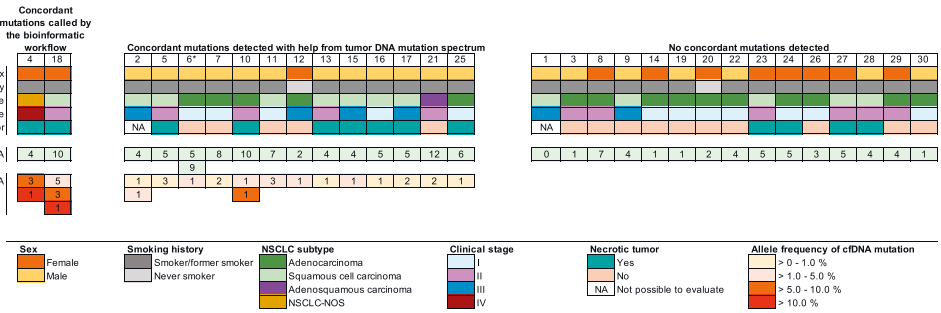


Fig. 1. Summary of patient characteristics and concordance in tumor DNA and matched plasma cell-free DNA (cfDNA). Matched tumor DNA, peripheral blood leukocyte DNA and cfDNA were sequenced from 30 patients with non-small cell lung cancer. Patients were divided into three categories. Cases with concordant mutations that were called by the bioinformatic workflow in cfDNA were classified in the first category (n = 2). Cases where concordant mutations were detectable in cfDNA only with prior knowledge from tumor DNA where classified in the second category (n = 13). Patients with no detected concordant mutations were classified in the last category (n = 14). One patient had no detectable mutations in tumor (Patient 1). *Patient 6 had two synchronous tumors that were sequenced separately. cfDNA; cell-free DNA, NSCLC-NOS; non-small cell lung cancer - not otherwise specified.

predictor of mutation detection in cfDNA. These tumors are more necrotic and thereby release more DNA (Stroun et al., 2001; Caruso et al., 2012). In line with this reasoning, we found that 69% of evaluable squamous cell carcinomas were necrotic compared to 20% of adenocarcinomas.

Although the large gene panel ensured mutation detection in most patients, applying large panels for cfDNA sequencing is challenging for the bioinformatic analysis. The main issue is the low fraction of circulating tumor DNA in cfDNA, and high haploid genome coverage is necessary to detect the variants with low AF. The average haploid genome coverage of most cfDNA samples in our study was not high (mean 207×). It is possible that size selection of the cfDNA NGS libraries excluded fragments from the tumor since circulating tumor DNA is shorter than other cfDNA (Mouliere et al., 2011). Tumor mutations were called only in the two cfDNA samples with relatively high fraction of tumor DNA. The median AF of the called tumor mutations was 5.5%, while the median AF of the mutations identified with help from the matching tumor DNA was 0.9%.

Variants were called in all cfDNA samples, but only in two patients the variants were confirmed in the matching tumor tissue. In the remaining samples, the called variants could not be confirmed in the matching tumor tissue. The origin of these variants is unknown. The same observation has been reported in other studies (Guo et al., 2018; Jamal-Hanjani et al., 2016). Although such variants have been attributed to tumor heterogeneity, a recent study found that most discordant mutations were technical artifacts (Stetson et al., 2019). Some suggest that such variants originate from clonal hematopoiesis of indeterminate potential (Hu et al., 2018).

In this study, no AF cut-off was applied in the bioinformatic filtering, but rather a criterium that the mutation must be present in ≥3 UMI families. This is stringent and a dynamic AF cut-off could be applied. However, finding the optimal cut-off is challenging, especially in data sets from large gene panels. Low AF cut-off setting generates big data sets with too many variants that makes it difficult to identify true tumor mutations. Suboptimal sequencing data with low coverage is even more difficult to interpret. High AF cut-off setting will miss calling of true tumor mutations and thus decrease the sensitivity of detection.

The low number of duplicates in each UMI family was a drawback in this study. Tumor mutations in cfDNA in most patients did not pass the quality filters with the set cut-off for number of duplicates (≥3) but were identified directly in the UMI sequences because the true tumor mutation profile was known. Therefore, through this study we showed that it was still possible to identify true tumor mutations in cfDNA by using the mutation information of matching tumor DNA in spite of suboptimal cfDNA samples and sequencing data.

Several factors can result in suboptimal cfDNA sequencing data, such as access to limited plasma quantity and variable plasma quality. This directly affects the amount and quality of input DNA for NGS library preparation, which further affects the NGS library quality. Low sequencing coverage may also lead to suboptimal data but increasing sequencing coverage may not always be an option due to limited resources.

In conclusion, we were able to identify tumor mutations in cfDNA from NSCLC patients using a commercially available, comprehensive NGS gene panel and bioinformatic workflow. We also show that it is possible to obtain mutation information from suboptimal cfDNA sequencing data by sequencing tumor DNA and PBL DNA from the same patients.

Acknowledgement

The library preparation and sequencing were provided by the Genomics Core Facility (GCF), Norwegian University of Science and Technology (NTNU). GCF is funded by the Faculty of Medicine and Health Sciences at NTNU and The Central Norway Regional Health Authority. This study was supported, in part, by The Central Norway Regional Health Authority and the Joint Research Committee between St. Olav's hospital and the Faculty of Medicine and Health Sciences, NTNU.

F.S, H.D and B.H.G designed and planned the study. S.G.F.W evaluated all histological material. A.L.O optimized experimental protocol and performed experiments together with GCF. A.L.O analyzed data and interpreted data with B.H.G, S.G.F.W and H.D. A.L.O wrote the manuscript and incorporated feedback from all authors. A.L.O is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

This work was supported by The Central Norway Regional Health Authority and the Joint Research Committee between St. Olav's hospital and the Faculty of Medicine and Health Sciences, NTNU.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yexmp.2019.104347.

References

Abbosh, C., Birkbak, N.J., Wilson, G.A., Jamal-Hanjani, M., Constantin, T., Salari, R., Le Quesne, J., Moore, D.A., Veeriah, S., Rosenthal, R., Marafioti, T., Kirkizlar, E.,

Watkins, T.B.K., McGranahan, N., Ward, S., Martinson, L., Riley, J., Fraioli, F., Al Bakir, M., Grönroos, E., Zambrana, F., Endozo, R., Bi, W.L., Fennessy, F.M., Sponer, N., Johnson, D., Laycock, J., Shafi, S., Czyzewska-Khan, J., Rowan, A., Chambers, T., Matthews, N., Turajlic, S., Hiley, C., Lee, S.M., et al., 2017. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature 545, 446–451.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–249.

Bettegowda, C., Sausen, M., Leary, R.J., Kinde, I., Wang, Y., Agrawal, N., Bartlett, B.R., Wang, H., Luber, B., Alani, R.M., Antonarakis, E.S., Azad, N.S., Bardelli, A., Brem, H., Cameron, J.L., Lee, C.C., Fecher, L.A., Gallia, G.L., Gibbs, P., Le, D., Giuntoli, R.L., Goggins, M., Hogarty, M.D., Holdhoff, M., Hong, S.-M., Jiao, Y., Juhl, H.H., Kim, J.J., Siravegna, G., Laheru, D.A., Lauricella, C., Lim, M., Lipson, E.J., Marie, S.K.N., Netto, G.J., et al., 2014. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci. Transl. Med. 6 (224ra24-224ra24).

Brierley, J., Gospodarowicz, M.K., Wittekind, C. (Eds.), 2017. TNM Classification of Malignant Tumours, 8th ed. John Wiley & Sons, Inc, Chichester, West Sussex, UK; Hoboken, NJ.

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

Chaudhuri, A.A., Lovejoy, A.F., Chabon, J.J., Newman, A., Stehr, H., Say, C., Aggarwal, S., Carter, J.N., West, R.B., Neal, J.W., Wakelee, H.A., Loo, B.W., Alizadeh, A., Diehn, M., 2016. CAPP-Seq circulating tumor DNA analysis for early detection of tumor progression after definitive radiation therapy for lung cancer. Int. J. Radiat. Oncol. 96, S41–S42.

Chen, K., Zhang, J., Guan, T., Yang, F., Lou, F., Chen, W., Zhao, M., Zhang, J., Chen, S., Wang, J., 2017. Comparison of plasma to tissue DNA mutations in surgical patients with non–small cell lung cancer. J. Thorac. Cardiovasc. Surg. 154, 1123–1131.e2.

Chen, K.-Z., Lou, F., Yang, F., Zhang, J.-B., Ye, H., Chen, W., Guan, T., Zhao, M.-Y., Su, X.-X., Shi, R., Jones, L., Huang, X.F., Chen, S.-Y., Wang, J., 2016. Circulating tumor DNA detection in early-stage non-small cell lung cancer patients by targeted sequencing. Sci. Rep. 6.

Dayem Ullah, A.Z., Oscanoa, J., Wang, J., Nagano, A., Lemoine, N.R., Chelala, C., 2018. SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. Nucleic Acids Res. 46, W109–W113.

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

Guo, Q., Wang, J., Xiao, J., Wang, L., Hu, X., Yu, W., Song, G., Lou, J., Chen, J., 2018. Heterogeneous mutation pattern in tumor tissue and circulating tumor DNA warrants parallel NGS panel testing. Mol. Cancer 17, 131.

Hu, Y., Ulrich, B.C., Supplee, J., Kuang, Y., Lizotte, P.H., Feeney, N.B., Guibert, N.M., Awad, M.M., Wong, K.-K., Jänne, P.A., Paweletz, C.P., Oxnard, G.R., 2018. False-positive plasma genotyping due to clonal hematopoiesis. Clin. Cancer Res. 24, 4437–4443.

Jamal-Hanjani, M., Wilson, G.A., Horswell, S., Mitter, R., Sakarya, O., Constantin, T., Salari, R., Kirkizlar, E., Sigurjonsson, S., Pelham, R., Kareht, S., Zimmermann, B., Swanton, C., 2016. Detection of ubiquitous and heterogeneous mutations in cell-free DNA from patients with early-stage non-small-cell lung cancer. Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. 27, 862–867.

Kumar, P., Henikoff, S., Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4, 1073–1081.

Mouliere, F., Robert, B., Arnau Peyrotte, E., Del Rio, M., Ychou, M., Molina, F., Gongora, C., Thierry, A.R., 2011. High fragmentation characterizes tumour-derived circulating DNA. PLoS One 6, e23418.

Ono, A., Isaka, M., Serizawa, M., Omae, K., Kojima, H., Nakashima, K., Omori, S., Wakuda, K., Kenmotsu, H., Naito, T., Murakami, H., Urakami, K., Nagashima, T., Sugino, T., Kusuhara, M., Takahashi, T., Yamaguchi, K., Ohde, Y., 2019. Genetic alterations of driver genes as independent prognostic factors for disease-free survival in patients with resected non-small cell lung cancer. Lung Cancer 128, 152–157.

Phallen, J., Sausen, M., Adleff, V., Leal, A., Hruban, C., White, J., Anagnostou, V., Fiksel, J., Cristiano, S., Papp, E., Speir, S., Reinert, T., Orntoft, M.-B.W., Woodward, B.D., Murphy, D., Parpart-Li, S., Riley, D., Nesselbush, M., Sengamalay, N., Georgiadis, A., Li, Q.K., Madsen, M.R., Mortensen, F.V., Huiskens, J., Punt, C., van Grieken, N., Fijneman, R., Meijer, G., Husain, H., Scharpf, R.B., Diaz, L.A., Jones, S., Angiuoli, S., Ørntoft, T., Nielsen, H.J., et al., 2017. Direct detection of early-stage cancers using circulating tumor DNA. Sci. Transl. Med. 9 (eaan2415).

Stetson, D., Ahmed, A., Xu, X., Nuttall, B.R.B., Lubinski, T.J., Johnson, J.H., Barrett, J.C., Dougherty, B.A., 2019. Orthogonal comparison of four plasma NGS tests with tumor suggests technical factors are a major source of assay discordance. JCO Precis Oncol. 1–9.

Stroun, M., Lyautey, J., Lederrey, C., Olson-Sand, A., Anker, P., 2001. About the possible origin and mechanism of circulating DNA apoptosis and active DNA release. Clin. Chim. Acta Int. J. Clin. Chem. 313, 139–142.

Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S.C., Kok, C.Y., Noble, K., Ponting, L., Ramshaw, C.C., Rye, C.E., Speedy, H.E., Stefancsik, R., Thompson, S.L., Wang, S., Ward, S., Campbell, P.J., Forbes, S.A., 2019. COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res. 47, D941–D947.

Travis, W.D., Brambilla, E., Burke, A., Marx, A., Nicholson, A. (Eds.), 2015. WHO Classification of Tumours of Lung, Pleura, Thymus and Heart, 4 ed. International Agency for Research on Cancer, Lyon, France.

Zhang, B., Niu, X., Zhang, Q., Wang, C., Liu, B., Yue, D., Li, C., Giaccone, G., Li, S., Gao, L., Zhang, H., Wang, J., Yang, H., Wu, R., Ni, P., Wang, C., Ye, M., Liu, W., 2019. Circulating tumor DNA detection is correlated to histologic types in patients with early-stage non-small-cell lung cancer. Lung Cancer. https://doi.org/10.1016/j.lungcan.2019.05.034.

Table S1. Patient characteristics of the 30 analyzed patients

Patient	Sex	Age at diagnosis	Smoker/former smoker	Histology*	Tissue specimen	Diameter of primary tumor (mm) †	pTNM 8	Clinical stage	Necrotic tumor	Days between biopsy/surgery to blood draw	Tumor cell content (%)
1	Male	61	Yes	SCC	Biopsy	63	pT3	3	Not evaluable	0	50
2	Male	70	Yes	SCC	Biopsy	60	T3N1	3	Not evaluable	19	60-70
3	Male	68	Yes	ADC	Resected	43	pT2bN0	2	No	155	70-90
4	Female	79	Yes	NSCLC-NOS	Biopsy	79	pT4NXM1c	4	Yes	0	25-50
5	Male	78	Yes	SCC	Resected	30	pT3N0	2	Yes	15	40-50
6‡	Male	61	Yes	ADC	Resected	12/22	pT2aN0/pT1bN0	1	No/No	36	40-50 / 10-20
7	Male	71	Yes	ADC	Resected	15	pT1bN0	1	No	15	70-80
8	Female	73	Yes	ADC	Resected	45	pT2bN0	2	No	52	50-60
9	Male	79	Yes	SCC	Resected	80	pT4N0 (cT3N0)	3	No	37	70-80
10	Male	66	Yes	ADC	Resected	25	pT2aN1	2	Yes	21	60-80
11	Male	65	Yes	SCC	Resected	17	pT1bN0	1	No	20	30-50
12	Female	81	No	ADC	Resected	18	pT1bN2	3	No	29	40-50
13	Male	65	Yes	SCC	Resected	50	pT2bN0	2	Yes	35	60-80
14	Female	57	Yes	ADC	Resected	19	pT2aN0	1	No	36	40-50
15	Male	68	Yes	SCC	Resected	65	pT3N2	3	Yes	1	50-60
16	Male	68	Yes	SCC	Resected	21	pT1cN0	1	Yes	14	50-60
17	Male	81	Yes	SCC	Resected	27	pT1cN2	3	Yes	0	70-80
18	Female	75	Yes	SCC	Resected	60	pT3N0	2	Yes	1	70-80
19	Male	58	Yes	ADC	Resected	14	pT1bN0	1	No	1	70-80
20	Female	76	No	ADC	Resected	14	pT1bN0	1	No	1	70-80
21	Male	72	Yes	ADSC	Resected	41	pT2bN0	2	No	1	60-80
22	Male	74	Yes	ADC	Resected	7	pT1aN0	1	No	1	5-10
23§	Female	80	Yes	SCC	Resected	18/7	pT3N0	2	Yes	1	80-90
24	Female	71	Yes	ADC	Resected	25	pT1cN0	1	Yes	1	70-80
25	Male	71	Yes	ADC	Resected	25	pT2aN0	1	Yes	22	70-80
26	Female	65	Yes	ADC	Resected	14	pT1bN0	1	No	1	50
27	Female	75	Yes	SCC	Resected	56	pT3N0	2	Yes	1	50-60
28	Male	60	Yes	SCC	Resected	52	pT3N0	2	Yes	1	50-70
29	Female	70	Yes	ADC	Resected	42	pT2bN0	2	No	1	80-90
30	Male	70	Yes	ADC	Resected	18	pT1bN0	1	No	1	80-90

*ADC; adenocarcinoma, SCC; squamous cell carcinoma, ADSC; adenosquamous carcinoma, NSCLC-NOS; non-small cell lung cancer - not otherwise specified

† Determined by pathologist

‡Tissue from two synchronous tumor were sequenced

§Tissue from two synchronous tumors were mixed and sequenced together

Table S2. Sequencing quality

Patient	Sample type	Sequencing coverage	Haploid genome coverage	Number of bases with < 100x UMI coverage	Percent of target area with < 100x UMI coverage	Median number of PCR duplicates/UMI family
1	cfDNA	2,117	443	64,654	8%	1
1	PBL DNA	912	621	41,070	5%	1
1	Tumor	1,992	808	47,932	6%	1
2	Tumor	1,231	329	264,412	32%	1
2	cfDNA	1,140	348	148,284	18%	1
2	PBL DNA	811	478	58,997	7%	1
3	cfDNA	2,731	124	732,043	87%	2
3	PBL DNA	946	678	41,219	5%	1
3	Tumor	3,806	813	24,632	3%	2
4	cfDNA	1,118	219	438,517	52%	2
4	Tumor	1,031	428	157,833	19%	1
4	PBL DNA	916	604	64,978	8%	1
5	cfDNA	1,312	190	470,008	56%	2
5	Tumor	1,032	428	103,400	12%	1
5	PBL DNA	893	673	40,629	5%	1
6	cfDNA	2,033	116	757,860	91%	2
6	PBL DNA	868	558	47,793	6%	1
6	TumorA	2,719	591	41,242	5%	2
6	TumorB	2,366	597	43,239	5%	2
7	cfDNA	2,501	175	554,739	66%	2
7	PBL DNA	903	528	46,986	6%	1
7	Tumor	3,111	1,252	43,684	5%	1
8	cfDNA	1,704	110	781,578	93%	2
8	Tumor	2,278	797	10,736	1%	1
8	PBL DNA	930	797	10,736	1%	1
9	cfDNA	1,055	286	162,226	19%	1
9	PBL DNA	926	679	37,630	4%	1
9	Tumor	3,319	1,024	11,456	1%	1
10	cfDNA	3,024	69	825,836	99%	1
10	Tumor	2,539	379	126,943	15%	2
10	PBL DNA	904	497	49,797	6%	1
11	cfDNA	3,201	86	812,104	97%	1
11	PBL DNA	735	458	76,109	9%	1
11	Tumor	3,514	577	33,341	4%	2
12	cfDNA	1,215	128	751,656	90%	2

12	PBL DNA	771	442	53,479	6%	1
12	Tumor	3,564	1,044	12,448	1%	1
13	Tumor	1,592	225	226,951	27%	2
13	cfDNA	2,673	299	452,337	54%	2
13	PBL DNA	556	381	120,698	14%	1
14	cfDNA	1,847	230	260,158	31%	2
14	PBL DNA	621	443	66,105	8%	1
14	Tumor	2,473	501	36,635	4%	2
15	cfDNA	2,515	167	593,913	71%	3
15	PBL DNA	675	457	68,600	8%	1
15	Tumor	3,389	604	30,180	4%	2
16	cfDNA	2,233	201	516,257	62%	2
16	Tumor	2,199	299	109,065	13%	2
16	PBL DNA	740	462	65,436	8%	1
17	cfDNA	1,936	183	526,152	63%	2
17	PBL DNA	646	405	85,119	10%	1
17	TumorB	2,026	769	30,362	4%	1
17	TumorA	3,218	1,084	12,475	1%	1
18	cfDNA	2,569	171	566,868	68%	2
18	PBL DNA	705	547	41,533	5%	1
18	Tumor	2,135	607	23,531	3%	2
19	cfDNA	1,681	273	241,077	29%	2
19	PBL DNA	680	380	83,743	10%	1
19	TumorB	1,655	666	31,579	4%	1
19	TumorA	2,168	815	8,355	1%	1
20	cfDNA	2,548	177	593,469	71%	2
20	PBL DNA	607	350	74,068	9%	1
20	TumorB	1,504	799	25,012	3%	1
20	TumorA	1,625	811	12,896	2%	1
21	cfDNA	1,972	134	727,892	87%	2
21	PBL DNA	519	377	97,869	12%	1
21	TumorB	1,594	785	23,742	3%	1
21	TumorA	1,600	849	11,248	1%	1
22	cfDNA	1,698	175	621,379	74%	2
22	TumorA	2,620	787	14,209	2%	1
22	TumorB	2,107	1,009	27,043	3%	1
22	PBL DNA	1,630	1,061	20,972	3%	1
23	PBL DNA	515	332	177,336	21%	1
23	cfDNA	2,106	356	71,152	9%	2

23	TumorA	2,303	603	21,454	3%	2
23	TumorB	2,027	780	20,549	2%	1
24	cfDNA	2,886	247	308,218	37%	2
24	PBL DNA	553	358	214,024	26%	1
24	TumorB	1,904	617	30,378	4%	1
24	TumorA	3,325	1,643	4,377	1%	1
25	cfDNA	2,733	252	259,305	31%	2
25	PBL DNA	596	306	140,477	17%	1
25	Tumor	3,689	1,190	13,614	2%	1
26	cfDNA	1,915	272	242,410	29%	2
26	PBL DNA	597	432	58,944	7%	2
26	Tumor	1,728	825	16,079	2%	1
27	PBL DNA	589	209	213,651	26%	2
27	cfDNA	1,415	265	279,228	33%	2
27	Tumor	2,771	649	14,854	2%	2
28	cfDNA	3,625	149	663,773	79%	2
28	PBL DNA	601	499	60,762	7%	1
28	Tumor	4,338	801	11,010	1%	2
29	cfDNA	2,508	196	514,687	62%	2
29	PBL DNA	573	363	73,310	9%	1
29	Tumor	2,493	728	13,765	2%	2
30	cfDNA	2,277	173	563,842	67%	2
30	PBL DNA	520	407	103,229	12%	1
30	Tumor	1,995	731	13,805	2%	1

Table S3. Detected tumor mutations in tumor DNA and cfDNA

Patient	Position, GRCh37	Gene	Mutation type	Reference allele	Mutant allele	Cosmic ID	Tumor*			cfDNA				
							No. of mutant UMI family	No. of total UMI family	Allele frequency (%)	No. of mutant reads	No. of mutant UMI family	No. of total UMI family	Allele frequency (%)	Forward/reverse balance†
2	2 : 140992455	*LRP1B*	Splice site	T	A		35	242	14.5	2	1	260	0.4	0.5
2	17 : 40481601	*STAT3*	coding nonsyn	C	A	COSM48777	42	292	14.4	11	4	185	2.2	0.5
2	15 : 45008528	*B2M*	coding nonsyn, frameshi	TC	T		27	73	37.0		0	117		
2	17 : 7577574	*TP53*	coding nonsyn	T	C	COSM118935	135	230	58.7	2	0	323		
3	12 : 25398284	*KRAS*	coding nonsyn	C	T	COSM1135366	146	966	15.1	2	0	177		
4	17 : 41234501	*BRCA1*	coding nonsyn	G	C	COSM7339774	62	746	8.3	106	19	346	5.5	0.5
4	3 : 105377911	*CBLB*	coding nonsyn	C	G		48	457	10.5	50	13	211	6.2	0.4
4	7 : 101740776	*CUX1*	coding nonsyn	A	G		35	375	9.3	57	13	125	10.4	0.5
4	X : 44949031	*KDM6A*	coding nonsyn	G	A	COSM4825087	95	582	16.3	132	24	196	12.2	0.5
5	7 : 55273024	*EGFR*	coding nonsyn	A	T		82	728	11.3	2	1	273	0.4	0.5
5	16 : 50785755	*CYLD*	coding nonsyn	G	T		103	907	11.4	5	1	239	0.4	0.0
5	17 : 7574017	*TP53*	coding nonsyn	C	A	COSM11411	111	395	28.1	7	2	233	0.9	0.0
5	2 : 48027323	*MSH6*	coding nonsyn	T	A		151	539	28.0		0	215		
5	3 : 178936092	*PIK3CA*	coding nonsyn	A	G	COSM764	254	2124	12.0		0	317		
6-1	3 : 47162102	*SETD2*	coding nonsyn	C	A		26	491	5.3		0	98		
6-1	3 : 89390991	*EPHA3*	coding nonsyn	G	T		25	466	5.4		0	57		
6-1	5 : 35873695	*IL7R*	coding nonsyn	G	T		68	926	7.3		0	145		
6-1	6 : 138201276	*TNFAIP3*	coding nonsyn	G	T		31	557	5.6		0	97		
6-1	11 : 119149239	*CBL*	coding nonsyn	G	A	COSM41789	98	989	9.9		0	33		
6-1	1 : 242048788	*EXO1*	coding nonsyn	GG	TT		297	996	29.8	4	1	99	1.0	0.5
6-2	1 : 27106528	*ARID1A*	coding nonsyn:stop-gai	G	T	COSM51450	192	605	31.7		0	88		
6-2	2 : 136872585	*CXCR4*	coding nonsyn	G	A		140	1178	11.9		0	149		
6-2	3 : 89480402	*EPHA3*	coding nonsyn	C	A		450	1227	36.7		0	172		
6-2	4 : 55570007	*KIT*	coding nonsyn	G	T		75	402	18.7		0	101		
6-2	4 : 66217232	*EPHA5*	coding nonsyn	C	A		230	1435	16.0		0	228		
6-2	7 : 140549922	*BRAF*	coding nonsyn	T	C		209	722	29.0		0	217		
6-2	8 : 41791486	*KAT6A*	coding nonsyn	T	C		123	892	13.8		0	115		
6-2	17 : 7578440	*TP53*	coding nonsyn	T	C	COSM10762	118	425	27.8		0	106		
7	2 : 141680621	*LRP1B*	coding nonsyn	C	G		969	2737	35.4	5	1	315	0.3	0.0
7	7 : 140439637	*BRAF*	coding nonsyn	C	G		642	2562	25.1	17	1	259	0.4	0.0
7	1 : 17349144	*SDHB*	coding nonsyn	G	A		97	649	15.0		0	143		
7	2 : 140997015	*LRP1B*	coding nonsyn	T	C		618	1653	37.4		0	171		
7	2 : 141660503	*LRP1B*	coding nonsyn	C	T	COSM1691115	457	1305	35.0		0	114		
7	7 : 151932951	*KMT2C*	coding nonsyn	C	A		493	2194	22.5		0	209		
7	9 : 21971186	*CDKN2A*	coding nonsyn	G	A	COSM12473	433	1020	42.5		0	234		
7	16 : 23647172	*PALB2*	coding nonsyn	C	A		455	1571	29.0		0	159		
8	2 : 141214043	*LRP1B*	coding nonsyn	G	T		41	562	7.3		0	93		
8	2 : 141762997	*LRP1B*	coding nonsyn	C	A		44	685	6.4		0	92		
8	10 : 8115975	*GATA3*	coding nonsyn	G	T		89	1065	8.4		0	85		
8	12 : 25398285	*KRAS*	coding nonsyn	C	A	COSM516	79	746	10.6		0	117		
8	17 : 29685497	*NF1*	intronic, splice?	G	T	COSM162887	37	555	6.7		0	93		
8	19 : 1220494	*STK11*	coding nonsyn	G	T	COSM48786	31	457	6.8		0	72		
9	3 : 30729974	*TGFBR2*	coding nonsyn:stop-gai	G	T	COSM6939927	247	753	32.8		0	285		
9	17 : 7578455	*TP53*	coding nonsyn	C	G	COSM43836	322	976	33.0		0	243		
9	19 : 11144110	*SMARCA4*	coding nonsyn	G	A		475	1337	35.5		0	518		
9	22 : 41565529	*EP300*	coding nonsyn	G	A	COSM122851	115	756	15.2		0	324		

10	2 : 141200172	*LRP1B*	coding nonsyn	C	A	COSM273655	148	928	16.0		0	118		
10	2 : 141083421	*LRP1B*	coding nonsyn	G	T	COSM6333908	72	435	16.6	3	1	75	1.3	0.5
10	5 : 176665350	*NSD1*	coding nonsyn	T	C		58	321	18.1	8	2	38	5.3	0.3
10	1 : 11186846	*MTOR*	coding nonsyn	G	T		37	158	23.4		0	29		
10	3 : 47125269	*SETD2*	coding nonsyn	T	C		100	556	18.0		0	107		
10	6 : 117630047	*ROS1*	coding nonsyn	T	A		51	385	13.3		0	37		
10	7 : 151864314	*KMT2C*	coding nonsyn	C	A		31	310	10.0		0	53		
10	17 : 7577120	*TP53*	coding nonsyn	C	T	COSM10660	68	280	24.3		0	68		
10	19 : 10602620	*KEAP1*	coding nonsyn	G	A	COSM1524062	67	274	24.5		0	44		
11	10 : 43606725	*RET*	coding nonsyn	C	A		55	533	10.3	4	1	59	1.7	0.0
11	6 : 152265476	*ESR1*	coding nonsyn	T	C		97	571	17.0	175	2	46	4.4	0.0
11	7 : 101870949	*CUX1*	coding nonsyn	G	T		57	215	26.5		0	58		
11	17 : 7578268	*TP53*	coding nonsyn	A	C	COSM44571	139	470	29.6		0	166		
12	17 : 7577096	*TP53*	coding nonsyn	T	G	COSM11665	312	985	31.7		0	116		
12	7 : 55259515	*EGFR*	coding nonsyn	T	G	COSM6224	575	2198	26.2	3	2	192	1.0	0.0
13	19 : 17945497	*JAK3*	coding nonsyn	G	A		125	245	51.0		0	154		
13	19 : 10610359	*KEAP1*	coding nonsyn	C	A	COSM6917923	187	351	53.3	4	2	207	1.0	0.0
13	1 : 36937737	*CSF3R*	coding nonsyn	G	A		34	126	27.0		0	88		
13	9 : 5089701	*JAK2*	coding nonsyn	C	T		38	117	32.5		0	121		
13	17 : 7577120	*TP53*	coding nonsyn	C	A	COSM10779	60	195	30.8		0	219		
14	3 : 89259484	*EPHA3*	coding nonsyn	G	A		58	857	6.8		0	264		
15	19 : 10602923	*KEAP1*	coding nonsyn	C	G	COSM2812630	58	202	28.7	1	1	89	1.1	0.0
15	19 : 10602493	*KEAP1*	coding nonsyn	C	T	COSM94572	80	314	25.5		0	179		
15	19 : 10602926	*KEAP1*	coding nonsyn	C	T	COSM6911905	59	203	29.1		0	95		
15	X : 48650295	*GATA1*	coding nonsyn	G	C		45	228	19.7		0	89		
16	5 : 176637210	*NSD1*	coding nonsyn	C	T	COSM1177184	46	145	31.7	8	1	131	0.8	0.5
16	5 : 149516571	*PDGFRB*	coding nonsyn	C	A	COSM7321876	30	111	27.0		0	201		
16	6 : 33287199	*DAXX*	coding nonsyn	C	G		111	298	37.3		0	111		
16	17 : 7578442	*TP53*	coding nonsyn	T	C	COSM10808	98	179	54.8		0	166		
16	X : 47422697	*ARAF*	coding nonsyn	G	C	COSM6966552	149	202	73.8		0	118		
17	2 : 141250263	*LRP1B*	splice site	C	A		240 / 141	854 / 503	28.1 / 28.0	4	1	177	0.9	0.5
17	7 : 151864464	*KMT2C*	splice site	C	A		161 / 113	882 / 460	18.3 / 24.6	4	1	108	0.9	0.0
17	3 : 105397359	*CBLB*	coding nonsyn	T	A		164 / 144	1098 / 825	14.9 / 17.5		0	157		
17	9 : 5123056	*JAK2*	coding nonsyn	T	C		163 / 131	476 / 422	34.2 / 31.0		0	89		
17	11 : 69466018	*CCND1*	coding nonsyn	A	G	COSM6981484	337 / 201	905 / 566	37.2 / 35.5		0	161		
18	20 : 36031760	*SRC*	coding nonsyn	A	T		109	490	22.2		0	93		
18	1 : 27100830	*ARID1A*	coding nonsyn	G	T		200	798	25.1	35	4	152	2.6	0.4
18	1 : 17355130	*SDHB*	coding nonsyn	G	T		148	623	23.8	53	8	269	2.7	0.3
18	13 : 32944549	*BRCA2*	coding nonsyn	A	G		204	314	65.0	47	3	110	2.7	0.5
18	15 : 66727501	*MAP2K1*	coding nonsyn	G	A		175	850	20.6	170	11	280	3.9	0.5
18	1 : 36932163	*CSF3R*	coding nonsyn	C	G		249	499	49.9	8	4	100	4.0	0.5
18	13 : 48923160	*RB1*	intronic, splice?	G	A	COSM3936343	215	328	65.6	144	10	194	5.2	0.5
18	17 : 29663750	*NF1*	coding nonsyn	C	A		254	649	39.1	218	13	173	7.5	0.5
18	8 : 38282205	*FGFR1*	coding nonsyn	T	A		263	1129	23.3	77	13	167	7.8	0.4
18	17 : 7578190	*TP53*	coding nonsyn	A	G	COSM10758	213	315	67.6	41	9	74	12.2	0.5
19	12 : 25398284	*KRAS*	coding nonsyn	C	A	COSM520	98 / 131	717 / 1039	13.7 / 12.6		0	323	0.0	
20	3 : 178952085	*PIK3CA*	coding nonsyn	A	G	COSM775	141 / 143	768 / 820	18.4 / 17.4		0	137		
20	7 : 55249013	*EGFR*	coding nonsyn	AA	GGGTT	COSM18431	366 / 490	1840 / 2525	19.9 / 19.4		0	214		
21	7 : 140477861	*BRAF*	coding nonsyn	T	C	COSM3715613	447 / 634	1351 / 1954	33.1 / 32.5	4	1	384	0.3	0.0

21	1 : 193205414	*CDC73*	coding nonsyn	G	T		360 / 377	1112 / 1011	32.4 / 37.3	8	1	256	0.4	0.5
21	1 : 156844765	*NTRK1*	coding nonsyn	A	T		332 / 306	1945 / 1774	17.1 / 17.3		0	191		
21	3 : 69985876	*MITF*	coding nonsyn	G	A		70 / 90	378 / 523	18.5 / 17.2		0	133		
21	4 : 55127323	*PDGFRA*	coding nonsyn	G	T		187 / 262	605 / 612	30.9 / 42.8		0	193		
21	4 : 55958873	*KDR*	coding nonsyn	C	A		229 / 374	744 / 931	30.8 / 40.2		0	208		
21	5 : 138223203	*CTNNA1*	coding nonsyn	G	T		280 / 292	743 / 771	37.7 / 37.9		0	117		
21	6 : 35430582	*FANCE*	coding nonsyn	A	T		200 / 139	822 / 608	24.3 / 22.9		0	90		
21	9 : 87322779	*NTRK2*	coding nonsyn	T	A		414 / 521	669 / 821	61.9 / 63.5		0	117		
21	12 : 121426766	*HNF1A*	coding nonsyn	C	T		355 / 306	1043 / 804	34.0 / 38.1		0	147		
21	17 : 7578458	*TP53*	coding nonsyn	G	C	COSM11087	511 / 278	1014 / 570	50.4 / 48.8		0	117		
21	18 : 42531680	*SETBP1*	coding nonsyn	C	A		456 / 377	1512 / 1224	30.2 / 30.8		0	104		
21	20 : 57429185	*GNAS*	coding nonsyn	G	A		320 / 206	1122 / 811	28.5 / 25.4		0	86		
22*	11 : 108098601	*ATM*	coding nonsyn	G	T		54 / 14	1929 / 1080	2.8 / 2.2	6	0	160		
22	12 : 25398284	*KRAS*	coding nonsyn	CC	AA	COSM512	27 / 15	1899 / 1003	1.4 / 1.5	0	0	182		
22	5 : 176520725	*FGFR4*	coding nonsyn	G	C		17 / 16	750 / 828	2.3 / 1.9	0	0	231		
22	2 : 141473585	*LRP1B*	coding nonsyn	C	A		28 / 11	1652 / 1268	1.7 / 0.9	2	0	292		
23	1 : 242015664	*EXO1*	coding nonsyn	G	A		160 / 90	445 / 795	36.0 / 11.3	0	0	364		
23	3 : 41266113	*CTNNB1*	coding nonsyn	C	G	COSM5679	61 / 98	326 / 1191	18.7 / 8.2	3	0	636		
23	17 : 7578176	*TP53*	intronic, splice	C	T	COSM6906	179 / 25	302 / 208	59.3 / 16.8	0	0	156		
23	19 : 10600447	*KEAP1*	coding nonsyn	G	A	COSM564847	176 / 68	817 / 932	21.5 / 7.3	1	0	601		
23	19 : 18266872	*PIK3R2*	coding nonsyn	C	G		90 / 34	472 / 320	19.1 / 10.6	0	0	262		
23	20 : 52194891	*ZNF217*	coding nonsyn	G	A		362 / 250	971 / 2081	37.3 / 12.0	2	0	805		
24	3 : 30715706	*TGFBR2*	coding nonsyn	G	C	COSM218768	255 / 99	1558 / 557	16.4 / 17.8	5	0	336		
24	12 : 25398284	*KRAS*	coding nonsyn	C	T	COSM1135366	294 / 174	1366 / 933	21.5 / 18.7	4	0	258		
24	17 : 7577551	*TP53*	coding nonsyn	C	A	COSM1646853	361 / 94	1824 / 478	19.8 / 19.7	2	0	222		
24	19 : 2191101	*DOT1L*	coding nonsyn	G	A		587 / 186	3220 / 1062	18.2 / 17.5	2	0	348		
25	X : 44918538	*KDM6A*	coding nonsyn	T	A		102	744	13.7		1	122	0.9	0.0
25	7 : 2946289	*CARD11*	coding nonsyn	C	G		134	566	23.7		0	115		
25	7 : 151833996	*KMT2C*	coding nonsyn	T	C		249	1915	13.00		0	335		
25	13 : 48942661	*RB1*	intronic, splice	A	G		398	1212	32.8		0	285		
25	17 : 7577108	*TP53*	coding nonsyn	C	A	COSM10749	302	1062	28.4		0	279		
25	19 : 10610283	*KEAP1*	coding nonsyn	C	T		248	909	27.3		0	277		
26	3 : 105459397	*CBLB*	coding nonsyn	C	G		167	1262	13.2		0	358		
26	5 : 112175390	*APC*	coding nonsyn:stop-gai	C	T	COSM13121	167	952	17.5		0	326		
26	7 : 55259524	*EGFR*	coding nonsyn	T	A	COSM6213	377	1421	26.5		0	355		
27	2 : 178098960	*NFE2L2*	coding nonsyn	C	G	COSM124736	137	675	20.3		0	233		
27	10 : 43595945	*RET*	coding nonsyn	G	A		165	846	19.5		0	346		
27	17 : 7577130	*TP53*	coding nonsyn	A	C	COSM44956	123	376	32.7		0	102		
27	19 : 1646356	*TCF3*	coding nonsyn	G	A		43	337	12.8		0	253		
27	X : 53239892	*KDM5C*	coding nonsyn	C	A		356	992	35.9		0	209		
28	2 : 141004717	*LRP1B*	coding nonsyn	C	G		34	572	5.9		0	87		
28	13 : 32914212	*BRCA2*	coding nonsyn	C	G		52	631	8.2		0	111		
28	17 : 7577141	*TP53*	coding nonsyn	C	T	COSM10867	38	482	7.9		0	78		
28	2 : 198267360	*SF3B1*	coding nonsyn	T	C	COSM131553	38	740	5.1		23	151	15.2	0.4
29	4 : 1803752	*FGFR3*	splice site	G	C		40	388	10.3		0	103		
29	12 : 25398285	*KRAS*	coding nonsyn	C	A	COSM516	303	763	39.7		0	179		
30	3 : 47058743	*SETD2*	3utr,coding nonsyn	A	G	COSM4969185	58	373	15.6		0	64		

* Seven tumor were sequenced in two replicates. Count, coverage and AF are shown for both replicates

† Ratio between the number of mutations detected in read 1 and read 2

Paper II

Experimental and Molecular Pathology 126 (2022) 104771

Contents lists available at ScienceDirect

Experimental and Molecular Pathology

journal homepage: www.elsevier.com/locate/yexmp

Fragmentation assessment of FFPE DNA helps in evaluating NGS library complexity and interpretation of NGS results

Anine Larsen Ottestad [a,b], Elisabeth F. Emdal [a,c], Bjørn H. Grønberg [a,b,*], Tarje O. Halvorsen [a,b], Hong Yan Dai [a,c]

[a] Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
[b] Department of Oncology, St. Olav's Hospital, Trondheim University Hospital, Trondheim, Norway
[c] Department of Pathology, Clinic of Laboratory Medicine, St. Olav's Hospital, Trondheim University Hospital, Trondheim, Norway

ARTICLE INFO

Keywords:
Next-generation sequencing
Formalin-fixed paraffin-embedded tissue
Library preparation
Quality assessment
DNA fragmentation

ABSTRACT

Formalin-fixed paraffin-embedded (FFPE) tissue remains the most common source for DNA extraction from human tissue both in research and routine clinical practice. FFPE DNA can be considerably fragmented, and the amount of DNA measured in nanograms may not represent the amount of amplifiable DNA available for next-generation sequencing (NGS). Two samples with similar input DNA amounts in nanograms can yield NGS analyses of considerably different quality. Nevertheless, many protocols for NGS library preparation from FFPE DNA describe input DNA in nanograms without indication of a minimum requirement of amplifiable genome equivalent DNA.

An important NGS quality metric is the library complexity, reflecting the number of DNA fragments from the original specimen represented in the final library. Aiming to illustrate the relationship between DNA fragmentation degree and library complexity, we assessed the fragmentation degree of 116 lung cancer FFPE DNA samples to calculate the amount of amplifiable input DNA used for library preparation. Mean unique coverage, coverage uniformity, and mean number of PCR duplicates with the same unique molecular identifier were used to evaluate library complexity.

We showed that the amount of amplifiable input DNA predicted library complexity better than the input measured in nanograms. The frequent discrepancy between DNA amount in nanograms and the amount of amplifiable DNA indicate that the fragmentation degree should be considered when performing NGS of FFPE DNA. Importantly, the fragmentation assessment may help when interpreting NGS data and be a useful tool for evaluating library complexity in the absence of unique molecular identifiers.

1. Introduction

Formalin-fixed paraffin-embedded (FFPE) human tumor tissue samples are collected for routine histopathological diagnostic procedures. They also represent a vast and valuable resource for molecular analyses and retrospective cancer genetic studies. However, the quality of DNA from FFPE samples varies largely when compared with DNA isolated from fresh-frozen tumor tissue. When preparing FFPE samples, formalin functions as a cross-linking agent for tissue fixation and stabilizes the tissue structure by creating covalent linkage between macromolecules, such as DNA-DNA, DNA-protein, and protein-protein. Reversing the formalin-formed cross-linking during DNA extraction causes fragmentation of FFPE DNA. In addition, formalin causes the release of purine bases from nucleic acids and induces DNA fragmentation (Do and Dobrovic, 2015; Srinivasan et al., 2002). Since the extent of fixation may vary among samples, the extent of fragmentation may also vary. The quality of FFPE DNA directly affects the quality metrics of downstream NGS analyses, such as library size, average read depth and uniformity (Robbe et al., 2018; Spencer et al., 2013). Although optimizing tissue fixation conditions and DNA extraction methods improve FFPE DNA quality (Einaga et al., 2017; Heydt et al., 2014; McDonough et al., 2019), the quality is still influenced by many stochastic factors, such as time until fixation, perioperative ischemic time, fixation time and size of tissue samples, storage time and extent of necrosis in tissue

* Corresponding author at: Department of Clinical and Molecular Medicine, NTNU, Kunnskapssenteret, Olav Kyrres gate 17, 7030 Trondheim, Norway.
E-mail address: bjorn.h.gronberg@ntnu.no (B.H. Grønberg).

https://doi.org/10.1016/j.yexmp.2022.104771
Received 10 September 2021; Received in revised form 13 March 2022; Accepted 9 April 2022
Available online 12 April 2022
0014-4800/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

samples (Bass et al., 2014). Many studies have demonstrated the feasibility of using FFPE DNA for next generation sequencing, including gene specific targeted sequencing, whole-exome and whole-genome sequencing (Carrick et al., 2015; Hedegaard et al., 2014; Kerick et al., 2011; Robbe et al., 2018), though in many clinical pathology departments, gene specific targeted NGS analyses of FFPE DNA are still the most common routine diagnostic molecular analyses.

In library preparation for targeted next-generation sequencing (NGS), the genomic regions of interest are enriched from input DNA by either hybridization-capture using gene-specific probes or PCR amplicon-based enrichment using gene-specific primers (Chang and Li, 2013). For both approaches, the final sequencing-ready library is generated by PCR enrichment. Thus, the final library includes PCR duplicates that provide no additional value in the data analysis unless they are identified by a common characteristic unique to each set of duplicates. One option is to utilize an inherent characteristic, such as fragment length, but this is not possible for amplicon-based libraries generated using opposing primer pairs. An alternative is to introduce a synthetic characteristic to the input DNA before PCR, such as a unique molecular identifier (UMI), that will be common for all subsequent PCR duplicates (Kinde et al., 2011)

Several NGS library preparation protocols used in routine diagnostics omit the incorporation of UMIs, possibly because this adds another step in the laboratory and requires specialized algorithms for data analysis. One consequence is that these unidentified duplicates contribute to the mean coverage that is used as a quality metric for NGS. However, the mean unique coverage is arguably a more precise quality metric because it reflects the number of unique human genome equivalents (hGEs) from the input DNA represented in the final library, i.e., the library complexity.

High library complexity is desirable to achieve high analytic sensitivity and specificity. McNulty *et al.* prepared NGS libraries from cell culture DNA which exhibits the similar quality as DNA from fresh frozen tissue. By varying the amount of input DNA measured in nanograms, they demonstrated that library complexity was enhanced by increasing the amount of input DNA (McNulty et al., 2020). However, if input FFPE DNA for library preparation is only measured in nanograms, the true amount of DNA available for subsequent PCR in NGS libraries may vary widely because the fragmentation degree of FFPE DNA may differ greatly. A study by McDonough et al. showed that NGS quality metrics such as raw base coverage varied widely among specimens although the same amounts of input DNA in nanogram were used for targeted library preparation (McDonough et al., 2019). Commercial kits for DNA quality assessment, such as the KAPA Human Genomic DNA Quantification and QC kit and a multiplex PCR assay (Life Science Innovations, Qualitative Multiplex PCR Assay) have been used prior to NGS library preparation in many studies (McDonough et al., 2019; Pel et al., 2018). However, publications rarely include descriptions of how the results of DNA quality assessments have been interpreted and used for NGS library preparation and eventually data interpretation.

The aim of this study was to demonstrate how a DNA fragmentation assay might be applied and how FFPE DNA fragmentation might affect subsequent NGS library complexity. We performed NGS analysis of 116 DNA samples extracted from FFPE lung cancer samples using the QIAseq Human Actionable Solid Tumor panel. The NGS libraries were prepared with UMIs. Thus, the number of UMIs for each sequenced region reflected the complexity of each library. We also assessed the fragmentation degree of input DNA and calculated the true quantity of DNA fragments as potential templates for amplification. Using this approach, we demonstrated that in the case of FFPE DNA samples, the quantity of input DNA according to the amount of amplifiable DNA fragments rather than amount in nanograms better reflected the number of potential hGE templates and thus provided better prediction of the complexity of NGS.

2. Material and methods

2.1. Samples and approval

In total, 116 tumor samples from 114 lung cancer patients diagnosed between 2007 and 2018 at St. Olav's University Hospital, Trondheim, Norway, were retrieved from Biobank1, our regional lung cancer biobank. The biobank was approved by the Norwegian Regional Committee for Medical and Health Research Ethics (REC) Central, the Norwegian Health Department, and the Norwegian Data Protection Authority. The REC Central also approved the present study. Patients had a median age of 68 (range 46–86), 48% were female, 92% had a performance status of 0–1 and 29% had stage IV disease. Most tumors were adenocarcinomas (n = 103), and the rest were adenosquamous carcinoma (n = 1), large cell neuroendocrine carcinoma (n = 2), small cell carcinoma (n = 1) and non-small cell lung cancer not otherwise specified (n = 9).

2.2. DNA extraction from FFPE tissue

DNA was extracted from archival FFPE tumor blocks. At the Department of Pathology at St. Olav's University Hospital, phosphate-buffered 4% formaldehyde solution (HistoLab Products AB, Gothenburg, Sweden) with a pH of 7.2–7.4 is used as fixation solution. For the large surgically resected samples, the fixation time is between 3 and 5 days; for small biopsies it is overnight for approximately 12–16 h. Fixation is then performed for another two hours in a tissue processor before paraffin embedding. Fixation of tiny needle biopsies are carried out directly in the tissue processor. Two to five tissue sections of 10 μm were cut from the areas with highest tumor cell density identified by an experienced lung cancer pathologist by regular light microscopy. The number of sections was determined empirically according to the size of the defined area and the tumor cell density. DNA was extracted using the QIAcube Connect (Qiagen, Valencia, CA) and the QIAamp DNA FFPE Tissue Kit (Qiagen, Valencia, CA) and then eluted in 200 uL of the supplied buffer. DNA concentration was measured fluorometrically by Qubit® (Thermo Fisher Scientific, Waltham, MA) using either the dsDNA BR or the HS Assay Kit depending on the yield.

2.3. Fragmentation assessment of FFPE DNA

Quantitative real-time PCR (qPCR) of *FCGR3b* with a fragment length of 300 bp was performed to assess the fragmentation degree in FFPE DNA. The *FCGR3b* gene was chosen for practical reasons, since it is already in routine use and validated at our pathology department. This gene is not known to be amplified in lung cancer, and differential expression or epigenetics would not influence the quantification by qPCR. In this study, we used quantification of the amplifiable copy number of *FCGR3b* not as an absolute number, but rather as a condition to determine the amount of amplifiable input DNA within a certain range. Only amplification with extremely high copy numbers will influence this range assessment, and *FCGR3b* is not known for this kind of amplification. The abundance of this fragment was measured relative to an unfragmented DNA control sample extracted from leukocytes from a healthy person. We assumed that DNA fragmentation caused by formalin fixation occurs randomly and, therefore, also within this gene. We assumed that the number of *FCGR3b* fragments at ≥300 bp present in the DNA sample was proportional to the number of hGEs with a fragment length of at least 300 bp. For comparison in a subset of samples, we also quantified the fragments of the gene *ALB* of ≥150 bp by qPCR.

qPCR was performed according to a protocol developed, validated, and used in diagnostic routine at our pathology department. Specifically, qPCR was performed using 10 ng DNA, 0.6 uM primer solution (for 150 bp or 300 bp), molecular grade water and iQ™ SYBR® Green Supermix (Bio-Rad, Hercules, CA) in a 25 μL reaction. Each assay contained triplicates of FFPE DNA, DNA isolated from peripheral blood as

control of non-fragmented DNA, and a non-template control. All runs were processed on a Bio-Rad® CFX96 using the following run program: 95 °C/10 min – 40 cycles of 94 °C/30 s, 56 °C/30 s, 72 °C/30 s– melting curve program: 95 °C to 60 °C, and increment of 0.5 °C for 1 s.

The difference in mean threshold value (ΔCt) between FFPE DNA and the blood DNA control was used to calculate the number of fragments with at least 150 bp/300 bp in the tumor DNA relative to the control. The fragmentation degree of tumor DNA was defined as $2^{\Delta Ct}$.

2.4. Next-generation sequencing and mutation detection

NGS libraries were prepared following the manufacturers' instructions using 1.7–250.8 ng FFPE DNA without considering the DNA fragmentation degree. Libraries were made using QIAseq Targeted DNA Human Actionable Solid Tumor Panel (Qiagen, Valencia, CA), which included UMIs. In brief, DNA was enzymatically fragmented, end-repaired, and A-tailed followed by ligation to a 5′ sequencing adapter that contained the UMI. The regions of interest were then selected by targeted PCR using an adapter primer and gene-specific primers that contained a universal primer sequence. The library was then amplified in a universal PCR using primers for the 5'adapter and a 3'primer complementary to the primer seat added in the targeted PCR. The 3′ primer also contained the 3′ sequencing adapter sequence. Libraries were quantified by KAPA Library Quantification kit (Roche, Switzerland) and pooled together in equimolar amounts before sequencing. 151 bp pair end sequencing was performed on the Illumina MiSeq or NextSeq platform (Illumina, San Diego, CA).

Data analysis was performed using the CLC Genomic Workbench version 12.0.2 (Qiagen, Valencia, CA) and a panel-specific workflow that utilized the UMIs. All reads passing the quality filters were used for downstream analyses. Mean read coverage was defined as the mean number of reads that covered each target position, without using the UMI information. Duplicates with the same UMI sequence were then grouped into a "UMI family", and the mean unique coverage was defined as the mean number of UMI families that covered each target position. Variants were called if they were present in 75% of the duplicates in a family. Variants below 5% allele frequency were discarded to avoid erroneously calling mutations that spontaneously arise in DNA over time.

2.5. Statistical analysis

Linear regression was used to explore the relation between mean unique coverage, total number of reads, amount of input DNA (Fig. 2A) and number of genome equivalents (Fig. 2B). R version 1.1.463 was used for statistical analyses, and figures were made using the ggplot2 package. The level of statistical significance was defined as $p \leq 0.05$.

3. Results

3.1. High variation in fragmentation degree of FFPE DNA

The concentration of DNA extracted from the 116 samples ranged from 0.103 ng/μL to 136.0 ng/μL (median 4.9 ng/μL). qPCR with a 300 bp fragment of the *FCGR3b* gene was then used to estimate the fragmentation degree in the FFPE samples compared to a control sample of DNA extracted from whole blood. Such DNA is minimally fragmented and considered to be of high quality compared to FFPE DNA. The quality of FFPE DNA was therefore defined as the fragmentation degree relative to the whole blood DNA control sample. The relative fragmentation degree was calculated using the difference in Ct value (ΔCt) between each FFPE sample and the control and the formula $2^{\Delta Ct}$. Thereby, a ΔCt value of 3.3 implied a 10-fold fragmentation degree of FFPE relative to the control. As shown in Fig. 1, the fragmentation degree of FFPE DNA ranged from 1 to >500-fold compared to the control. Sixteen samples were > 500-fold fragmented and the sample with the worst quality was 6339-fold fragmented.

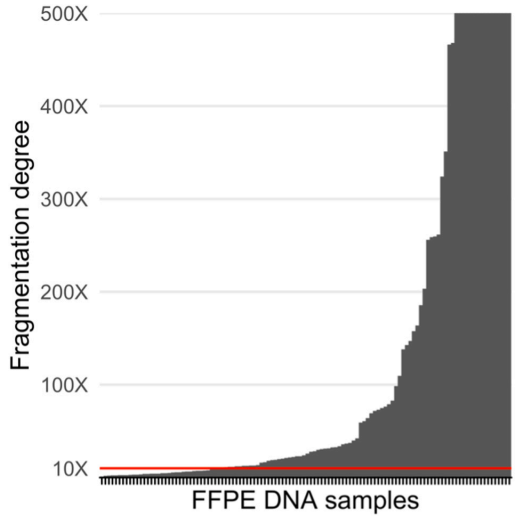


Fig. 1. The fragmentation degree of 116 lung tumor FFPE samples relative to a control of genomic DNA extracted from leukocytes from a healthy person. The y-axis shows the values of $2^{\Delta Ct}$ and is capped at 500-fold. Seventy-two percent of the samples were more than 10-fold fragmented than the control, as indicated by the red line. Based on experience, FFPE derived DNA is usually 10-fold fragmented. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

An additional qPCR assay was performed with a 150 bp fragment in the gene *ALB* in a subset of the FFPE DNA samples (n = 41), and the fragmentation degree varied from 1- to 19-fold (median 3-fold). The samples that were more than 10-fold fragmented at 300 bp (10- to 520-fold) varied between 2- and 19-fold fragmented at 150 bp.

3.2. The number of amplifiable hGEs in input FFPE DNA can predict the library complexity

NGS libraries were prepared using 1.7–250.8 ng input DNA and QIAseq Targeted DNA Human Actionable Solid Tumor Panel. To evaluate the impact of DNA fragmentation on the library complexity, we used the fragmentation degree to calculate the number of potential hGE templates present in the input for each library. The UMIs were used to evaluate the library complexity.

First, we examined the relationship between input DNA amount and the mean unique coverage in the panel target region. Overall, the mean unique coverage ranged from 16× to 3098× (median 326×) in the 116 libraries. A high total number of reads per library did not result in a high mean unique coverage $R^2 = 0.00137$, $p = 0.694$) (point colors in Fig. 2). Furthermore, a higher input DNA amount in nanograms did not systematically increase the mean unique coverage (Fig. 2A), though the correlation between input DNA and mean unique coverage was statistically significant ($R^2 = 0.277$, $p < 0.0001$), mainly because the highest input DNA amounts resulted in high mean unique coverage; 10 out of 13 libraries made with more than 200 total ng input DNA had mean unique coverage >1000×. Fig. 2A shows that high fragmentation degree decreased the mean unique coverage among libraries made from the approximately same input amount in ng. In line with this observation, a higher number of amplifiable hGEs in the input DNA generally increased the mean unique coverage ($R^2 = 0.410$, $p < 0.0001$) (Fig. 2B).

Second, we examined the relationship between input DNA amount and the number of duplicates/UMI family. The libraries made from lower amounts of DNA in nanograms contained a high number of

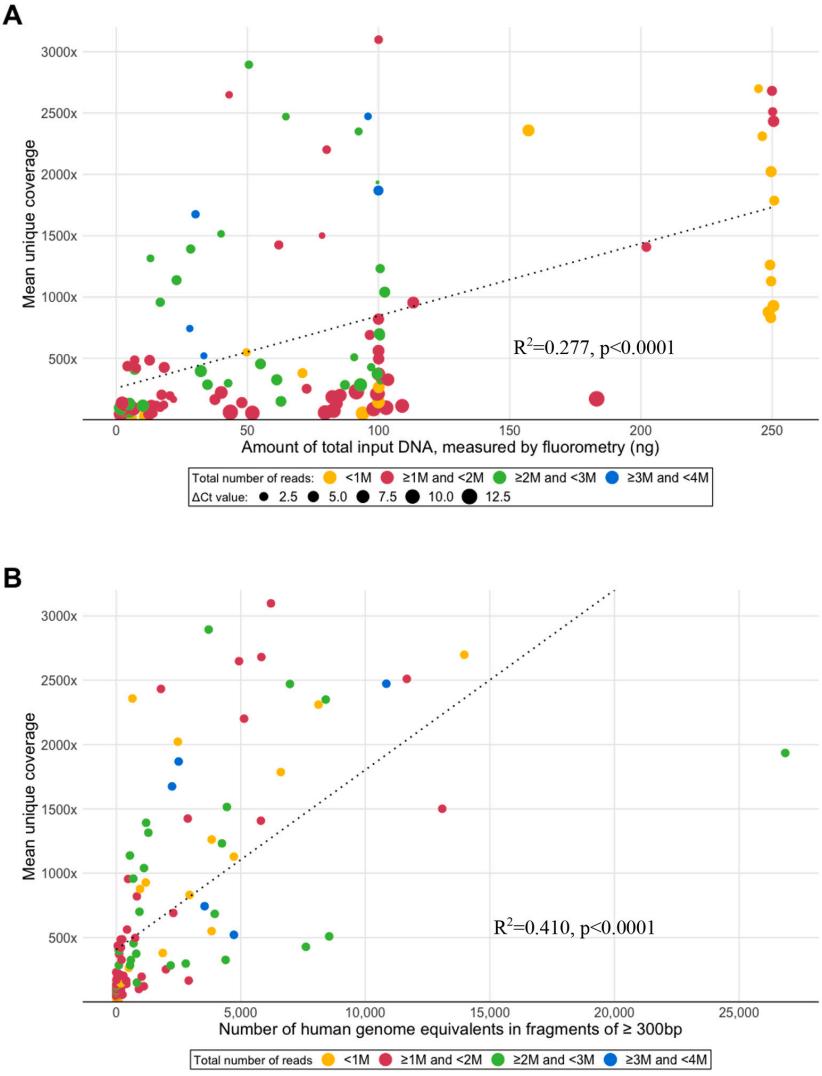


Fig. 2. Mean unique coverage for each of the 116 libraries versus A) the amount of input DNA measured by fluorometry and B) the amount of amplifiable hGE templates measured by qPCR. All reads passing the quality filters were analyzed, and the color shows the number of reads analyzed in each sample. In fig. A, the point size increases with increased fragmentation degree of the input DNA.

duplicates/UMI family (Fig. 3A). However, Fig. 3A shows that the number of duplicates/UMI family varied among samples with the same amount of DNA in nanograms, while it was more uniform among samples with the same input DNA calculated as the number of amplifiable hGEs (Fig. 3B).

We observed that all samples with >200 ng total DNA input had few duplicates per UMI family (1.1–2.8). These libraries did not have sufficient duplicates per UMI family for error correction, which indicates that an input DNA of more than 200 ng potentially causes ineffective PCR reactions.

Third, we examined the relationship between input DNA amount and uniformity of the mean unique coverage of each gene in the target panel. To exemplify, in Fig. 4 we present the coverage uniformity in two libraries that were prepared using 100 ng DNA input, and both libraries generated approximately 1.7 million reads. The input DNA used to prepare the first library (Fig. 4A) was 5-fold fragmented and the input DNA for the second library (Fig. 4B) was 520-fold fragmented relative to the control. This corresponds to an estimated input of 6200 and 58 potential hGE templates, respectively. The coverage uniformity was superior in the library made from the highest number of potential templates (Fig. 4A).

3.3. DNA fragmentation assessment helps to evaluate the NGS analytic sensitivity

We examined the impact of input DNA amount on the detection of mutations in six representative libraries made from variable DNA amounts (Table 1; results from all 116 samples are presented in the

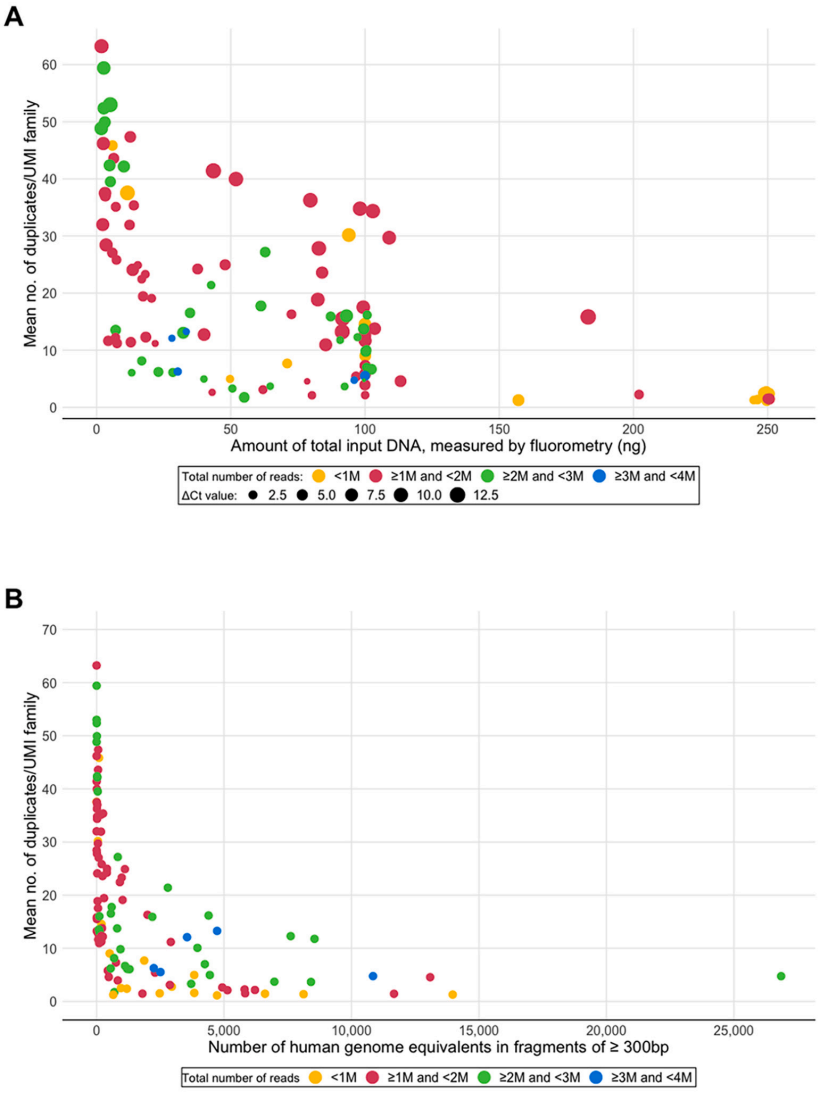


Fig. 3. Mean number of duplicates per UMI family in each of the 116 libraries versus A) the amount of total input DNA and B) the amount of amplifiable input DNA. All reads passing the quality filters were analyzed, and the color shows the number of reads analyzed in each sample. In fig. A, the point size increases with increased fragmentation degree of the input DNA.

supplementary table). The libraries 237 and 841 were prepared from approximately 100 ng DNA. The mean read coverage in the panel target region (calculated without using the UMIs) was comparable between the two samples. This similarity indicates comparable analytical sensitivity. However, the estimated number of input hGEs for each library differed considerably (21 in 237 vs. 26,855 in 841). As a result, the mean number of duplicates per UMI family in library 237 was much higher (34.4 compared to 4.7 in library 841). Therefore, the validity of a low frequency variant in a library such as library 237 in our cohort should be carefully evaluated, especially if UMIs are not incorporated.

Fragmentation assessment further enabled evaluation of whether a mutation-negative sample was likely a true negative. Data analysis of libraries 29 and 837 shown in Table 1 resulted in no detected mutations. Considering the number of input hGEs, rather than the mean read coverage, it is likely that sample 837 is a true negative sample, while we cannot rule out that the analysis of sample 29 represents a false negative result. Without UMIs or awareness of the DNA fragmentation degree, it would not be possible to accurately evaluate an apparently negative result.

We observed that the mean unique coverage was higher than the estimated input of hGE templates in some libraries (Table 1). This discrepancy suggests that a significant number of hGE templates of shorter fragment lengths than 300 bp were available for amplification in these samples. When using the additional assay with a 150 bp fragment to analyze a subset of the FFPE DNA samples, we observed that the number of 150 bp templates was up to a 10-fold higher than the number of 300 bp templates (data not shown). On the other hand, e.g. libraries 841 and 837 had lower unique coverage than the estimated input of hGE

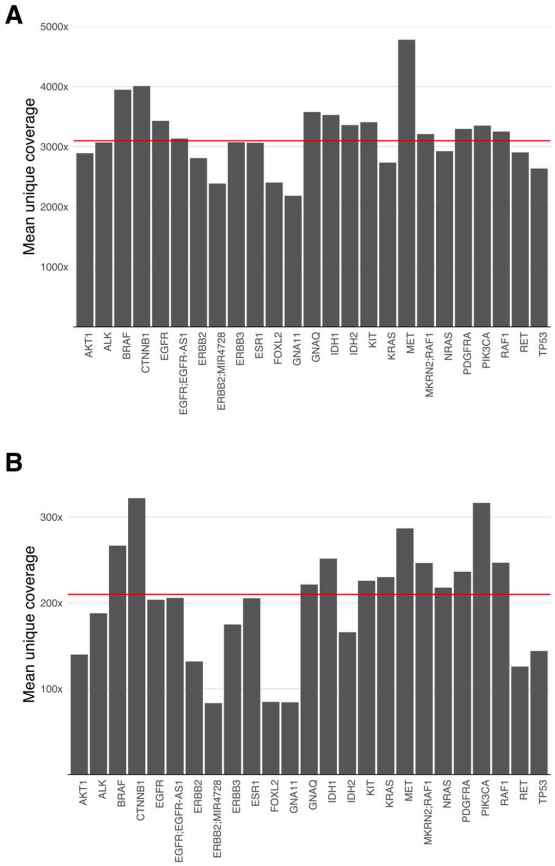


Fig. 4. The coverage uniformity in two libraries made from 100 ng input DNA. The graphs show the mean unique coverage for each gene covered by the QIAseq Human Actionable Solid Tumor panel. The red line indicates the mean unique coverage for the whole target region in each library. The fragmentation degree varied in the input DNA used to prepare the libraries. The library in fig. A) was prepared using 6200 potential hGE templates. The library in fig. B) was prepared using 58 potential hGE templates. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

templates; this suggests that too much input DNA may decrease the number of duplicates per UMI family and consequently reduces complexity.

4. Discussion

The challenge of using FFPE DNA for NGS analyses in clinical practice is ensuring a sufficient amount and quality of input DNA. There is often less input DNA available than optimal. It is, however, possible to generate an NGS library from as little as a few nanograms of FFPE DNA by using amplicon-based enrichment. For such samples, ensuring sufficient library complexity is necessary for precisely interpreting whether variants with low frequency are truly present or not.

In this study, we calculated the potentially amplifiable hGEs for NGS library by assessing the fragmentation degree of FFPE DNA using a qPCR method. We incorporated UMIs and constructed NGS libraries from 116 lung cancer FFPE DNA samples. We did not exclude the samples with lower amounts of available input DNA as recommended by the manufacturer. Using UMIs enabled us to trace the DNA fragment in the original input DNA and evaluate the complexity of each library. We then demonstrated that, for libraries prepared with FFPE DNA, increasing input DNA amount would help to increase such complexity, but only when adjusting for amplifiable hGEs in each sample.

By comparing the NGS results of libraries constructed with different amounts of cell line derived DNA, McNulty *et al.* showed that higher input DNA amounts provide higher NGS library complexity (McNulty et al., 2020). In our study, we found similar associations when the input DNA amount from FFPE tissue was measured as the number of potential hGE templates. Libraries made from higher amounts of input hGE had higher library complexity in terms of higher mean unique coverage, higher coverage uniformity and fewer duplicates per UMI family. However, if input FFPE DNA was fragmented to such a degree that few hGEs were available, the library complexity did not necessarily increase with higher input FFPE DNA amount measured in nanograms. Libraries made from a low number of hGEs generated a comparable number of total reads as libraries made from a high number of hGEs, but many of the reads were duplicates belonging to the same UMI families. Duplicates are necessary for error correction. However, Xu *et al.* concluded that four duplicates per UMI family are sufficient (Xu et al., 2017), suggesting that very high numbers of duplicates do not increase the library complexity.

Others have also concluded that the DNA amount from FFPE samples measured in nanograms by either spectrophotometer or by fluorometer may not always represent the amount of amplifiable input DNA (hGEs) available for library preparation. Heydt *et al.* compared and evaluated the impact of five different DNA quantification methods on downstream amplicon-based NGS performance in order to find the best method to assess the quantity of input FFPE DNA (Heydt et al., 2014). Both spectrophotometric and fluorescent dye-based quantification systems and a qPCR method were used, and they found that the DNA concentration varied widely when using different methods.

Heydt *et al.* constructed 24 libraries from two samples that were divided into four sets of different input DNA amounts. Each set consisted of three solutions with the same amount of DNA calculated by three different methods. In contrast to our results, they found that the mean read coverage and number of called variants were comparable independent of the quantification method. Their results are not necessarily comparable to our findings since they did not use molecular barcodes to label their input DNA or assess mean unique coverage; instead, they used read coverage as the quality metric. As we have shown, PCR can generate high amounts of duplicates even from low amounts of input DNA and consequently, high mean coverage.

In this study, most FFPE DNA samples (72%) were more than 10-fold fragmented compared to the non-fragmented control DNA. Furthermore, the fragmentation degree of FFPE DNA varied considerably between samples, underscoring the need to evaluate fragmentation of each sample. We believe that unknown and low quantities of amplifiable input DNA might be one reason for failed NGS which is frequently reported in studies involving clinical FFPE tissue samples (Flaherty et al., 2020; Middleton et al., 2020; Stockley et al., 2016).

We observed that the mean unique coverage varied between libraries generated from similar amounts of hGEs in input DNA. One explanation might be the different fragmentation degree of fragments shorter than 300 bp. It may also indicate that fragmentation alone might not affect complexity (Hedegaard et al., 2014), and that other FFPE-related DNA modification factors might impede efficient library generation.

Our study demonstrated several benefits of assessing the available amount of amplifiable input DNA for NGS. First, we show that the number of amplifiable hGE can vary considerably between FFPE DNA samples with the same DNA amount in nanograms. Second, we show that the amount of hGEs better predicts library complexity than the amount of input DNA in nanogram. Third, assessing fragmentation is valuable when interpreting NGS data, especially for the samples with low yields and poor quality, since the risk of both false positive and false

Table 1
NGS quality and mutations detected in six representative samples. Full overview of all 116 samples is listed in the Supplementary Table.

Sample ID	Total DNA concentration (ng/uL)	Fragmentation degree relative to the control	Total DNA input amount for NGS (ng)	No. of potential hGE templates	Mean read coverage in the panel target region	Mean unique coverage in the panel target region	Number of duplicates/ UMI family	Mutations detected	Mutant allele frequency (unique count/ unique coverage)
37	0.103	466×	1.7	1	13,750×	97×	48.8	*TP53* NP_001119586.1:p. Arg273Leu	56.3% (27/48)
29	0.136	260×	2.3	3	11,356×	136×	32.0	No mutation detected	N/A
237	34.3	1520×	102.9	21	9216×	98×	34.4	*ERBB3* NP_001973.2:p. Arg103Cys *KRAS* NP_203524.1:p. Gly13Cys *TP53* NP_001263625.1:p. Glu248*	7.9% (11/139) 13.1% (21/160) 55.4% (62/112)
841	6.6	1×	99.6	26,855	13,236×	1935×	4.7	*RAF1* NP_002871.1:p. Ser259Phe	76.2% (2073/2721)
826	24.0	31×	249.6	2476	3973×	2022×	1.5	*EGFR* NP_005219.2: p.Leu858Arg	17.8% (440/2470)
837	39.4	7×	250.2	11,667	4552×	2511×	1.4	No mutation detected	N/A

hGE; human genome equivalent, NGS; next-generation sequencing, UMI; unique molecular index,

negative variant calling increases. Therefore, a successful NGS library requires input DNA amount that is neither too low nor too large, and with a predefined amount of amplifiable hGE.

The most obvious approach to overcome challenges due to high fragmentation is to add more input DNA, but this is not always possible. For example, biopsies from lung cancer tumors are often small, especially those that are obtained through bronchoscopy. Our results indicate that when more DNA cannot be analyzed, it is important to be aware of how the hGE content in input DNA might influence the library complexity and potentially help in interpreting the variant call and avoid false positives or negatives.

CRediT authorship contribution statement

Anine Larsen Ottestad: Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration, Funding acquisition. **Elisabeth F. Emdal:** Validation, Investigation, Resources, Data curation. **Bjørn H. Grønberg:** Resources, Writing – original draft, Writing – review & editing, Funding acquisition. **Tarje O. Halvorsen:** Software, Formal analysis, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition. **Hong Yan Dai:** Conceptualization, Methodology, Validation, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgements

The library sequencing was provided by the Genomics Core Facility (GCF), Norwegian University of Science and Technology (NTNU). GCF is funded by the Faculty of Medicine and Health Sciences at NTNU and The Central Norway Regional Health Authority. This study was supported by The Central Norway Regional Health Authority. The funding source did not have any role in design, conduct of the study, interpretation, or publication of results.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yexmp.2022.104771.

References

Bass, B.P., Engel, K.B., Greytak, S.R., Moore, H.M., 2014. A review of Preanalytical factors affecting molecular, protein, and morphological analysis of formalin-fixed, paraffin-embedded (FFPE) tissue: how well do you know your FFPE specimen? Arch. Pathol. Lab. Med. 138, 1520–1530. https://doi.org/10.5858/arpa.2013-0691-RA.

Carrick, D.M., Mehaffey, M.G., Sachs, M.C., Altekruse, S., Camalier, C., Chuaqui, R., Cozen, W., Das, B., Hernandez, B.Y., Lih, C.-J., Lynch, C.F., Makhlouf, H., McGregor, P., McShane, L.M., Phillips Rohan, J., Walsh, W.D., Williams, P.M., Gillanders, E.M., Mechanic, L.E., Schully, S.D., 2015. Robustness of next generation sequencing on older formalin-fixed paraffin-embedded tissue. PLoS One 10, e0127353. https://doi.org/10.1371/journal.pone.0127353.

Chang, F., Li, M.M., 2013. Clinical application of amplicon-based next-generation sequencing in cancer. Cancer Gene Ther. 206, 413–419. https://doi.org/10.1016/j.cancergen.2013.10.003.

Do, H., Dobrovic, A., 2015. Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. Clin. Chem. 61, 64–71. https://doi.org/10.1373/clinchem.2014.223040.

Einaga, N., Yoshida, A., Noda, H., Suemitsu, M., Nakayama, Y., Sakurada, A., Kawaji, Y., Yamaguchi, H., Sasaki, Y., Tokino, T., Esumi, M., 2017. Assessment of the quality of DNA from various formalin-fixed paraffin-embedded (FFPE) tissues and the use of this DNA for next-generation sequencing (NGS) with no artifactual mutation. PLoS One 12, e0176280. https://doi.org/10.1371/journal.pone.0176280.

Flaherty, K.T., Gray, R.J., Chen, A.P., Li, S., McShane, L.M., Patton, D., Hamilton, S.R., Williams, P.M., Iafrate, A.J., Sklar, J., Mitchell, E.P., Harris, L.N., Takebe, N., Sims, D.J., Coffey, B., Fu, T., Routbort, M., Zwiebel, J.A., Rubinstein, L.V., Little, R. F., Arteaga, C.L., Comis, R., Abrams, J.S., O'Dwyer, P.J., Conley, B.A., for the NCI-MATCH team, 2020. Molecular landscape and actionable alterations in a genomically guided cancer clinical trial: National Cancer Institute Molecular Analysis for Therapy Choice (NCI-MATCH). J. Clin. Oncol. 38, 3883–3894. https://doi.org/10.1200/JCO.19.03010.

Hedegaard, J., Thorsen, K., Lund, M.K., Hein, A.-M.K., Hamilton-Dutoit, S.J., Vang, S., Nordentoft, I., Birkenkamp-Demtröder, K., Kruhøffer, M., Hager, H., Knudsen, B., Andersen, C.L., Sørensen, K.D., Pedersen, J.S., Ørntoft, T.F., Dyrskjøt, L., 2014. Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human Cancer and Normal tissue. PLoS One 9, e98187. https://doi.org/10.1371/journal.pone.0098187.

Heydt, C., Fassunke, J., Künstlinger, H., Ihle, M.A., König, K., Heukamp, L.C., Schildhaus, H.-U., Odenthal, M., Büttner, R., Merkelbach-Bruse, S., 2014. Comparison of pre-analytical FFPE sample preparation methods and their impact on massively parallel sequencing in routine diagnostics. PLoS One 9, e104566. https://doi.org/10.1371/journal.pone.0104566.

Kerick, M., Isau, M., Timmermann, B., Sültmann, H., Herwig, R., Krobitsch, S., Schaefer, G., Verdorfer, I., Bartsch, G., Klocker, H., Lehrach, H., Schweiger, M.R.,

2011. Targeted high throughput sequencing in clinical cancer settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity. BMC Med. Genet. 4, 68. https://doi.org/10.1186/1755-8794-4-68.

Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K.W., Vogelstein, B., 2011. Detection and quantification of rare mutations with massively parallel sequencing. Proc. Natl. Acad. Sci 108, 9530–9535. https://doi.org/10.1073/pnas.1105422108.

McDonough, S.J., Bhagwate, A., Sun, Z., Wang, C., Zschunke, M., Gorman, J.A., Kopp, K. J., Cunningham, J.M., 2019. Use of FFPE-derived DNA in next generation sequencing: DNA extraction methods. PLoS One 14, e0211400. https://doi.org/10.1371/journal.pone.0211400.

McNulty, S.N., Mann, P.R., Robinson, J.A., Duncavage, E.J., Pfeifer, J.D., 2020. Impact of reducing DNA input on next-generation sequencing library complexity and variant detection. J. Mol. Diagn. 0 https://doi.org/10.1016/j.jmoldx.2020.02.003.

Middleton, G., Fletcher, P., Popat, S., Savage, J., Summers, Y., Greystoke, A., Gilligan, D., Cave, J., O'Rourke, N., Brewster, A., Toy, E., Spicer, J., Jain, P., Dangoor, A., Mackean, M., Forster, M., Farley, A., Wherton, D., Mehmi, M., Sharpe, R., Mills, T.C., Cerone, M.A., Yap, T.A., Watkins, T.B.K., Lim, E., Swanton, C., Billingham, L., 2020. The National Lung Matrix Trial of personalized therapy in lung cancer. Nature 583, 807–812. https://doi.org/10.1038/s41586-020-2481-8.

Pel, J., Leung, A., Choi, W.W.Y., Despotovic, M., Ung, W.L., Shibahara, G., Gelinas, L., Marziali, A., 2018. Rapid and highly-specific generation of targeted DNA sequencing libraries enabled by linking capture probes with universal primers. PLoS One 13, e0208283. https://doi.org/10.1371/journal.pone.0208283.

Robbe, P., Popitsch, N., Knight, S.J.L., Antoniou, P., Becq, J., He, M., Kanapin, A., Samsonova, A., Vavoulis, D.V., Ross, M.T., Kingsbury, Z., Cabes, M., Ramos, S.D.C., Page, S., Dreau, H., Ridout, K., Jones, L.J., Tuff-Lacey, A., Henderson, S., Mason, J., Buffa, F.M., Verrill, C., Maldonado-Perez, D., Roxanis, I., Collantes, E., Browning, L., Dhar, S., Damato, S., Davies, S., Caulfield, M., Bentley, D.R., Taylor, J.C., Turnbull, C., Schuh, A., 2018. Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 genomes project. Genet. Med. 20, 1196–1205. https://doi.org/10.1038/gim.2017.241.

Spencer, D.H., Sehn, J.K., Abel, H.J., Watson, M.A., Pfeifer, J.D., Duncavage, E.J., 2013. Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens. J. Mol. Diagn. 15, 623–633. https://doi.org/10.1016/j.jmoldx.2013.05.004.

Srinivasan, M., Sedmak, D., Jewell, S., 2002. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. Am. J. Pathol. 161, 1961–1971. https://doi.org/10.1016/S0002-9440(10)64472-0.

Stockley, T.L., Oza, A.M., Berman, H.K., Leighl, N.B., Knox, J.J., Shepherd, F.A., Chen, E. X., Krzyzanowska, M.K., Dhani, N., Joshua, A.M., Tsao, M.-S., Serra, S., Clarke, B., Roehrl, M.H., Zhang, T., Sukhai, M.A., Califaretti, N., Trinkaus, M., Shaw, P., van der Kwast, T., Wang, L., Virtanen, C., Kim, R.H., Razak, A.R.A., Hansen, A.R., Yu, C., Pugh, T.J., Kamel-Reid, S., Siu, L.L., Bedard, P.L., 2016. Molecular profiling of advanced solid tumors and patient outcomes with genotype-matched clinical trials: the Princess Margaret IMPACT/COMPACT trial. Genome Med. 8, 109. https://doi.org/10.1186/s13073-016-0364-2.

Xu, C., Nezami Ranjbar, M.R., Wu, Z., DiCarlo, J., Wang, Y., 2017. Detecting very low allele fraction variants using targeted DNA sequencing and a novel molecular barcode-aware variant caller. BMC Genomics 18. https://doi.org/10.1186/s12864-016-3425-4.

Paper III

Cancer Treatment and Research Communications 29 (2021) 100471

Contents lists available at ScienceDirect

Cancer Treatment and Research Communications

journal homepage: www.sciencedirect.com/journal/cancer-treatment-and-research-communications

Associations between tumor mutations in cfDNA and survival in non-small cell lung cancer☆

Anine Larsen Ottestad [a,b], Hong Yan Dai [a,c], Tarje Onsøien Halvorsen [a,b], Elisabeth Fritzke Emdal [c], Sissel Gyrid Freim Wahl [a,c], Bjørn Henning Grønberg [a,b,*]

[a] Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway
[b] Department of Oncology, Cancer Clinic, St. Olav's Hospital, Trondheim University Hospital, 7006, Trondheim, Norway
[c] Department of Pathology, Clinic of Laboratory Medicine, St. Olav's Hospital, Trondheim University Hospital, 7006 Trondheim, Norway

ARTICLE INFO

Keywords:
Next-generation sequencing
Prognosis
Progression-free survival
ctDNA
cfDNA

ABSTRACT

Introduction: Studies have indicated that detection of mutated *KRAS* or *EGFR* in circulating tumor DNA (ctDNA) from pre-treatment plasma samples is a negative prognostic factor for non-small cell lung cancer (NSCLC) patients. This study aims to investigate whether this is the case also for NSCLC patients with other tumor mutations.
Methods: Tumor tissue DNA from 107 NSCLC patients was sequenced and corresponding pre-treatment plasma samples were analyzed using a limited target next-generation sequencing approach validated in this study. Patients without detected mutations in tumor samples were excluded from further analyses.
Results: Mutations were detected in tumor samples from 71 patients. Median age was 68 years, 51% were female, and 88% were current/former smokers, 91% had adenocarcinoma, 4% had squamous cell carcinoma and 6% had other NSCLC. The distribution between stage I, II, III and IV was 33%, 8%, 30%, and 29%, respectively. Between one and three tumor mutation(s) were detected in ctDNA from corresponding plasma samples. Patients with detected ctDNA had shorter PFS (9.6 vs. 41.3 months, HR: 2.9, 95% CI: 1.6–5.2, $p = 0.0003$) and OS (13.6 vs. 115.0 months, HR: 4.0, 95% CI: 2.1–7.6, $p = 0.00002$) than patients without detected ctDNA. ctDNA remained a significant negative prognostic factor for OS (HR: 2.5, 95% CI: 1.1-5.7, p=0.0327), but not PFS, in the multivariable analyses adjusting for baseline patient and disease characteristics including stage of disease.
Conclusions: This study adds further evidence supporting that detectable tumor mutations in cfDNA is associated with a worse prognosis in NSCLC harboring a variety of tumor mutations.

Introduction

The treatment for patients with non-small cell lung cancer (NSCLC) is mainly recommended based on assessment of TNM stage of disease, molecular markers, WHO performance status, and comorbidities [1, 2]. Even if the TNM staging system is based on growing databases and has become more detailed in recent years, patients with the same TNM stage receiving similar treatment still have different outcomes [3].

Circulating tumor DNA (ctDNA) originates in tumor cells and leaks into the circulation [4]. Studies indicate that detection of ctDNA by identification of tumor mutations in plasma collected before treatment is a negative prognostic factor for patients with lung adenocarcinoma [5–7]. Most studies, including one by our group [8], have investigated NSCLC patients with mutations in *KRAS* or *EGFR*. Notably, most NSCLC patients harbor other mutations than *KRAS* or *EGFR*, and the mutation spectrum is heterogeneous between patients [9,10]. ctDNA accounts for a minority of total cell-free DNA (cfDNA) found in plasma and requires sensitive methods for its detection [11]. Droplet digital polymerase chain reaction (ddPCR) is a highly sensitive method for detecting mutations in cfDNA, but it is laborious to analyze samples with different genes. Next-generation sequencing (NGS) allows analysis of any number of genes in each sample and enables analysis of different genes in several samples simultaneously. However, NGS is rather expensive since the detection of low-frequency mutations requires high genome coverage. The cost also depends on the number of target regions, i.e., the number of genes or mutations screened. To keep the costs at an acceptable level,

☆ This work was supported by the Liaison Committee for Education, Research, and Innovation in Central Norway.
* Corresponding author at: Department of Clinical and Molecular Medicine, NTNU, 7491 Trondheim, Norway.
E-mail address: bjorn.h.gronberg@ntnu.no (B.H. Grønberg).

https://doi.org/10.1016/j.ctarc.2021.100471

Available online 29 September 2021
2468-2942/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

we tested and validated an approach to limit the number of target regions in cfDNA to the region(s) found to contain a mutation in the patients' tumor tissue DNA.

The main aim of this study was to investigate whether detectable ctDNA in a pretreatment plasma sample was associated with progression-free survival (PFS) and overall survival (OS) in NSCLC patients with a minimum of one mutation in their tumor DNA using this targeted NGS approach.

Methods

Patients and approval

This study was approved by the Regional Committee for Medical and Health Research Ethics (REC) in Central Norway. Patients were included in our regional research biobank, Biobank1, approved by the REC in Central Norway, the Ministry of Health and Care Services, and the Norwegian Data Protection Authority. Participants are over 18 years old and have given written informed consent.

One hundred and seven patients registered as having adenocarcinoma, diagnosed between March 2007 and April 2018, were included. These were the patients registered with adenocarcinomas in our biobank from which tumor tissue and pretreatment plasma samples were available for analyses. Patients with a tumor mutation in *KRAS* codon 12/13 included in another study were excluded from this study [10]. We considered this sample size and follow-up time sufficient for this exploratory study. All tumor specimens were reviewed and classified according to the WHO 2015 classification system. The disease stage was assessed according to the Eighth Edition of the TNM Classification of lung cancer [14].

Test and validation of limited target next-generation sequencing

We used NGS to analyze mutations in plasma cfDNA that were previously found in the patient's tumor DNA. A limited number of target regions in cfDNA covering the positions of the identified tumor mutation (s) were sequenced by selecting a set of region-specific primers for each patient. Before analyzing patient cfDNA samples, we tested if this approach was feasible. The sensitivity of the NGS approach was investigated by sequencing constructed DNA solutions with different mutant allele frequencies of a *KRAS* mutation. ddPCR was used to validate NGS results from the constructed DNA solutions and nine cfDNA samples from *KRAS* mutated NSCLC patients.

We obtained a *KRAS* G12C positive formalin-fixed paraffin-embedded (FFPE) tumor sample with a known mutant allele frequency (MAF) of 9%. We composed four *KRAS* G12C DNA solutions with defined MAF of 1.15%, 0.17%, 0.016%, and 0.0016%, respectively, by diluting the tumor DNA with peripheral blood DNA from the same patient. In this way, we artificially constructed DNA solutions imitating cfDNA samples with known concentrations of ctDNA.

NGS libraries were made using 40–48 ng DNA and reagents from QIAseq Human Targeted DNA panel (Qiagen, Hilden, Germany) except the primer mix. In brief, DNA was fragmented, end-repaired, and a-tailed followed by ligation to a 5′ adapter. Adapters contained a unique molecular index (UMI) that provided a unique tag for all original DNA fragments. The ligated fragments were purified, and the region of interest was selected by PCR using an adapter primer and a 1 nM solution of region-specific primers with a 5′ universal sequence. The concentration of gene-specific primers highly exceeded the concentration of input genomic DNA in the hybridization reaction, ensuring that complementary DNA was formed from all available genomic templates. The PCR products were then purified and amplified in a second PCR using an adapter primer and a primer complementary to the universal sequence. Libraries were sequenced using the MiSeq platform (Illumina, San Diego, CA).

NGS mutation analysis was performed using CLC Biomedical Workbench v.20.0 (Qiagen, Hilden, Germany) with a ready-to-use workflow for QIAseq Targeted DNA panels. Following mapping reads to the genome, any reads outside the region of interest were excluded. Two or more reads with the same UMI were grouped into a "UMI family." Single reads with no duplicates were discarded. Variants were called if 75% of duplicates in a UMI family contained the variant. Only the position of the mutation of interest was considered. A solution was classified as "mutation detected" if *KRAS* G12C was detected in at least one big UMI family, defined as a family made from $\geq$ 4 duplicates. ddPCR (Bio-Rad Laboratories, Hercules, CA) was performed using 40–48 ng constructed cfDNA. NGS was carried out in duplicates and these duplicate solutions were analyzed by ddPCR in quadruplicates. Patient cfDNA was analyzed in triplicates by ddPCR.

Tumor DNa sequencing and region selection

Tumor DNA was isolated from diagnostic FFPE tumor tissue using QIAamp DNA FFPE Tissue Kit (Qiagen, Hilden, Germany) and sequenced using QIAseq Human Actionable Solid Tumor panel (Qiagen, Hilden, Germany). This panel covered the whole coding region of the genes *ERBB2, PIK3CA,* and *TP53*, the exons of *BRAF, EGFR, KIT, KRAS, NRAS,* and *PDGFRA*, and hotspots in *AKT1, ALK, CTNNB1, ERBB3, ESR1, FOXL2, GNA11, GNAQ, IDH1, IDH2, MET, RAF1*, and *RET*. Theoretically, this panel enables the detection of at least one mutation in about 65% of adenocarcinoma tumors and more than 80% of squamous cell carcinoma tumors [9,10].

Bioinformatic analysis was performed using CLC Genomic Workbench 20.0 (Qiagen, Hilden, Germany) with a ready-to-use workflow for QIAseq Targeted DNA panels. Variants were classified as a mutation if the variant was non-synonymous with a MAF of at least 5%. Patients without detected tumor mutation were excluded from further analyses.

For each tumor mutation, we identified the primers used to amplify the region with the mutation. A 5′ universal sequence was added to all primers, and the primers were synthesized by Eurogentec (Liège, Belgium).

Patient cfDNA extraction

Plasma samples from the 71 patients with at least one detected tumor tissue mutation were analyzed. The median time from blood draw to tissue biopsy was one day (range 0–213). At blood draw, plasma was prepared from 10 mL whole blood with EDTA or citrate anticoagulant. Within two hours of sampling, the blood samples were either centrifuged once at 2500x g for 10 min or first at 1500x g for 15 min, and then at 10,000x g for 10 min. Plasma was transferred to cryotubes and stored at -80 °C. cfDNA was isolated using QIAamp Circulating Nucleic Acid kit (Qiagen, Hilden, Germany) from 1.6 to 6 mL plasma and eluted in 50 uL of the supplied buffer. DNA concentration was measured by Qubit® (Thermo Fisher Scientific, Waltham, MA), using the dsDNA HS Assay Kit.

cfDNA analyses by limited target NGS

NGS libraries were made using a variable amount of cfDNA (5.3–72.5 ng) and as described for the constructed samples. In 63 samples, primers were added to target one region (i.e., one mutation). Two regions were targeted in eight samples and three regions were targeted in three samples. The goal was to achieve the same genomic coverage at each target region in all samples. Therefore, when libraries were pooled before sequencing, we doubled the library amount from the eight samples with two targets. Similarly, we tripled the amount of the libraries with three targets. The number of libraries in each pool was adjusted to generate 700–800,000 reads per target using the MiSeq v3 platform (Illumina, San Diego, CA). NGS data were analyzed as described for the constructed DNA solutions.

Statistics

PFS was defined as the number of days from lung cancer diagnosis until progression or death of any cause. OS was defined as the number of days from diagnosis until death of any cause. Patients were treated and followed according to local routines. The median follow-up time for PFS and OS were estimated using the reverse Kaplan–Meier method, and the median PFS and OS were estimated using the Kaplan-Meier method. Survival analyses were performed using Cox proportional hazard models. The multivariable model was adjusted for sex, age, WHO performance status (PS), and disease stage. The threshold for statistical significance was set at 0.05. All statistical analyses were performed using R (version 3.6.1).

Results

Patient characteristics

A summary of patient characteristics is presented in Table 1. The median age was 68 years (range 48–86), 54 (51%) were female, and 94 (88%) were current or former smokers, 97 patients (91%) had adenocarcinoma, four (4%) had squamous cell carcinoma, one (1%) had adenosquamous carcinoma, and five (5%) had NSCLC not otherwise specified. Thirty-five patients (33%) had stage I disease, nine (8%) stage II, 32 (30%) stage III, and 31 (29%) stage IV. Individual patient characteristics are presented in Table A1.

Validation of limited target NGS

The strategy for ctDNA detection was to use NGS to only analyze for mutations found in the patients' tumor DNA. The performance of this limited target NGS approach was explored by constructing four solutions made to mimic cfDNA with theoretical MAF of *KRAS* G12C at 1.15%, 0.17%, 0.016%, and 0.0016% that were analyzed by both NGS and ddPCR. Comparable results were obtained in the solutions with MAFs 1.15% and 0.17% (Fig. 1A). In addition, the mutation in the solution with MAF 0.0016% was detected using ddPCR. While the input DNA amount was similar in both NGS and ddPCR, the number of unique human genome equivalents (hGEs) analyzed by NGS was 5600–8070 compared to 10,580–17,774 by ddPCR. The partitioning of DNA in ddPCR likely contributed to superior sensitivity.

We proceeded to analyze cfDNA from nine patients with *KRAS*-mutated tumors using both NGS and ddPCR, and comparable results were obtained by the two techniques. An excellent correlation was observed between the observed MAFs by NGS and ddPCR in the constructed solutions and patient samples together (adjusted $R^2 = 0.9944$, $p = 2.2 \times 10^{-16}$) (Fig. 1B). Concordant results between NGS and ddPCR were observed in all samples except one, where a mutation was detected with MAF 0.17% by NGS but not detected by ddPCR. This may be attributed to the different strategies for target amplification. In NGS, amplification required only one target-specific primer to bind to DNA while ddPCR required two. Consequently, short DNA fragments that did not contain both primer sites would not be amplified and analyzed.

Although we did not validate the efficiency for the primers of non-KRAS genes for the NGS analysis of cfDNA specifically, we reasoned from NGS data of tumor DNA that all available genomic templates in cfDNA were converted to complementary DNA when an excess of gene-specific primers were used in the first step of the hybridization reactions. The primers used in the following PCR reactions were the same for the KRAS and non-KRAS genes which ensured similar sensitivity of mutation detection for all genes.

Table 1
Patient characteristics.

	All patients ($n = 107$) Number (%)	Patients included for cfDNA analyses ($n = 71$) Number (%)	Patients with detected ctDNA ($n = 29$) Number (%)	Patients without detected ctDNA ($n = 42$) Number (%)
Age (median, range)	68 (48–86)	68 (50–86)	66 (50–81)	69 (54–86)
Sex				
Female	54 (50)	38 (54)	16 (55)	22 (52)
Male	53 (50)	33 (46)	13 (45)	20 (48)
Smoking history				
Never-smoker	13 (12)	10 (14)	3 (10)	7 (17)
Current or former smoker	94 (88)	61 (86)	26 (90)	35 (83)
Histology				
Adenocarcinoma	97 (91)	65 (92)	26 (90)	39 (93)
Squamous cell carcinoma	4 (4)	3 (4)	2 (7)	1 (2)
Adenosquamous carcinoma	1 (1)	1 (1)	0	1 (2)
NSCLC—NOS	5 (5)	2 (3)	1 (3)	1 (2)
PS				
0	59 (55)	41 (58)	13 (45)	28 (67)
1	40 (37)	26 (37)	12 (41)	14 (33)
2	8 (8)	4 (6)	4 (14)	0
Disease stage				
I	35 (33)	24 (34)	1 (3)	23 (55)
II	9 (8)	7 (10)	2 (7)	5 (12)
III	32 (30)	20 (28)	15 (52)	5 (12)
IV	31 (29)	20 (28)	11 (38)	9 (21)

NSCLC—NOS: non-small cell lung cancer – not otherwise specified, PS: WHO performance status.

Tumor DNA mutation detection and region selection

At least one tumor mutation was detected in 71 patients (66%). One mutation was detected in 60 patients, two mutations in nine patients, and three mutations in two patients. The 85 mutations were detected in the genes *TP53* (54%), *EGFR* (14%), *KRAS* (12%), *PIK3CA* (9%), *BRAF* (4%), *ERBB2* (1%), *ALK* (1%), *ERBB3* (1%), *NRAS* (1%), *PDGFRA* (1%) and *RAF1* (1%). The specific tumor mutations are listed in Table A.2. For each tumor mutation, we identified the panel primers flanking the mutated region. Between one and nine target-specific primers were selected for each patient.

Detection of ctDNA by limited target NGS

NGS libraries were made using 1620–21,982 hGEs (median 6482) and target-specific primers determined by the tumor DNA sequencing. One mutation (i.e., one region) was targeted in cfDNA from 61 patients, two mutations in seven patients, and three mutations in three patients. (Accidently, only one region was targeted in the plasma sample from patient 45 with two detected mutations in the tumor).

The number of UMI families that covered the position of the tumor mutation ranged from 266 to 5955 (median 973). One UMI family represents one hGE from the original sample. On average, the number of UMI families was 22% of the number of hGE used for library preparation (range 2−71%). The loss of hGEs was greater than the loss observed in the constructed solutions (50%). cfDNA is more fragmented and consists of shorter fragments than DNA from whole blood that was used to construct the artificial cfDNA. Therefore, fewer DNA molecules may have been available for the first PCR with gene-specific primer binding in the patient cfDNA samples which contributes to a greater loss.

At least one tumor mutation first detected in tumor DNA was also detected in corresponding cfDNA samples from 29/71 patients (41%). In total, 32 such mutations were detected with MAFs between 0.05% and 65.7% (median 1.8%). The mutation with MAF 65.7% was an exon 20 insertion in *EGFR* in a sample where another mutation in *TP53* with MAF 6.3% was detected. We believe that the high MAF was most likely caused by an an amplification of the *EGFR* gene in the tumor, since this

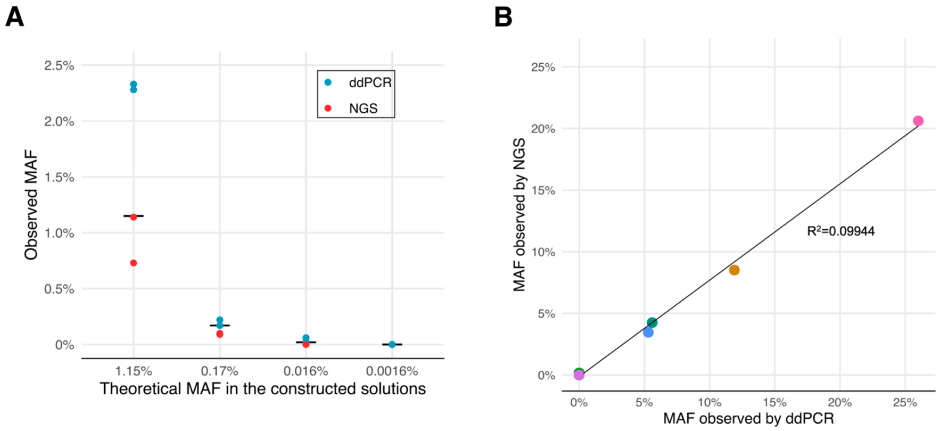


Fig. 1. To test the limited target NGS method, we constructed cfDNA solutions and patient cfDNA samples that were analyzed by both NGS and ddPCR. **(A)** Four DNA solutions were constructed to mimic cfDNA with KRAS G12C mutation with MAF 1.15%, 0.17%, 0.016%, and 0.0016%. Each solution was analyzed in duplicates. The horizontal line represents the theoretical MAF. **(B)** MAF of KRAS codon 12 mutations in nine patient cfDNA samples analyzed by both NGS and ddPCR. The mutation was detected in five samples and undetected by both technologies in four samples. cfDNA: circulating cell-free DNA, ddPCR: droplet digital polymerase chain reaction, NGS: next-generation sequencing, MAF: mutant allele frequency.

mutation has consistently been reported to be a tumor-associated mutation in NSCLC and is often amplified (although a germline mutation cannot be ruled out since we did not sequence normal DNA from our patients). For the other samples, the MAFs ranged from 0.05% to 24.5%. Individual mutation data are listed in Table A.2.

In 8/10 cfDNA samples where more than one region was targeted, either all or none of the tumor mutations were detected. In 2/10 patients (patients 21 and 22), only one mutation was detected in each sample. ctDNA was detected in 1/24 (4%) patients with stage I, 2/7 (29%) with stage II, 15/20 (75%) with stage III, and 11/20 (55%) with stage IV disease.

Association between ctDNA detection and progression-free survival (PFS)

Median follow-up time for PFS was 88.7 months (95% CI: 45.2–105.9) and 23 patients were alive and relapse-free at the time of data analysis in November 2020. Overall, the median PFS was 17.5 months (95% CI: 7.6–126.2). Median PFS was significantly shorter for patients with detected ctDNA than for those without detected ctDNA (9.6 months vs. 41.3 months, HR: 2.9, 95% CI: 1.6–5.2, $p = 0.000325$) (Fig. 2A). In the multivariable analysis, PS and disease stage, but not detectable ctDNA, were significantly associated with PFS (Table 2). In terms of two-year PFS, 83% of patients with detected ctDNA relapsed or died within two years, compared to 38% of patients without detected ctDNA.

Association between ctDNA detection and overall survival (OS)

Median follow-up time for OS was 65.6 months (95% CI 45.2–106.0) and 28 patients were alive at the time of data analysis. Overall, the median OS was 27.5 months (95% CI: 13.0–126.2). Median OS was significantly shorter for patients with detected ctDNA than for patients without (13.6 months vs. 115.0 months (HR: 4.0, 95% CI: 2.1–7.6, $p = 0.0000201$) (Fig. 2B). The multivariable analysis showed that detected ctDNA, stage IV and PS 2 were significant, negative prognostic factors (Table 2).

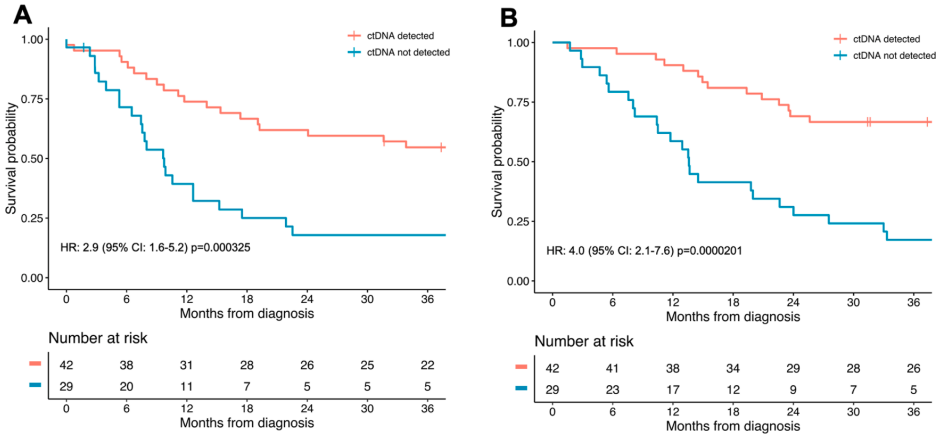


Fig. 2. Kaplan-Meier plot for **(A)** progression-free survival and **(B)** overall survival. CI: confidence interval, ctDNA: circulating tumor DNA, HR**:** hazard ratio.

Table 2
Cox multivariable model for PFS and OS. Statistically significant values are given in bold.

	Progression-free survival		Overall survival	
	Hazard-ratio (95% CI)	*p*-value	Hazard-ratio (95% CI)	*p*-value
Sex (male vs female)	0.94 (0.52–1.70)	0.827	0.91 (0.48–1.74)	0.785
Age	1.02 (0.98–1.06)	0.227	1.03 (0.99–1.07)	0.160
Stage II vs I	1.07 (0.28–4.13)	0.920	1.16 (0.21–6.33)	0.865
Stage III vs I	2.68 (0.97–7.35)	0.0562	3.13 (0.90–10.91)	0.0727
Stage IV vs I	7.00 (2.81–17.40)	**0.0000286**	9.56 (3.19–28.67)	**0.0000555**
PS (2 vs 1/0)	13.07 (2.59–66.00)	**0.00185**	10.31 (2.09–50.92)	**0.00421**
Detection of ctDNA	1.58 (0.75–3.32)	0.227	2.49 (1.08–5.74)	**0.0327**

CI: confidence interval, ctDNA: circulating tumor DNA, PS: WHO performance status.

Discussion

In this study of 107 patients with all stages of NSCLC, we detected tumor mutation(s) in 66% of the samples, and when sequencing corresponding pre-treatment plasma samples from these 71 patients, we detected the same mutation(s) in 41% of patients. We found that detection of ctDNA in plasma was significantly associated with shorter PFS and OS in the univariable analyses, and ctDNA remained a significant negative prognostic factor for OS in the multivariable analyses.

There are several other studies of the prognostic role of ctDNA in NSCLC. Pavan et al. found that TP53 mutations in plasma detected by NGS, the most commonly detected mutation in ctDNA in our cohort, negatively affected survival both in NSCLC patients who received immune checkpoint inhibitor therapy and those who did not [13]. Michaelidou et al. used ddPCR to analyze *KRAS* mutations in cfDNA from 114 advanced NSCLC patients with tumor *KRAS* status either mutated, wild-type, or unknown, and found detection of mutated *KRAS* in ctDNA to be significantly associated with both PFS and OS [5]. In a previous study by our group, we used ddPCR to analyze matching tumor and plasma samples from 60 patients with known *KRAS* mutations and found that detectable *KRAS* in cfDNA was significantly associated with both PFS and OS [8]. Peng et al. found similar associations by sequencing both tumor tissue and cfDNA from 77 patients with resectable NSCLC using a 127-gene panel [6]. Mardinian et al. used NGS to analyze *KRAS* mutations both in tumor tissue and cfDNA from 433 patients with various cancer types, including NSCLC, and found a significant association between ctDNA detection and shorter OS [7]. Interestingly, they also showed that the value of ctDNA as a prognostic marker was greater when *KRAS* mutation was detected in both tumor tissue and cfDNA, compared to either sample type alone.

However, results are not uniform across all studies. In a study of 58 *KRAS*-mutated NSCLC patients, detection of *KRAS* in cfDNA was not associated with shorter PFS [12], while two recent studies found similar associations as we report here [5,6]. These studies also observed an independent prognostic association between ctDNA detection and PFS. Differences in patient selection with respect to histology and disease stage, and frequency of computed tomography evaluation might explain why results differ.

The main limitation of all these studies, including ours, is the sample sizes, and in particular the number of patients with low disease stage and detectable ctDNA was too low to draw firm conclusions. Another limitation is that we only detected tumor mutations in 71 of the 107 patients included in our study. A broader NGS panel or sampling tissue from different parts of the tumors might increase the mutation detection rate, but the latter is limited by the access to routinely obtained tissue. Analyzing a panel of genes in cfDNA alone could overcome the above-mentioned challenges associated with tumor tissue analysis. On the other hand, limiting the cfDNA analysis to known tumor mutations reduces the chance of detecting false positives in cfDNA. The NGS approach applied in this study was a cost-efficient method for analyses of tumor mutations in cfDNA. In the study by Peng et al. using a 127-gene panel, 1.2 mutations were on average detected in the tumor samples. cfDNA was analyzed for the matching mutations using the same 127-gene panel, which demonstrates that large panel sequencing of cfDNA generates a myriad of uninformative data [6].

Another potential limitation is that we did not assess total tumor volume, which has been shown to be associated with presence of ctDNA [14]. Thus, we cannot rule out that presence of ctDNA is a surrogate marker for large total tumor volume in our cohort. On the other hand, the impact of tumor volume varies with TNM stage and treatment, and tumor volume is not routinely assessed in the clinic.

This was a retrospective study and patients were included in the biobank over a long time period, and there was a large variation in time from plasma samples were collected until biopsies were obtained. There were major changes in diagnostic workup during this period (e.g. PET CT for staging of disease and reflex-testing for EGFR-mutations and ALK-rearrangements were introduced during this period). Furthermore, there was no standardized schedule for follow-up or imaging. In general, patients with advanced disease are followed more closely than patients who have undergone potentially curative treatment. We also did not adjust for treatment. Treatment is strongly correlated with disease stage, each treatment group was very small, there were major changes in treatment policy and the number of available therapies increased rapidly during the period patients were included in the biobank. The influence of treatment is especially relevant for those who have targetable mutations, but not all patients respond to targeted therapy, and there is a large variation in treatment response and response duration for all administered therapies. These differences in diagnostic workup, follow-up and treatment may explain why detection of ctDNA did not remain a significant prognostic factor for PFS in the multivariable analysis, and why there was no statistically significant difference in PFS or OS between the few stage II patients and stage I patients in the multivariable analyses in Table 2, and only a trend towards differences between stage III and stage I patients.

A challenge for any detection technology is the low amount of cfDNA that is available from a plasma sample. This is especially an issue when there is additional loss of DNA in preparation for NGS. We observed a 50% loss in the solutions of constructed cfDNA, and a bigger loss in the patient cfDNA samples (median 78% loss). In contrast, we observed almost no loss in ddPCR. When analyzing the constructed solutions, we observed comparable performance of the NGS approach to ddPCR but cannot rule out that the loss in preparation for NGS led to a lower detection rate than if we had used ddPCR for analyzing the patient samples. The most important reason for using NGS is the ability to detect various mutations in several samples simultaneously. Except for *KRAS* and *EGFR*, only six pairs of patients in our study shared the same point mutation. Furthermore, the cfDNA amount is limited by the available plasma volume. Using an NGS panel enables analysis of several mutations in cfDNA without requiring larger plasma volumes and increases the likelihood of detecting at least one mutation [15]. This was the case for two patients in this study, in which one out of two or three mutations was detected in cfDNA.

Our results add further evidence supporting that detection of tumor mutations in cfDNA is associated with a worse prognosis in NSCLC. A prominent feature of our study is that it suggests that this association is not limited to *EGFR* or *KRAS*. Furthermore, it shows that a small, customized NGS panel may be used for the analysis of cfDNA, which has important implications for feasibility in routine clinical practice. The use of a customized NGS panel increases the sensitivity of detecting ctDNA and reduces the risk of false positives, but the requirement of analyzable

tumor samples limits the use of this approach. Larger prospective clinical trials are necessary to fully explore the clinical value of cfDNA analyses, and several other issues need to be addressed; in general, there is a lack of standardized methods for cfDNA analyses, plasma collection/processing/storage, data interpretation and definition of relevant mutations. Finally, the mechanisms explaining why ctDNA is a negative prognostic factor should be explored.

In conclusion, we found that detectable ctDNA was a negative prognostic factor in NSCLC patients with various tumor mutation spectrums.

CRediT authorship contribution statement

Anine Larsen Ottestad: Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft. **Hong Yan Dai:** Conceptualization, Methodology, Writing – original draft. **Tarje Onsøien Halvorsen:** Formal analysis, Writing – review & editing. **Elisabeth Fritzke Emdal:** Investigation, Resources, Writing – review & editing. **Sissel Gyrid Freim Wahl:** Investigation, Writing – review & editing. **Bjørn Henning Grønberg:** Conceptualization, Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: A.L.O.: the institution has received grants from AstraZeneca, grants from Roche, during the conduct of the study, H.Y.D.: none T.O.H.: personal fees from Pfizer, personal fees from Pierre Fabre, outside the submitted work, E.F.E.: none, S.G.F.W.: none, B.H.G.: personal fees from MSD, grants and personal fees from AstraZeneca, personal fees from Debiopharm, personal fees from Pfizer, personal fees from Janssen-Cilag, personal fees from Sanofi, personal fees from Takeda, personal fees from Novartis, grants and personal fees from Roche, personal fees from Pierre Fabre, outside the submitted work.

Acknowledgement

The next-generation sequencing was provided by the Genomics Core Facility (GCF), Norwegian University of Science and Technology (NTNU). GCF is funded by the Faculty of Medicine and Health Sciences at NTNU and Central Norway Regional Health Authority. We thank Biobank1 (St. Olav's Hospital) for collecting and storing the plasma samples.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ctarc.2021.100471.

References

[1] N.H. Hanna, A.G. Robinson, S. Temin, et al., Therapy for stage IV non–small-cell lung cancer with driver alterations: ASCO and OH (CCO) joint guideline update, J. Clin. Oncol. 39 (9) (2021) 1040–1091, https://doi.org/10.1200/JCO.20.03570.

[2] N.H. Hanna, B.J. Schneider, S. Temin, et al., Therapy for stage IV non–small-cell lung cancer without driver alterations: ASCO and OH (CCO) joint guideline update, J. Clin. Oncol. 38 (14) (2020) 1608–1632, https://doi.org/10.1200/JCO.19.03022.

[3] P. Goldstraw, K. Chansky, J. Crowley, et al., The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer, J. Thorac. Oncol. 11 (1) (2016) 39–51, https://doi.org/10.1016/j.jtho.2015.09.009.

[4] M. Stroun, J. Lyautey, C. Lederrey, A. Olson-Sand, P. Anker, About the possible origin and mechanism of circulating DNA apoptosis and active DNA release, Clin. Chim. Acta Int. J. Clin. Chem. 313 (1–2) (2001) 139–142, https://doi.org/10.1016/s0009-8981(01)00665-9.

[5] K. Michaelidou, C. Koutoulaki, K. Mavridis, et al., Detection of KRAS G12/G13 mutations in cell free-DNA by droplet digital PCR, offers prognostic information for patients with advanced non-small cell lung cancer, Cells. 9 (11) (2020) 2514, https://doi.org/10.3390/cells9112514.

[6] M. Peng, Q. Huang, W. Yin, et al., Circulating tumor dna as a prognostic biomarker in localized non-small cell lung cancer, Front. Oncol. 10 (2020), https://doi.org/10.3389/fonc.2020.561598.

[7] K. Mardinian, R. Okamura, S. Kato, R. Kurzrock, Temporal and spatial effects and survival outcomes associated with concordance between tissue and blood KRAS alterations in the pan-cancer setting, Int. J. Cancer 146 (2) (2020) 566–576, https://doi.org/10.1002/ijc.32510.

[8] Wahl S.G.F., Dai H.Y., Emdal E.F., et al. Prognostic value of absolute quantification of mutated KRAS in circulating tumour DNA in lung adenocarcinoma patients prior to therapy. J. Pathol. Clin. Res.. Published online January 27, 2021:cjp2.200. doi: 10.1002/cjp2.200.

[9] The Cancer Genome Atlas Research Network, Comprehensive genomic characterization of squamous cell lung cancers, Nature 489 (7417) (2012) 519–525, https://doi.org/10.1038/nature11404.

[10] The cancer genome atlas research network. comprehensive molecular profiling of lung adenocarcinoma, Nature 511 (7511) (2014) 543–550, https://doi.org/10.1038/nature13385.

[11] C. Fiala, E.P. Diamandis, Utility of circulating tumor DNA in cancer diagnostics with emphasis on early detection, BMC Med. 16 (1) (2018), https://doi.org/10.1186/s12916-018-1157-9.

[12] E. Zulato, I. Attili, A. Pavan, et al., Early assessment of KRAS mutation in cfDNA correlates with risk of progression and death in advanced non-small-cell lung cancer, Br. J. Cancer 123 (1) (2020) 81–91, https://doi.org/10.1038/s41416-020-0833-7.

[13] A. Pavan, A.B. Bragadin, L. Calvetti, et al., Role of next generation sequencing-based liquid biopsy in advanced non-small cell lung cancer patients treated with immune checkpoint inhibitors: impact of STK11, KRAS and TP53 mutations and co-mutations on outcome, Transl. Lung Cancer Res. 10 (1) (2021) 202–220, https://doi.org/10.21037/tlcr-20-674. Jan.

[14] T. Ohira, K. Sakai, J. Matsubayashi, et al., Tumor volume determines the feasibility of cell-free DNA sequencing for mutation detection in non-small cell lung cancer, Cancer Sci. 107 (11) (2016) 1660–1666, https://doi.org/10.1111/cas.13068.

[15] B.R. McDonald, T. Contente-Cuomo, S.-.J. Sammut, et al., Personalized circulating tumor DNA analysis to detect residual disease after neoadjuvant therapy in breast cancer, Sci. Transl. Med. 11 (504) (2019) eaax7392, https://doi.org/10.1126/scitranslmed.aax7392.

Corrected figures for paper III –

Associations between tumor mutations in cfDNA and survival in non-small cell lung cancer

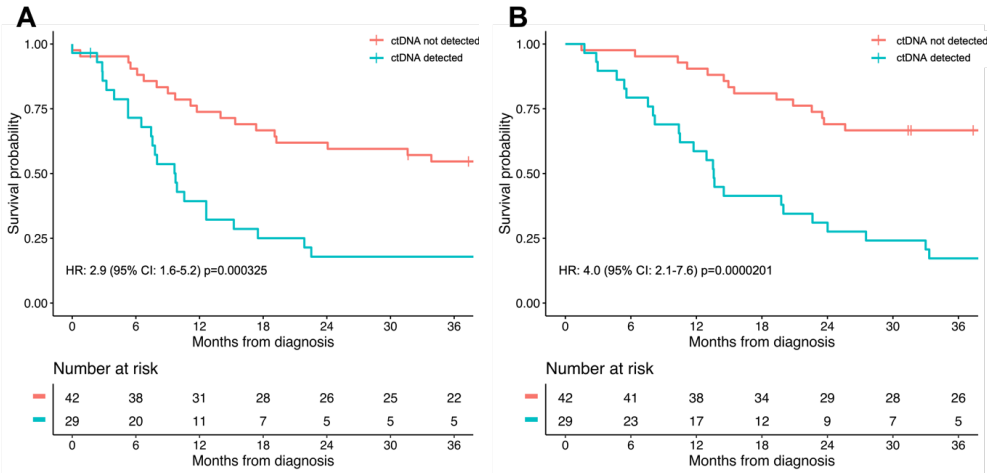


Figure 2: Kaplan-Meier plot for (A) progression-free survival and (B) overall survival. The color legend was incorrect (opposite) in the published version of the article. CI: confidence interval, ctDNA: circulating tumor DNA, HR: hazard ratio

Table A1. Detailed patient characteristics

Patient	Sex	Age at diagnosis	Histology	Disease stage	Smoker/ former smoker	WHO performance status	No. of mutations detected in tumor DNA
1	Male	62	Adenocarcinoma	IVA	Yes	1	0
2	Female	56	Adenocarcinoma	IB	Yes	0	1
3	Female	79	Squamous cell carcinoma	IIIA	Yes	2	1
4	Male	79	Adenocarcinoma	IVB	Yes	1	1
5	Male	66	Adenocarcinoma	IVB	Yes	1	1
6	Male	79	Squamous cell carcinoma	IVA	Yes	1	0
7	Female	62	Adenocarcinoma	IIB	Yes	0	0
8	Female	60	NSCLC-NOS	IIIA	Yes	1	1
9	Female	60	Adenocarcinoma	IIIA	Yes	1	0
10	Male	68	Adenocarcinoma	IB	Yes	0	1
11	Female	66	Adenocarcinoma	IIIB	Yes	1	0
12	Female	71	Adenocarcinoma	IA	Yes	1	2
13	Male	66	Adenocarcinoma	IVA	Yes	2	1
14	Female	62	Adenocarcinoma	IVA	No	0	2
15	Male	77	Adenocarcinoma	IVB	Yes	2	0
16	Male	60	Adenocarcinoma	IA	Yes	0	1
17	Male	60	Adenocarcinoma	IA	No	0	1
18	Male	73	Adenocarcinoma	IVB	Yes	2	0
19	Female	79	Adenocarcinoma	IIIB	Yes	2	0
20	Male	60	Adenocarcinoma	IVB	Yes	0	1
21	Female	65	Adenocarcinoma	IB	Yes	1	2
22	Female	66	Adenocarcinoma	IIIA	Yes	1	3
23	Male	61	Adenocarcinoma	IIIB	Yes	1	0
24	Male	61	Adenocarcinoma	IVB	Yes	1	0
25	Male	48	Adenocarcinoma	IIB	Yes	0	0
26	Female	57	Adenocarcinoma	IVA	Yes	0	0
27	Male	78	Adenocarcinoma	IIIB	Yes	0	0
28	Female	62	Adenocarcinoma	IVA	Yes	1	0
29	Male	62	Adenocarcinoma	IB	Yes	0	2
30	Female	71	Adenocarcinoma	IVB	Yes	1	1
31	Female	66	Adenocarcinoma	IVA	Yes	1	1
32	Male	69	Adenocarcinoma	IVA	Yes	0	0
33	Male	70	Adenocarcinoma	IIIB	Yes	0	0
34	Male	77	Adenocarcinoma	IVA	Yes	0	1
35	Male	63	Adenocarcinoma	IIIB	Yes	1	0
36	Female	68	Adenocarcinoma	IB	Yes	0	1
37	Male	81	NSCLC-NOS	IB	Yes	0	1
38	Male	79	Squamous cell carcinoma	IIA	Yes	0	1
39	Female	56	Adenocarcinoma	IA	No	0	1
40	Female	70	Adenocarcinoma	IA	Yes	0	1
41	Female	81	Adenocarcinoma	IA	Yes	0	1
42	Male	71	Adenocarcinoma	IA	Yes	0	0
43	Female	72	Adenocarcinoma	IVA	No	0	1
44	Male	81	Adenocarcinoma	IA	Yes	1	1
45	Female	57	Adenocarcinoma	IVA	No	0	2
46	Female	81	Adenocarcinoma	IIIA	Yes	1	1
47	Male	80	Adenocarcinoma	IIIA	Yes	0	0
48	Female	58	Adenocarcinoma	IIA	Yes	0	1
49	Female	63	Adenocarcinoma	IA	Yes	0	1
50	Female	81	Adenocarcinoma	IIIA	Yes	1	0
51	Male	69	Squamous cell carcinoma	IIIB	Yes	0	1
52	Male	73	Adenocarcinoma	IA	Yes	0	0

53	Female	63	Adenocarcinoma	IIIB	Yes	0	1
54	Male	58	Adenocarcinoma	IVA	Yes	0	1
55	Female	72	Adenocarcinoma	IVB	Yes	2	2
56	Male	74	Adenocarcinoma	IVA	Yes	2	1
57	Male	73	Adenocarcinoma	IIIB	Yes	0	1
58	Female	78	NSCLC-NOS	IVA	Yes	0	0
59	Male	75	Adenocarcinoma	IVA	Yes	1	1
60	Female	66	Adenocarcinoma	IA	No	2	0
61	Female	59	Adenocarcinoma	IVB	Yes	0	1
62	Male	80	Adenocarcinoma	IIIA	Yes	1	1
63	Female	70	Adenocarcinoma	IB	Yes	0	0
64	Male	82	Adenocarcinoma	IVA	Yes	1	0
65	Male	55	Adenocarcinoma	IIIA	Yes	0	0
66	Female	80	Adenocarcinoma	IVA	No	0	1
67	Male	58	Adenocarcinoma	IA	No	0	1
68	Male	67	Adenocarcinoma	IVA	Yes	1	2
69	Female	72	Adenocarcinoma	IVA	Yes	0	1
70	Female	65	NSCLC-NOS	IIIB	Yes	1	0
71	Male	52	Adenocarcinoma	IIIA	Yes	0	1
72	Male	67	Adenocarcinoma	IA	Yes	0	0
73	Female	81	Adenocarcinoma	IIIA	No	1	1
74	Female	64	Adenocarcinoma	IB	Yes	1	0
75	Female	50	Adenocarcinoma	IVA	Yes	1	1
76	Male	72	Adenocarcinoma	IIIA	Yes	1	1
77	Male	79	Adenocarcinoma	IIIA	Yes	1	1
78	Male	63	Adenocarcinoma	IVA	Yes	0	1
79	Female	69	Adenocarcinoma	IB	Yes	0	1
80	Female	56	Adenocarcinoma	IIIB	No	0	1
81	Male	81	Adenocarcinoma	IIIA	Yes	1	2
82	Female	77	Adenocarcinoma	IIB	Yes	1	1
83	Female	66	Adenocarcinoma	IIIA	Yes	1	1
84	Female	70	Adenocarcinoma	IB	Yes	0	1
85	Male	78	NSCLC-NOS	IIIB	Yes	1	0
86	Female	72	Adenocarcinoma	IA	No	0	0
87	Female	74	Adenocarcinoma	IA	Yes	1	0
88	Female	65	Adenocarcinoma	IIIA	Yes	0	1
89	Female	53	Adenocarcinoma	IA	Yes	0	0
90	Female	77	Adenocarcinoma	IA	No	0	0
91	Female	70	Adenosquamous carcinoma	IA	Yes	1	1
92	Male	70	Adenocarcinoma	IIB	Yes	0	1
93	Female	66	Adenocarcinoma	IIB	Yes	0	1
94	Male	72	Adenocarcinoma	IIIA	Yes	0	1
95	Male	65	Adenocarcinoma	IIB	Yes	1	1
96	Female	54	Adenocarcinoma	IB	Yes	0	1
97	Female	72	Adenocarcinoma	IIIA	Yes	1	1
98	Male	76	Adenocarcinoma	IB	Yes	1	1
99	Male	86	Adenocarcinoma	IA	Yes	1	1
100	Male	58	Adenocarcinoma	IB	Yes	0	3
101	Male	57	Adenocarcinoma	IB	Yes	0	1
102	Female	56	Adenocarcinoma	IIIA	No	0	3
103	Female	63	Adenocarcinoma	IA	Yes	1	1
104	Male	72	Adenocarcinoma	IVA	Yes	0	0
105	Male	66	Adenocarcinoma	IB	Yes	0	0
106	Female	70	Adenocarcinoma	IIIA	Yes	0	1
107	Male	59	Adenocarcinoma	IIB	Yes	0	1

NSCLC-NOS: non-small cell lung cancer - not otherwise specified

Table A2. Outcomes and mutations detected in tumor tissue DNA and circulating tumor DNA

Patient	Time from diagnosis until death (months)	Death before data analysis. 1=yes, 0=no	Time from diagnosis until progression or death (months)	Relapse/ death before data analysis. 1=yes, 0=no	Time from blood draw until treatment (days)	Mutated gene	Amino acid change	Tumor MAF (unique coverage)	ctDNA MAF (unique coverage)
2	72.7	1	38.4	1	43	*PIK3CA*	NP_006209.2:p.Glu545Gln	12,28% (57)	0% (772)
3	5.4	1	0.0	1	1	*TP53*	NP_001119586.1:p.Arg273Leu	56,25% (48)	8,32% (997)
4	1.5	1	0.8	1	16	*ERBB2*	NP_001276865.1:p.Thr902Ile	16,95% (59)	0% (5955)
5	6.4	1	5.3	1	27	*ALK*	NP_004295.2:p.Pro1027Leu	13,22% (121)	0% (2590)
8	8.0	1	7.4	1	68	*TP53*	NP_001263625.1:p.Glu185Asp	69,27% (872)	0,37% (803)
10	151.1	0	151.1	0	19	*TP53*	NP_001119586.1:p.Ser215Ile	30,48% (807)	0% (755)
12	14.5	1	14.0	1	20	*BRAF*	NP_004324.2:p.Lys483Glu	11,17% (197)	0% (4569)
12						*TP53*	NP_001263625.1:p.Arg209Leu	10,84% (83)	0% (2716)
13	3.0	1	2.3	1	14	*TP53*	NP_001263625.1:p.Gly115Val	56,98% (172)	4,52% (663)
14	11.7	1	5.3	1	13	*EGFR*	NP_005219.2:p.Val769_Asp770insAlaSerVal	27,93% (111)	9,32% (3625)
14						*TP53*	NP_001263626.1:p.Leu189Phe	16,07% (56)	0,31% (4869)
16	126.2	1	126.2	1	45	*KRAS*	NP_203524.1:p.Gly12Cys	6,83% (161)	0% (971)
17	136.1	0	136.1	0	83	*EGFR*	NP_005219.2:p.Leu858Arg	28,43% (3626)	0% (576)
20	4.7	1	3.2	1	27	*KRAS*	NP_203524.1:p.Gln61His	16,3% (816)	0,83% (478)
21	24.0	1	12.6	1	122	*PDGFRA*	NP_001334759.1:p.Trp572*	21,95% (41)	0% (649)
21						*PIK3CA*	NP_006209.2:p.Asp133Asn	11,21% (116)	0,12% (784)
22	72.0	1	60.8	1	63	*ERBB3*	NP_001973.2:p.Arg103Cys	7,91% (139)	0% (646)
22						*KRAS*	NP_203524.1:p.Gly13Cys	13,12% (160)	0% (433)
22						*TP53*	NP_001263625.1:p.Glu248*	55,36% (112)	3,24% (648)
29	115.0	1	91.6	1	38	*PIK3CA*	NP_006209.2:p.Glu81Lys	38,33% (120)	0% (973)
29						*TP53*	NP_001263625.1:p.Ala122Thr	33,33% (36)	0% (856)
30	10.4	1	9.6	1	16	*TP53*	NP_001263625.1:p.Gln292His	17,32% (976)	2,30% (434)
31	13.0	1	6.1	1	211	*TP53*	NP_001263625.1:p.Tyr195Asn	80,51% (118)	0% (266)
34	2.8	1	2.8	1	24	*TP53*	NP_001263625.1:p.Tyr195*	67,86% (56)	10,3% (368)
36	110.8	0	110.8	0	27	*KRAS*	NP_203524.1:p.Gln61His	10,59% (321)	0% (620)
37	11.2	1	11.2	1	64	*TP53*	NP_001119586.1:p.Tyr220Cys	44,34% (539)	0% (617)

38	15.5	1	15.4	1	27	*TP53*	NP_001119586.1:p.Ser215Gly	11,43% (70)	0% (373)
39	105.6	0	105.6	0	66	*EGFR*	NP_005219.2:p.Glu746_Ala750del	32,2% (879)	0% (1299)
40	106.4	0	106.4	0	34	*EGFR*	NP_005219.2:p.Val769_Asp770insAlaSerVal	20,2% (203)	0% (2364)
41	106.0	0	106.0	0	40	*TP53*	NP_001119586.1:p.Tyr234Cys	20,63% (126)	0% (1060)
43	20.9	1	0.0	1	69	*BRAF*	NP_004324.2:p.Asn486_Pro490del	23,08% (546)	0% (3005)
44	97.5	0	97.5	0	213	*TP53*	NP_001119586.1:p.Tyr220Cys	33,33% (285)	0% (2285)
45	19.3	1	11.7	1	36	*EGFR*	NP_005219.2:p.Glu746_Ala750del	9,35% (845)	0% (489)
45						*TP53*	NP_001119586.1:p.Cys176fs	9,90% (404)	Not analyzed*
46	13.5	1	9.9	1	103	*TP53*	NP_001263625.1:p.Cys236Gly	6,03% (1443)	0,97% (3090)
48	102.1	0	102.1	0	25	*KRAS*	NP_203524.1:p.Glu31Gln	5,66% (53)	2,75% (981)
49	101.1	0	101.1	0	33	*EGFR*	NP_005219.2:p.Leu858Arg	10,77 % (2274)	0% (890)
51	14.5	1	3.9	1	39	*TP53*	NP_001263625.1:p.Tyr197Cys	95,51% (178)	13,6% (351)
53	13.7	1	9.7	1	61	*PIK3CA*	NP_006209.2:p.Gly451Val	24,29% (5357)	0,73% (3279)
54	33.3	1	21.9	1	41	*TP53*	NP_001263625.1:p.Gly115Val	8,89% (3421)	0,40% (494)
55	1.7	1	1.7	1	6	*EGFR*	NP_005219.2:p.Val769_Asp770insGlyValAlaSerVal	76,92% (4904)	65,70% (3989)
55						*TP53*	NP_001263625.1:p.Arg234Cys	38,59% (311)	6,33% (852)
56	5.6	1	2.9	1	21	*KRAS*	NP_203524.1:p.Gly12Cys	21,02% (547)	5,17% (4118)
57	94.7	0	94.7	0	35	*TP53*	NP_001263690.1:p.Lys312Glu	13,19% (91)	0,21% (2325)
59	23.5	1	6.7	1	44	*TP53*	NP_001119586.1:p.Ile251Phe	32,78% (180)	0% (549)
61	20.0	1	15.2	1	10	*EGFR*	NP_005219.2:p.Leu747_Pro753delinsSer	71,79% (312)	24,5% (5948)
62	27.5	1	22.5	1	49	*TP53*	NP_001119586.1:p.Thr155Pro	28,04% (107)	5,68% (440)
66	40.5	1	31.6	1	13	*PIK3CA*	NP_006209.2:p.Thr1025Ala	8,95% (190)	0% (754)
67	88.7	0	88.7	0	20	*EGFR*	NP_005219.2:p.Leu747_Thr751del	29,41% (17)	0% (474)
68	8.2	1	8.0	1	5	*KRAS*	NP_203524.1:p.Gln61His	23,46% (2506)	25,00% (1381)
68						*TP53*	NP_001119586.1:p.Ala276Pro	17,31% (1571)	0,40% (641)
69	14.9	1	5.5	1	92	*TP53*	NP_001263689.1:p.Arg298Leu	27,35% (117)	0% (786)
71	12.9	1	6.5	1	35	*TP53*	NP_001263625.1:p.Cys203Phe	32,1% (648)	7,63% (720)
73	22.6	1	12.6	1	29	*EGFR*	NP_005219.2:p.Leu858Arg	25,16% (2198)	1,27% (708)
75	13.6	1	5.3	1	21	*KRAS*	NP_203524.1:p.Gly13Cys	53,67% (6273)	2,30% (1477)
76	33.0	1	10.6	1	50	*TP53*	NP_001263625.1:p.Glu165*	23,93% (117)	0,30% (2655)
77	7.6	1	7.6	1	34	*TP53*	NP_001119586.1:p.Leu194Arg	60,36% (666)	0,27% (749)
78	65.6	0	65.6	0	21	*TP53*	NP_001119586.1:p.Arg175His	87,21% (602)	0% (406)
79	62.3	0	24.1	1	1	*EGFR*	NP_005219.2:p.Leu858Arg	72,39% (10276)	0% (985)
80	23.7	1	9.0	1	18	*TP53*	NP_001119586.1:p.Thr155Pro	28,12% (32)	0% (488)

81	19.8	1	17.5	1	49	*KRAS*	NP_203524.1:p.Gln61His	8,86% (429)	0,05% (2133)
81						*TP53*	NP_001119586.1:p.Arg158Leu	9,58% (261)	0,16% (2528)
82	58.1	0	58.1	0	1	*KRAS*	NP_203524.1:p.Gly12Val	12,76% (721)	0% (3850)
83	57.9	0	57.9	0	27	*TP53*	NP_001263625.1:p.Lys125Glu	12,5% (1240)	0% (686)
84	57.0	0	19.1	1	1	*TP53*	NP_001263625.1:p.Arg119His	10,4% (125)	0% (1529)
88	10.5	1	7.8	1	28	*TP53*	NP_001263625.1:p.Ser67fs	39,71% (1171)	0,25% (1579)
91	22.6	1	8.0	1	0	*TP53*	NP_001119586.1:p.Arg283fs	13,38% (1203)	0% (1406)
92	25.6	1	19.2	1	1	*TP53*	NP_001119586.1:p.Val274Phe	65,67% (67)	0% (1621)
93	46.7	0	46.7	0	1	*TP53*	NP_001119586.1:p.Arg280fs	27,11% (225)	0,25% (785)
94	10.3	1	9.7	1	2	*BRAF*	NP_004324.2:p.Gly464Val	9,00% (689)	0% (1154)
95	45.2	0	45.2	0	1	*NRAS*	NP_002515.1:p.Gln61Arg	15,9% (346)	0% (3265)
96	46.7	0	46.7	0	1	*TP53*	NP_001119586.1:p.Val274Phe	27,47% (1576)	0% (779)
97	46.2	0	41.3	1	1	*TP53*	NP_001119586.1:p.Arg158Leu	31,86% (970)	0% (3623)
98	44.1	0	33.8	1	1	*TP53*	NM_001126114.2:c.*190del	7,77% (386)	0% (2036)
99	43.8	0	43.8	0	1	*TP53*	NP_001119586.1:p.Lys120fs	23,24% (2143)	0% (1503)
100	42.2	0	42.2	0	1	*PIK3CA*	NP_006209.2:p.Thr1025Ile	8,38% (167)	0% (324)
100						*TP53*	NP_001119586.1:p.Gly154Val	7,4% (338)	0% (778)
100						*TP53*	NP_001119586.1:p.His168Leu	20,45% (264)	0% (715)
101	40.4	0	40.4	0	33	*EGFR*	NP_005219.2:p.Leu858Arg	17,81% (2470)	0% (1237)
102	42.2	0	42.2	0	1	*PIK3CA*	NP_006209.2:p.Pro539Arg	18,00% (1367)	1,19% (670)
102						*PIK3CA*	NP_006209.2:p.His1047Arg	18,52% (1582)	0,9% (702)
102						*TP53*	NP_001119586.1:p.Arg248Trp	24,81% (673)	0,6% (706)
103	37.4	0	37.4	0	1	*TP53*	NP_001119586.1:p.Arg267Pro	31,9% (1627)	0% (1075)
106	31.7	0	31.7	0	1	*RAF1*	NP_002871.1:p.Ser259Phe	76,19% (2721)	0% (1850)
107	31.4	0	17.3	1	1	*TP53*	NM_001276696.1:c.*280_*283dup	7,72% (933)	0% (868)

MAF; mutant allele frequency

*This mutation was not analyzed by a mistake

Paper IV

Title

Associations between detectable circulating tumor DNA and tumor glucose uptake measured by ^{18}F-FDG PET/CT in early-stage non-small cell lung cancer

Authors:

Anine Larsen Ottestad* (1,2), Håkon Johansen (3), Tarje Onsøien Halvorsen (1,2), Hong Yan Dai (1,4), Sissel Gyrid Freim Wahl (1,4), Elisabeth Fritzke Emdal (4), Bjørn Henning Grønberg (1,2)

Author information:

1) Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), 7030, Trondheim, Norway

2) Department of Oncology, St. Olavs Hospital, Trondheim University Hospital, 7030, Trondheim, Norway

3) Department of Radiology and Nuclear Medicine, St. Olavs Hospital, Trondheim University Hospital, 7030 Trondheim, Norway

4) Department of Pathology, Clinic of Laboratory Medicine, St. Olavs Hospital, Trondheim University Hospital, 7030 Trondheim, Norway

*Corresponding author

Abstract

Background

The low level of circulating tumor DNA (ctDNA) in the blood is a well-known challenge for the application of liquid biopsies in early-stage non-small cell lung cancer (NSCLC) management. Studies of metastatic NSCLC indicate that ctDNA levels are associated with tumor metabolic activity as measured by ^{18}F-fluorodeoxyglucose positron emission tomography (^{18}F-FDG PET/CT). This study investigated this association in NSCLC patients considered for potentially curative treatment and explored whether the two methods provide independent prognostic information.

Method

Patients with stage I-III NSCLC who had routinely undergone an ^{18}F-FDG PET/CT scan and exploratory ctDNA analyses were included. Tumor glucose uptake was measured by maximum standardized uptake value (SUVmax), metabolic tumor volume (MTV), and total lesion glycolysis (TLG) from the ^{18}F-FDG PET/CT scans. ctDNA detectability and quantity, using variant allele frequency, were estimated by tumor-informed ctDNA analyses.

Results

In total, 63 patients (median age 70 years, 60 % women, and 90 % adenocarcinoma) were included. The tumor glucose uptake (SUVmax, MTV, and TLG) was significantly higher in patients with detectable ctDNA (n=19, $p<0.001$). The ctDNA quantity correlated with MTV (Spearman's $\rho=0.53$, $p=0.021$) and TLG (Spearman's $\rho=0.56$, $p=0.013$) but not with SUVmax (Spearman's $\rho=0.034$, $p=0.15$). ctDNA detection was associated with shorter OS independent of MTV (HR: 2.70, 95% CI: 1.07-6.82, $p=0.035$) and TLG (HR: 2.63, 95% CI: 1.06-6.51, $p=0.036$). Patients with high tumor glucose uptake and detectable ctDNA had shorter OS and PFS than those without detectable ctDNA, though these associations were not statistically significant ($p>0.05$).

Conclusion

There was a positive correlation between plasma ctDNA quantity and MTV and TLG in early-stage NSCLC patients. Despite the correlation, the results indicated that ctDNA detection was a negative prognostic factor independent of MTV and TLG.

Keywords:

^{18}F-FDG PET/CT, circulating tumor DNA, non-small cell lung cancer, glucose metabolism, liquid biopsy

List of abbreviations

ctDNA: circulating tumor DNA

^{18}F-FDG PET/CT: ^{18}F-fluorodeoxyglucose positron emission tomography

MTV: metabolic tumor volume

NSCLC: non-small cell lung cancer

OS: overall survival

PFS: progression-free survival

TLG: total lesion glycolysis

SUVmax: maximum standardized uptake value

Declarations

Ethics approval and consent to participate

This study was approved by the Central Norway Regional Committee for Medical and Health Research (REK Central, with approval number: 2015/356). Patients gave written informed consent which allowed for collection of biological material and information from their medical records. We confirm that all experiments were performed in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable

Availability of data and materials

The Norwegian Health Research Act prevents us from sharing any clinical data, this includes information about the individual patient, the ^{18}F-FDG PET/CT scans and ctDNA BAM files.

Competing interests

Not applicable

Funding

Funding was received from Central Norway Regional Health Authority (2018/42794) and The Cancer Foundation St. Olavs Hospital. The funding bodies had no role in the project beyond providing financial support.

Authors' contributions

B.H.G designed the study. H.J interpreted the ^{18}F-FDG PET/CT scans. S.G.F.W performed histological examination of tissue specimens and collected clinical data. E.F.E, A.L.O and H.Y.D analyzed and interpreted the ctDNA samples. A.L.O and T.O.H performed statistical analyses. A.L.O and B.H.G prepared the manuscript and all authors read and approved the manuscript.

Acknowledgements

Not applicable

Authors' information

Not applicable

Footnotes

Not applicable

Introduction

The analysis of tumor-specific mutations in circulating tumor DNA (ctDNA) can provide diagnostic and prognostic information in non-small cell lung cancer (NSCLC).[1] The low ctDNA quantity in plasma is, however, a well-known limitation of the utility of ctDNA analysis in NSCLC patients receiving potentially curative treatment.[2]

Previous studies, including studies on NSCLC, have reported that high ctDNA quantity is associated with high tumor metabolic activity, which can be estimated by ^{18}F-fluorodeoxyglucose positron emission tomography (^{18}F-FDG PET/CT).[1, 3–12] By this method, the glucose uptake level in the tumor region is semi-quantified as the standardized uptake value (SUV), and used to identify malign lesions based on their higher-than-normal SUV. ^{18}F-FDG PET/CT is routinely used to accurately assess the extent of disease in NSCLC patients eligible for potentially curative therapy, though the high normal glucose uptake in the brain limits the ability to detect brain metastases. In addition, the metabolic tumor volume (MTV) and total lesion glycolysis (TLG) can also be derived from the ^{18}F-FDG PET/CT scans. The highest SUV in the lesion (SUVmax), MTV, and TLG are candidate prognostic factors in NSCLC.[13]

The association between the glucose uptake level and ctDNA quantity is interesting for two reasons. First, it may contribute to a better understanding of what characterizes patients with detectable ctDNA and thus, identify those who might benefit from ctDNA analyses. Second, ctDNA analyses and ^{18}F-FDG PET/CT-derived parameters might provide overlapping prognostic information. This is especially relevant for early-stage NSCLC for which ^{18}F-FDG PET/CT is routinely performed, potentially limiting the prognostic value of ctDNA analyses. On the other hand, ctDNA might support findings on ^{18}F-FDG PET/CT scans and aid the interpretation, especially when lesions with low ^{18}F-FDG uptake are seen. Few studies have investigated the association between ctDNA detection and glucose uptake in early-stage NSCLC.

This study explored associations between tumor glucose uptake (measured by SUVmax, MTV, and TLG) and both ctDNA detectability and quantity in patients considered for potentially curative

120 treatment. Furthermore, we explored whether ctDNA detection and ^{18}F-FDG PET/CT-derived
121 parameters were independent prognostic factors.

122 **Methods**

123 Study population, approvals, and data collection

124 Biological material was retrieved from the regional research biobank, Biobank1, approved by the
125 Norwegian Regional Committee for Medical and Health Research Ethics in Central Norway, the
126 Ministry of Health and Care Services, and the Norwegian Data Protection Authority. Biobank
127 participants were 18 years or older and gave written informed consent. Patients were treated and
128 followed according to local routines.

129 The present study included patients with stage I-III NSCLC from three previous studies on
130 ctDNA who had available ^{18}F-FDG PET/CT scans obtained during their diagnostic workup. Clinical
131 data were collected from the patients' hospital medical records, which included accurate survival data.
132 The disease stage was assessed according to TNM v8.[14]

133 ctDNA-analyses

134 ctDNA data was available from three previous studies.[15–17] In one cohort, tumor tissue DNA was
135 screened for a pathogenic mutation in the gene *Kirsten Rat Sarcoma Viral Oncogene Homolog*
136 (*KRAS*).[18] Tumor tissue DNA in the other cohorts were screened for pathogenic mutations in 22[17]
137 or 275[16] genes using next-generation sequencing (NGS). Tumor-informed analyses of ctDNA were
138 performed in these studies using digital droplet polymerase chain reaction (ddPCR)[18] or NGS,[16,
139 17] and detection was defined as identifying ≥ one tumor-specific mutation(s) in ctDNA. ctDNA was
140 quantified using the variant allele frequency.

141

142 ^{18}F-FDG PET/CT scans

143 ^{18}F-FDG PET/CT was not available at our hospital until autumn 2013. Thus, patients underwent ^{18}F-
144 FDG PET/CT at three different hospitals: Haukeland University Hospital, Bergen (n=3), Oslo

145 University Hospital, Oslo (n=9), and St. Olav's Hospital, Trondheim (n=51). All hospitals used
146 scanners from Siemens Healthcare (Erlangen, Germany), specifically the Biograph 40 in Bergen, the
147 Biograph 64 in Oslo and the Biograph mCT 64 in Trondheim. The European Association of Nuclear
148 Medicine (EANM) granted an EANM Research Ltd. (EARL) ^{18}F-FDG PET/CT accreditation in
149 September 2015 for the ^{18}FDG PET/CT scanner at St. Olavs University Hospital. The EARL
150 accreditation status for the other centers between 2011-13 is unknown.

151 Image reconstruction was performed with iterative reconstruction, point-spread-function (PSF),
152 decay-, attenuation-, and scatter-correction. Time-of-flight (TOF) was used when available. Different
153 matrix sizes were applied at different sites. All examinations were done following the EANM procedure
154 guidelines for tumor imaging version 2.0.[19] Patients fasted at least four hours (median 14h) before
155 administration of 4 MBq ^{18}F-FDG/kg. Blood glucose level was 4.4-9.5 mmol/L (median 5.6 mmol/L),
156 and the interval between ^{18}F-FDG administration and the start of the acquisition was 51-159 minutes
157 (median 60 minutes). A low-dose CT for attenuation correction and anatomical localization was done
158 in the same session.

159 Datasets were transferred from the hospital's picture archiving and communication systems and
160 reprocessed using standard clinical software (AW Server 3.2 Ext. 3.0, General Electric Company) by a
161 nuclear medicine physician (HJ). The physician was blinded for the ctDNA data but not the previous
162 ^{18}F-FDG PET/CT reports. A 3D isocontour model with a threshold of SUV of 2.5 was used when
163 computing MTV and TLG (=product of MTV and SUVmean). MTV and TLG were calculated
164 manually in separate sessions for each lesion when ^{18}F-FDG uptake from lesions conflated. The highest
165 value of SUVmax in any lesion was used for each patient. For both MTV and TLG, the sum of all
166 lesions was used for statistical analyses. The raw PET data were not available and thus, the original
167 AC-PET reconstructions were used to assess MTV and TLG.

168 Statistics

169 SUVmax was compared between patients with and without detectable ctDNA using the Mann-Whitney
170 U test since the values were not normally distributed. Spearmans' correlation was used to investigate
171 the correlation between SUVmax and the ctDNA quantity, measured by the highest variant allele

172 frequency in cases of >1 variant. Logistic multivariable regression models included SUVmax
173 (continuous), histology, and disease stage to investigate the association between tumor glucose uptake
174 and ctDNA detection. All analyses were repeated for MTV and TLG.

175 PFS was defined as the time from lung cancer diagnosis until progression or death of any cause,
176 and OS was defined as the time from diagnosis until death of any cause. The median follow-up times
177 for PFS and OS were estimated using the reverse Kaplan-Meier method, and the median PFS and OS
178 were estimated using the Kaplan-Meier method. The impacts of ctDNA detectability and SUVmax on
179 OS and PFS were estimated by univariable Cox proportional hazard models. Multivariable models for
180 PFS and OS included SUVmax as a continuous factor and ctDNA detection. The combined prognostic
181 value of ctDNA detection and tumor glucose uptake was explored in patients with high SUVmax (>
182 median value in our cohort) by comparing outcome between those with and without detectable ctDNA
183 using the Kaplan-Meier method. All analyses of OS and PFS were repeated for MTV and TLG.

184 Statistical analyses were performed using R (version 3.6.1) with 0.05 as the threshold for
185 statistical significance.

186 **Results**

187 Patient characteristics

188 In total, 63 patients diagnosed between July 2009 and May 2018 met the eligibility criteria for the
189 present study (Table 1). The median age was 70, 38 (60 %) were female, and 59 (94 %) were smokers
190 or former smokers. Fifty-seven patients (90 %) had adenocarcinoma, two (3%) had squamous cell
191 carcinoma, three (5 %) had NSCLC not otherwise specified, and one (2%) had large cell neuroendocrine
192 carcinoma. Twenty-eight patients (44 %) had stage I disease, 12 (19 %) stage II, and 23 (37 %) stage
193 III. ctDNA was detected in plasma from 19 patients (30 %). Patients with detectable ctDNA had higher
194 disease stage and median MTV, TLG, SUVmax, and lower surgical rate than those without detectable
195 ctDNA. Otherwise, patient characteristics were similar between the two groups.

196

Table 1: Patient characteristics

	All patients		ctDNA detected		ctDNA not detected	
Total	63		19		44	
Age (median)	70	(52-83)	68	(52-83)	70	(52-81)
Sex						
Female	38	60 %	11	58 %	27	61 %
Male	25	40 %	8	42 %	17	39 %
Smoking status						
Smoker/former smoker	59	94 %	17	89 %	42	95 %
Never smoker	4	6 %	2	11 %	2	5 %
Histology						
Adenocarcinoma	57	90 %	16	84 %	41	93 %
Non-adenocarcinoma	6	10 %	3	16 %	3	7 %
WHO performance status						
0	37	59 %	11	58 %	26	59 %
1	23	37 %	8	42 %	15	34 %
2	2	3 %	0	0 %	2	5 %
3	1	2 %	0	0 %	1	2 %
Disease stage						
I	28	44 %	0	0 %	28	64 %
II	12	19 %	5	26 %	7	16 %
III	23	37 %	14	74 %	9	20 %
Treatment						
Surgery	48	76 %	8	42 %	40	91 %
Curative radiotherapy with or without chemotherapy	11	17 %	9	47 %	2	5 %
Palliative therapy	4	6 %	2	11 %	2	5 %
ctDNA detection method						
NGS*	33	52 %	13	68 %	20	45 %
ddPCR	30	48 %	6	32 %	24	55 %
^{18}F-FDG PET/CT parameters (median)						
MTV (cm^3)	7.5		61.2		3.7	
TLG (g/mL x cm^3)	39.1		460.5		14.8	
SUVmax (g/mL)	11.8		19.2		9.1	

*One patient was included from a study analyzing ctDNA by a 275 NGS gene panel, and cfDNA from the other 32 patients were analyzed by patient-specific NGS panels. ddPCR: droplet digital polymerase chain reaction, NGS: next-generation sequencing, MTV: metabolic tumor volume, SUV: standardized uptake value, TLG: total lesion glycolysis

197

198 Tumor glucose uptake in patients with and without detectable ctDNA

199 Patients with detectable ctDNA had significantly higher MTV ($p<0.001$), TLG ($p<0.001$), and SUVmax

200 ($p<0.001$) than patients without detectable ctDNA (Figure 1).

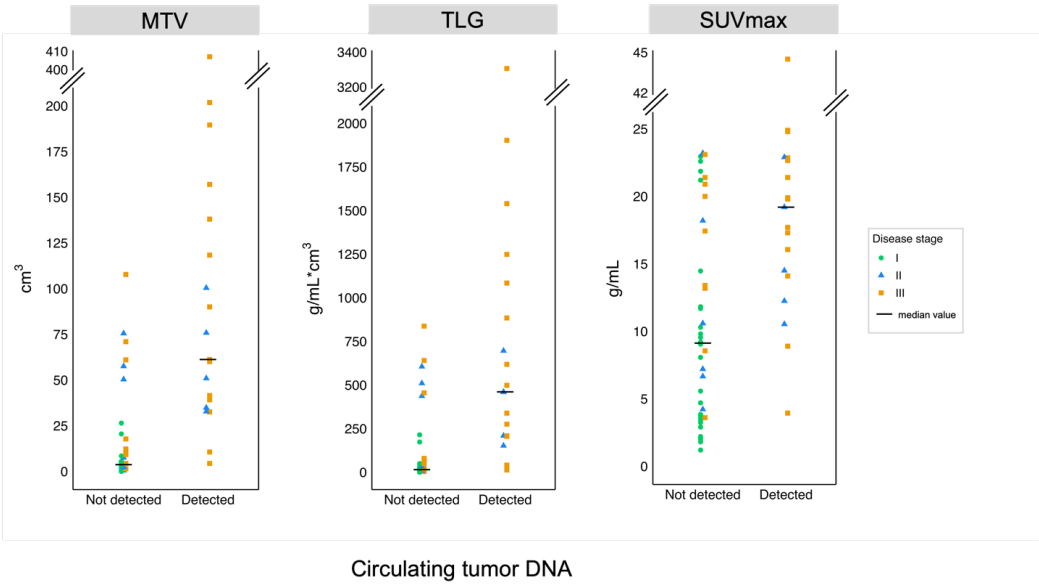


Figure 1. MTV, TLG and SUVmax derived from ^{18}F-FDG PET/CT scans from patients with and without detectable ctDNA. ctDNA: circulating tumor DNA, MTV: metabolic tumor volume, TLG: total lesion glycolysis, SUVmax: maximum standardized uptake value

The ctDNA quantity correlated with MTV (Spearman's ρ=0.53, p=0.0211) and TLG (Spearman's ρ=0.56, p=0.0127), but not with SUVmax (Spearman's ρ=0.34, p=0.15) in patients with detectable ctDNA (Figure 2).

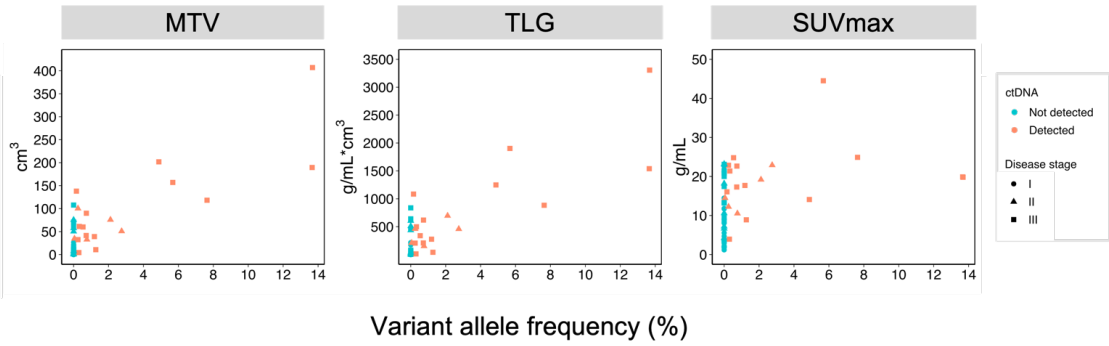


Figure 2. MTV, TLG, SUVmax and the ctDNA quantity, measured as the highest variant allele frequency.

205 Tumor glucose uptake as a predictor of ctDNA detection

206 Higher MTV (OR 19.3, 95% CI: 5.4-116.9, p<0.001), TLG (OR 9.0, 95% CI: 3.4-33.8, p<0.001), and
207 SUVmax (OR 47.3, 95% CI: 5.2-937.4, p=0.0030) were associated with ctDNA detection in univariable
208 logistic regression analyses. MTV (OR 1.03, 95% CI: 1.01-1.05, p=0.019) remained associated with
209 ctDNA detection independent of disease stage and histology in multivariable analysis (Table 2).
210 Similarly, TLG (OR 1.00, 95% CI: 1.00-1.01, p=0.038) was independently associated with ctDNA
211 detection, while SUVmax was not (OR: 1.07, 95% CI: 0.98-1.20, p=0.19). Since ctDNA was not
212 detected in stage I patients, sensitivity analyses including only stage II-III patients were performed, with
213 similar results (Table S1).

214

Table 2: Multivariable logistic regression models with ctDNA detection as response

	MTV			TLG			SUVmax		
	OR	95% CI	*p*	OR	95% CI	*p*	OR	95% CI	*p*
Stage I	1.00			1.00			1.00		
Stage II	8.75E+07	0-NA	0.99	1.14E+08	0-NA	0.99	1.75E+08	0-NA	0.99
Stage III	1.32E+08	0-NA	0.99	1.76E+08	0-NA	0.99	3.00E+08	0-NA	0.99
Adenocarcinoma	1.00			1.00			1.00		
Non-adenocarcinoma	3.59	0.29-87.72	0.34	3.74	0.3-90.89	0.32	2.72	0.28-61.19	0.42
MTV	1.03	1.01-1.05	**0.019**	-	-	-	-	-	-
TLG	-	-	-	1.00	1.00-1.01	**0.038**	-	-	-
SUVmax	-	-	-	-	-	-	1.07	0.98-1.20	0.19

Statistically significant p-values are shown in bold. CI: confidence interval; ctDNA: circulating tumor DNA; MTV: metabolic tumor volume; OR: odds ratio, SUVmax: maximum standardized uptake value; TLG: total lesion glycolysis

215 Tumor glucose uptake and ctDNA detection as prognostic factors

216 The median follow-up time for OS was 57.0 months (95% CI: 50.7-64.0), and 33 patients were alive at
217 the time of analysis. Overall, median OS was not reached (95% CI: 39.9 months - not reached [NR]).
218 Higher MTV (HR 1.00, 95% CI: 0.995-1.00, p=0.017), TLG (HR: 1.00, 95% CI: 0.995-1.00, p=0.017),
219 and SUVmax (HR: 1.05, 95% CI: 0.948-1.02, p=0.004) were associated with worse OS in univariable
220 analyses (Table 3).

221

222

Table 3: Univariable Cox proportional hazard analyses of OS

	HR	95% CI	*p*
ctDNA detected	3.13	1.46-6.73	**0.0034**
MTV	1.00	1.00-1.01	**0.017**
TLG	1.00	1.00-1.00	**0.013**
SUVmax	1.05	1.02-1.09	**0.0036**

Statistically significant p-values are shown in bold. CI: confidence interval; ctDNA: circulating tumor DNA; HR: hazard ratio, MTV: metabolic tumor volume, SUVmax: maximum standardized uptake value; TLG: total lesion glycolysis

Multivariable analyses (Table 4) showed that ctDNA detection was associated with worse OS independently of MTV (HR: 2.70, 95% CI: 1.07-6.82, p=0.035) and TLG (HR: 2.63, 95% CI: 1.06-6.51, p=0.036), but not SUVmax (HR: 2.30, 95% CI: 0.977-5.42, p=0.056). The ^{18}F-FDG PET/CT-derived parameters were not independently associated with OS in the same models (MTV, HR: 1.00, 95% CI: 0.996-1.01, p=0.55. TLG, HR: 1.00, 95% CI: 1.00-1.00, p=0.43. SUVmax, HR: 1.03, 95% CI: 0.995-1.08, p=0.087.)

Table 4: Multivariable Cox proportional hazard analyses of OS

	MTV			TLG			SUVmax		
	HR	95% CI	*p*	HR	95% CI	*p*	HR	95% CI	*p*
ctDNA not detected	1.00			1.00			1.00		
ctDNA detected	2.70	1.07-6.82	**0.035**	2.63	1.06-6.51	**0.036**	2.30	0.977-5.42	0.056
MTV	1.00	0.996-1.01	0.55						
TLG				1.00	1.00-1.00	0.43			
SUVmax							1.03	0.995-1.08	0.087

Statistically significant p-values are shown in bold. CI: confidence interval; ctDNA: circulating tumor DNA; HR: hazard ratio, MTV: metabolic tumor volume, SUVmax: maximum standardized uptake value; TLG: total lesion glycolysis

The median follow-up time for PFS was 57.0 months (95% CI: 50.7-65.6), and 29 patients were alive and progression-free at the time of analysis. The median PFS was 61.8 months (95% CI: 19.1-NR). ctDNA detection and higher MTV, TLG, and SUVmax were significantly associated with worse PFS in univariable Cox proportional hazard analyses (p<0.05, Table S2). None of the factors were independently associated with PFS in multivariable analyses (Table S3).

<u>Combined prognostic value of ctDNA analyses and ^{18}F-FDG PET/CT-derived parameters.</u>

Among patients with MTV above median value, those with detectable ctDNA had shorter OS than those without detectable ctDNA (median 20.4 months vs. NR, Figure 3), though the difference was not

240 statistically significant (HR: 2.2, 95% CI: 0.8-6.2, p=0.15). Similar difference was observed among
241 patients with TLG above median value (median 20.4 months vs. NR, HR: 2.2, 95% CI: 0.8-6.2, p=0.15)
242 and SUVmax above median value (median 20.4 months vs. NR, HR: 1.9, 95% CI: 0.7-5.4, p=0.202).
243 Similarly, among patients with MTV, TLG or SUVmax above median value, those with detectable
244 ctDNA had shorter PFS than those without detectable ctDNA, though the differences were not
245 statistically significant (Figure S1). There were too few patients (n<4) with detectable ctDNA among
246 those with MTV, TLG, or SUVmax below median values to perform such analyses.

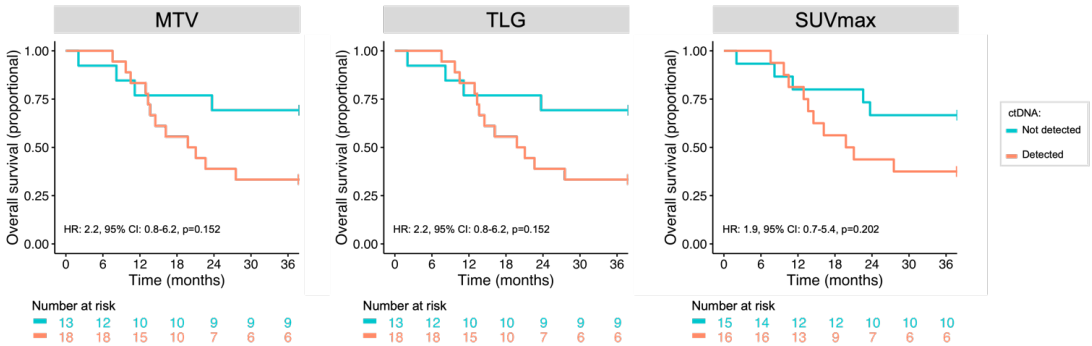


Figure 3. Kaplan-Meier plots showing OS for patients with A: MTV, B: TLG and C: SUVmax above the median value and split on ctDNA status.

247

248

Discussion

In this study of patients with stage I-III NSCLC considered for potentially curative therapy, we found that tumor glucose uptake was significantly higher in patients with detectable ctDNA, and the ctDNA quantity correlated with MTV and TLG. Nevertheless, ctDNA detection was a negative prognostic factor for OS independently of the ^{18}F-FDG PET/CT-derived parameters.

There is limited research on the association between ^{18}F-FDG PET/CT and ctDNA detection in early-stage NSCLC. Our results reflect those of Chabon *et al.*, who included 85 early-stage NSCLC patients and found that those with detectable ctDNA had higher MTV and that ctDNA quantity correlated with MTV.[12] In addition, they found that ctDNA detection was a negative prognostic factor independent of both MTV and disease stage. Another study of 92 patients enrolled in the TRACERx study found that high ^{18}F-FDG avidity, defined as the ratio between tumor and mediastinal SUVmax, predicted ctDNA detection.[11] That study did not investigate whether the prediction was independent of other patient and disease characteristics. While SUVmax was significantly higher in patients with detectable ctDNA in our study, it was not independently associated with ctDNA detection after adjusting for disease stage and histology.

Several studies have investigated the association between ctDNA characteristics and all three ^{18}F-FDG PET/CT-derived parameters in advanced NSCLC.[1, 4, 6, 7] Although these are mostly small studies (n=37-128) with methodological variations, they indicate a positive correlation between ctDNA release and glucose uptake. In agreement with our results, one study reported a correlation between ctDNA quantity and MTV and TLG but not with SUVmax.[4] Other studies found a positive correlation with SUVmax as well,[6, 7] such as the recent study of Jee and colleagues which included 128 advanced NSCLC patients, the largest study to date.[1]

A few studies on advanced NSCLC found no correlation between ctDNA release and tumor metabolic activity.[7, 8, 20] Common for these studies is that they analyzed the total cfDNA quantity rather than the mutant ctDNA fraction. Notably, González de Aledo-Castillo *et al.* observed that cfDNA at 100-250 bp length, which includes the typical length of ctDNA, correlated with glucose uptake, while total cfDNA quantity did not.[8] The total cfDNA quantity may be influenced by non-cancer related factors.[21]

We and others observed cases of undetectable or low ctDNA quantity but high glucose uptake and vice versa.[1, 3, 5, 6, 8, 12] Although the two variables usually correlate, evidence from studies on

279 NSCLC suggests that ctDNA analysis and ^{18}F-FDG PET/CT may provide independent prognostic
280 information.[1, 3, 5, 12] Studies of metastatic cancers, including NSCLC, have also indicated a
281 combined value of the two analyses.[1, 3, 9, 10] For example, Jee *et al.* demonstrated that ctDNA
282 detection was a negative prognostic factor both for patients with high and low glucose uptake.[1] The
283 high number of stage I patients without detectable ctDNA was the reason for not investigating the
284 prognostic role of detectable ctDNA among our patients with low glucose uptake levels.

285 The results of this study must be interpreted in the context of its limitations and strengths. This
286 was a retrospective study of mostly adenocarcinoma patients, of which 56 % had a *KRAS* mutation,
287 compared to ~38% in the Norwegian lung adenocarcinoma population.[18] The sample size did not
288 allow adjusting for other established prognostic factors such as disease stage, performance status, or
289 therapy in the multivariable OS and PFS analyses. Additionally, tumor characteristics associated with
290 ctDNA detectability, such as proliferation rate, the extent of necrosis, and vascular infiltration, were
291 not assessed.

292 The ^{18}F-FDG PET/CT scans were performed on several scanners and sites using various
293 dosages of ^{18}F-FDG/kg, though 80% of scans were performed on the scanner at our site using the same
294 protocol. The values of the ^{18}F-FDG PET/CT-derived parameters are, in principle, dependent on the
295 characteristics of the PET/CT camera, reconstruction parameters, matrix size, and PSF, which could
296 cause a batch effect. Although raw data was not available from the other hospitals and different
297 parameters may have influenced the calculated values, we do not believe that a potential batch effect
298 has significantly influenced the overall result. The variation in MTV and TLG was less than 5% when
299 values assessed locally were compared with centrally reconstructed parameters, and measures were
300 taken to compensate for the partial volume effects, including the tissue fraction effects. Importantly, no
301 tumors in our study were <3 mm, limiting the risk of SUVmax underestimation. It is unclear which ^{18}F-
302 FDG PET/CT variable holds the most prognostic information, but most other studies have used the
303 same variables as we did.[1, 3, 4, 6, 7] Finally, we did not have information about co-existing conditions
304 (e.g. sarcoidosis) or medications that might have influenced the ^{18}F-FDG uptake.

305 The lack of standardized methods for ctDNA detection and accurate quantification is a general
306 challenge for ctDNA research. Using the variant allele frequency for ctDNA quantification is common
307 but not optimal since the frequency depends on the total cfDNA quantity.

308 The high sensitivity and specificity are strengths of the tumor-informed approach applied for
309 ctDNA analysis in this study. Nevertheless, we cannot rule out the possibility of false negatives.
310 Another study detected ctDNA in 45 % of early-stage NSCLC patients and noted that the likelihood of
311 ctDNA detection increased with the number of analyzed mutations.[12] The tumor DNA screening
312 limited the number of mutations for ctDNA analysis, and there was only knowledge of one mutation in
313 the *KRAS*-positive cohort. Other contributing factors to the low detection rate may be the high
314 proportion and stage I patients and adenocarcinomas, which probably release less ctDNA than other
315 types of NSCLC.[11]

316 The inclusion of early-stage patients was, in our opinion, the main feature of our study.
317 Although many lower-stage NSCLC patients are cured by surgery and radiotherapy, the relapse rates
318 are still relatively high. The effectiveness of adjuvant *EGFR* tyrosine kinase inhibitors and
319 immunotherapy has recently been demonstrated, but similar to adjuvant chemotherapy, the absolute
320 survival benefit is limited. There is an unmet need for tools identifying those with the highest risk of
321 relapse who should be offered such adjuvant therapy to reduce the number of patients receiving
322 unnecessary medication.[22, 23] One might argue that the clinical implications of the prognostic
323 information of ctDNA detection and ^{18}F-FDG PET/CT variables are fewer since most of these patients
324 receive treatment anyway. In that setting, biomarkers predicting outcomes of specific treatments is more
325 important.

326 The question remains what is the relationship between ctDNA release and tumor metabolism.
327 It is important to remember that ^{18}F-FDG PET/CT estimates the level of tumor glucose uptake and
328 cannot be used to explain the metabolic state of the tumor. Elevated glucose uptake can reflect an
329 elevated level of aerobic metabolism or a shift towards anaerobic metabolism due to hypoxia or the
330 Warburg effect. Studies on NSCLC have indicated that squamous cell carcinomas are associated with
331 anaerobic metabolism and adenocarcinomas with aerobic metabolism.[24] According to a recent

publication, elevated glucose uptake might not reflect elevated metabolic activity at all.[25] The authors demonstrated that solid tumors in mice have a high glucose uptake without an increase in energy production and suggested that this is tolerated by the tumor cells by shutting down energy-costly tissue-specific processes. It is yet to be understood whether tumors in these different metabolic states have a similar rate of ctDNA release and whether the prognostic meaning of its release remains the same.

Conclusion

We found a positive correlation between plasma ctDNA quantity and tumor glucose uptake measured by ^{18}F-FDG PET/CT in early-stage NSCLC patients. Nevertheless, the result indicated that ctDNA analysis provided independent prognostic information from ^{18}F-FDG PET/CT and larger studies are needed to investigate if there is a combined prognostic value of the two analyses. Furthermore, there is a need for a better understanding of the mechanism behind ctDNA release and the biological rationale behind the potential prognostic impact since it cannot be explained by the tumor glucose uptake alone.

References

1. Jee J, Lebow ES, Yeh R, Das JP, Namakydoust A, Paik PK, et al. Overall survival with circulating tumor DNA-guided therapy in advanced non-small-cell lung cancer. Nat Med. 2022;28:2353–63.
2. Abbosh C, Birkbak NJ, Swanton C. Early stage NSCLC — challenges to implementing ctDNA-based screening and MRD detection. Nat Rev Clin Oncol. 2018;15:577–86.
3. Hyun MH, Lee ES, Eo JS, Kim S, Kang EJ, Sung JS, et al. Clinical implications of circulating cell-free DNA quantification and metabolic tumor burden in advanced non-small cell lung cancer. Lung Cancer. 2019;134:158–66.
4. Fiala O, Baxa J, Svaton M, Benesova L, Ptackova R, Halkova T, et al. Combination of Circulating Tumour DNA and ^{18} F-FDG PET/CT for Precision Monitoring of Therapy Response in Patients With Advanced Non-small Cell Lung Cancer: A Prospective Study. Cancer Genomics - Proteomics. 2022;19:270–81.
5. Winther-Larsen A, Demuth C, Fledelius J, Madsen AT, Hjorthaug K, Meldgaard P, et al. Correlation between circulating mutant DNA and metabolic tumour burden in advanced non-small cell lung cancer patients. Br J Cancer. 2017;117:704–9.
6. Lam VK, Zhang J, Wu CC, Tran HT, Li L, Diao L, et al. Genotype-Specific Differences in Circulating Tumor DNA Levels in Advanced NSCLC. J Thorac Oncol. 2021;16:601–9.
7. Morbelli S, Alama A, Ferrarazzo G, Coco S, Genova C, Rijavec E, et al. Circulating Tumor DNA Reflects Tumor Metabolism Rather Than Tumor Burden in Chemotherapy-Naive Patients with Advanced Non–Small Cell Lung Cancer: ^{18} F-FDG PET/CT Study. J Nucl Med. 2017;58:1764–9.
8. González de Aledo-Castillo Jm, Casanueva-Eliceiry S, Soler-Perromat A, Fuster D, Pastor V, Reguart N, et al. Cell-free DNA concentration and fragment size fraction correlate with FDG PET/CT-derived parameters in NSCLC patients. Eur J Nucl Med Mol Imaging. 2021;48: 3631–3642
9. Woff E, Kehagias P, Vandeputte C, Ameye L, Guiot T, Paesmans M, et al. Combining ^{18} F-FDG PET/CT–Based Metabolically Active Tumor Volume and Circulating Cell-Free DNA Significantly Improves Outcome Prediction in Chemorefractory Metastatic Colorectal Cancer. J Nucl Med. 2019;60:1366–72.
10. Delfau-Larue M-H, van der Gucht A, Dupuis J, Jais J-P, Nel I, Beldi-Ferchiou A, et al. Total metabolic tumor volume, circulating tumor cells, cell-free DNA: distinct prognostic value in follicular lymphoma. Blood Adv. 2018;2:807–16.
11. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature. 2017;545:446–51.
12. Chabon JJ, Hamilton EG, Kurtz DM, Esfahani MS, Moding EJ, Stehr H, et al. Integrating genomic features for non-invasive early lung cancer detection. Nature. 2020;580:245–51.
13. Kaira K, Higuchi T, Naruse I, Arisaka Y, Tokue A, Altan B, et al. Metabolic activity by 18F–FDG-PET/CT is predictive of early response after nivolumab in previously treated NSCLC. Eur J Nucl Med Mol Imaging. 2018;45:56–66.
14. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WEE, et al. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. J Thorac Oncol. 2016;11:39–51.
15. Wahl SGF, Dai HY, Emdal EF, Ottestad AL, Dale VG, Richardsen E, et al. Prognostic value of absolute quantification of mutated KRAS in circulating tumour DNA in lung adenocarcinoma patients prior to therapy. J Pathol Clin Res. 2021;7:209-219

406 16. Ottestad AL, Wahl SGF, Grønberg BH, Skorpen F, Dai HY. The relevance of tumor
407 mutation profiling in interpretation of NGS data from cell-free DNA in non-small cell lung
408 cancer patients. Exp Mol Pathol. 2019;:104347.
409 17. Ottestad AL, Dai HY, Halvorsen TO, Emdal EF, Wahl SGF, Grønberg BH. Associations
410 between tumor mutations in cfDNA and survival in non-small cell lung cancer. Cancer Treat
411 Res Commun. 2021;29:100471.
412 18. Wahl SGF, Dai HY, Emdal EF, Berg T, Halvorsen TO, Ottestad AL, et al. The Prognostic
413 Effect of KRAS Mutations in Non-Small Cell Lung Carcinoma Revisited: A Norwegian
414 Multicentre Study. Cancers. 2021;13:4294.
415 19. Boellaard R, Delgado-Bolton R, Oyen WJG, Giammarile F, Tatsch K, Eschner W, et al.
416 FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med
417 Mol Imaging. 2015;42:328–54.
418 20. Nygaard AD, Holdgaard PC, Spindler K-LG, Pallisgaard N, Jakobsen A. The correlation
419 between cell-free DNA and tumour burden was estimated by PET/CT in patients with advanced
420 NSCLC. Br J Cancer. 2014;110:363–8.
421 21. Kananen L, Hurme M, Bürkle A, Moreno-Villanueva M, Bernhardt J, Debacq-Chainiaux
422 F, et al. Circulating cell-free DNA in health and disease — the relationship to health
423 behaviours, ageing phenotypes and metabolomics. GeroScience. 2023;45:85–103.
424 22. Wu Y-L, Tsuboi M, He J, John T, Grohe C, Majem M, et al. Osimertinib in Resected *EGFR*
425 -Mutated Non–Small-Cell Lung Cancer. N Engl J Med. 2020;383:1711–23.
426 23. Felip E, Altorki N, Zhou C, Csőszi T, Vynnychenko I, Goloborodko O, et al. Adjuvant
427 atezolizumab after adjuvant chemotherapy in resected stage IB–IIIA non-small-cell lung
428 cancer (IMpower010): a randomised, multicentre, open-label, phase 3 trial. The Lancet.
429 2021;398:1344–57.
430 24. Schuurbiers OCJ, Meijer TWH, Kaanders JHAM, Looijen-Salamon MG, de Geus-Oei L-
431 F, van der Drift MA, et al. Glucose Metabolism in NSCLC Is Histology-Specific and Diverges
432 the Prognostic Potential of 18FDG-PET for Adenocarcinoma and Squamous Cell Carcinoma.
433 J Thorac Oncol. 2014;9:1485–93.
434 25. Bartman CR, Weilandt DR, Shen Y, Lee WD, Han Y, TeSlaa T, et al. Slow TCA flux and
435 ATP production in primary solid tumours but not metastases. Nature. 2023;614: 349–357

ISBN 978-82-326-7354-4 (printed ver.)
ISBN 978-82-326-7353-7 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU
Norwegian University of
Science and Technology