Clinical and research applications of circulating tumor DNA detection, monitoring, and dynamics in cancer patients

---

By

Christopher Boniface

B.A., Reed College, 2006

A DISSERTATION

Presented to the Department of Biomedical Engineering of the
Oregon Health & Science University School of Medicine
in partial fulfillment of the requirements for the degree
of Doctor of Philosophy

March 2023

School of Medicine

Oregon Health & Science University

Certificate of Approval

_____

This is to certify that the PhD dissertation of

Christopher Boniface

has been approved

_____

Dr. Paul Spellman (Mentor/Advisor)

_____

Dr. Summer Gibbs (Chair)

_____

Dr. Julia Maxson (Member)

_____

Dr. Reid Thompson (Member)

_____

Dr. Andrew Adey (Member)

Table of Contents

Dedications

This work is dedicated to the memory of my sister, Erika, who faced her own early death from cancer with bravery, humility, and laughter, and who showed a ten-year-old me that love and grief are two sides of the same coin. Slay those dragons!

This work is also dedicated to the memories of my grandfather, Tom – whose brilliance helped inspire my love of science and nature from a very young age – and grandmother, Sybil – whose grace and humility helped me understand the virtues of self-humbling and gratitude.

Acknowledgements

None of this work would have been possible without the support, love, and care of my incredible wife, Emily. She has been a rock for me for more than twenty years and without her I wouldn't have even considered that any of this was possible. Thank you for everything!

I would like to thank my two wonderful children, Tristan and Gwen, who every day provide me with inspiration and motivation toward being my best self and making the world they inherit a better place. *They have taught me how precious life truly is!*

I am eternally grateful to my parents, Anna and Richard, for loving me unconditionally, always being there for me, and teaching me to love the grandeur of the natural world. I wish to thank my step-parents, Paul and Michelle, for tolerating me in my youth and being great friends to me ever since. I would particularly like to thank my step-dad, Paul, for providing me with a fantastic example of determination and commitment to one's craft.

I would also like to thank current and former members of Dr. Paul Spellman's lab and my colleagues: First, Katie Johnson-Camacho, who got me an interview as a research assistant for Dr. Spellman – I couldn't imagine myself here now if it wasn't for that conversation in 2014. Dr. Tim Butler, Dr. Spellman's first grad student, who is an inspiration, teacher, and friend. If it weren't for his patience while I stared blankly at him after he'd finished explaining something (for the third time), and his endless help in the lab, I would never have made it this far. Dr. Myron Peto, who always stopped what he was doing to help me with some coding question and has also been a kind friend. Kami Chiotti, who has also helped me with code more than a few times – but more

importantly, been a lovely friend and colleague. Carol Halsey, who provided endless support navigating IRB and patient specimen acquisition, not to mention great conversation and lots of laughs. Taylor Kelley and Ramtin Rahmani, for tirelessly consenting and drawing blood from patients for these studies and for whom the entire second and third chapters of this dissertation would not be possible. Dr. Katie Baker and Dr. Theresa Lusardi, whose work and support on the study behind the third chapter were also key to its completion. Dr. Brett Johnson for his support, trust, and guidance for the SMMART projects while I was a research associate.

I would like to give a special thanks to Dr. Nima Nabavizadeh whose ideas, funding awards, and support also made this work possible. In the last 5 years, Nima's mentoring and friendship have been invaluable to me.

I would like to give special thanks to Dr. Joe Gray – without his confidence and support I may not have pursued my PhD.

I would like to thank the members of my dissertation committee, Dr. Summer Gibbs, Dr. Julia Maxson, and Dr. Reid Thompson, whose support and guidance have been invaluable. I would also like to thank Dr. Andrew Adey for serving on my final oral exam committee.

Final, I would like to give and extra special thanks to my mentor, Dr. Paul Spellman, who continually pushed me to be a better scientist, and who also kept my technical and scientific reveries grounded in reality. Perhaps more importantly, Paul has also been a friend and an inspiring role model.

List of Abbreviations[1]

AML	acute myeloid leukemia
bam	binary (compressed) sequence alignment file (sam)
BBB	blood-brain barrier
BL	baseline
Bx	biopsy
CAPP-Seq	cancer personalized profiling by deep sequencing
cfDNA[2]	cell-free DNA
CH	clonal hematopoiesis
ChemoRT, CRT	chemotherapy plus radiation therapy
CHIP	clonal hematopoiesis of indeterminate potential
CNV	copy number variant
CR	complete response
CRC	colorectal cancer
CTC	circulating tumor cells
ctDNA	circulating tumor DNA
CV	coefficient of variation

[1] Although I define each abbreviation at least once as they appear in the text, I have intentionally repeated some of these definitions in each chapter, and often in each section, to provide unfamiliar readers with a smoother reading experience.

[2] **I have intentionally chosen not to use the common abbreviation for cell-free, "cfDNA," in my work to avoid confusion with the abbreviation for circulating tumor DNA (ctDNA)**, which is a subclassification of the former. Both abbreviations are ubiquitous in the literature on these subjects, but as this work is mostly focused on ctDNA, I thought it best to avoid using the term "cfDNA" altogether (except for in Fig. 1.6 where space constraints made that difficult).

ddPCR	droplet-digital PCR
DIDA-Seq	dual-index degenerate adaptor sequencing (dual-UMI sequencing)
DNA	deoxyribonucleic acid
EBRT	external beam radiation therapy
EC	esophageal cancer
FFPE	formalin-fixed paraffin embedded
FOLFIRI	folinic acid, fluorouracil and irinotecan
FOLFOX	folinic acid, fluorouracil and oxaliplatin
FRP	false-positive rate
GBM	glioblastoma
GE	genome equivalents (typically diploid, unless written "hGE")
hGE	haploid genome equivalents
ICI	immune checkpoint inhibitor
iDES	integrated digital error suppression
Indels	insertions and deletions
IRB	institutional review board
kb/Mb	kilobases and megabases
Lung-CLiP	lung cancer likelihood in plasma
MAF (also VAF)[3]	mutant allele fraction, synonymous with: variant allele fraction, mutant allele frequency, & mutant allele frequency

[3] The four terms "mutant allele frequency," "mutant allele fraction," "variant allele frequency," and "variant allele fraction," which are abbreviated as "MAF" and "VAF," **are all interchangeable** as they relate to ctDNA. In this work I try to use only "mutant" and "MAF" with at least one exception found in Fig. 1.6. MAF and VAF are presented as fractions (*i.e.*, 0-1), unless otherwise indicated as percent (with "%").

MCED	multi-cancer early detection
MRD	minimal residual disease (synonymous with molecular residual disease)
MRI	magnetic resonance imaging
mut	mutation
NGS	next-generation sequencing
NOM	non-operative management
NSCLC	non-small cell lung cancer
PCR	polymerase chain reaction
pCR	pathologic complete response
PET/CT, PET-CT	positron-emission tomography and computer tomography
18F-FDG-PET/CT	18F-fluorodeoxyglucose PET/CT
RC	rectal cancer
RT	radiation therapy
SNV	single nucleotide variant (somatic mutation)
SNP	single nucleotide polymorphism (germline)
SSCS	single-stand consensus sequence
TIL	tumor-infiltrating lymphocyte
TKI	tyrosine kinase inhibitor
TME	total mesorectal excision
TNM	tumor, node, metastasis (staging method)
Tx	treatment
UMI	unique molecular index
WBC	white blood cell

WES	whole exome sequencing
WGS	whole genome sequencing
WT	wild-type (also, reference allele)

List of Tables

List of Figures

Dissertation Abstract

Liquid biopsy provides a minimally-invasive alternative to solid tissue biopsy for assessing tumor-derived molecules such as circulating tumor DNA (ctDNA). Detection and characterization of ctDNA in the blood shows promise in early cancer screening and diagnosis, and has prognostic value for risk-stratification at all stages of disease. ctDNA can also be used for minimally-invasive tumor genotyping, treatment monitoring, and pre-clinical detection of minimal residual disease and recurrence. Although once predicted as a panacea for these applications, the clinical implementation of ctDNA assays has met with significant technical hurdles. Moreover, poorly-understood biological mechanisms and a lack of standardized methods contribute to measurement variability. The work that follows is intended to (1) highlight the need for a deeper understanding of the biological underpinnings of ctDNA release and kinetics, and (2) test the limits of patient-specific and patient-agnostic ctDNA detection technologies in the clinical setting.

In Chapter II, "Blood, Toil, and Taxoteres: Biological Determinants of Treatment-Induced ctDNA Dynamics for Interpreting Tumor Response," I present my peer-reviewed work that explored our understanding of ctDNA origins and dynamics circa 2021. Because the factors that drive ctDNA abundance in some patients and not others were (and remain) unclear, I surveyed the available literature in which ctDNA dynamics were reported during therapy with the hope of revealing something, not just about treatment efficacy, but about the underlying biological mechanisms of ctDNA shedding. While interesting patterns exist in these data, without a better understanding of the complex biology of ctDNA, we will not realize the true potential of this information-rich biomarker.

In Chapter III, "The Feasibility of Patient-Specific ctDNA Monitoring for Subclinical Disease in Esophageal and Rectal Cancer," I present our peer-reviewed study published in 2021 wherein we tested the feasibility of patient-specific ctDNA monitoring for non-operative management of neoadjuvantly-treated esophageal and colorectal cancer patients. We found that ctDNA was detectable after treatment in patients who later recurred or had residual disease at the time of surgery. Conversely, patients without detectable ctDNA after treatment had a complete response at the time of surgery with no recurrence. These results suggested that patient-specific ctDNA assays could improve non-operative management of these devastating cancers.

Finally, in Chapter IV, "Mixed ctDNA dynamics and decreased detection rates in early-stage lung cancer patients during radiation treatment," I present a previously-unpublished clinical study in early-stage lung cancer patients, where ctDNA abundance is particularly low and biopsies are often unobtainable. We tested the hypothesis that the radiation therapy routinely used to treat suspicious lung nodules would induce ctDNA shedding and thereby improve detection rates. I adapted our patient-specific assay from Chapter III and crafted a novel computational workflow to identify low-abundance ctDNA in a patient-agnostic manner. Remarkably, we found that ctDNA detection rates decreased significantly days after ablative radiation, but overall ctDNA abundance and dynamics varied between patients during treatment. Further studies with larger cohorts and more frequent sampling will help clarify the relationship between these dynamics and radiation treatment.

Taken as a whole, this body of work adds novel and useful information to the field of ctDNA research and advances the clinical utility of liquid biopsy.

Chapter I: Introduction to circulating tumor DNA in liquid biopsy

1.1 Cell-free DNA and circulating tumor DNA (ctDNA)

In the last decade there has been a revolution in personalized cancer care. Targeted therapy, molecular genotyping, and improved risk stratification at all stages of disease have reduced treatment toxicity, improved patient care, and reduced mortality in nearly every type of cancer. Perhaps the holy grail of this revolution is comprehensive disease detection, characterization, and monitoring by minimally-invasive liquid biopsy, yet it remains elusive. Tumor cells are not isolated from their host's tissue or its physiology. They require nutrients, generate waste, and excrete a myriad of small molecules, as all living cells do, potentially providing a wealth of information about the tumor that is there for the taking. Therein lies the promise of liquid biopsy: minimally-invasive access to tumor-derived analytes with negligible impact to the patient – we simply need to be clever enough to detect and decode them. The most generalizable of liquid biopsy analytes is tumor-derived cell-free DNA, called circulating tumor DNA (ctDNA). This introduction will discuss cell-free DNA generally as well as the utility, challenges, and limitations of ctDNA as a biomarker. More specific discussion of ctDNA origins and the mechanisms that drive its abundance will be described later in Chapter II.

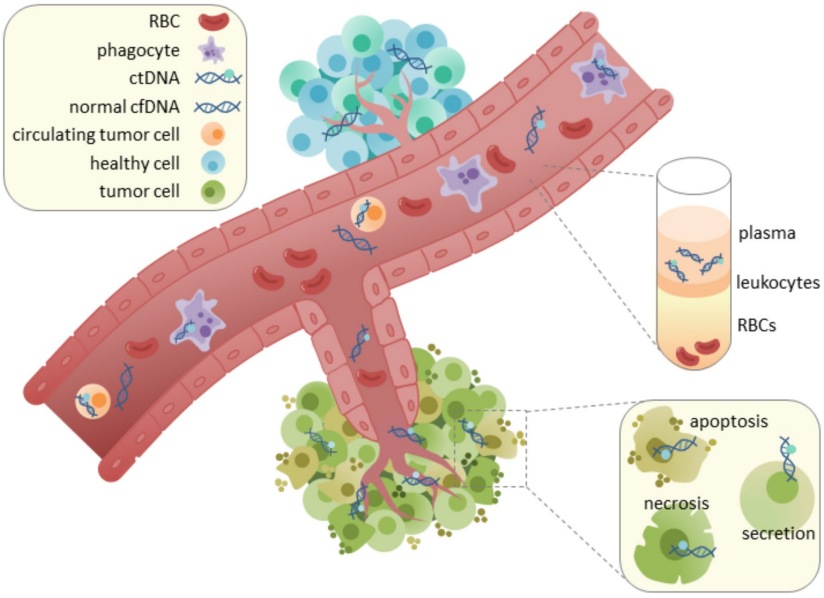


Figure 1.1. The physiology of cell-free DNA and circulating tumor DNA (ctDNA) in the blood.

Three major types of cell-free DNA shedding are shown (apoptosis, necrosis, and secretion). Blood components after fractionation by centrifugation are shown where cell-free DNA is found the "plasma" fraction and white blood cell DNA from the buffy-coat fraction (labeled "leukocytes"). Adapted from https://en.wikipedia.org/wiki/Circulating_tumor_DNA with https://creativecommons.org/licenses/by-sa/4.0/deed.en

Cell-free DNA is extracellular DNA found in nearly every bodily fluid and is thought to arise from multiple sources including normal cellular apoptosis (see Fig. 1.1). The work presented here is strictly concerned with cell-free DNA and ctDNA as it is found in the blood, which is primarily derived from hematopoietic cells [1] (see Fig. 1.2). First discovered in human plasma in 1948 by French researchers P. Mandel & P. Métais [2], cell-free DNA has long been thought to originate from the normal process of cell death and turnover in healthy tissue. It typically bears the ~160bp fragmentation pattern recognized as a hallmark of apoptosis resulting from cleavage of nuclear DNA around nucleosomes. A smaller fraction of high-molecular weight (>5k-10kb) species are also

sometimes observed and thought to be the product of other mechanisms, such as cell death via necrosis and pyroptosis, erythroid enucleation, or more exotic mechanisms like NETosis, cell-signaling, or micronuclei formation (see Aucamp *et al*. for a detailed review, [3]). So-called *secretion* or *active release* of cell-free DNA, which does not require cell death, has been theorized with some experimental support, but this term is misleading as it suggests energy-dependent mechanisms that have yet to be elucidated. Mounting evidence also suggests that apoptotic-like DNA fragmentation patterns can arise from necrosis-specific DNases and extracellular DNase activity in the blood [4-7]. This implies that cell-free DNA resulting from non-apoptotic mechanisms may appear fragmented in the same way as apoptotically-derived cell-free DNA, which upends many assumptions that still persist in the field to this day. For our purposes it is sufficient to summarize the origins of cell-free DNA, and therefore ctDNA – regardless of fragment size – into two categories: (a) the result of a cell's death or (b) originating from a cell that is still capable of division and propagation. The various mechanisms that fall under these categories are relevant throughout this work and are discussed in greater detail in Chapter II specifically as they pertain to ctDNA.

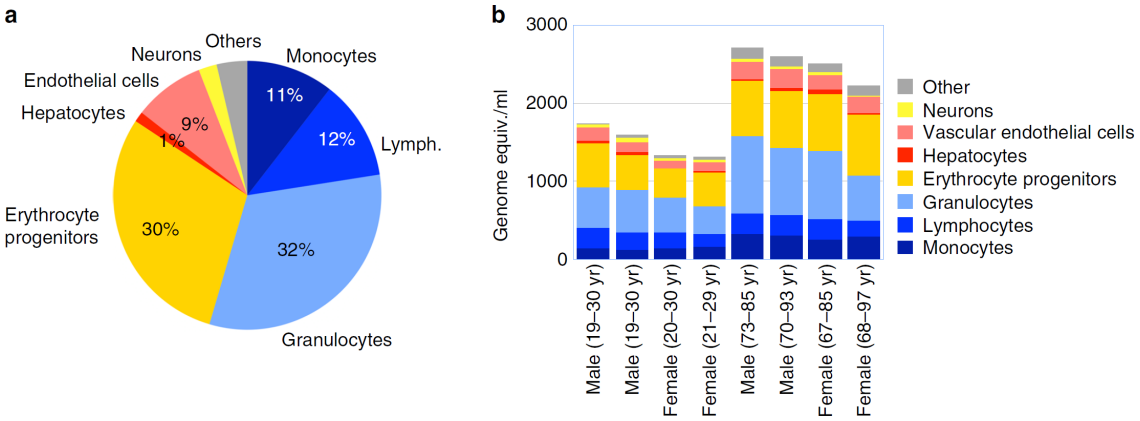


Figure 1.2. The tissue origins of cell-free DNA in healthy blood.
(**a**) Cellular sources of cell-free DNA in eight healthy individuals as predicted by methylation signature. (**b**) Distribution of data shown in **a** for each individual showing gender and age group ("young" 19-30 y/o, left four; "old" 67-93 y/o, right four). Adapted with permission from Moss et al. 2018 [1].

The rapid rate of clearance of cell-free DNA from circulation demands consideration in assay development, as it is anywhere from 15-120 minutes; similar rates have also been found for ctDNA [8]. This high turnover rate makes it ideal for real-time sampling but also contributes to significant longitudinal variability. In healthy individuals, the concentration of cell-free DNA is typically between 5-10 ng [or about 1,500 to 3,000 haploid genome equivalents (hGEs)] per ml plasma but can also vary over 10-fold in 24 hours [9]. One study found this variability to be as much as 1.4-fold in a 75-min period [10] and two other studies found a coefficient of variation (CV) of 30% over the course of three days in both cancer patients and healthy controls [11, 12]. The challenge then becomes understanding how the biology responsible for normal variation in cell-free DNA levels might impact variability in ctDNA and which end of the process is most determinant: production or clearance. One of those studies, by Hojbjerg and colleagues [12], looked at both total cell-free DNA and ctDNA levels over three consecutive days in stage IV lung cancer patients with stable disease and no concurrent

treatment. They found similar variability in total cell-free DNA and ctDNA, but 0.1- to 0.9-fold change (8%-57% CV) over three days in the ctDNA fraction (*i.e.*, mutant allele fraction, MAF, defined as ctDNA/total cell-free DNA). A study by Wagner and colleagues at OHSU [13] found no significant changes in diurnal abundance of cell-free DNA, but significant differences between individuals. Finally, there is also evidence that cell-free DNA abundance is linked to genetic and environmental similarities and also increases with age [14]. Again, it is unclear if these results are a consequence of production rates or clearance rates and they highlight one of the primary challenges in using ctDNA as a generalizable biomarker (see section 1.4 below and Chapter II). Consequently, understanding the sources of day-to-day and inter-patient variability is key to its widespread clinical utilization.

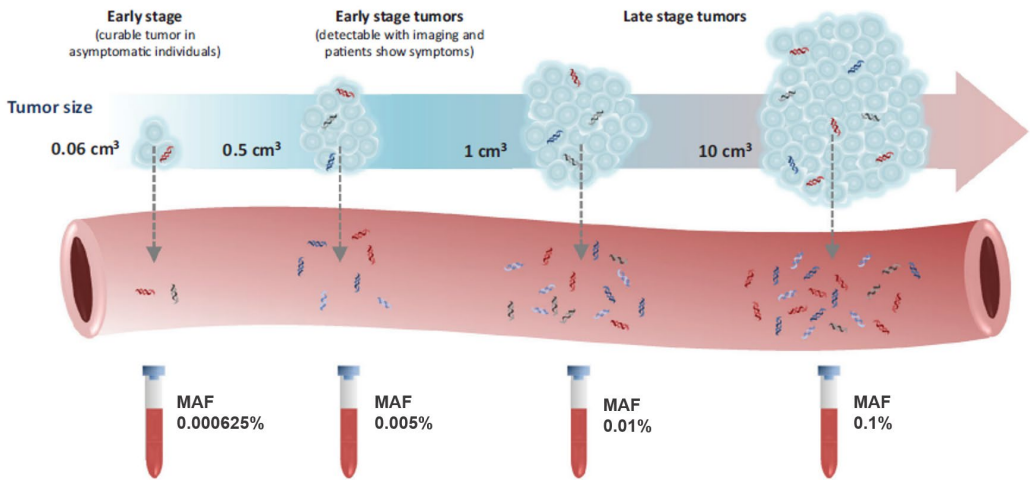


Figure 1.3. Expected ctDNA abundance by tumor size.
Literature-base estimation of ctDNA fraction in cell-free DNA (mutant allele frequency, MAF) as a function of tumor size. More advanced disease with greater tumor burden are generally associated with increased ctDNA abundance. Adapted with permission from Bronkhorst et al. 2019 [15].

1.2 Research and clinical applications of ctDNA detection and analysis

Despite these challenges, the effort to utilize ctDNA analysis for various clinical purposes has marched on. As with so many such technological advances, we often figure out how to exploit something before we even begin to understand how it works. The clinical utility of ctDNA is often defined in four unique applications that are often in line with disease stage: (1) early-detection screening; (2) tumor-burden and minimal residual disease (MRD) detection; (3) tumor genotyping/genetic profiling; and (4) tumor evolution and mechanisms of resistance. Typically, ctDNA abundance increases with tumor burden (see Fig. 1.3), therefore as we progress through these applications and disease severity, assay utility and reliability usually also improve. These four applications are summarized in this section.

Early-detection screening is intended to detect cancer in either general or high-risk populations. These assays – called MCED (multi-cancer early-detection) when applicable to more than one cancer type – are not to be confused with genetic screening which can identify inherited mutations, such as mutant BRCA1/2, that might confer increased future risk of disease but don't indicate the current presence of cancer. Instead, early-detection screening cell-free DNA tests look for tumor-specific features, like methylation status or somatic mutations (*i.e.*, ctDNA), as an indication of an existing tumor. In high-risk populations such as life-long smokers, these assays could be imagined as a companion diagnostic used in combination with an annual low-dose CT screening for lung cancer. Unfortunately, early-stage tumors often do not shed enough ctDNA to be detectable by this method (more on this in Chapter IV). To try to improve sensitivity, a pan-cancer screening MCED assay ([16], https://grail.com/) is available to clinicians that uses machine learning to detect tumor-derived ctDNA from differential

methylation patterns in next-generation sequencing (NGS) data. Although this has greatly improved assay sensitivity, there are still significant challenges in certain cancers at early-stages (Fig. 1.4).

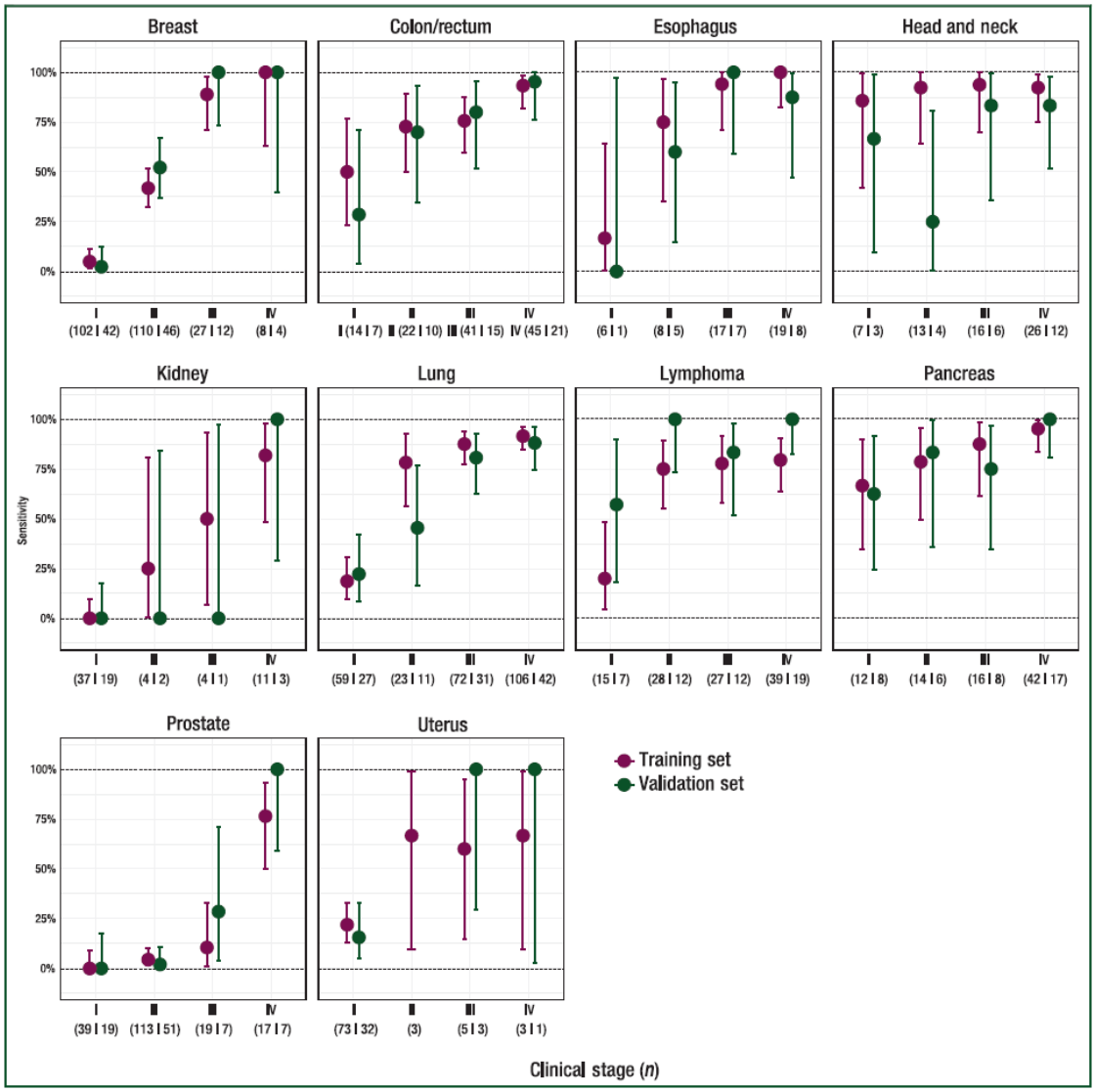


Figure 1.4. ctDNA detection rates using a methylation-base assay in multiple cancer type.

Results of a validation study for the methylation-based multi-cancer early detection (MCED) assay developed by GRAIL (https://grail.com/) showing assay sensitivity by stage in 12 cancer types. Assay sensitivity is assumed to be a direct result of ctDNA abundance. Many cancers have detection rates far below 50% in stage I disease (*e.g.* breast, kidney, lung, prostate, and uterus). Adapted with permission from Klein et al. 2020 [16] under https://creativecommons.org/licenses/by-nc-nd/4.0/

In the research setting, ctDNA abundance is associated with **tumor burden** but often with many exceptions, as will be discussed in section 1.4 and Chapter II (see Fig. 1.5 for the association in lung cancer patients and also note the variability). Clinical ctDNA assays evaluating **MRD status** and **tumor burden** typically use a patient-informed set of mutations (from sequencing of a tumor biopsy) to assess the presence and abundance of ctDNA, respectively. These tests are commercially available and use NGS-based approaches similar to those developed by our lab [17] to detect and quantify ctDNA. MRD testing is useful for measuring the efficacy and completeness of treatments such as chemotherapy and surgery, particularly in the neoadjuvant setting, or during surveillance for recurrence as shown in Chapter III. When serial testing is done, the resulting data can also be used to estimate the increase or decrease in tumor size or growth rate. With *a priori* knowledge of patient's tumor genotype (*i.e.*, mutations), these assays readily outperform imaging-based modalities with significant clinical lead time for detection. Several cases of improved clinical lead time using this method are also presented in Chapter III.

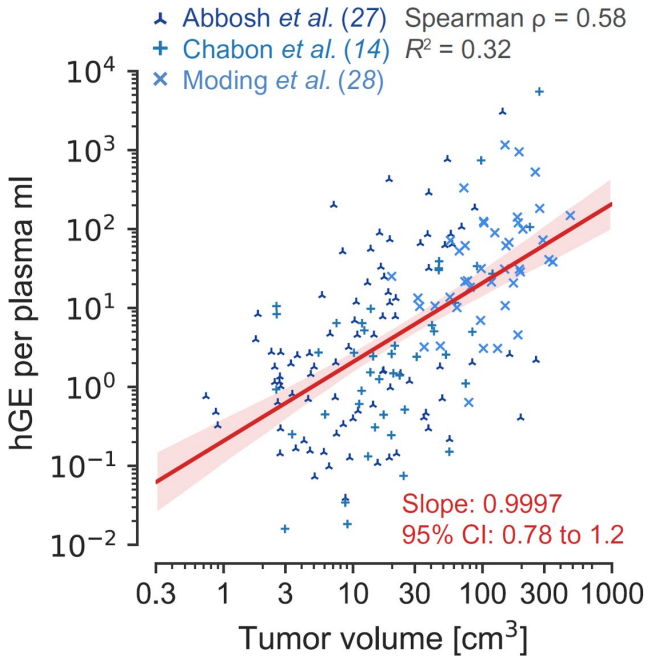


Figure 1.5. Association of ctDNA abundance and tumor volume in lung cancer. ctDNA abundance expressed as haploid genome equivalents (hGE), or mutant copies, per ml of plasma as a function of tumor volume in 176 non-small cell lung cancer patients from 3 studies. The authors used linear regression to predict 0.21 hGE per plasma ml for a tumor volume of 1 cm^3. Note the high degree of variability, where large tumors may not shed much ctDNA and small tumors can shed more than expected. From Avanzini et al. 2020 [18]. Reprinted with permission from AAAS.

Tumor genotyping by liquid biopsy is based on the idea that genetic profiling of the ctDNA provides a real time picture of a patient's disease that is representative of the dominant tumor cell population (*i.e.*, clones) and sometimes subpopulations or subclones – either from a single heterogeneous tumor, or multiple tumors in different locations. In principle, it is superior to a solid tissue biopsy, which may be subject to intra- and inter-tumor location or sampling bias, and is highly representative of the overall disease [19]. This application typically involves either whole exome sequencing (WES) or a disease-appropriate multi-gene panel and can provide a comprehensive picture of a patient's disease as compared to sequencing solid tissue, particularly when

it is more advanced and there are multiple tumors, or when tumor location prevents safe biopsy. Moreover, serial sampling can be done in short intervals to provide repeat snapshots that are not possible with solid tissue biopsies. This approach is particularly attractive when monitoring for the emergence of actionable mutations for targeted therapies such as ALK fusions and EGFR mutations in lung cancer [20].

The fourth category of applications uses serial sampling to track **tumor evolution** and detect **mechanisms of resistance**. This is done by longitudinal interrogation of patient-specific mutations alone or in-combination with WES or a smaller multigene panel. If multiple subclones with selective potential have been identified in a single tumor or multiple tumors, serial monitoring of the relative abundance of these subclones can provide a picture of the fitness and relative contribution of each subpopulation (Fig. 1.6). A broader panel can also be useful in detecting previously undetected mutations *de novo* that might arise by selective advantage. These assays may have particular value when done during targeted therapy to determine which populations thrive under that selective pressure and which do not. Not only does this provide valuable near-real-time information to clinicians that may lead to improved disease management, it has tremendous implications for our understanding of tumorigenesis and tumor evolution. Obtaining such data would not be feasible by solid tissue biopsy in human subjects or in animal models.

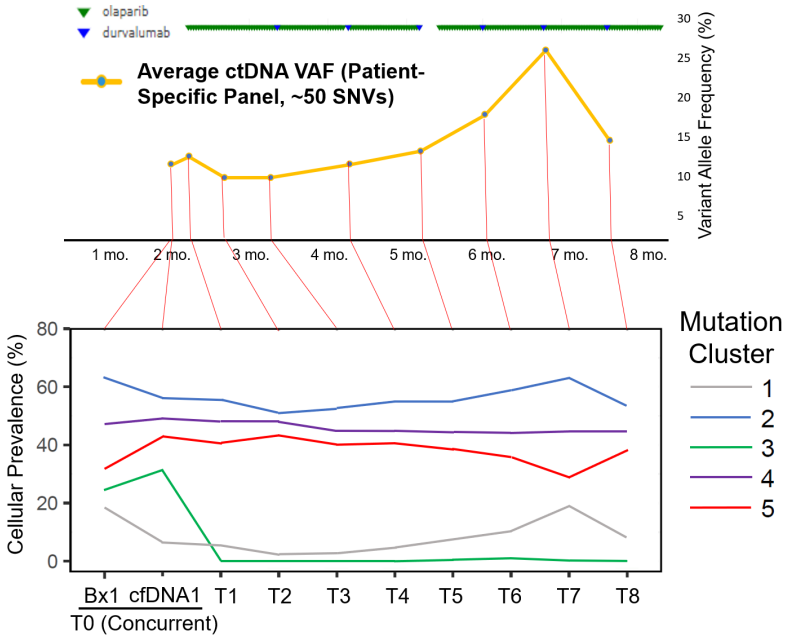


Figure 1.6. Subclonal inference using ctDNA dynamics during treatment. ctDNA abundance dynamics using whole-exome sequencing (for Bx1 and cfDNA1) and a patient-informed mutations panel (n≈50 SNVs, for times T1-T8) in a metastatic breast cancer patient during treatment (top). Read counts of ~50 individual mutations (not shown) at each time point were input into a clustering algorithm to identify potential subclonal patterns (bottom). The resulting trajectories of the 5 mutation clusters suggest that overall increases in ctDNA abundance (top) is the result of relative increased representation of subclones associated with cluster 1 and possible clonal outgrowth. (SMMART patient, data not previously published)

1.3 Methods for evaluating ctDNA

1.3.1 Distinguishing features of ctDNA

Various methods of detecting and quantifying ctDNA have been developed, each with its own benefits and drawbacks. In theory, any feature of cell-free DNA that distinguishes it as tumor-derived can be used for this task. The majority of our focus thus far has been on tumor-specific (*i.e.*, somatic) mutations such as single-nucleotide

variants (SNVs), small insertions or deletions (indels), or gene fusions. However, larger chromosomal abnormalities, focal or whole-chromosome amplifications or losses (copy number variations or CNVs), epigenetic methylation states, and even small variations fragment size have all been exploited for evaluating ctDNA. Most of these features can be detected using well-characterized next-generation sequencing (NGS) techniques, but other methods are more specialized for certain features and sensitivities. Due to the stochastic nature of sampling large numbers of individual molecules present in cell-free DNA in search of ctDNA, the likelihood of detection and the accuracy of quantification are heavily dependent on the number of reporters interrogated – particularly when sampling for rare events (see section 1.4.1 and Figs. 1.7B and 1.7C). The number of mutations we can interrogate is limited by the number of mutations present in the tumor as well as the method by which we go looking for them in cell-free DNA. Because of these constraints, the development of methods such as differential methylation profiling, which can interrogate 100,000s of potential tumor-specific markers in a single genome (*i.e.*, methylation sites), may prove a more fruitful approach. Other methods that utilize copy number variations (CNVs) or fragment size require ctDNA fractional abundance of 1%-3%, which usually limits their use to more advanced disease [21, 22].

The detection of ctDNA using genetic mutations such as SNVs, indels, and CNVs is the most established and well documented approach. The original work presented in this dissertation is primarily focused on these methods and, henceforth, ctDNA detection will be discussed in the context of SNVs, indels, and CNVs, unless otherwise indicated.

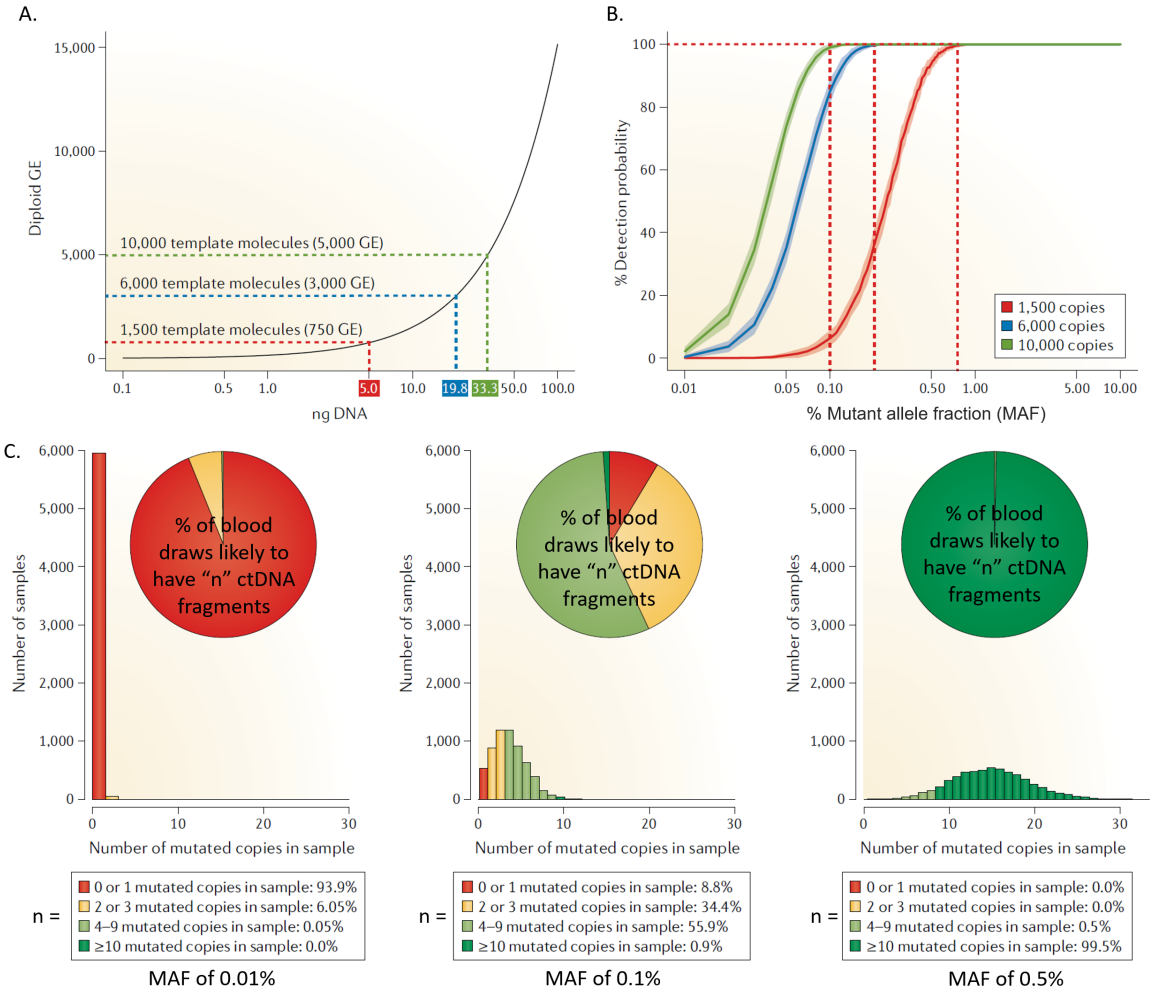


Figure 1.7. The impact of stochastic effects on ctDNA detection in cell-free DNA. (**A**) Number of diploid genome equivalents (*i.e.*, 2 x haploid GE) present in a given amount of cell-free DNA. (**B**) The relationship between the cumulative probability of detecting 10 ctDNA hGEs in a sampling of 1,500, 6,000, and 10,000 cell-free DNA hGEs as a function of percent MAF. The red lines indicate the minimum MAF that can be reliably (10 molecules at 100%) detected in a given amount of DNA (hGEs) interrogated. (**C**) Poisson distributions generated by simulated sampling of cell-free DNA populations of 6,000 hGEs with ctDNA present at a given MAF (0.01%, 0.1%, and 0.5%). Pie charts show the probability of sampling "n" ctDNA fragments; where n = 0 or 1, 2 or 3, 4 to 9, or ≥10 ctDNA hGEs at each MAF. Adapted with permission from Heitzer et al. 2019 [23].

1.3.2 Patient-informed and patient-agnostic ctDNA assays

All four general applications described in section 1.2 fall somewhere on a spectrum between *patient-informed* or *patient-agnostic*. The patient studies presented in Chapters II and III utilize both of these approaches for different purposes. Further explanation of the underlying principles and methods for each are provided here.

Patient- or tumor-informed assays require prior knowledge of genetic information about the tumor(s) in the patient being tested, typically from one or more solid tissue biopsies or resected tumors. Genotyping of this tissue for somatic mutations or other genetic aberrations can inform patient-specific enrichment panels or indicate which mutations to monitor in a generalized panel. These panels are used as part of DNA sequencing library preparation to limit the library to only those that are of interest. The benefit of this *target enrichment* approach is that, unlike whole genome sequencing or whole exome sequencing, we restrict valuable sequencing reads to those regions of the genome that we expect to find a mutation. In this way, instead of sequencing 10 hGEs of the entire coding region (*i.e.*, 10X depth), which is ~50Mbases of sequencing each, we can sequence just the genomic positions with potential mutations, say a 5kb panel, to 100,000X with the same amount of sequencing. By reducing our search space by 10,000-fold, we increase our sensitivity in equal measure. This approach makes detecting one mutant hGE (*i.e.*, ctDNA) mingling with 100,000 healthy hGEs of cell-free DNA much more feasible because we can sample more of them. Moreover, by interrogating more than one tumor-derived mutation, *e.g.* 10 which is easily covered in a 5kb panel, we can improve the probability from 1 in 100k to 10 in 100k, or 1 in 10k. Patient- or tumor-informed ctDNA assays have been used extensively in research and are currently available for detecting MRD throughout treatment and surveillance in some

cancers (https://www.natera.com). Of course, these assays require a sample of solid tumor tissue from which to build the assay panel.

Patient-agnostic approaches are intended to detect ctDNA without foreknowledge of patient-specific somatic mutations or with the hope of discovering previously undetected mutations. As discussed in section 1.2 above, cancer screening of general or high-risk populations requires this approach, but suitable assays have yet to be developed in all but a few limited settings. In the setting of non-small cell lung cancer (NSCLC), a patient-agnostic genomic target enrichment panel was developed by Newman and colleagues [24-26] with the intent of detecting ctDNA in high-risk individuals not already diagnosed with cancer. Although it has yet to be tested in a large clinical trial, they report disease stage-dependent detection rates ranging from 40%-98%, stage I to IV, respectively [26]. Their approach sequences bulk cell-free DNA from each patient to sampling depths of ~5,000X at specific genomic positions defined by their custom enrichment panel. The panel targets regions of the genome found to have recurrent mutations in thousands of NSCLC patients, as independently reported in publicly available datasets (*i.e.*, The Cancer Genome Atlas, https://www.cancer.gov/tcga). In Chapter IV, I combine this target enrichment panel with a sequencing method that colleagues and I previous developed [17] to improve ctDNA detection rates in early-stage NSCLC during treatment. The most challenging aspect of patient-agnostic applications using NGS, particularly in early-stage disease, is distinguishing true mutations from sequencing errors – as will be discussed in section 1.4 below. Assuming a tumor is present, without the ground truth of tumor-specific mutations, very low-frequency ctDNA molecules harboring mutations are difficult to confidently distinguish from background error. Sequencing errors can result in potential false-positives in healthy individuals and mis-identification of tumor mutations in cancer patients.

1.4 Challenges to ctDNA as a biomarker

Unfortunately, the challenges of implementing cDNA detection, quantification, and characterization assays in research and clinical applications are not trivial. In the following sections I will discuss the various technical and biological dilemmas that have plagued widespread utilization of ctDNA analysis as well as some of the approaches that have been developed to overcome them – and more specifically, those used in the research presented in this work.

1.4.1 Biological constraints and pitfalls

The greatest biological constraint to ctDNA analysis is the relatively low amount of DNA material available in a given blood draw. ctDNA abundance is inherently limited by the number of tumor cells in the body. Using data from lung cancer patients, various studies have estimated the expected total amount of ctDNA in circulation and its fractional abundance in cell-free DNA given the size of a tumor [18, 27]. These studies calculated that a typical stage I tumor (~1 cm^3 and ~$1x10^9$ cells) would contribute ≤1000 fragmented haploid genomes (aka, haploid genome equivalents, or hGE) to circulation at any given sampling, resulting in a fractional abundance (MAF) ranging between 0.002% and 0.03%. We must then consider the probability that these molecules are captured in a 10-20 ml normal blood draw, which may only consist of 5,000 total hGE (see Fig. 1.7A), with the average total volume of blood in circulation of 5,000 ml; Fig. 1.7C shows a probabilistic model for detecting ctDNA at three different MAFs (0.01%, 0.1%, and 0.5%) using 20ng of input cell-free DNA (~6,000 hGE). Simulated sampling of this population reveals a Poisson distribution where the probability of detecting 10 or more ctDNA molecules drops from 99.5% with a MAF of 0.5% (Fig. 1.7C, right) to 0.0% with a MAF of 0.01% (Fig. 1.7C, left). Moreover, because no analyte preparation can be 100%

efficient, there is a practical limit of detection that is dictated by biology as we test for smaller and smaller tumors. This sets a hard limit on any ctDNA assay, regardless of how technically sensitive it is. Fig. 1.7B demonstrates the minimum MAFs that can be reliably detected for a given amount of cell-free DNA interrogated.

Because cell-free DNA can theoretically be derived from any source in the body (including foreign sources such as microbes and viruses), extra care must be taken to mask their sequences from the data when present. Since sequencing data is mapped to a human reference genome, foreign DNA is typically not present in the resulting alignment file, however somatic mutations can also occur in non-cancerous tissue. This raises the larger question of when a mutated cell(s) is considered a cancer, but for our purposes we will limit ourselves to cell populations that are lethal cancer. Clonal hematopoiesis (CH) is the most significant contributor of somatic mutations to the cell-free DNA milieu. As discussed in section 1.1 (see Fig. 1.2), blood cells are thought to be the primary source of cell-free DNA in circulation. As we age, the process of CH decreases the genetic diversity of blood cells by selective pressure and can eventually lead to hematopoietic malignancies. CH of indeterminate potential, or CHIP, is the term used to describe these populations after they reach a fractional abundance of 2% or more, but CH commonly persists at lower fractional abundances. These somatic mutations are readily detected in cell-free DNA and can be mistaken for ctDNA in the absence of proper experimental controls. The best way to mask these potential false-positives from ctDNA analysis is to sequence the fraction of white blood cells (WBCs) isolated from the same blood sample as the cell-free DNA. As described in Chapter IV, by identifying the mutations with sequencing reads in both cell-free DNA and WBC tissue compartments, CH can typically be detected and omitted from analysis. Unfortunately, this has yet to become standard practice in the field, although more studies are including it in their analysis.

1.4.2 Technical hurdles and limitations

The technical challenges described below are only compounded by the biological limitations outlined in section 1.4.1 above. However, the primary technical challenge of detecting low-MAF ctDNA using next-generation sequencing (NGS) may be false-positives resulting from random polymerase errors. Enzymes such a *Taq* polymerase are used in the amplification steps of sequencing library preparation and during the sequencing-by-synthesis process used by NGS to generate sequencing reads. During DNA extension, even high-fidelity polymerases incorrectly incorporate new nucleotides at a rate of ~1.8×10^{-4} errors/base/doubling [28], or 1 error in every 5.5kb for each doubling. PCR amplification steps done during library preparation can involve 10-20 cycles in multiple steps, which would translate to 10-20 doublings given 100% efficiency. The cumulative error rate of NGS (post-library preparation) varies between sequencing platforms. For our purposes, we will consider the Illumina NovaSeq 6000, which is the current standard for high-throughput short-read sequencing and has an independently-reported median error rate of 0.109 (s.d. = 0.35), or 1 error in every 917 bases (s.d. = 285) [29]. In my experience, the actual error rates of traditional NGS vary from 1 error in 0.1kb-1kb, which limits detection of true mutation in cell-free DNA to 1% or 0.1% MAF. The use of molecular barcoding or unique molecular indexing (UMI) techniques has dramatically improved these error rates, up to 1000-fold in the case of duplex sequencing [30]. As illustrated in Fig. 1.8, by encoding a unique oligonucleotide barcode onto each input DNA molecule, prior to PCR amplification, the sequence of the original, biologically-unique cell-free DNA fragment can be reconstructed from raw sequencing data. As PCR amplification geometrically increases the relatively small number of input DNAs, these barcodes get propagated, as new errors do as well, such that the final library contains many copies of each original input molecule with its own UMI. Software

is then used to aggregate sequencing reads having same UMI and genomic mapping position to generate a consensus sequence where errors introduced during amplification steps are removed. Both studies presented in this work use a dual-UMI method previously developed in our lab, called “DIDA-Seq” (shown simplified in Fig. 1.8), for patient-informed ctDNA monitoring at sensitivities of one ctDNA molecule in 10k to 50k cell-free DNA molecules [17, 31-34].

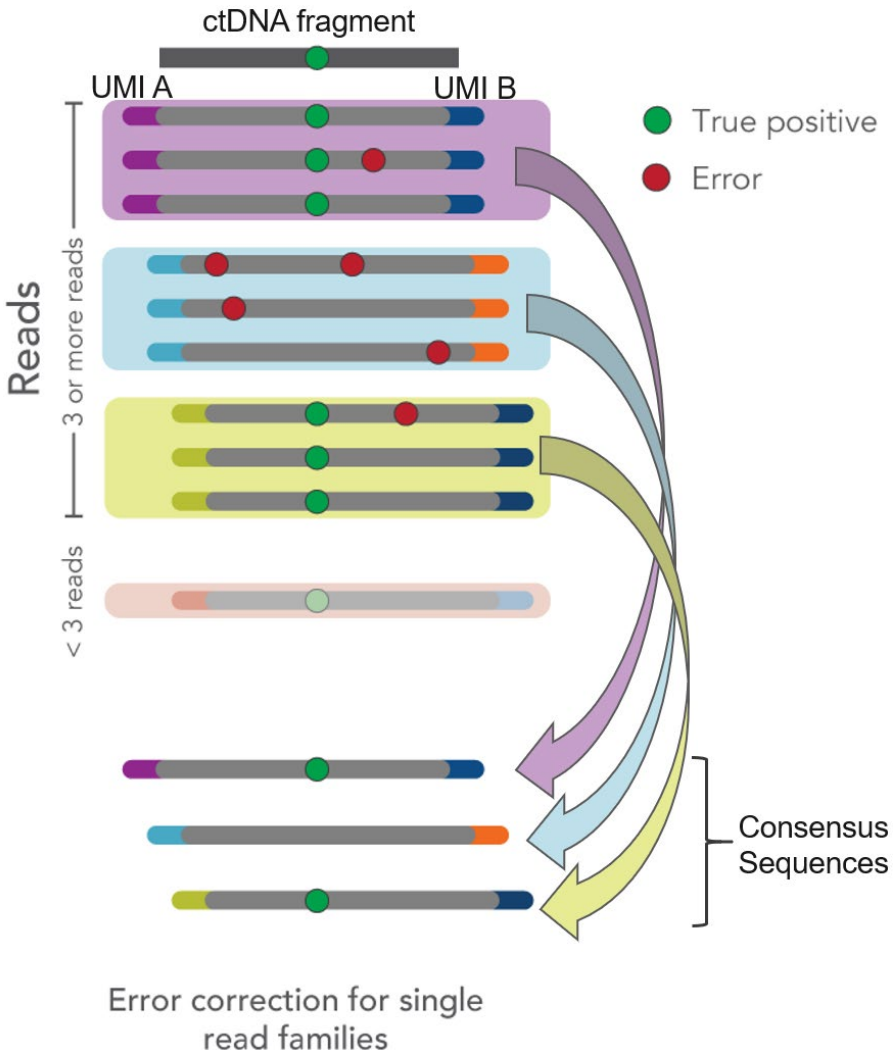


Figure 1.8. Schematic depicting a dual unique molecular index (UMI)-based error-correction sequencing method.

Bar-coding of input cell-free DNA fragments (top) with unique UMIs ("UMI A" and "UMI B") occurs by adaptor ligation (not shown) during library preparation. *In silico* consensus-making from matching UMI families (purple, blue, and yellow) with 3 or more PCR duplicates (middle) produces 3 single-stranded consensus sequences (SSCSs; bottom) that reconstruct the unique DNA input molecule. Errors introduced during library preparation (red dots) are removed during consensus making and true positives are retained (green dots). Adapted with permission from https://www.idtdna.com/pages/support/faqs/how-do-i-use-umis-for-error-correction

Another technical challenge involves stereotypical errors that are introduced *ex vivo* in the form of mechanical or oxidative damage DNA damage that occurs prior to library amplification steps. These errors can be difficult to detect and remove by UMI-based consensus-making. Previous studies have linked 8-oxguanine conversion and cytosine deamination, which result in the C>T and G>A nucleotide substitutions, to mechanical fragmentation (*i.e.*, sonication) and the high incubation temperatures required for target enrichment by hybridization capture (discussed in section 1.3.2) [25, 26]. These same studies presented computational methods for identification and removal of these errors, which are used and discussed in Chapter IV of this work.

In Fig. 1.9, I summarize general workflows for ctDNA analysis and show the steps in which nucleotide substitutions and other errors can be introduced, creating the potential for false positives in the analysis. I also illustrate the major bottlenecks and efficiency losses which reduce sample complexity and thus analytical sensitivity.

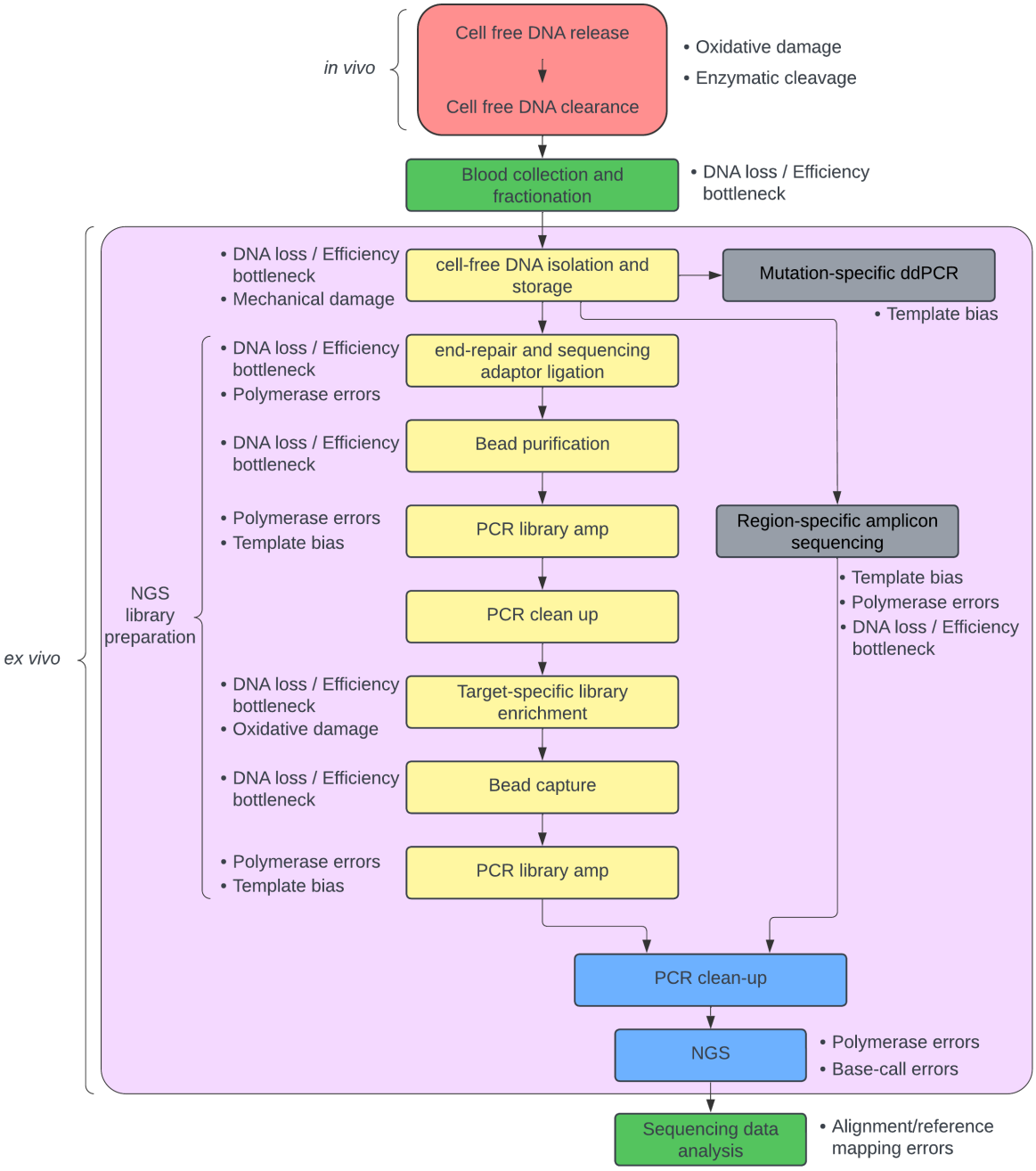


Figure 1.9. The fate of cell-free DNA from its release to sequencing and analysis. Each step is annotated as appropriate with error introduction, including type, and efficiency losses. Amplicon-based sequencing and ddPCR are shown in dark grey for reference. Note that their use is limited by multiplexing complexity and only fixed genomic positions with usable priming sites can be used.

1.4.3 Unexplained sources of variability

Finally, perhaps the most mysterious and intractable problem with ctDNA as a biomarker is its variability, not just by tumor location or type. Largely assumed to be biologically driven, ctDNA levels can vary dramatically and often inexplicably between patients with similar disease and stage (see Fig. 1.10) [35, 36], and even in the same patient on different days (as described in section 1.1 above). Even in patients with large tumors and high tumor burden from advanced disease or metastasis, ctDNA levels can be unexpectedly low or undetectable (see Fig. 1.5) [23, 37]. Various physiological factors and mechanisms have been proposed to explain this variability, although definitive data remain scant. **Chapter II of the work that follows will outline these possibilities and the existing evidence to support or refute them.**

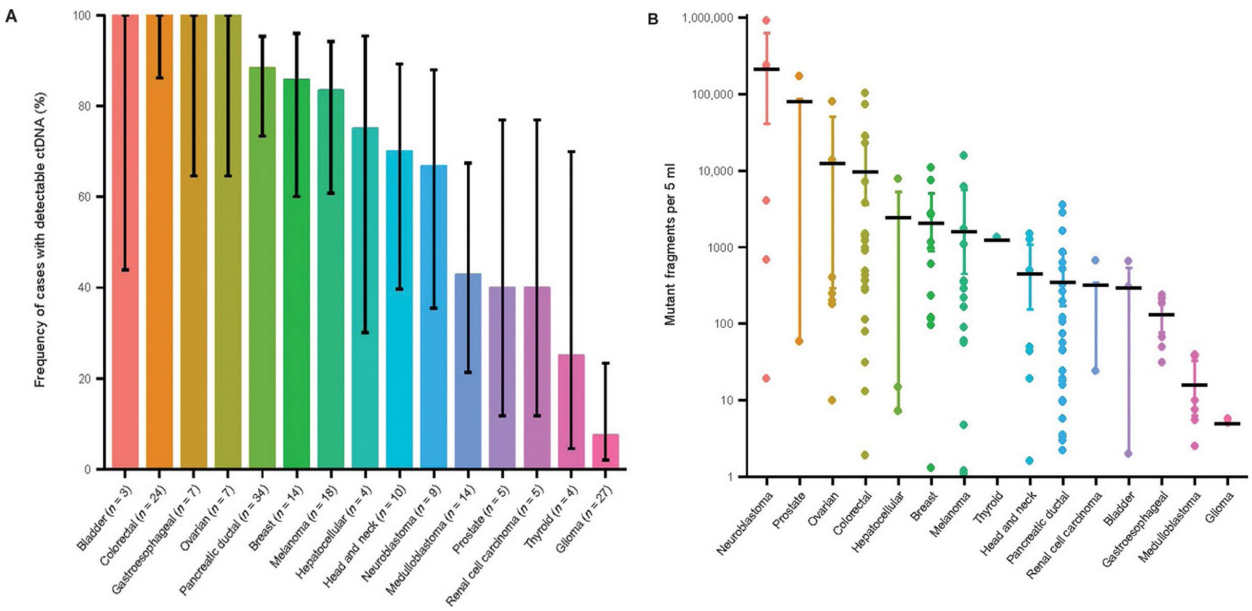


Figure 1.10. ctDNA measured in various cancer types.
ctDNA detection rates (**A**) and abundance (**B**) in 15 cancer types. Adapted with permission from Bettegowda et al. 2014 [35].

Chapter II: Blood, Toil, and Taxoteres: Biological Determinants of Treatment-Induced ctDNA Dynamics for Interpreting Tumor Response

Christopher T. Boniface[1,2]* and Paul T. Spellman[1,2]*

[1] Knight Cancer Institute, Oregon Health & Science University, Portland, OR, United States

[2] Cancer Early Detection Advanced Research Center, Knight Cancer Institute, Oregon Health & Science University, Portland, OR, United States

* Corresponding authors (boniface@ohsu.edu and spellmap@ohsu.edu)

Adapted from Boniface et al. 2022 [38]

Abstract

Collection and analysis of circulating tumor DNA (ctDNA) is one of the few methods of liquid biopsy that measures generalizable and tumor specific molecules, and is one of the most promising approaches in assessing the effectiveness of cancer care. Clinical assays that utilize ctDNA are commercially available for the identification of actionable mutations prior to treatment and to assess minimal residual disease after treatment. There is currently no clinical ctDNA assay specifically intended to monitor disease response during treatment, partially due to the complex challenge of understanding the biological sources of ctDNA and the underlying principles that govern its release. Although studies have shown pre- and post-treatment ctDNA levels can be prognostic, there is evidence that early, on-treatment changes in ctDNA levels are more accurate in predicting response. Yet, these results also vary widely among cohorts, cancer type, and treatment, likely due to the driving biology of tumor cell proliferation, cell death, and ctDNA clearance kinetics. To realize the full potential of ctDNA monitoring in cancer care, we may need to reorient our thinking toward the fundamental biological underpinnings of ctDNA release and dissemination from merely seeking convenient clinical correlates.

2.1 Background

Circulating tumor DNA (ctDNA) is extracellular DNA in plasma that originates from tumor cells and has emerged as a useful biomarker in minimally-invasive liquid biopsy [39-41]. ctDNA abundance shows broad correlation with tumor burden and generally reflects the tumor DNA content such that clinical assays are commercially available for detection of molecular/minimal residual disease (MRD) and tumor

mutational profiling [27, 42, 43]. However, there are currently no ctDNA-based assays approved for serial monitoring during treatment to assess immediate tumor response and treatment efficacy.

Serial ctDNA monitoring during treatment can provide insight into underlying biological factors that can potentially be used to predict response, treatment efficacy, and long-term outcomes [17, 31, 44-46]. In practice however, ctDNA levels can appear erratic across time points and are often inconsistent between patients with similar disease and treatment. This variability may be partially the result of disparate sampling frequency (often within the same study), extraction methods, and analytical approaches between studies. More likely, this variation is driven by factors that have yet to be elucidated and may vary between patients, such as individual host physiology, tumor location, tumor biology, and treatment modality.

Evidence that ctDNA concentration is more dependent on tumor cell replication rates than simply on tumor volume also suggests that understanding tumor biology and patient physiology are necessary to guide proper interpretation of ctDNA dynamics [47, 48]. In some studies, early spikes in ctDNA shortly after treatment may predict a favorable clinical response, in keeping with the hypothesis that shedding is directly associated with treatment-induced tumor cell death [49-51]. It is unclear however, how soon after treatment initiation this spike must occur in such cases, emphasizing the importance of collection timing. Nevertheless, early and rapid ctDNA clearance during treatment has consistently been shown to correlate with objective response and outcome [52-55]. Evidence supports the idea that ctDNA release is clearly a byproduct of tumor cell proliferation, though whether this is through increased cell turnover and higher death rates or active release during cellular expansion is still an open question.

This review seeks to discuss the biological sources of variability in ctDNA abundance, with the hope that thoughtful analysis and a mechanistic understanding of

ctDNA release will allow improved approaches to ctDNA interpretation in clinical response and progression.

2.1.1 A note on ctDNA detection and its implications for this review

Typically, ctDNA is detected and characterized using methods such as droplet digital polymerase chain reaction (ddPCR) or next-generate sequencing (NGS) (reviewed by Heitzer et al. [40]) either targeting genomic positions based on *a priori* knowledge of tumor mutations or by calling mutations *de novo* at novel sites and recurrent hotspots. Although both *a priori* and *de novo* approaches to ctDNA detection assume that molecules harboring alternate alleles are tumor-derived, the later approach allows for ctDNA detection without *a priori* knowledge of tumor-specific mutations but is subject to much more uncertainty. The detection of tumor-derived copy-number aberrations in cell-free DNA is also possible and is dependent solely on read counts to detect gains or losses found in tumor cells [21]. *De novo* mutations can be called from whole-genome and whole-exome sequencing, or smaller panels targeting just a few sites or genes known to harbor recurrent mutations. Panels intended to detect mutations at these canonical sites are often less informative about passenger mutations, secondary drivers, and subclonal populations. The same may be true for tumor-informed and patient-specific panels depending on the breadth of the panel used and the sampling bias of the original tumor tissue used to design the panel, particularly in the case of high intratumoral heterogeneity. These various approaches further confound our ability to compare results across studies, patient populations, and cancers. This problem is particularly true for studies where ctDNA was characterized by the prevalence of a single mutation in a single gene, where subclonal populations driven by other genetic

aberrations may be under selective pressure during treatment. Of course, with a broader analytical space comes greater cost and complexity and additional challenges for implementation of accurate ctDNA assays in a clinical setting. For example, a simple ddPCR or amplicon-sequencing test to detect the presence of low-abundance EGFR mutations in the cell-free DNA of lung cancer patients is much cheaper and simpler to validate and execute in a diagnostic laboratory than a whole-exome or a multi-gene sequencing panel with similar accuracy and sensitivity. However, such an assay may not be representative of the entire tumor cell population, particularly if there are treatment-resistant subclones that harbor distinct genotypes. Consequently, careful evaluation of single-target vs multi-target approaches is necessary.

Although the data available to assess ctDNA abundance as it relates to clinical observations, treatment response, and outcome consist primarily of mutant-allele detection and prevalence estimates, evidence suggests approaches like methylation profiling by whole-genome or targeted bisulfite sequencing may be more sensitive and are not dependent on the presence of genetic aberrations [42]. A serial comparison of single-nucleotide variants (SNVs) and methylation profiles in EGFR T790M-positive advanced cancer patients found that methylation levels closely followed SNV mutant allele frequency and both were predictive of long-term treatment response [56]. Silva and colleagues [57] reported changes in cell-free DNA methylation over time that were associated with therapy response and progression in prostate cancer patients. Few studies, however, have assessed methylation dynamics in cell-free DNA with high-frequency sample collection during early phases of treatment, and therefore the data presented here are biased toward somatic mutations as a means of ctDNA detection and characterization. It remains to be seen how broad the search space needs to be in order to effectively monitor tumor cell populations by ctDNA, and which reporter (*i.e.*,

mutations, methylation, etc.) will be the most informative, but may vary by patient, tumor, and treatment.

2.2 Biological factors that most affect ctDNA abundance

In order to utilize ctDNA monitoring during treatment we must understand the various factors that impact ctDNA concentration over time. Multiple sources of ctDNA have been suggested including apoptosis, necrosis, and so called “active/passive release” (reviewed by Aucamp et al. 2018) [3]. Apoptosis, necrosis, and other forms of tumor cell death result in ctDNA release into interstitial space where it moves to the lymphatic system and blood circulation [45, 58]. It is also hypothesized that extracellular DNA can be released from living cells in various contexts that are both energy-dependent and independent, and range in mechanism from shedding of mis-segregated DNA to intercellular signaling [59]. Once ctDNA enters circulation, it is subject to further degradation by DNases in the blood and is putatively removed by the liver, spleen, and/or kidneys within 30 to 120 minutes [60]. Changes in the balance of these processes due to treatment are assumed to be reflected in ctDNA dynamics, which, if correctly interpreted, may inform us about a patient’s disease state and response to therapy (Fig. 2.1). Furthermore, disruption of biological homeostasis resulting from disease and treatment can increase overall levels of cell-free DNA, decreasing the relative abundance of ctDNA and thus impacting assay sensitivity. This can also make interpreting data from studies that simply report mutant allele frequency challenging without accounting for such changes in total cell-free DNA. The following sections address how the unique biology of a cancer and host physiology influences ctDNA abundance in the blood stream.

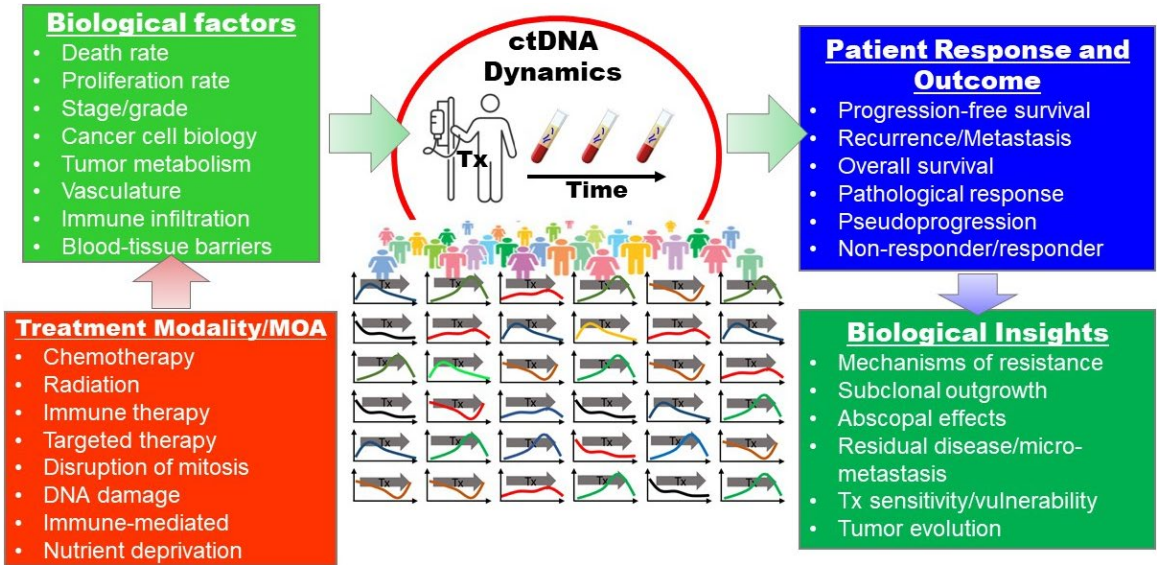


Figure 2.1. Biological and clinical insights from ctDNA dynamics during cancer treatment.

The relationships between treatment, biological factors, clinical indicators and outcomes, and potential insights with regard to serial ctDNA monitoring during treatment. (MOA = mechanism of action; Tx = treatment)

2.2.1 Cancer type and biology

Even before our ability to distinguish ctDNA molecules amongst cell-free DNA, there appeared to be a clear relationship between cell-free DNA abundance and disease state. The more severe the patient's cancer, the more cell-free DNA present in their blood (reviewed by de Miranda et al 2021 [61], Grabuschnig et al. 2020 [59], Aucamp et al. 2018 [3]), suggesting that disease burden impacts cell-free DNA homeostasis. Later work has shown that ctDNA levels, independent of non-tumor, cell-free DNA, vary dramatically across disease type, tumor location, and stage [12, 35, 62-64]. *In vitro* assays have helped isolate the mechanistic variables involved in cell-free DNA release,

particularly with regard to known apoptotic and necrotic processes. These experiments found that cell-free DNA release can vary significantly between cell-lines with different phenotypes and histologies [65-70].

Observations in human cohorts also correlate ctDNA with tumor histology, grade, and stage. Various studies in neoadjuvantly-treated breast cancer patients found that ctDNA levels and mutations were significantly different between breast cancer subtypes [17, 46, 71, 72]. Expression levels of the proliferation-associated nuclear protein, Ki-67, have also been directly associated with ctDNA characteristics in breast and lung cancers [73-75]. Studies assessing ctDNA in non-small cell lung cancer (NSCLC) patients found that ctDNA concentration was correlated with tumor stage, histology, and degree of cytological atypia [26, 76, 77].

2.2.2 Tumor volume, growth rate, and metabolism

Cell death has historically been considered the largest contributor to ctDNA and tumor volume the most reliable predictor of ctDNA abundance. More recently, various mechanisms and conditions have been proposed in which living tumor cells, particularly during mitosis, could shed DNA in both an energy-dependent and independent manner [59, 65]. Therefore, it is likely the complex interplay between tumor cell proliferation and death that determines ctDNA abundance. This relationship has significant implications for how we should think about ctDNA measurements in the context of tumor volume, growth rate, and metabolism. For example, one can easily imagine a scenario where proliferation and death rates both increase but are in balance resulting in increased ctDNA shedding but no net change in tumor volume [18]. Furthermore, although ctDNA may be hypothetically representative of the entire cancer cell population, it is likely subject to significant composition bias from differential cell turnover rates across

subclones [41]. Examining these variables *in vitro* and *in vivo* can shed light on which processes contribute more to ctDNA abundance.

Tumor growth rate and metabolism are often inferred by measuring tumor glucose uptake. Studies in metastatic melanoma patients found a strong correlation between the tumor PET avidity (a measure of cellular glucose uptake) and ctDNA abundance, independent of tumor volume [78]. These results are supported by studies in resected NSCLC, where a correlation was found between increased mitotic rates and higher ctDNA levels measured 24 hours prior to surgery, as well as increased levels of the proliferation marker Ki-67 [73-75, 77]. Indeed, *in vitro* studies have consistently found that large amounts of cell-free DNA can accumulate in the media of actively proliferating cell populations independent of apoptosis or necrosis [65, 79-82]. DNA fragments resulting from mis-segregation events during mitosis were found to be released by actively proliferating cancer cells via the creation of micronuclei [83, 84], but their relative contribution to overall ctDNA abundance *in vivo* remains unclear. Similar to the DNA products of necrosis, it has been assumed that these fragments would appear distinct from apoptotic ctDNA given their larger size and that their contribution would therefore be obvious. However, evidence is emerging that cleavage of larger DNA fragments by extracellular DNases may also occur [80]. These DNA fragments might have an apoptotic fragmentation pattern, yet be generated by non-apoptotic mechanisms, such as release during proliferative states.

The relationship between tumor cell proliferation and death is not independent. In healthy tissues, cell density homeostasis is achieved by compensating for cell death with an appropriate rate of cell proliferation. This process is known as “apoptosis-induced proliferation” (reviewed by Heitzer et al. 2020 [40] and Ryoo et al. 2012 [85]), but it is unclear how significant a role it plays in tumors. The consequences for ctDNA shedding could be straightforward, where increased cell death leads to increase cell birth and so

on, and both processes result in increased ctDNA levels. However, positive feedback mechanisms like this may be tissue-dependent and could be dysregulated in cancer, complicating interpretation of ctDNA dynamics.

2.2.3 Tumor vasculature, blood vessel proximity, and hypoxia

Tumor vascularization and proximity to major blood vessels are also features of tumor physiology that might be expected to significantly impact ctDNA levels. Blood flow to a tumor is the direct means by which ctDNA enters circulation and it affects the metabolic activity of a tumor by providing oxygen and nutrients [86]. As a tumor grows its vasculature becomes more irregular and dysfunctional leading to reduced oxygen levels, hypoxia and necrosis. Necrosis, reduced nutrient levels, and limited accesses to wider blood circulation all potentially effect ctDNA abundance in unique ways. Vasculature can also impact drug delivery and efficacy, which may also affect ctDNA shedding. It is unclear how much ctDNA abundance is dependent on direct access to blood vessels. Interstitial ctDNA is assumed to passively enter circulation through nearby blood vessels, but other processes like macrophage clearance of dead and dying cells (see "Immune Response" and "Immunotherapy" sections below) may also play a role in transporting ctDNA from areas with poor vasculature to the bloodstream [45, 58, 60].

Results from studies directly comparing tumor vascularization, angiogenesis, and ctDNA abundance, are inconsistent between studies. Post-excision pathology by Abbosh et al. [27] found lymphovascular invasion to be predictive of ctDNA detection in early-stage lung cancer. Two other studies in lung cancer found vascular invasion to be marginally [77], or not at all [76] correlated with ctDNA detection. Interestingly, when only looking at patients with EGFR mutations, ctDNA was significantly correlated with

vascular invasion in the former study by Cho and colleagues [77]. In liver cancer, microvascular invasion was correlated with preoperative ctDNA levels [87]. In recent preliminary data collected from a large cohort of colorectal cancer (CRC) patients, ctDNA was found to be strongly associated with lymphovascular invasion [88]. In neurological malignancies, which typically have less detectable ctDNA, Nabavizadeh and colleagues [89] found that tumor vessel size was correlated with detectable ctDNA. Notably, previous work by the same group and others found that the amount of microvascular proliferation was not significantly correlated to ctDNA in glioblastoma (GBM) specifically [90, 91]. The nature of such studies makes it challenging to discern if these correlations are independent of tumor stage and volume. Proving a causal link may only be possible with further evaluation of preclinical models, tumor pathology, and imaging.

Hypoxia is in many ways a measure of tumor cell access to functional vasculature [92]. As a tumor grows, cells become more isolated from functional vasculature, despite upregulated angiogenesis that is characteristic of many cancers. This process selects for cells that are more tolerant of low-oxygen conditions while the remaining population become necrotic [93]. There is a clear link between hypoxia and necrosis and some studies have suggested that ctDNA is primarily derived from necrotic processes [39, 66, 94, 95]. This suggests that as a tumor grows and vasculature becomes more distant and dysfunctional, wider ctDNA abundance could either increase due to further necrosis, or decrease due to reduced access to that vasculature. Since both forces are not equal in all tumors, the overall effect on ctDNA levels from this process may not be neutral. *In vitro* experiments with CRC cells have found that hypoxic conditions induced cell-free DNA production during the first 24 hours but decreased dramatically over the following 48-72 hours [68]. These results are also consistent with previous findings in both tumor-injected and tumor-free mice where hypoxia induced cell-

free DNA release [96]. Deprivation of the metabolite, folate, has been found to induced double-strand DNA breaks and mis-segregation events, which may also lead to ctDNA shedding in nutrient-starved tumors as well [97].

2.2.4 Organ encapsulation

The free movement of cell-free DNA between tissue and blood may be restricted in some organs. Blood-tissue barriers have been identified throughout the body, such as the thymus, testes, retina, and intestines, but it is unclear what role they might play in cell-free DNA exchange [98]. The blood-brain barrier (BBB) is often cited as the primary reason that neurological malignancies, particularly gliomas, produce less detectable ctDNA than other cancer types [35]. In a 2018 review on ctDNA kinetics, Khier and Lohan speculate that physiological barriers, like the BBB, restrict the movement of cell-free DNA throughout the body while also acknowledging the exception of placental cell-free DNA, which has been shown to move quite freely throughout the mother [45, 99]. Notably, disruption of the BBB that results in increased permeability and risk of metastasis also resulted in increased levels of ctDNA in patients with GBM [89]. Several studies have shown that disrupting the BBB in animal models using focused ultrasound techniques leads to increase blood levels of ctDNA and other biomarkers [100, 101]. Therefore, it is possible that changes in tissue-blood barrier permeability, particularly during treatment, might significantly affect ctDNA dynamics.

2.2.5 Immune response

Although there may be a significant role for inflammation and infection (*e.g.*, sepsis) in cell-free DNA release, this review is primarily interested in the extent to which they directly impact ctDNA release from cancer cells. Early studies exploring the origins of cell-free DNA found that macrophages may play a significant role in cell-free DNA

release though phagocytosis of dead and dying cells [102]. Phagocytes have been shown to digest apoptotic cells and release the resulting cell debris and fragmented DNA [3, 102, 103]. In the GBM study mentioned earlier, ctDNA levels were strongly associated with the density of macrophages around the tumor [89]. In healthy individuals cell turnover is a tightly regulated process where apoptotic cells are quickly removed by phagocytes, however, this process appears to be dysfunctional in tumors resulting in excess cell debris (including DNA) that accumulates locally and in circulation [60, 95]. The extent to which ctDNA levels might be directly affected by tumor cell targeting and/or clearance by immune cells is still an open question. Evidence for this phenomenon however, might be found in studies where ctDNA levels spike within 2 weeks of immune-therapy initiation in metastatic melanoma patients, if and only if, the tumors were responsive [104].

2.2.6 Cell-free DNA clearance

Cell-free DNA digestion and clearance, whether achieved locally via phagocytosis or in circulation via the liver, spleen, and kidneys, is influenced by a number of factors [45, 60, 105, 106]. As described above, cell-free DNA clearance *in situ* is potentially dependent on interstitial diffusion and the presence of phagocytic cells, however, once it is in circulation its half-life is determined by extracellular DNase activity and organ function [45]. ctDNA half-life in the blood ranges from 30 to 120 minutes [60] making blood collection timing critical. The decreased levels of DNase activity observed in the blood of cancer patients potentially explains the accompanying increase in cell-free DNA levels from disruption of homeostasis [107, 108]. Studies have also suggested that cell-free DNA clearance and half-life is dependent on proper liver and kidney function suggesting that treatment toxicity in cancer patients could affect ctDNA clearance rates and abundance [109, 110]. The role of renal function in ctDNA clearance

is not well understood, however, based on experiments assessing cell-free DNA levels in urine [60]. The presence of cell-free DNA in urine implies involvement of the kidneys in clearance from circulation, however, patients with chronic renal failure were found not to have increased levels of cell-free DNA in their plasma [111]. Methylation profiling has suggested that cell-free DNA present in urine is derived from white blood cells, kidney cells and urinary tract cells, but data from stem cell transplant patients found that the majority of this DNA was from the renal system itself and not plasma [112, 113].

2.3 The effect of treatment on ctDNA abundance

The effect of treatment on tumor cell proliferation and death, and thus ctDNA dynamics, is dependent on its mechanism of action, efficacy, and tumor biology. Considering the factors described above that influence ctDNA abundance, it is not surprising that there are many discernable differences between the ctDNA dynamics of responders and non-responders during treatment. Predicting their timing and trajectories is not so simple, particularly when considering the short half-life of ctDNA. We might at least expect that ctDNA dynamics should reflect treatment response depending on the mechanism of action of a given treatment, but the timing of those effects is still unclear (Fig. 2.2). The correlation between *in vitro* and *in vivo* models of treatment-induced cell death remains largely unclear, and is likely dependent on a variety of factors including the treatment and tissue of interest. Despite the paucity of data, it is possible that tumor cell death can occur within hours of treatment and therefore ctDNA levels may rapidly increase as well [67, 114-116]. Serial ctDNA monitoring has been done with collection times ranging from minutes, hours or days after administration of treatment, to weeks and months. Some guidance may be gleaned from these studies, but the optimal time for sampling may be unique to cancer type and therapy and may need to be determined

empirically. The following sections outline the expected effects of cancer treatment modalities on ctDNA dynamics, and what existing evidence, if any, tells about these hypotheses.

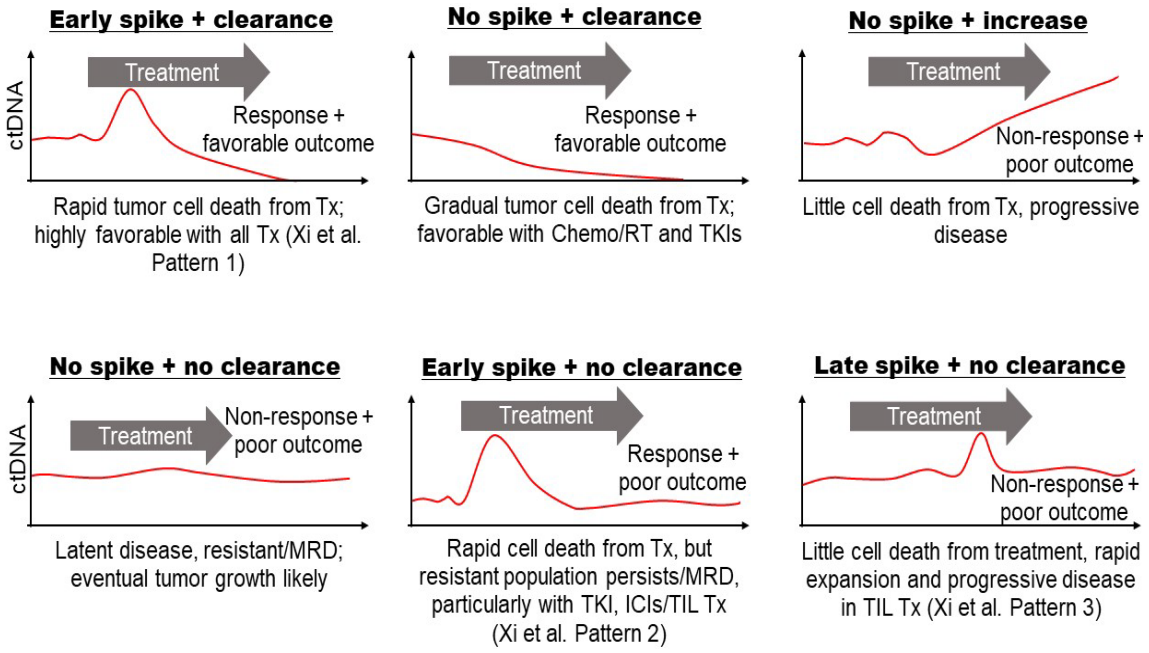


Figure 2.2. Hypothetical ctDNA dynamics from blood sampled frequently before, during, and after treatment.
Characteristic ctDNA dynamics are depicted as observed in various studies throughout the text or otherwise hypothesized. The possible driving tumor biology and outcomes are described for each plot. Response patterns defined by Xi et al. [49] for metastatic melanoma patients undergoing tumor infiltrating lymphocyte (TIL) immunotherapy are indicated when relevant. (Tx = treatment; TKI = tyrosine kinase inhibitors; ICI = immune checkpoint inhibitors; MRD = molecular/minimal residual disease; ChemoRT = chemoradiation therapy)

2.3.1 Chemotherapy and radiation

Cytotoxic chemotherapies and radiation therapy (RT) are often used independently or in combination as first-line treatment in many cancers. Chemotherapy

agents function by disrupting mitosis or causing DNA damage leading to cell-cycle arrest, mitotic catastrophe, and apoptosis. Radiation therapy kills cells by DNA damage as well, but it also elicits an immune response and vascular damage, which can result in subsequent rounds of tumor cell death. Mitotic arrests and DNA damage are thought to cause tumor cell death within 6 to 72 hours of administration *in vivo*, so detecting ctDNA shedding in response to effective treatment may require immediate sampling [116]. Unfortunately, very few studies sample ctDNA within the first 72 hours of chemotherapy. It has also been suggested that treatment-induced mitotic catastrophe can cause delays in cell death from chemotherapy and RT for up to a week [67, 91, 117]. These various mechanisms of action may result in multiple shedding events, where one may be more informative about treatment response over another.

Limited studies assessing ctDNA levels immediately after treatment are conflicting. In castration-resistant prostate cancer patients receiving docetaxel-based therapy, early ctDNA levels were found to increase rapidly within 1 hour of administration with a corresponding decrease in total cell-free DNA [118]. This observation is consistent with increased tumor cell sensitivity to cytotoxic agents compared to healthy tissue. Contrary to these findings, however, CRC patients receiving FOLFOX did not exhibit a spike in ctDNA at any point within the first 48 hours of treatment, despite high-resolution sampling at 3, 9, 18, 23, 26, 42, and 47 hours [119]. Another study involving metastatic CRC patients receiving FOLFIRI looked at ctDNA levels before and 7 days after each of the first two treatment cycles and again at progression [120]. Interestingly, this study found that temporary increases in ctDNA while on treatment were predictive of progressive disease and worse survival rates, and suggested ctDNA monitoring within the first week of treatment to evaluate treatment efficacy. While ctDNA levels were decreased at the time of radiological assessment compared to baseline for all patient, patients with temporary spikes in ctDNA appeared to have more sustained ctDNA

burden than those with favorable response (see Fig. 2.2. “Early spike + no clearance”). Clonal composition was also found to vary during treatment suggesting a differential response to treatment among tumor cell subpopulations. Increases in cell-free DNA methylation levels of the tumor suppressor genes, APC and RASSF1A, 24 hours after receiving cisplatin-based chemotherapy were correlated with improved tumor response and overall outcome in advanced lung cancer patients [121]. The same study showed that methylation levels of those genes in lung cancer cells also peaked 24 hours after cisplatin exposure, however, it is unclear if the hypermethylated DNA was tumor-derived in patients. Notably, this study also found that elevated methylation of APC and/or RASSF1A in tumor-bearing mice were associated with tumor cell death as determined by biopsy shortly after treatment and blood collection. This finding might suggest that the methylated cell-free DNA originated from these dying tumor cells. In pancreatic cancer patients sampled for 4 weeks following treatment with gemcitabine, decreases in ctDNA were correlated with tumor response [122]. In our work and others’, decreases in or complete clearance of ctDNA levels during low-resolution sampling (*i.e.,* weeks to months) of neoadjuvantly-treated breast cancer patients were associated with pathological complete response at the time of surgery [17, 31, 46].

In patient cohorts receiving combination chemoradiation therapy (CRT), decrease and clearance of ctDNA after 3 to 4 weeks was associated with tumor response in oropharyngeal and lung cancer [123, 124]. Studies employing high-frequency sampling at time points within hours of treatment are sparse, however, recent data from Breadner et al. [125] found that ctDNA abundance increased in 77% of stage III/IV non-small cell lung cancer patients shortly after receiving CRT with peak abundance observed 7 hours after chemotherapy initiation and 2 days after the first fraction of radiation. Early spikes in ctDNA were seen in some patients receiving CRT for treatment of locally advanced head and neck cancer, but were not correlated with

response [51]. Rather, overall decreases in ctDNA at later time points, which were not unique to patients who had early peaks, were more predictive of outcome.

Some of the earliest observations of cell-free DNA by Leon and colleagues [126] occurred in patients receiving RT alone, finding that general cell-free DNA levels decreased after treatment. Aucamp et al. [3] speculate that the reason for this may have been the coincident destruction of phagocytes needed to generate cell-free DNA. Another consideration that may confound ctDNA measurement from irradiated tumors is that, although mitotic arrest and catastrophe are the primary means by which RT is thought to kill cancer cells, they have also been found to result in mis-segregation events [127, 128]. As discussed above, this process can result in releasing of DNA from living cells, providing another potential source of ctDNA that does not coincide with cell death. A recent study found that irradiation of head and neck cancer and NSCLC cell lines induced cell-free DNA shedding after 6 to 24 hours in culture [67]. The same study found that ctDNA levels increased within 24 hours and peaked 96 to 144 hours after 20Gy of irradiation in xenograft mouse models. Interestingly, the authors found that treatment-induced senescence that was overcome with the senolytic drug, Navitoclax, lead to apoptosis and increased ctDNA release. In human subjects, investigation of RT alone in NSCLC has shown mixed results. Walls and colleagues [129] found that 3 of 5 patients had decreased ctDNA levels 3-days after their first RT fraction, while the remaining 2 had increased levels of some tumor-derived variants, but not others. Preliminary data from our lab [34] and another study from Chen et al. [130] found that ctDNA levels were elevated 24-48 hours after the first dose of stereotactic ablative radiotherapy. These three studies in NSCLC patients used varying doses of radiation per fraction (2.75Gy, 12Gy, and ~12.5Gy, respectively), which along with sampling time differences, may account for the discrepancy. Chaudhuri and colleagues [131] also reported that mid-RT ctDNA levels in 13 NSCLC patients were correlated with outcomes at 2 years. A recently

published study by our lab found that a metastatic breast cancer patient had increased ctDNA levels while receiving palliative radiation therapy. Deep sequencing using a 53-mutation panel representative of both clonal and subclonal mutations, which were previously identified from WES of multiple tumors, revealed differential response in ctDNA levels for various subclones with sample collection every 48-hours during RT [32]. Differential response in subclonal ctDNA abundance is suggestive of varying sensitivity in irradiated tumor cell subpopulations and/or an abscopal response. First observed by Dr. R.H. Mole in 1953, the abscopal effect is the shrinkage of a distant, untreated tumor in response to RT of another tumor. It is thought that the destruction of cells in the irradiated tumor elicits an immune response that affects the non-radiated tumors elsewhere in the body [132]. In our case, radiation of a single lesion may have induced immune-mediated responses and ctDNA shedding from distant metastatic sites that harbored subclonal tumor cell populations. It is also worth noting that CTCs may also be sources of ctDNA and CTC release timing may also be similar to tumor ctDNA shedding [133]. In preliminary data in head and neck cancer patients treated with RT, 3 of 11 patients had increased circulating tumor cell (CTC) counts after the first fraction of RT, and 5 of 6 patients had increases CTCs after 2 weeks into therapy [134].

2.3.2 Immunotherapy

ctDNA monitoring during treatment with immune checkpoint inhibitors (ICIs) has shown promise in multiple cancer types. A pan-cancer analysis done by Zhang and colleagues [52] found that changes in ctDNA levels during ICI treatment may be predictive of benefit. Patients in this study with increases in ctDNA levels during treatment had worse outcomes as compared to those that did not. Furthermore, patients with ctDNA clearance after detectable pre-treatment levels had the best progression-free and overall survival. A recent study by Herbretreau and colleagues [104] in patients with

metastatic melanoma found that significant increases in ctDNA levels during the first 2 to 4 weeks of anti-PD1 (with or without anti-CTLA4) allowed early and highly-specific identification of treatment-resistant patients. Furthermore, ctDNA levels that rapidly decreased after starting PD-1 inhibitors were highly predictive of responses consistent with pseudoprogression [135, 136]. When compared to changes in ctDNA levels later in treatment, regardless of early changes, increases beyond 12 weeks were not necessarily predictive of non-response, further suggesting that early sampling is more informative [137]. In NSCLC patients treated with ICI, decreases in ctDNA 2 weeks after treatment initiation were strongly correlated with radiographic response and progression-free survival [138]. A study investigating early response to tumor infiltrating lymphocyte (TIL) immunotherapy in metastatic melanoma patients identified three patterns of ctDNA dynamics that could be used to stratify patients by overall survival [49]. Patients with an early spike in ctDNA within 5-10 days of treatment followed by clearance showed a statistically significant survival outcome over patients who had early peaks but latent ctDNA burden, or no peaks with or without clearing (see Fig. 2.2). The study's authors speculate that early spike it ctDNA was in part due to the newly-transferred lymphocytes "identifying their targets and are effective in killing [them]."

2.3.3 Targeted therapy

Given the clinical implications of tumor heterogeneity, one of the most significant unanswered questions in ctDNA analysis is whether ctDNA observed during therapy is more representative of resistant or responsive tumor cell populations. In a study of lung cancer patients undergoing EGFR tyrosine kinase inhibitor (TKI) therapy, ctDNA sampled 2 weeks after treatment initiation revealed activating mutations not previously detected in the tumor biopsies [139]. Another study of lung cancer patients on TKIs found that clearing of ctDNA within days of treatment was associated with response,

whereas sudden increases in ctDNA load later in treatment correlated with rapid tumor progression and poor outcome [140]. In lung cancer patients receiving either anti-EGFR or HER2 therapies, increases in ctDNA abundance were observed within 4-12 hours after initiation of treatment while total cell-free DNA was relatively constant [115]. Phallen and coauthors point out that this timeframe is consistent with other studies in which apoptosis is observed in vitro within 6 to 48 hours of treatment with EGFR TKIs. Patients in this study with an initial radiographic response all had ctDNA abundance eventually decrease by more than 95% within the first 19 days of treatment. Interestingly, baseline levels in a study of ALK-fusion positive lung cancer patients, pre-treatment ctDNA levels were not correlated with treatment response, yet changes in ctDNA during treatment with ALK TKIs were associated with progression [141].

Increases in ctDNA abundance of therapy-sensitive clones corresponds with response, however, increases in ctDNA abundance of therapy-resistant clones can also portend clinical progression [40]. Outgrowth of subclonal tumor cell populations that are resistant to targeted therapy can be directly observed in allele-specific ctDNA dynamics. For example, genomic changes conferring resistance to targeted therapy in prostate cancer patients were detected by increasing fractions of the resistance-associated allele in several studies [142, 143].

2.4 Discussion

Host physiology and tumor biology affect ctDNA abundance while changes in ctDNA levels during treatment may indicate disease response. Cancer type and stage appear to have the most dramatic impact on ctDNA abundance, and significant decreases in, or clearance of ctDNA early in treatment seems to be predictive of response and improved outcomes. The association of treatment response and overall

decreases in ctDNA levels during treatment is consistent with the hypothesis that tumor burden and tumor growth rate are reflected in ctDNA dynamics. Very early changes (1 to 3 hours) in ctDNA levels have been hypothesized to reflect treatment response as well, but this appears to be less generalizable. For example, we might expect effective chemotherapy to induce ctDNA shedding immediately, and early, high-frequency sampling to detect it, yet observations between NSCLC, CRC and prostate cancer patients sampled within 1-3 hours of chemotherapy were inconsistent [118, 119, 144]. Unfortunately, there are a lack of studies sampling within this timeframe. Immunotherapy may be less fast-acting than chemotherapy given the time required for the body to prepare a successful immune response. PET/CT imaging has shown tumor responses with 4-6 weeks of treatment with ICIs in melanoma patients [145]. Sample collection at 2 weeks following treatment found changes in ctDNA that correlated with outcome in ICI therapy of melanoma patients, but earlier time points were not collected [104]. It possible that changes in ctDNA in response to treatment existed sooner, again, earlier, high-frequency sampling is needed to test such hypotheses. TKI-induced cell death appears to occur within 6 to 48 hours of exposure *in vitro* [114]; a similar timeframe as cell death from cytotoxic agents. Evidence presented in this review suggests that ctDNA dynamics might reflect TKI-induced cell death in this timeframe in NSCLC patients more consistently than during chemotherapy.

One potential use of ctDNA monitoring during treatment that has been explored by our lab and others [34, 125], is to induce ctDNA shedding from either inaccessible tumors or suspicious lesions for evaluation. Radiation treatment seems particularly suited for this task, however any method of perturbation that elicits ctDNA shedding could be used. For example, such approaches could improve the detection rates of ctDNA assays like CAPP-Seq and Lung-CLiP [26] in lung cancer patients or where low-dose CT is already in use for screening high-risk populations. Compression of breast

tissue during mammography has been shown to temporarily increase ctDNA abundance, which could be leveraged for minimally-invasive biopsy or early detection [75]. Other work has explored the use of ultrasound to elicit better movement of blood biomarkers across the BBB in preclinical brain tumor models [100, 101]. Again, a clear understanding of early ctDNA dynamics in response to tumor perturbations is crucial before such approaches can be implemented in the clinic.

Variability in ctDNA measurements between patients and studies has been a challenge for serial monitoring. Efforts have been made to assess the biological variability of both cell-free DNA and ctDNA between measurements taken over short intervals [12, 13]. Several groups have attempted to standardize criteria for evaluating differences between pre-treatment and on-treatment ctDNA levels. O'Leary and colleagues [146] created a "circulating tumor DNA ratio" or CDR, defined simply as the ratio of on-treatment to pre-treatment levels, to evaluate treatment response in metastatic breast cancer. Herbreteau, Kruger and colleagues [104, 122] defined a "quantitative biological response and progression criteria" where patients were stratified by increases or decreases in ctDNA during treatment as compared to baseline. This approach also recognized the variability in accuracy at each time point when evaluating significance between measurements at different time points. Out of similar concern, our lab has also developed a Bayesian approach for testing statistical significance between ctDNA mutations in NGS data from serial collections [31]. Given the variability in coverage of NGS data, which determines the limit of detection, the allele-specific background error rate, and the stochastic nature of mutant read detection by NGS, sophisticated methods may be required to account for these uncertainties in ctDNA evaluation.

Finally, attempts to model ctDNA shedding have shown promise in predicting ctDNA abundance based on tumor size and growth rate. Avanzini and colleagues [18]

presented a stochastic mathematical framework based on observations in lung cancer patients which could extrapolate ctDNA copy counts using tumor cell proliferation rate, death rate, shedding probability, clearance rate, and starting tumor volume. Such modeling can reveal unexpected behavior that may be informative of real-world scenarios. For example, in simulations the authors unexpectedly found that a slow growing cancer generated more ctDNA molecules than a faster-growing cancer of the same size when the faster growth was achieved by proportional increases and decreases in birth and death rates, respectively. However, if a faster growth rate is achieved by increased birth rate and a stable death rate, the difference in ctDNA release was negligible. By integrating these variables over time, similar models might be useful in predicting changes in ctDNA abundance from changes in birth and death rates resulting from treatment.

2.5 Conclusion

The utilization of ctDNA in assessing treatment response will require a better understanding of the biological factors involved. We believe that ctDNA monitoring has the potential to truly revolutionize personal medicine in cancer care but there remain significant challenges that must first be overcome.

Chapter III: The Feasibility of Patient-Specific ctDNA Monitoring for Subclinical Disease in Esophageal and Rectal Cancer†

Christopher Boniface[1,2], Christopher Deig[3], Carol Halsey[4], Taylor Kelley[4], Michael B. Heskett[4], Charles R. Thomas Jr.[3], Paul T. Spellman[1,4] and Nima Nabavizadeh[3*]

[1] Cancer Early Detection Advanced Research (CEDAR) Center, Division of Oncological Sciences, Knight Cancer Institute, Oregon Health & Science University (OHSU), 2720 SW Moody Ave., Portland, OR 97201, USA

[2] Department of Biomedical Engineering, Oregon Health & Science University (OHSU) School of Medicine, 3181 SW Sam Jackson Park Rd, Portland, OR 97239, USA

[3] Department of Radiation Medicine, Oregon Health & Science University (OHSU), 3181 SW Sam Jackson Park Rd, KPV4, Portland, OR 97239, USA

[4] Department of Molecular and Medical Genetics, Oregon Health & Science University (OHSU) School of Medicine, 3181 SW Sam Jackson Park Rd, Portland, OR 97239, USA

* Correspondence: nabaviza@ohsu.edu; Tel.: +1-503-494-8756

Adapted from Boniface et al. 2021 [31].

† Originally published with the title "The Feasibility of Patient-Specific Circulating Tumor DNA Monitoring throughout Multi-Modality Therapy for Locally Advanced Esophageal and Rectal Cancer: A Potential Biomarker for Early Detection of Subclinical Disease"

Abstract

As non-operative management (NOM) of esophageal and rectal cancer is becoming more prevalent, blood-biomarkers such as circulating tumor DNA (ctDNA) may provide clinical information in addition to endoscopy and imaging to aid in treatment decisions following chemotherapy and radiation therapy. In this feasibility study, we prospectively collected plasma samples from locally advanced esophageal (n = 3) and rectal cancer (n = 2) patients undergoing multimodal neoadjuvant therapy to assess the feasibility of serial ctDNA monitoring throughout neoadjuvant therapy. Using the Dual-Index Degenerate Adaptor-Sequencing (DIDA-Seq) error-correction method, we serially interrogated plasma cell-free DNA at 28–41 tumor-specific genomic loci throughout therapy and in surveillance with an average limit of detection of 0.016% mutant allele frequency. In both rectal cancer patients, ctDNA levels were persistently elevated following total neoadjuvant therapy with eventual detection of clinical recurrence prior to salvage surgery. Among the esophageal cancer patients, ctDNA levels closely correlated with tumor burden throughout and following neoadjuvant therapy, which was associated with a pathologic complete response in one patient. In this feasibility study, patient- and tumor-specific ctDNA levels correlated with clinical outcomes throughout multimodality therapy suggesting that serial monitoring of patient ctDNA has the potential to serve as a highly sensitive and specific biomarker to risk-stratify esophageal and rectal cancer patients eligible for NOM. Further prospective investigation is warranted.

3.1 Introduction

As non-operative management (NOM) of locally-advanced esophageal and rectal cancer following chemotherapy and radiation therapy is more widely adopted [147, 148], a sensitive and specific biomarker of sub-clinical tumor burden has the potential of further assisting in the assessment of patients best suited for an upfront non-operative approach or for detecting early sub-clinical recurrences in patients in need of salvage surgery in the surveillance period [149]. Circulating tumor DNA (ctDNA) has been extensively investigated for its diagnostic and prognostic utility as such a biomarker [41, 150-153]; however, the longitudinal application of ctDNA monitoring throughout multi-modality therapy (systemic therapy, radiotherapy, and/or surgery) has been less extensively studied [150]. Undetectable ctDNA following definitive treatment has been shown to be associated with pathologic complete response (pCR) and improved outcomes, particularly in the neoadjuvant setting for breast cancer [154, 155]; however, the feasibility of serial ctDNA measurements in the neoadjuvant setting for rectal and esophageal cancer has not been previously reported.

Our ctDNA monitoring technique, called DIDA-Seq (dual-indexed, degenerate adaptor-sequencing), combines unique-molecular indexing (UMI)-based error correction with custom hybridization capture at many genomic loci of somatic variants previously identified by whole-exome sequencing of the patient's tumor tissue. This method allows the detection of ctDNA in the blood with very high accuracy (1 error in 10k-50k observations) and sensitivity (0.005%-0.02% minimum variant allele frequency [17]). This study aims to test the feasibility of using patient and tumor-specific ctDNA monitoring to assess residual disease in five esophageal and rectal cancer patients during and after multimodal neoadjuvant therapy (Fig. 3.1).

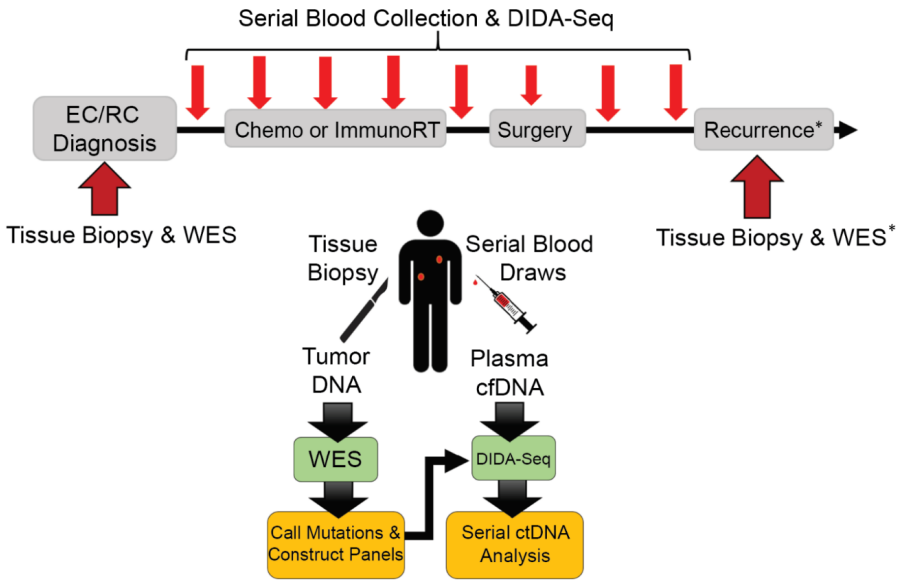


Figure 3.1. Patient treatment and sample collection schema for blood draws and solid tissue biopsies.
Solid tissue biopsies were collected after initial diagnosis of esophageal or rectal cancer (ER/RC) and prior to treatment for whole-exome sequencing (WES). Blood was collected prior to treatment and then at ~1-month intervals during treatment and surgery, and ~3-month intervals during follow-up monitoring. Mutations were called between solid tissue biopsy WES and matched buffy coat WES and used to construct patient-specific sequencing library enrichment panels. Cell-free DNA (cfDNA) isolated from blood draws was sequenced using Dual-Index Degenerate Adaptor-Sequencing (DIDA-Seq) at sites identified in each patient's tumor biopsy to retrospectively determine circulating-tumor DNA (ctDNA) prevalence. *Patient 3 had a solid tissue biopsy of a metastasis which was also analyzed by WES and variants were included in their patient-specific panel.

3.2 Materials and Methods

3.2.1 Patient Enrollment, Tissue Processing, and DNA Extraction

In this feasibility study, human specimens and data (including blood, tumor tissue, and clinical information) were prospectively acquired from participants with

locally-advanced esophageal (*n* = 3) or rectal cancer (*n* = 2) undergoing definitive multimodal therapy after their informed written consent (Oregon Health & Science University, IRB# 10163). Plasma samples were collected at baseline, throughout therapy and surveillance. Biopsy tissue was collected at diagnosis and recurrence (Fig. 3.1). Blood draws were serially collected and fractionated for cfDNA isolation using the "double spin" method (≤40 mL, a range of 6–40 mL, in 5 × 6 mL or 4 × 10 mL, purple-capped EDTA tubes) [17, 41]. DNA was extracted from FFPE, plasma, and buffy coat using commercially available kits (see below). Within 6 h of collection, blood plasma was isolated by first spinning whole blood at 1000 g for 10 min, separating the top plasma layer into 1-mL aliquots, then spinning those aliquots at 15,000 xg for 10 min, transferring the supernatant to cryovials, and storing at −80 °C. Fixed formalin paraffin-embedded biopsies and tumor-tissue were collected, and DNA extraction was carried out using QIAgen FFPE DNA extraction kit (QIAGEN). DNA was extracted from plasma and buffy coat using Macherey-Nagel NucleoSnap and QIAgen Blood and Tissue kits, respectively. All DNA extractions were quantified using the Qubit 3 fluorometric quantification system (ThermoFisher Scientific) and size distribution was checked with a BioAnalyzer 2100 (Agilent Technologies). DNA isolated from FFPE samples and buffy coat were fragmented by sonication to 150bp using a Covaris E220 prior to library preparation (cfDNA was not fragmented prior to ligation).

3.2.2 Whole-Exome Sequencing Library Preparation

Whole-exome sequencing (WES) libraries were prepared from tissue biopsies using 100–500 ng of sonicated FFPE of buffy coat DNA and the KAPA Hyper-Prep Kit (KAPA Biosystems) with the Agilent SureSelect XT Target Enrichment System and Human All Exon V5 capture baits (Agilent Technologies). Next generation sequencing was carried out using the Illumina NextSeq 500 platform by our institution's Massively

Parallel Sequencing Shared Resource to an average, de-duplicated depth of 329X and 121X for tumor and buffycoat matched-normal libraries, respectively (Appendix A Table 5.1).

3.2.3 Somatic Mutation Calling and Design of Tumor-Specific Capture Panels

FastQ data files were aligned and processed using BWA MEM (0.7.12, GATK, Broad Institute). Somatic variants were called using aligned BAM files and MuTect (1.1.4, GATK, Broad Institute, MA) between tumor and the patient's matched normal from blood buffy coat [156]. All WES BAM files can be found in the Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) under the BioProject accession number PRJNA637431 [157]. Single nucleotide variant (SNV) calls were filtered out if they were present in the dbSNP database (www.ncbi.nlm.nih.gov/projects/SNP). SNVs were filtered by frequency (requiring >1% mutant allele frequency and >3 supporting reads in the tumor, and <2% mutant allele frequency in the matched normal) and depth (requiring ≥30X coverage in the tumor and ≥14X coverage in the matched normal) and were further assessed and hand-curated using Oncotator [158] and IGV [159] software. For tumor-specific capture targets, approximately 50 SNVs were chosen for each patient based on inferred clonality, sequence context, and potential functional impact. To address concerns over properly representing cell subpopulations, intronic mutations were included in each panel. Tumor-specific hybrid capture panels were constructed by querying the human reference genome (GRCh37/hg19) for the 120bp surrounding the target loci of interest. The resulting nucleotide sequences were submitted to IDT DNA to generate biotinylated bait oligos using the NGS Discovery Pools tool

(https://www.idtdna.com/). Mutation sites and bait oligo sequences are described in Boniface et al. (2021) Supplementary Table S1 [31].

3.2.4 DIDA-Seq Library Preparation and Sequencing

DIDA-Seq error-correction libraries were prepared similarly to what is previously described and sequenced on Illumina platforms. Briefly, 30–100 ng of cell-free DNA was input into the Kapa Biosystems Hyper Prep kit with custom DIDA-Seq adaptors followed by hybridization capture using the IDT xGen Hybridization and Wash Kit using a single, 18 h capture incubation step instead of the double-incubation steps previously described [17, 41]. Libraries were sequenced on either the Illumina HiSeq 2500, paired-end 100 bp, with dual 14-bp indexing cycles or the Illumina NextSeq 500, paired-end 70 bp with dual 14-bp indexing cycles. All DIDA-Seq BAM files can be found in the Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) under the BioProject accession number PRJNA637431 [157].

3.2.5 Evaluation of Tumor-Specific Capture Panel Performance and ctDNA Prevalence

The error-correction pipeline for analyzing DIDA-Seq data was based on the duplex sequencing pipeline with substantial modification to be compatible with our data [30]. The DIDA-Seq computational pipeline was implemented as previously described [17] and the mutant allele frequency (MAF) was determined for each mutation at each time point by dividing the number of mutant error-corrected (i.e., consensus) reads by the total number of consensus reads at that site and multiplying by 100 (note that all MAFs are reported as percentages in Appendix A Tables). The aggregate MAFs for each time point were calculated by summing the mutant consensus reads at all sites

interrogated, dividing that by the total number of consensus reads across all sites and multiplying by 100. Each hybrid capture panel was evaluated using unrelated patient cfDNA samples as negative controls. We sequenced each patient time-point library to a mean, consensus read depth of 5.2kX (range = 159X to 23.4kX) per site-of-interest. We sequenced each negative control library to an average per-site consensus read depth of 43.6kX coverage (range = 3.9kX–127kX) with an average per-site error rate of 0.0067% or 1 error in 15k site-of-interest observations (range = 1 error in 2.7k to 125k site-of-interest observations) providing an average limit of detection of 0.016% MAF (i.e., the mean of the lowest statistically significant MAF from Patients 2–5, see Appendix A Table 5.3). When we aggregated negative control site-of-interest consensus read counts for each panel, we calculated an average per-panel error rate of 0.0057%, or 1 error in 17.7k observations (bases) with a range of 1 in 12.5k to 22.9k based on the assumption that mutant consensus reads found in the negative control were caused by PCR or sequencer error (see Appendix A Table 5.2 for panel-specific error rates). We compared the mutation-specific MAF in the patient's plasma at each time-point to the MAF of the same site in the set of pooled negative controls using the Weitzman overlapping coefficient [160] (see Section 3.2.6 below). A *p*-value was generated for each site, as well as all sites aggregated by tumor-specific panel, using the overlap coefficient between the beta distributions of the sample and the negative control read counts as described below. Any individual site with greater than 0.05% MAF in the negative controls was omitted from evaluation of ctDNA levels in the respective target patient. Data points having a *p*-value of 0.05 or less were considered significantly different from the negative controls, effectively determining our lower limit of detection given the total sequencing depth at each time point. To correct for differences in cell-free DNA concentration between blood draws, the aggregate MAF was converted into haploid genome equivalents per mL (hGE/mL) of plasma by the following equation:

$$Mutant\ genomes\ per\ mL\ plasma\ \ or\ \ \frac{mutant\ hGE}{mL\ plasma} =$$

$$\frac{cfDNA\ concentration\left(\frac{ng}{ml\ plasma}\right)}{0.003\left(\frac{ng}{genome}\right)} * mutant\ allele\ frequency$$

3.2.6 Significance Tests for ctDNA Measurements

The significance of ctDNA measurements (i.e., mutant consensus reads) at each time point, as compared to the panel's negative control, was determined prior to conversion to haploid genome equivalents per ml (hGE/mL) plasma and is dependent on the sequencing depth at each site at that time point. A Bayesian approach was used to test the null-hypothesis that the sample MAF and negative control MAF were generated from the same distribution. This statistical approach was used because we assumed that a higher sample size (i.e., deeper sequencing) confers a more accurate parameter estimate (i.e., 100 mutant reads in 100,000 is more accurate than 1 in 100). Therefore, a Beta distribution was created for the sample and for the negative control (Eqs. (A1) and (A2)), setting the "a" and "b" parameter values to the number of variant reads and number of reference reads, respectively. Next, the Weitzman overlapping coefficient [160] (Eq. (A3)) was used to measure the similarity between the sample and negative control distributions to create a significance value. In cases where the number of mutant consensus reads was greater than zero but the estimated *p*-value was also greater than 0.05, we determined the minimum number of mutant consensus reads (given the number of total consensus reads), for which the cumulative binomial distribution is greater than or equal to the error rate of the given sites as determined by the negative control (error-rate = mutant consensus reads in negative control/total consensus reads in the negative control). If the observed number of mutant consensus reads exceeded this

value, we considered it to be marginally significant and therefore above the limit of detection (see Appendix A Table 5.3 for *p*-values and binomial test results for each panel at each time point). Note that the overlapping coefficient method can result in low *p*-values (<<0.05) if the sample MAF ≈ negative control MAF and the depth of the negative control is much greater (>100-fold) than the depth of the sample. In such cases, the ctDNA measurement was considered below the limit of detection if the MAF of the sample was equal to or less than that of the negative control (e.g., Appendix A Table 5.3, Patient 2, month "0.0").

$$X_{sample} \sim Beta\left(a_{sample}, b_{sample}\right) \text{ (A1)}$$

$$X_{neg_{ctrl}} \sim Beta\left(a_{neg_{ctrl}}, b_{neg_{ctrl}}\right) \text{ (A2)}$$

$$\int \min\left[f_{neg_{ctrl}}(x), f_{sample}(x)\right] dx \text{ (A3)}$$

3.3 Results

3.3.1 Elevated ctDNA Levels are Associated with Recurrence in Rectal Adenocarcinoma with Clinically-Useful Lead Time

Patient 1 is a 33-year-old female who presented with cT3N1M0 distal rectal adenocarcinoma and enrolled on an unrelated phase II trial evaluating the efficacy of total neoadjuvant therapy (eight cycles of FOLFOX chemotherapy and long-course chemoradiation) followed by non-operative management for clinical complete responders based on MRI and endoscopy (NCT02008656) [157]. Whole exome sequencing (WES) of a pretreatment tissue biopsy revealed 81 non-synonymous single nucleotide variants (SNVs, Boniface et al. (2021) Supplementary Table S1 [31].).

DIDA-Seq of 28 loci was used to monitor ctDNA levels throughout the patient's treatment course. ctDNA levels decreased fivefold during four months of total neoadjuvant therapy (Fig. 3.2A). She was without clinically detectable disease for six months following total neoadjuvant therapy and proceeded with NOM per the trial; however, ctDNA levels remained elevated. Eleven months following total neoadjuvant therapy, endoscopic surveillance revealed a biopsy-confirmed recurrence and the patient underwent salvage total mesorectal excision (TME).

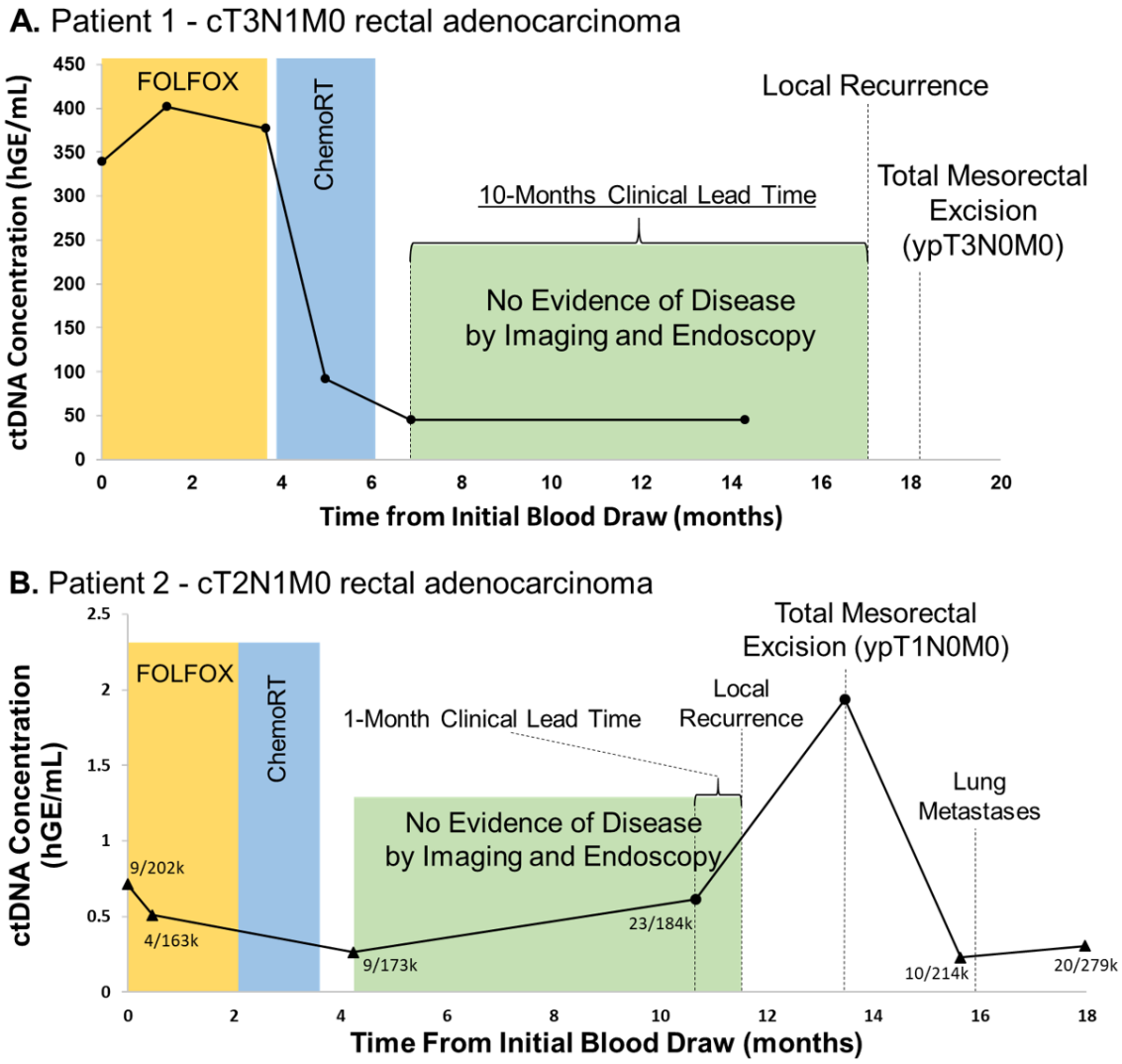


Figure 3.2. Rectal cancer patients with detectable post-treatment ctDNA eventually had local recurrence.

Serial ctDNA levels were retrospectively analyzed using DIDA-Seq and using patient-specific capture panels. Aggregate mutant allele frequency (MAF) was converted to haploid genome equivalents per ml (hGE/mL) of plasma and plotted over treatment course. (**A**) A 28-site capture panel was used for Patient 1 and (**B**) a 35-site capture panel was used for Patient 2. Statistical significance, as compared to a negative control, was determined at each time point. ctDNA values not significantly different from negative controls are indicated (triangle) and aggregate mutant reads/total reads are reported. Statistical significance was determined prior to converting aggregate MAF to hGE/mL plasma.

Patient 2 is a 59-year-old male who presented with cT2N1M0 mid-rectal adenocarcinoma and enrolled on the aforementioned phase II study. WES of this patient's tumor biopsy found 106 total non-synonymous SNVs and 35 sites were used to assess ctDNA levels in blood draws. At baseline and following total neoadjuvant therapy, ctDNA levels were not considered significantly above negative control, however mutant reads were present (Fig. 3.2B). Similarly, this patient also proceeded with NOM given clinical complete response seen on endoscopy and imaging. However, ctDNA levels were detectable eight months following the completion of total neoadjuvant therapy, further increased one month prior to biopsy-proven local recurrence and continued to rise until the time of salvage TME. Following TME, ctDNA levels again returned to below the limits of detection in spite of later oligometastatic progression. Unfortunately, the performance of this patient's capture panel in the negative control was the lowest of all five panels which resulted in decreased overall sensitivity at the time of oligometastatic progression (see Appendix A Table 5.2, "Aggregate error rate (%)").

3.3.2 ctDNA Levels are Associated with Tumor Burden and Progression in Oligometastatic Esophageal Cancer

Patient 3 is a 72-year-old male with oligometastatic esophageal cancer who presented with metastatic disease 2 years prior and had received extensive therapy under an immunotherapy trial. Given oligoprogression at the primary site only (distal esophagus), tumor board recommendations were for the patient to undergo neoadjuvant therapy prior to esophagectomy, at which time he was enrolled on our feasibility study. WES revealed significant intertumoral heterogeneity with only 45% of mutations shared and panel sites were selected to represent both shared and private mutations. Using DIDA-Seq, we assessed 17 mutations found only in the primary tissue biopsy and 14

mutations shared between that tumor and a subsequent metastasis (Fig. 3.3). Increasing ctDNA levels throughout neoadjuvant therapy were consistent with clinical non-response. ctDNA levels became undetectable post-esophagectomy but were again elevated seven months following surgery, concordant with clinical progression.

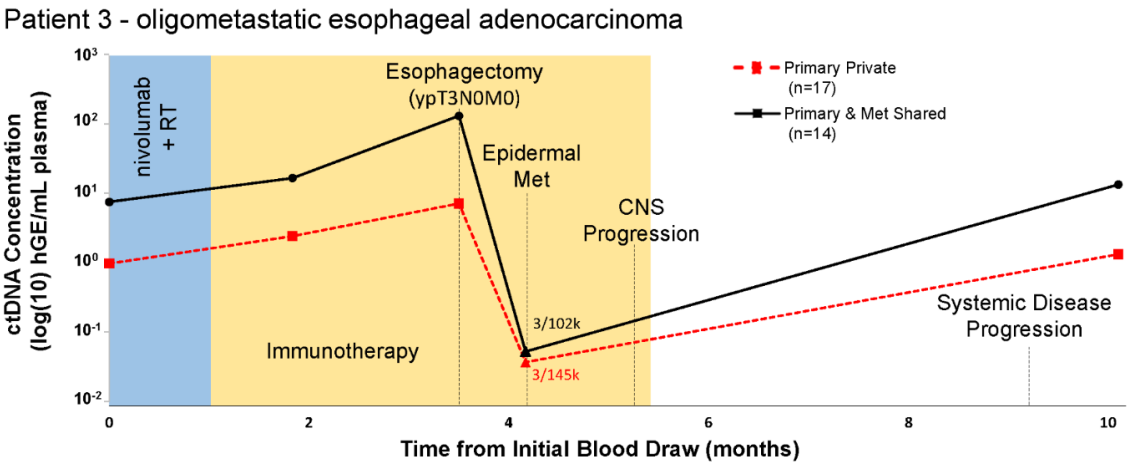
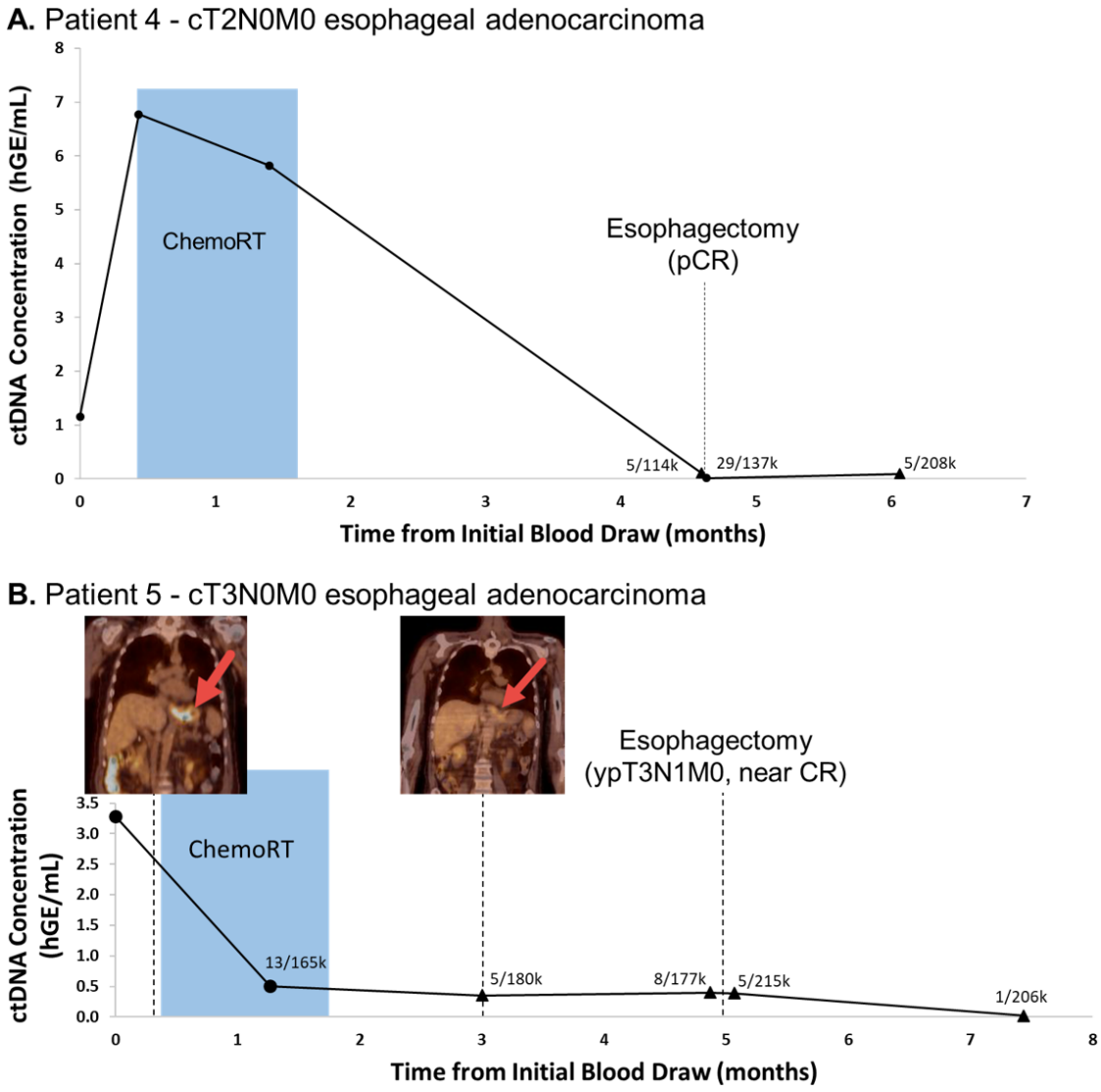


Figure 3.3. Oligometastatic esophageal adenocarcinoma cancer patient with primary-only oligoprogression.

Patient 3 had elevated ctDNA levels associated with systemic disease progression. Whole-exome sequencing of both the primary tissue biopsy and a subsequent metastatic dermal lesion revealed a high mutation burden and 45% overlap in mutation profiles. Serial ctDNA levels were retrospectively analyzed using DIDA-Seq and using patient-specific capture panels. Aggregate mutant allele frequency (MAF) was converted to haploid genome equivalents per ml (hGE/mL) of plasma and plotted in log10-scale over treatment course. Plot shows ctDNA monitoring using mutations either private to the primary tissue biopsy (n = 17, solid black line) or shared between the primary tissue biopsy and the biopsy of the metastatic dermal lesion (n = 14, dashed red line). Statistical significance, as compared to a negative control, was determined at each time point. ctDNA values not significantly different from negative controls are indicated (triangle) and aggregate mutant reads/total reads are reported. Statistical significance was determined prior to converting aggregate MAF to hGE/mL plasma.

3.3.3 Undetectable ctDNA is Associated with Pathologic Complete Response (pCR) Following Tri-Modality Therapy for Esophageal Adenocarcinoma

Patient 4 is a 61-year-old male with a history of cT2N0M0 distal esophageal adenocarcinoma who underwent neoadjuvant chemoradiation and esophagectomy. WES revealed 585 non-synonymous mutations and 39 sites were interrogated in blood draws by our capture panel. ctDNA levels declined during neoadjuvant therapy, associated with reduced tumor size and avidity on PET-CT, and were near the limit of detection (i.e., indeterminate as compared negative control values, see Methods) with 5 mutant reads in 114k total reads immediately prior to surgery, and 29 mutant reads in 137k total reads immediately following surgery (Fig. 3.4A). Surgical pathology confirmed a pCR and ctDNA levels remained undetectable as compared to the negative control at final follow-up 6 weeks following his esophagectomy.



(figure caption on next page)

(continued from figure on previous page)

Figure 3.4. Esophageal adenocarcinoma cancer patients had significant declines in ctDNA during and following neoadjuvant chemoradiation.
Patient 4 (**A**) had surgical confirmation of pathologic complete response (pCR) and Patient 5 (**B**) had near complete response (CR). Serial ctDNA levels were retrospectively analyzed using DIDA-Seq and using patient-specific capture panels. Aggregate mutant allele frequency (MAF) was converted to haploid genome equivalents per ml (hGE/mL) of plasma and plotted over treatment course. A 39-site capture panel was used for Patient 4 and a 41-site capture panel was used for Patient 5. 18F-FDG-PET/CT showed reduced tumor size and avidity (red arrows) corresponding to near complete response in Patient 5 (B, inset). Statistical significance, as compared to a negative control, was determined at each time point. ctDNA values not significantly different from negative controls are indicated (triangle) and aggregate mutant reads/total reads are reported. Statistical significance was determined prior to converting aggregate MAF to hGE/mL plasma.

Patient 5 is a 69-year-old male with cT3N0M0 distal esophageal adenocarcinoma who received neoadjuvant chemoradiation prior to esophagectomy with surgical pathology confirming a near-complete response. WES found 135 non-synonymous mutations and 41 genomic sites were included in the ctDNA panel. As with Patient 4, ctDNA levels in this patient were elevated prior to treatment, but quickly fell below the limit of detection during chemoradiation with concurrent reduced tumor size and avidity on PET-CT. ctDNA levels remained statistically insignificant at final follow-up 10 weeks following esophagectomy with no clinical evidence of disease at that time (Fig. 3.4B). Eight months later, the patient was found to have a malignant pleural effusion; however, plasma was unable to be collected to evaluate the recurrence of ctDNA.

3.4 Discussion

Here, we have demonstrated the feasibility of using patient- and tumor-specific ctDNA monitoring throughout neoadjuvant therapy and surveillance, identifying that such an assay may have the potential to detect sub-clinical disease and more precisely select candidates for organ preservation or those who may benefit from early salvage resection. Given the morbidity and mortality of large oncologic surgeries, notably esophagectomy, non-operative management for complete responders to neoadjuvant therapy is intriguing and is an active area of investigation [147, 161]. Current standard of care for locally-advanced esophageal or rectal cancer consists of neoadjuvant therapy followed by planned surgical resection irrespective of response or biomarker readout. Up to 50% of esophageal squamous cell carcinoma patients exhibit a pCR following neoadjuvant chemoradiation. This has been consistently shown to predict for better disease-free survival and overall survival [162-167] with a meta-analysis identifying a 33–36% overall survival benefit when a pCR is achieved [168]. Given the morbidity and mortality associated with esophagectomy [169-171], avoidance of resection is desirable in those who are at low risk for having residual disease. Furthermore, there is growing evidence in the rectal cancer literature that regimented clinical assessment of patients following neoadjuvant chemoradiation can potentially identify those who are clinical complete responders, allowing avoidance of immediate surgery [172-174]. A multicenter U.S. trial recently presented preliminary findings testing this hypothesis and found that a watch and wait strategy in a large proportion of patients achieving pCR after neoadjuvant therapy resulted in organ preservation without compromising survival [161].

Many providers are reluctant to adopt this approach broadly given the poor sensitivity and specificity of clinical response assessments. Current post-neoadjuvant clinical assessment for both esophageal and rectal cancers consists only of direct

endoscopic visualization and anatomic/functional imaging (CT, PET/CT, and MRI). These tests have difficulty differentiating small regions of treatment-related inflammation or fibrosis from persistent tumor and vice-versa. Multiple studies have examined the concordance rates between these tests and pathology specimens, none of which have exhibited sufficient sensitivity or specificity to accurately identify true complete responders. In rectal cancer, functional MRI has shown great promise with a substantial improvement in sensitivity and specificity (~85% for both) [175]. However, in esophageal cancer assessment of complete response is considerably poor where a combination of endoscopic ultrasound and PET/CT yields only a specificity of 30% [176]. Moreover, as lymph node metastases are still identified in up to 8% of patients with pCR of the primary tumor [177], a more robust and unambiguous biomarker for assessment of complete clinical response is needed and will drastically impact treatment decision making.

There are limited published data on ctDNA quantification during and after neoadjuvant therapy and its correlation with treatment response and suitability for surgery [150]. ctDNA has been shown useful in the detection of minimal residual disease following breast conservation therapy for women with early-stage breast cancer, with detection of ctDNA in plasma after completion of curative therapy predicting metastatic relapse with high accuracy [178]. In a similar study for Stage II and III rectal cancer patients receiving tri-modality therapy with planned surgery, the presence of tumor-specific ctDNA during post-neoadjuvant chemoradiation was highly predictive for disease recurrence despite adjustment for stage, CEA levels, and use of adjuvant therapy [179]. Additionally, in a heterogeneous cohort of esophageal cancer patients receiving chemoradiation either in the neoadjuvant or definitive setting, post-chemoradiation panel-based mutation detection of ctDNA was associated with tumor progression, metastasis, and shorter survival [180]. The results and feasibility of our

patient- and tumor-specific ctDNA assay in this cohort of patients adds to this body of literature and the impact of ctDNA as a useful response assessment biomarker.

There are some limitations to our ctDNA methodology, however. The DIDA-Seq method we have utilized achieves high sensitivity by sequencing select sites to great depth with UMI-based error correction. Consequently, three limiting factors must be considered: 1) hypermutated source tissue, 2) tumor heterogeneity, and 3) variability in performance between selected loci. In Patient 3, mutations shared between the primary and subsequent metastasis were 20-fold more prevalent than those private to the primary and therefore easier to detect. However, a clinical application of our assay for monitoring ctDNA would typically be limited to the mutations found only in the initial tissue biopsy. This highlights the importance of designing patient panels that are representative of both treatment-responsive and treatment-resistant cancer cell populations. Furthermore, poor site selection may contribute to high, panel-specific error-rates as seen in Patient 2, which had the worst performing panel of all five patients (Appendix A Table 5.3). For example, it is possible that mutant reads found in this patient at time points prior to surgery, which were determined to be below the panel's limit of detection, were indeed true positives and thus would have provided additional clinical lead time. As sequencing costs decrease, it may be feasible to routinely monitor cfDNA for every mutation identified by exome- or whole-genome sequencing of tumor biopsies, potentially mitigating such issues.

3.5 Conclusion

In this feasibility study, patient- and tumor-specific ctDNA analysis throughout multi-modality therapy for esophageal and rectal cancer patients was shown to be feasible and potentially useful in the assessment of treatment response which would

have particular utility in watch and wait and organ preservation strategies. Further investigation with a larger and more homogenous cohort is warranted.

3.6 Author Contributions

CB authored this report, produced all figures, carried out all laboratory work for data generation, and conducted all data analysis. CD edited this report and aggregated de-identified patient clinical information and imaging. NN, CH, and TK identified and consented potential study candidates. CH and TK procured patient biological samples. MBH assisted in statistical analysis and provided code. CRT, PTS, and NN conceived this project, edited, and gave final approval for this manuscript. PTS and NN procured funding.

3.7 Funding

NN received funding for this project through an American Society for Clinical Oncology Young Investigator Award (https://www.asco.org/) and a Radiological Society of North America Seed Grant (https://www.rsna.org/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

3.8 Data availability statement

The data used for analysis in this report consisted of de-identified patient clinical information and next-generation sequencing data using both custom capture and whole-exome capture library preparations. Patient clinical data are presented in the report and sequencing data (BAM files) are available on the Sequence Read Archive website

(www.ncbi.nlm.nih.gov/sra) under the BioProject accession number PRJNA637431 (https://www.ncbi.nlm.nih.gov/sra/PRJNA637431).

3.9 Patient Consent

In this feasibility study, human specimens and data (including blood, tumor tissue, and clinical information) were prospectively acquired from participants with locally-advanced esophageal (n = 3) or rectal cancer (n = 2) undergoing definitive multimodal therapy after their informed written consent (Oregon Health & Science University, IRB# 10163).

Chapter IV: Mixed ctDNA dynamics and decreased detection rates in early-stage lung cancer patients during radiation treatment

Christopher Boniface[1], Kathryn Baker[2], Christopher Deig[3], Carol Halsey[4], Taylor Kelley[5], Ramtin Rahmani[5], Garth Tormoen[6], Paul Spellman[1], Nima Nabavizadeh[6]

[1] Cancer Early Detection Advanced Research Center, Knight Cancer Institute, Oregon Health & Science University, Portland, OR 97239, USA

[2] Natera, Inc, Austin, TX, USA

[3] RadiantCare Radiation Oncology, Lacey, WA, USA

[4] Center for Women's Health, Oregon Health & Science University, Portland, OR 97239, USA

[5] School of Medicine, Oregon Health & Science University, Portland, OR 97239, USA

[6] Department of Radiation Medicine, Oregon Health & Science University, Portland, OR 97239, USA

Abstract

Quantification and detection of circulating tumor DNA (ctDNA) has been used to identify the presence of cancers. Ablative radiation therapy kills tumor cells to reduce tumor burden and it follows that these dying tumor cells could lead to increased ctDNA abundance. We carried out deep, error-corrected sequencing of cell-free DNA collected serially from 12 stage I, and 2 stage II/III non-small cell lung cancer (NSCLC) patients undergoing external-beam radiation treatment (EBRT) after initial diagnosis. We found that ctDNA detection rates (*i.e.*, the number of mutations passing filters) decreased at the first on-treatment blood draw as compared to baseline (43% to 7%). Total ctDNA abundance decreased in 6 patients and increased in 5 patients between those same blood draws, with one patient showing evidence of tumoral heterogeneity. Both patients with stage II/III disease had the largest increases in ctDNA abundance from baseline. Multiple blood draws improved ctDNA detection from 43% to 50% with a second blood draw and to 71% with 4 blood draws. Additionally, EGFR mutations were detectable in 6 patients during EBRT that were not detected prior to treatment. Taken together, these results provide a counterpoint to previous work that showed improved ctDNA detection after radiation therapy in more advanced disease.

4.1 Introduction

Non-small cell lung cancer (NSCLC) constitutes 20% of new cancer cases and nearly one quarter of all cancer deaths in the US [181]. Early detection and intervention significantly improve patient outcome and long-term survival, yet over 75% of patients have regional or metastatic disease at the time of diagnosis. Recent efforts to expand access to low-dose CT screening for at-risk populations has been shown to improve

survival rates by detecting cancer at earlier stages, but this comes with significant risk of increases in over-diagnosis and over-treatment [182, 183].

Tumor mutation profiling in NSCLC is also becoming more valuable as new treatments are approved for genetic alterations in genes such as EGFR. So-called 'actionable' mutations can confer resistance to first line treatment and become dominant in the primary tumor and metastases at later stages [27, 184, 185]. Although the clinical use of liquid biopsy is expanding, molecular-based treatment guidance has traditionally relied on solid-tissue biopsies for tumor profiling. Yet, solid-tissue biopsy is invasive, risky, and often insufficient or impossible in early-stage NSCLC patients. Consequently, many patients suspected of early-stage NSCLC are treated based on imaging diagnosis alone, without biopsy conformation of suspicious lung nodules. Liquid biopsy is therefore uniquely suited as a companion diagnostic to CT screening and in cases where biopsy confirmation of an imaging-based diagnosis is not available. It is also useful for tumor mutation profiling in leu of a solid-tissue biopsy. However, such an assay remains challenging to implement due to the low fractional abundance of tumor-derived biomarkers, such as circulating tumor DNA (ctDNA), in early-stage NSCLC.

ctDNA detection rates in stage I NSCLC patients range between 25-60% using next-generation sequencing (NGS) of commonly mutated genes [24, 26, 186]. Early-stage tumors (~1 cm^3) are estimated to contribute as little as 1-2 haploid genome equivalents per 15 ml plasma or <0.1% of total cell-free DNA, which is below the error-rate of traditional NGS [18, 74]. Molecular barcoding and computational strategies have been developed to discriminate between tumor-derived mutations and background-error in NGS [25, 30]. However, identifying tumor-derived ctDNA at mutant allele fractions (MAFs) below 0.1%-1% without prior genotyping of the tumor continues to be unreliable and presents unique challenges. Furthermore, as cell-free DNA is thought to be primarily

derived from white blood cells (WBCs), somatic mutations derived from clonal hematopoiesis (CH) can confound ctDNA detection [1].

Our lab and others have observed increased ctDNA abundance after exposure to ionizing radiation in animal models and in cancer patients [32, 34, 67, 125, 129, 130]. We hypothesized that ablative doses of radiation, such as those used in hypofractionated external-beam radiation treatment (EBRT), would lead to increased ctDNA shedding due to cytotoxicity and could therefore improve the detection rates of ctDNA-based assays in early-stage NSCLC. Furthermore, this increase in ctDNA shedding could improve liquid biopsy-based tumor genotyping. The routine radiation treatment of patients presenting suspicious lung nodules at our institution presented an opportunity to test this hypothesis. We sought to leverage the possible ctDNA enrichment effects of EBRT with computation-based error reduction workflows to improve ctDNA detection and characterization.

4.2 Methods

4.2.1 NSCLC patient and healthy control study consent, treatment, and sample acquisition

Human specimens and data (including whole blood and clinical information) were prospectively acquired from participants undergoing first-line external-beam radiation after diagnosis of NSCLC by imaging (n = 14, 12 x stage I, 1 x stage II, and 1 x stage III) after their informed written consent (Oregon Health & Science University Institutional Review Board Study #10163, first approved 19 October 2017, see Table 4.1). All patients were treated with 48-60 Gy external-beam radiation in 4-8 fractions over 10-12 days (Fig. 4.1). Blood samples were collected in EDTA tubes prior to treatment

(baseline, BL), within 21-192 (median = 47) hours following at least 1 treatment, and at various time during EBRT. Three patients had draws following the completion of treatment at 48- to 96-hour intervals (patients tb184, tb187, and tb196). Blood was collected from 10 healthy individuals between 25-55 years of age after their informed written consent under the same IRB.

						Tumor Size (longest axis, mm)				
Patient	Age/Sex	Smoking Hx (PY)	NSCLC Stage*	TNM*	EBRT Regime	PreTx	PostTx (~3 mo)	Delta	Follow-up (~3 mo)**	Months to progression
tb169	79/M	60	IA2	cT1bN0M0	7.5 Gy x 8 frac.	18	***	ND	***	NA
tb174	66/M	50	IA2	cT1bN0M0	12 Gy x 5 frac.	16	16	0	Stable	19
tb179	72/M	50	IA2	cT1bN0M0	12 Gy x 5 frac.	12	9	3	Stable	NA
tb181	71/M	65	IA1	cT1aN0M0	11 Gy x 5 frac.	12	6	6	PR	NA
tb182	68/M	50	IA3	cT1cN0M0	11 Gy x 5 frac.	27	23	4	Stable	NA
tb184	68/M	40	IA1	cT1aN0M0	11 Gy x 5 frac.	20	11	9	PR	7
tb187	58/M	45	IA2	cT1bN0M0	11 Gy x 5 frac.	16	NM	ND	CR	NA
tb188	71/M	60	IA2	cT1bN0M0	11 Gy x 5 frac.	14	12	2	Stable	NA
tb190	79/M	42	IA2	cT1bN0M0	11 Gy x 5 frac.	15	14	1	Stable	NA
tb191	66/M	147	IA2	cT1bN0M0	11 Gy x 5 frac.	12	12	0	Stable	NA
tb194	61/M	25	IIB	cT3N0M0	11 Gy x 5 frac.	10	13	-3	Stable	NA
tb196	70/M	56	IIIA	cT4N0M0	4 Gy x 15 frac.	53	19	34	PR	6
tb197	77/M	8	IA1	cT1aN0M0	12 Gy x 4 frac.	29	NM†	ND	NM†	NA
tb199	64/M	25	IA1	cT1aN0M0	11 Gy x 5 frac.	23	NM†	ND	NM†	NA

Table 4.1: Patient demographics and clinical details.

* 8th Edition [21]; **RECIST 1.1 criteria (https://project.eortc.org/recist/wp-content/uploads/sites/4/2015/03/RECISTGuidelines.pdf); *** data unavailable; † not measure (NM) due to pneumonitis; PY = pack-year; ND = no data; Tx = radiation treatment; PR = partial response; CR = complete response; NA = not applicable; frac. = treatment fractions

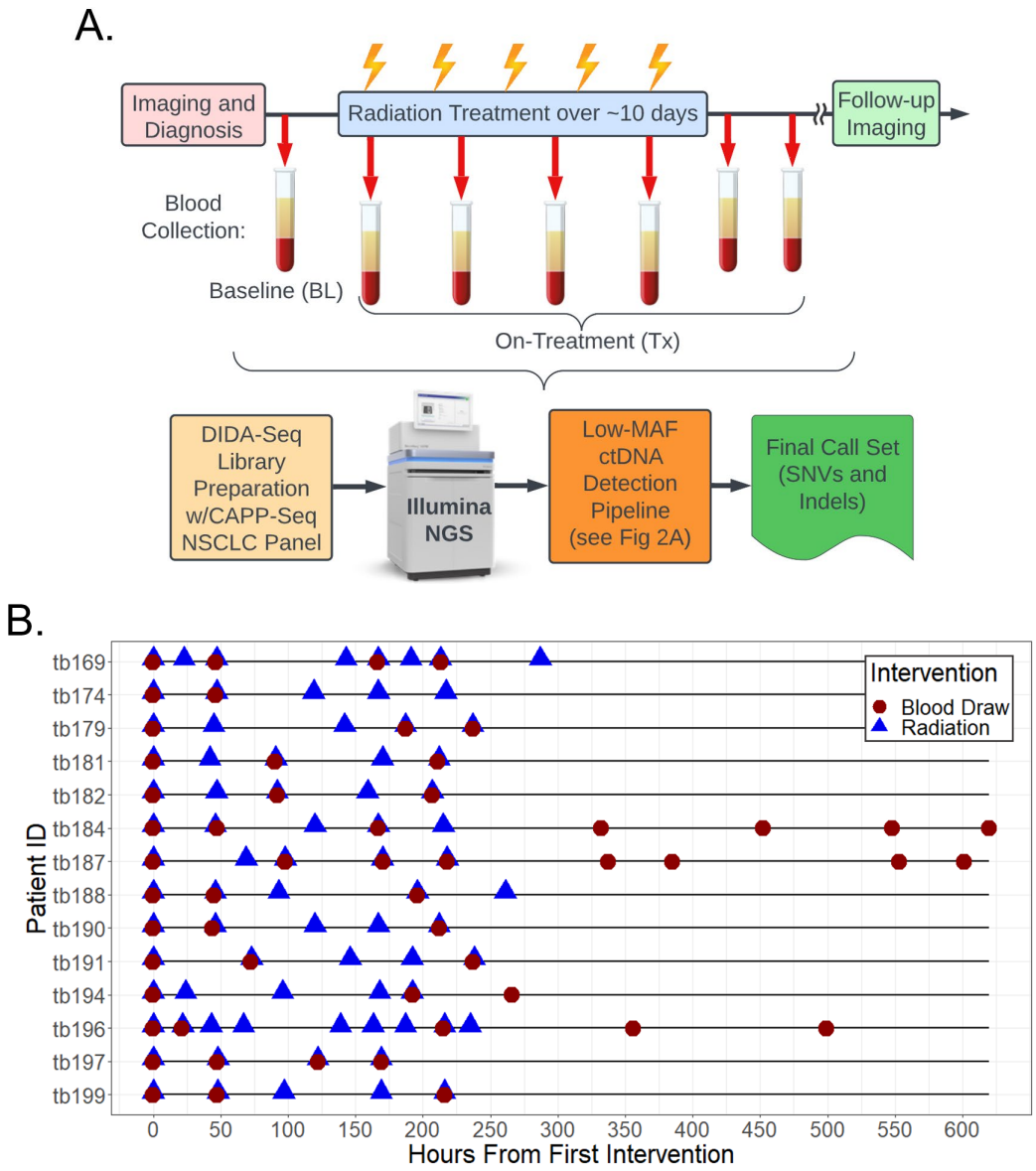


Figure 4.1. Idealized sample collection and data analysis schematic with actual cohort treatment and blood draw time lines.

An idealized time line for patient blood sample collection during external beam radiation treatment (EBRT) is shown (**A**) as well as the general workflow for library preparation, sequencing, and analysis (**B**). Note that EBRT and blood collection timing and number of draws varied between patients in our study. Post-treatment blood collection occurred in 3 patients and follow-up imaging varied from 6 to 8 weeks following the final EBRT fraction. [DIDA-Seq, dual-indexed degenerate adaptor sequencing [17]; CAPP-Seq, cancer personalized profiling by deep sequencing [24]); NSCLC, non-small cell lung cancer; NGS, next-generation sequencing; MAF, mutant allele fraction; SNV, single nucleotide variant; Indel, insertion or deletion].

4.2.2 Clinical measurements

We collected the following clinical data on NSCLC patients in this study which can be found in Table 4.1: smoking history (pack-years), NSCLC and TNM stage at diagnosis (8th Edition [187]), treatment regime, tumor size (longest axis) at diagnosis and at ~3 month follow-up imaging, RECIST tumor status at follow-up imaging (RECIST 1.1 criteria), and number of months from diagnosis to progression for up to 2-years of surveillance (until study completion).

4.2.3 Sample processing and DNA isolation

Blood draws (20-30 ml) were fractionated within 1 hour of collection for cfDNA isolation using a double spin method [41]. Briefly, blood plasma was isolated by first spinning whole blood at 1,000 x g for 10 min at 4°C, separating the top plasma layer into ~1-ml aliquots, spinning those aliquots at 15,000 x g for 10 min. at 4°C, and transferring the supernatant to cryovials for storage at -80°C. The buffy coat fraction was aliquoted (200 µl) after the initial spin and stored at -80°C. Cell-free DNA was isolated from 5-10ml of plasma using Macherey-Nagel NucleoSnap (Macherey-Nagel GmbH & Co., Duren, Germany) and white blood cell (WBC)-derived DNA was extracted using QIAgen Blood and Tissue kits (QIAGEN, Redwood City, CA, USA). All DNA extractions were quantified using the Qubit 3 (Thermo Fisher Scientific, Waltham, MA, USA) and size distribution was confirmed using the Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). Prior to library preparation DNA isolated from WBCs was fragmented by sonication to ~150 bp using a Covaris E220 (Covaris Inc., Woburn, MA, USA). Cell-free DNA was not sonicated prior to library preparation.

4.2.4 DIDA-Seq NGS library preparation, hybridization capture enrichment, and sequencing

Dual-index degenerate adaptor sequencing (DIDA-Seq) libraries were prepared as previously described [31] using a hybrid capture target enrichment panel adapted from the lung cancer-specific CAPP-Seq panel developed by Newman and colleagues (2014) [24]. Briefly, cell-free DNA or sonicated WBC DNA was end-repaired and A-tailed prior to a 15-minute ligation at 20°C to Illumina-compatible sequencing adaptor oligos containing multiplexing and UMI sequences (IDT USA, Coralville, IA, USA) using KAPA HyperPrep Reagents (KAPA Biosystems, Capetown, South Africa). Ligated DNA was PCR amplified using Illumina library amplification primers for 8-10 cycles and enriched by 18-hour hybridization capture using the xGen Hybridization and Wash kit (IDT USA, Coralville, IA, USA) with biotinylated oligos targeting regions described by Newman and colleagues [24]. Post-capture sequencing libraries were further PCR amplified for 8-10 cycles. Next-generation sequencing was carried out on the Illumina NovaSeq S4 platform using paired-end, 150 bp reads.

4.2.5 DIDA-Seq alignment, consensus-making, tag-swap filtering, and read processing

The mutation calling workflow was optimized to identify single nucleotide variations (SNVs) and insertion-deletions (indels) as depicted in Fig. 4.2 and Appendix C Fig. 7.1. Raw fastq files were demultiplexed to unmapped bam files (ubam) using the *Demux* demultiplexing tool (Fulcrum Genomics, http://www.fulcrumgenomics.com) with no index mismatches allowed. Alignment and processing for ubam files was based on the GATK4 Best Practices Small-Variant Discovery Workflow (GATK, Broad Institute, Cambridge, MA, USA) with adaptations for consensus-making steps. These steps were

carried out using available tools and custom scripts [BWA-MEM v0.7.12 using hg38 [188], SAMtools v1.3 [189], and Picard tools v2.25.1 [190]]. The FGbio tool *CallMolecularConsensusReads* was used to identify PCR duplicates based on read-pair UMIs and mapping positions (UMI families). A detailed description of this algorithm is described in the FGbio tools documentation (http://fulcrumgenomics.github.io/FGbio/tools/latest/). Briefly, the tool first converts the base-quality scores (the MAPQ sam field) to error probabilities. Consensus reads then are called within UMI families using the individual base error probabilities to resolve inter-family base-call mismatches. Finally, a posterior error rate is assigned to each consensus base call as a consensus base quality score. The FGbio *FilterConsensusReads* tool was used to filter consensus sequences with the following settings: *--min-reads* 3, *--max-read-error-rate* 0.0005, *--max-read-error-rate* 0.01, *--max-no-call-fraction* 0.2, *--min-base-quality* 40. Output consensus bam files were filtered for tag-swapping events (defined as two or more consensus sequences aligning to the same start and stop coordinates and possessing the same p5 or p7 UMI sequence but not both). A custom script was used to identify these events and output the consensus read with the lowest read error rate (the *cE* sam field) and the most PCR duplicates per pseudo-family (with random tie-breaks). The resulting consensus reads were written to a final bam file for error rate analysis, background polishing, and mutation calling. Base calls were extracted from bams at every genomic position of the target enrichment panel using the *bam-readcount* utility (https://github.com/genome/bam-readcount) unless otherwise specified (*i.e., iDES* background polishing).

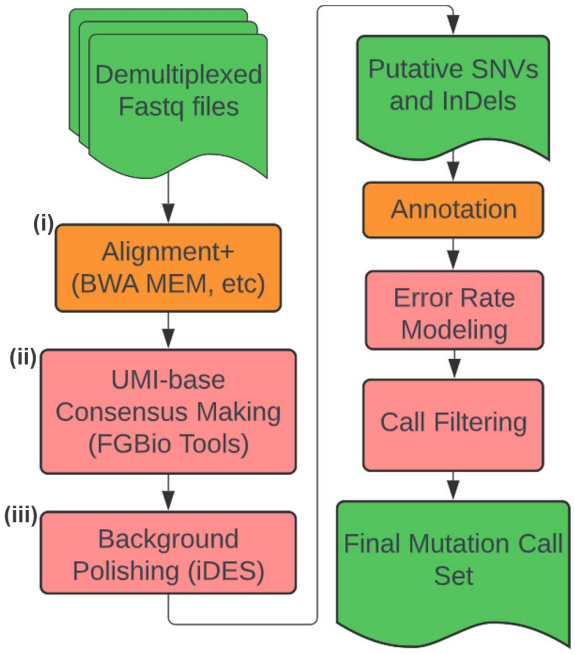


Figure 4.2. Condensed schematic of low-MAF *de novo* mutation calling pipeline. A final mutation call set is generated for each input library (demultiplexed fastqs) after fastq alignment (**i**), consensus making (**ii**), background polishing (**iii**), and filtering steps (3rd party tools used as indicated, see Appendix C Fig. 7.1 for details). Red boxes indicate key points of pipeline optimization.

4.2.6 Background polishing of single-nucleotide variants (SNVs)

For SNVs, background polishing was carried out on read count files generated for each library using the *integrated digital-error suppression* (*iDES*) tools developed by Newman and colleagues [25] with default settings. We generated the *iDES* background error database used for subsequent background polishing with pre-consensus bams from the WBC DNA libraries, as they had the greatest impact on error reduction. Resulting read counts were later used to generate library-specific error rates for each substitution-type during call filtering (see section *Methods: 4.2.9*). To generate a list of putative SNV calls, read count files were filtered to include only alternate base calls supported by at least 2 reads and 50X total coverage.

4.2.7 Selection of putative insertions and deletions (indels)

Background error polishing was not carried out for detection of indels and we assumed that most recurrent artifacts causing indels (such as poorly mapped reads) would be present in multiple patient samples and therefore flagged during exclusion list generation and annotation steps later in our workflow. Instead, pre-polished read count files were filtered to exclude SNVs, and remaining indels were required to have ≥2 supporting reads and >50X coverage.

4.2.8 Generation of exclusion lists and calculation of alternative allele background frequencies

We catalogued alternate alleles with the potential to give rise to false positives using the criteria described below ("hotspots" as defined by Newman *et al* [24] were ignored). We combined read count files with putative SNVs and indels from plasma and WBC samples of NSCLC and healthy subjects, and then filtered them to exclude alternate alleles with <5 supporting reads. We catalogued likely SNPs (≥40% MAF) found in any library and all alternative alleles found in healthy cell-free DNA. Positions with alternate alleles present in multiple patients and with multiple alternative alleles found in a single patient were catalogued. All catalogued positions (*i.e.*, exclusion lists) were later used during annotation and filtering of putative mutation calls.

Alternate allele background frequencies were calculated for the "Overlap Method" test described below (*Methods 4.2.9*) from all healthy cell-free DNA read count files. For each position in the target enrichment capture space, we recorded the number of supporting reads for all alleles (reference and alternate) and calculated the alternate allele fractions if present.

4.2.9 Annotation of putative SNVs and indels

Read count files of putative SNVs and indels were annotated with amino-acid changes using the *Annovar* utility [191] (version 2019Oct24, https://annovar.openbioinformatics.org). Putative calls were also annotated with hotspot status and exclusion list flags catalogued above. We defined a hotspot as a mutation described in previous work by Newman and Colleagues [24] or reported as "Pathogenic" in the *ClinVar* [192] database (version 2023-01-21, https://ftp.ncbi.nlm.nih.gov/pub/clinvar/tab_delimited/variant_summary.txt.gz).

4.2.10 Evaluation and filtering of putative SNVs and indels

We considered all non-synonymous putative SNVs and indels and removed all non-hotspot calls flagged as a SNP, present in healthy cfDNA, present in multiple patients, or with multiple alleles during annotation. To account for the variability in depth at each position in the target capture space, we used a Bayesian approach with the previously described Overlap Method [31, 160] when evaluating putative calls. Briefly, for each putative mutation, beta distributions were generated to test the null hypothesis that the putative MAF and the MAF of the same allele (if present) in healthy background frequency database arose from the same distribution. We omitted putative mutations if the calculated p-value exceeded 0.05 or 0.1, for non-hotspots and hotspots, respectively, in order to favor sensitivity over specificity for known or previous documented mutations. If the number of alternate read counts and total depth at the given position in the aggregate healthy controls was 0 and >50,000X, respectively, then the candidate mutation was retained. To mitigate contributions to cell-free DNA from clonal hematopoiesis, mutations having >1 supporting read in their matched WBC library

were omitted from further consideration (this filter was not used for calls analyzed for clonal hematopoiesis, *Results 4.3.3*).

4.2.11 Final call evaluation by error probability as a function of number of supporting reads

To account for variability in the error rate of each base substitution type in each library, we evaluated putative mutations based on the number of supporting reads. We assumed that as the number of supporting reads for an alternate allele increases, the likelihood of error decreases. We modeled this behavior independently for each base substitution type in each library by exponential regression considering only positions having >1,000X depth and with alternate alleles supported by <6 reads (see Fig. 4.3). Putative SNVs having alternate read counts that were less than the minimum number required, as calculated from a given error rate threshold, were omitted from further consideration. For indels, we used the median value of all 3 substitution types matching the reference allele of the indel. Note that we evaluated hotspots using less stringent error rate thresholds (~10-fold less) in order to give more statistical weight and flexibility in optimizing filtering.

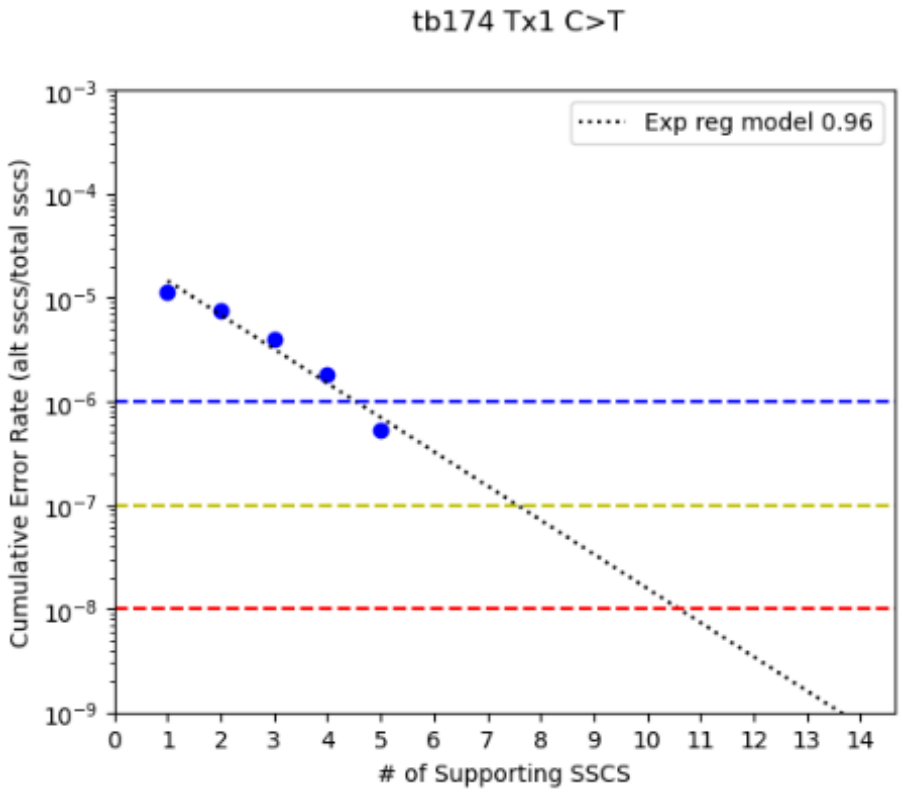


Figure 4.3. Example data and exponential regression showing the relationship between the error rate and the number of supporting reads for C>T substitutions. For each library, the number of supporting single-stranded consensus sequences (SSCSs) for a given base substitution type (A>C, A>G, …T>C, 24 total types) was counted across the entire library and plotted against the cumulative error rate, defined as the total alternate bases/total bases sequenced for each reference type, A, C, G, and T. We modeled this relationship by exponential regression and use this model to estimate the minimum number of supporting reads required to achieve a predefined error rate threshold, which was determined by either a probabilistic approach or empirically using depth-matched healthy controls. The intersection of the model [*i.e.*, x = [ln(y) – a]/b, where y=error rate and x=number of supporting reads] with the desired error rate threshold (solve for x) gives the minimum number of SSCSs needed to achieve that error rate. Colored dash lines reflect error rate thresholds of 1E-6 (blue), 1E-7 (yellow), and 1E-8 (red), which if specified, would require >5, >8, and >11 supporting SSCSs, respectively, for any C>T mutation called in this library. An example is shown using the sequencing library from Tx1 blood draw of patient tb174.

4.2.12 Calculation of the minimum error rate threshold for filtering mutation calls and subsampling of bam files to normalize ctDNA detection rates at BL and Tx1

We used error rate thresholds determined by either (a) fitting an exponential regression to the arrays *Y*, *X*, where *Y=1/[total reference base positions*$_{A,C,T,G}$ *x mean depth]* and *X = mean depth = 1,2,3,...,13000*, and solving for y (*i.e.*, error rate threshold) given *X* (*i.e.*, mean depth of the call library) (see Fig. 4.4) or (b) empirically, where we calculated the error rate threshold for which the false-positive rate (FPR) was 0 (*i.e.,* no calls passing filters) in the healthy, cell-free DNA libraries subsampled to match the depth of the library being analyzed. The former approach was used for ctDNA characterization and dynamics (see *Results 4.3.3*) and resulted in an FPR=0.1, and the latter approach was used for more stringent ctDNA detection rate analysis (see *Results 4.3.4*) as described in *Methods 4.2.12-13* below.

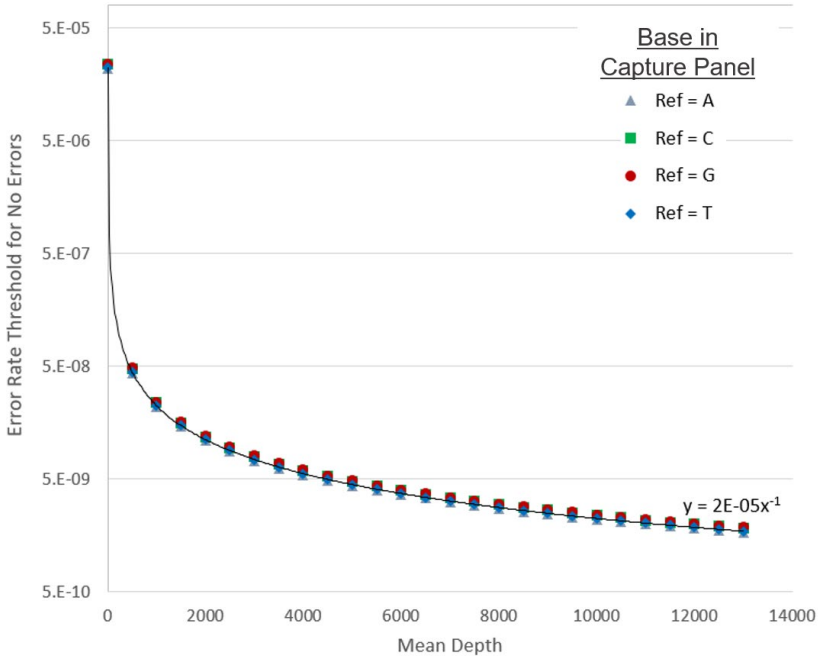


Figure 4.4. Relationship between error rate and on-target mean depth.
The maximum error rate for "no errors" was determined as 1/(the total number of reference bases sequenced + 1) (y) to achieve a given mean depth (x) for each base type (A, C, G, and T) in the CAPP-Seq [24] target capture space. To calculate error rate thresholds used in ctDNA dynamics analysis (defined as the maximum error rate tolerable) we input the mean library depth (x) and solved for y. For example, at a library mean depth of 5000X, we expect ~124M adenines to be sequenced in the capture space, therefore, using the fit equation $y=2x10^{-5}x^{-1}$, where x=5000, we calculate that an error rate threshold of 1/124M or ~ 8.05E-9 is needed to achieve less than 1 A>N errors in 124M adenines. **This probabilistic approach was used to filter putative mutation calls in full-sized sequencing libraries (*i.e.*, not subsampled) and achieved a false positive rate of 10% when used to calculated error rate thresholds for filtering calls in healthy controls (n=10).**

For comparison of ctDNA detection rates across time points from the same patient, final bam files were subsampled using the *samtools subsample* tool (Samtools v1.13) with a random seed and read output fraction. To assess the effect of EBRT on ctDNA detection rates, we limited our analysis to the BL and the first on-treatment draw, called "Tx1" (see *Results 4.3.4*). For each patient, we subsampled the bam file with the

greater mean-depth, such that the final mean-depth was within 10X of the lesser mean-depth library for that patient (see Appendix C Fig. 7.1). Parity depth between bams was readily achieved for each patient by matching the total paired-end read count of the subsampled library to that of the target depth library. Healthy control libraries were each subsampled to match the parity mean depth of the patient being assessed. The resulting set of subsampled healthy controls was used to empirically determine an error rate threshold that achieved a false positive rate (FPR) of zero. Putative call filtering was then carried out using the respective threshold for each patient as appropriate.

4.2.13 Characterization of mutations and ctDNA dynamics at BL and throughout treatment

For mutation characterization (see *Results 4.3.3*) and ctDNA dynamics (see *Results 4.3.5*), we included blood draws collected after Tx1 and bams were not subsampled prior to mutation calling and filtering. For each library, we used the error rate threshold predicted by an exponential regression model of error rate as a function of mean depth (see *Methods: 4.2.11* and Fig. 4.4). Mutation calls that were unique to draws other than baseline and Tx1 are indicated in Fig. 4.7 and Fig. 4.11 and were not included in the evaluation of ctDNA detection rates. All subsampled and full-sized bam files were subject to mutation calling and filtering using identical workflows as outlined above (see Appendix A Fig. 7.1).

4.2.14 Sequencing error rate calculations

To measure the error reduction achieved by UMI-base consensus making and background polishing steps (see *Results 4.3.2*), we generated base call counts (pileups) at every position targeted by the capture enrichment panel with depth >1,000X and MAF

<20% for all libraries sequenced at three stages of our workflow: post-alignment, post-consensus making, and post-background polishing. We summed the base calls for every position and allele in each library and calculated the cumulative error rate for each base substitution type as the number of non-reference bases divided by the total number of reference and non-reference bases.

4.2.15 ctDNA abundance calculation

To account for variability in the amount of cell-free DNA in circulation at any given draw when measuring ctDNA abundance, we converted MAFs to haploid mutant genome equivalents per ml of plasma (hGE/ml plasma) as in our previous work [31]. MAFs were calculated using read counts acquired from full-sized bam files. Total ctDNA abundance at BL and Tx1 (see *Results 4.3.3*) was calculated using summed alternate read counts and position depth for all mutations called at those time points.

4.2.16 Detection of gene fusions

We assessed full-sized and subsampled libraries for the presence of gene fusions within the target enrichment capture space using two software packages, *GeneFuse* and *FACTERA* [193, 194] using default settings.

4.2.17 Statistical methods, coding, and figure generation

For each putative mutation, the allele fractions and depths were compared between negative control and NSCLC libraries using the Overlap Method as described above (*Methods 4.2.9*) and in previous work [31]. The resulting Weitzman Coefficient was used as the p-value in each comparison. We carried out all regression analyses using the *SciPy Stats* python package [195]. For ctDNA detection, the number of

mutation calls passing filters was evaluated between BL and Tx1 draws using a paired, two-tailed Student's t-test. The same method was used to compare ctDNA abundance between BL and Tx1 (see *Results 4.3.5*). We used a two-tailed proportions z-score test (significance = 0.05) to compare the rate of detection between BL and Tx1 draws (see *Results 4.3.4*).

For evaluation of clinical parameters and ctDNA measurements we carried out linear regression analysis on multiple predictors and outcomes using regression analysis to generate a Pearson correlation coefficient and found no significant linear correlation between each series. We were not powered to make statistical evaluations for categorical variables, such as tumor response (RECIST 1.1 Criteria) at 3-month follow-up, and progression status at study completion.

All custom scripts and executables were written in Python 3.10.6 (https://www.python.org/). All data figures were generated using Python *matplotlib* (v3.6.3, https://matplotlib.org/) and *seaborn* packages (v0.11.2, https://seaborn.pydata.org/); and R (v4.1.1, https://cran.r-project.org/) using the *GenVisR* package (v1.30.0, http://bioconductor.org/packages/release/bioc/html/GenVisR.html, [196]).

4.3 Results

4.3.1 Study design, cohort details and sample collection

Newly-diagnosed NSCLC patients were consented for this study (n=12, n=1, and n=1, stage I, II, and III, respectively) and samples were collected during fractionated EBRT (Fig. 4.1 and Table 4.1). Each blood draw coinciding with an EBRT fraction was acquired 30-60 minutes prior to receiving radiation on the day of treatment. We were

unable to sample patients at matching intervals, consequently the first on-treatment blood sample (Tx1) was collected after a single fraction of EBRT in 8 patients, after two fractions in 4 patients, and after 3 or 4 fractions in the remaining 2 patients. All but 2 patients (tb174 and tb194) had at least one additional collection during treatment. Three patients also had blood collected at multi-day intervals for ~2 weeks after receiving the final EBRT fraction (tb184, tb187, and tb196) to survey post-treatment ctDNA dynamics.

4.3.2 Development and optimization of low-MAF *de novo* mutation calling pipeline

Fractional abundance of ctDNA (i.e., MAF) in early-stage NSCLC is typically at or below the background error rate of traditional NGS and the sensitivity of conventional mutation calling pipelines [26, 27, 131, 197]. In order to detect ingle nucleotide variants (SNVs) and short insertion/deletion evens (indels) below 1% MAF without foreknowledge of tumor genotype, we combined our previously published DIDA-Seq UMI-based workflow [17, 31, 32] with the CAPP-Seq NSCLC selector panel and in-silico background error-suppression methods, previously published by Newman and colleagues [24, 25]. We developed a hybrid computational workflow and expanded our mutation call filtering approach to minimize false positives (Fig. 4.2 and Appendix C Fig. 7.1).

First, to more accurately construct consensus reads and better leverage base quality score data for downstream call filtering, we replaced our majority-rules consensus maker with several FGbio tools (Fulcrum Genomics, http://www.fulcrumgenomics.com), which generate and filter consensus reads in an individual base quality score-aware manner. To optimize our computational workflow and call filtering algorithm, we sequenced an average of 79.9 ng (range 40ng-70ng, s.d. 39.5ng) of cell-free DNA and ~500ng of white blood cell (WBC) DNA isolated from 10-20

mL of blood collected from healthy adults (n=10) to a median on-target, single-stranded consensus sequence (SSCS) depth of 7535X (IQR: 6817X-8853X). To reduce the incidence of stereotypical errors introduced during library preparation, such as 8-oxogaunine conversion (G>T and C>A), we next employed the background polishing method for SNVs developed by Newman and colleagues [24]. In combination, these steps reduced the error rates of our sequencing data by 50- to 1000-fold (Fig. 4.5A). Notably, WBC libraries had slightly higher pre-consensus error rates for G>T/C>A substitutions, likely due to DNA sonication during library preparation, however, cell-free DNA libraries saw a greater reduction in these error rates from both mitigation methods.

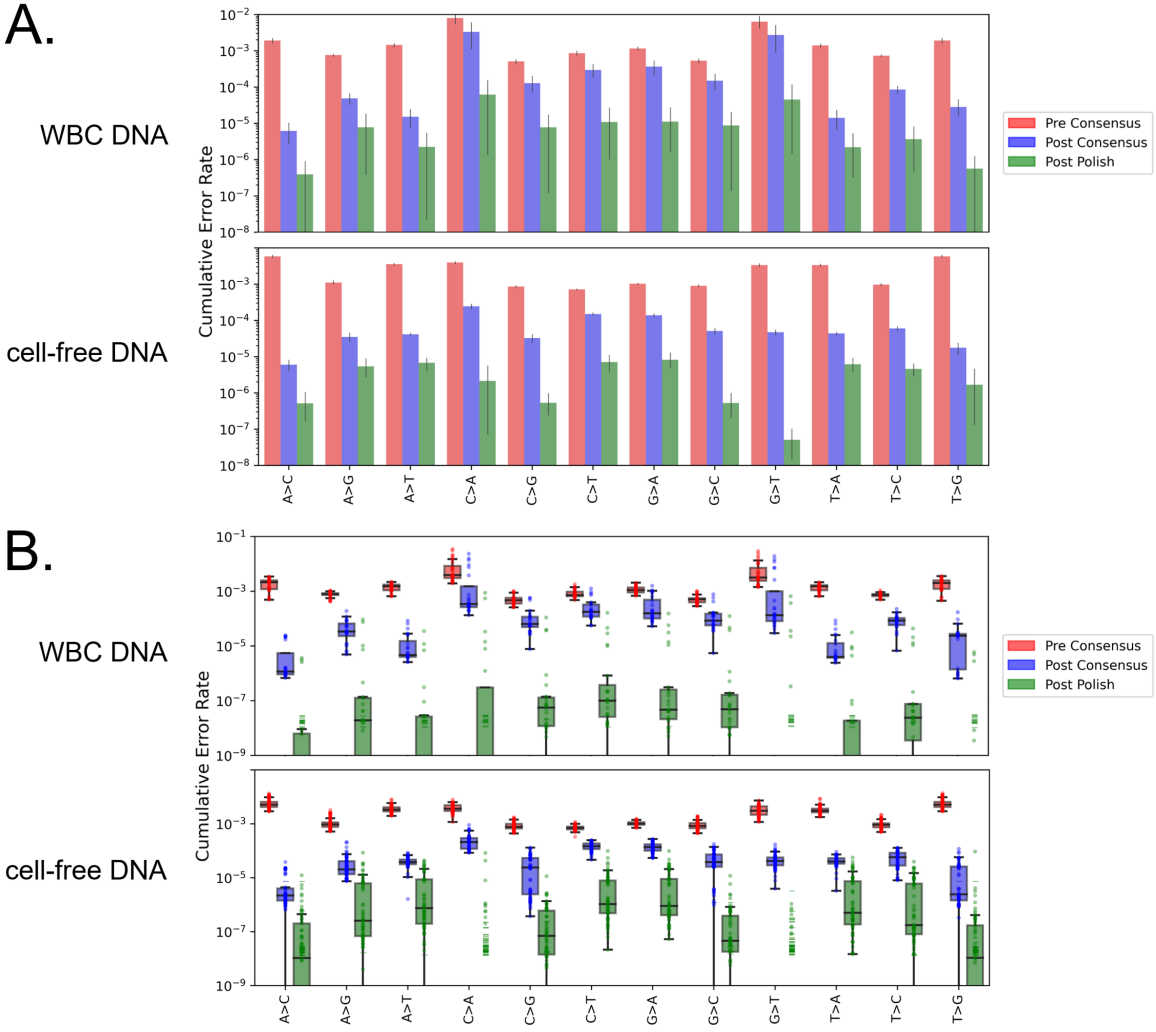


Figure 4.5. Error rate distributions for cell-free DNA and WBC DNA at key stages of error reduction.

Cumulative error rates (total alternate bases/total bases sequenced for each reference type, A, C, G, and T) are shown (**A**) with means and s.d. for each base substitution type for all cell-free DNA libraries and all WBC DNA libraries at three stages (**i-iii**) of the error reduction pipeline shown in Fig. 4.2 and Appendix C Fig. 7.1. The distribution of cumulative error rates among libraries for each base substitution type is plotted (**B**) for three stages (**i-iii**) of the pipeline**.** In cases where there were no base substitutions of a given type in a library, a dash is plotted (in leu of a point) to indicate the total number of reference bases sequenced in a library [y = 1 / (total reference bases sequenced +1)]. This notation provides an estimate of the sampling limit when no errors were detected and is not intended to imply a lower limit of detection.

The remaining non-synonymous SNVs or indels where annotated and filtered to suppress false positives using catalogued positions (*i.e.*, exclusion lists). We omitted mutation calls if there were ≥2 supporting consensus sequences in the patient-matched WBCs to avoid calls resulting from clonal hematopoiesis. We assessed the position-specific background error of each call position using a Bayesian approach with the healthy control libraries as our prior. We found significant variability in the error rates across individual libraries (Fig. 4.5B) and assumed that this would confound a generalized approach for putative call assessment. Therefore, we developed an algorithm that modeled the likelihood of error as a function of the number of supporting reads. For each mutation the model predicts the minimum number of supporting SSCSs needed to exceed a predetermined background error rate given the base substitution type. Using the healthy control libraries, we determined these thresholds empirically as the error rates for which the number of mutations passing filter (*i.e.*, false-positive rate, FPR) was zero (see Appendix B Table 6.2) or probabilistically from the on-target mean depth of the given library (see *Methods 4.2.11*). We subjected all cell-free DNA and WBC DNA to our optimized mutation calling workflow. Final bam files (*i.e.*, post-consensus making, etc.) had significant variability in on-target coverage as shown in Fig. 4.6. This was due to varying DNA amounts available for library construction, adaptor ligation efficiency, which all contribute to library complexity, as well as varying sequencing depth, and on-target efficiency during hybridization capture enrichment.

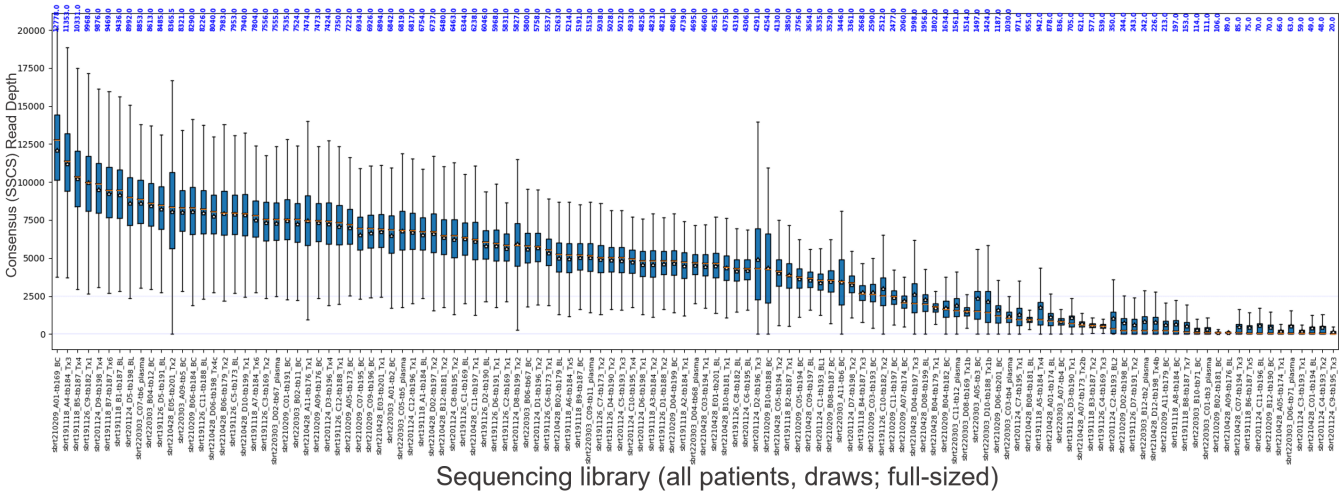


Figure 4.6. On-target depth distributions of sequencing libraries used in this study.

For each library, read counts at genomic positions covered in the CAPP-Seq [24] target enrichment hybridization capture panel (~100kb) were used to plot total coverage across the panel (depth). Read counts were derived from final full-sized bams comprised of single-stranded consensus sequences (SSCSs) aligning to positions in the CAPP-Seq panel and represent unique input DNA molecules. Mean depths are shown as gold stars and median values are shown with orange dashes and annotated in blue (top).

4.3.3 ctDNA-derived mutations detected in NSCLC patients and the impact of clonal hematopoiesis (CH)

The limit of detection of ctDNA using NGS is determined by the number of unique molecules sequenced at a given genomic position. Putative non-synonymous mutations from all NSCLC patient cell-free DNA libraries were filtered as described (*Methods 4.2.9-10*) and are shown in Fig. 4.7 and Appendix B Table 6.3 We detected 34 non-synonymous SNVs and indels in 19 genes across the enrichment capture space, with the majority in the NSCLC-associated EGFR and TP53 genes.

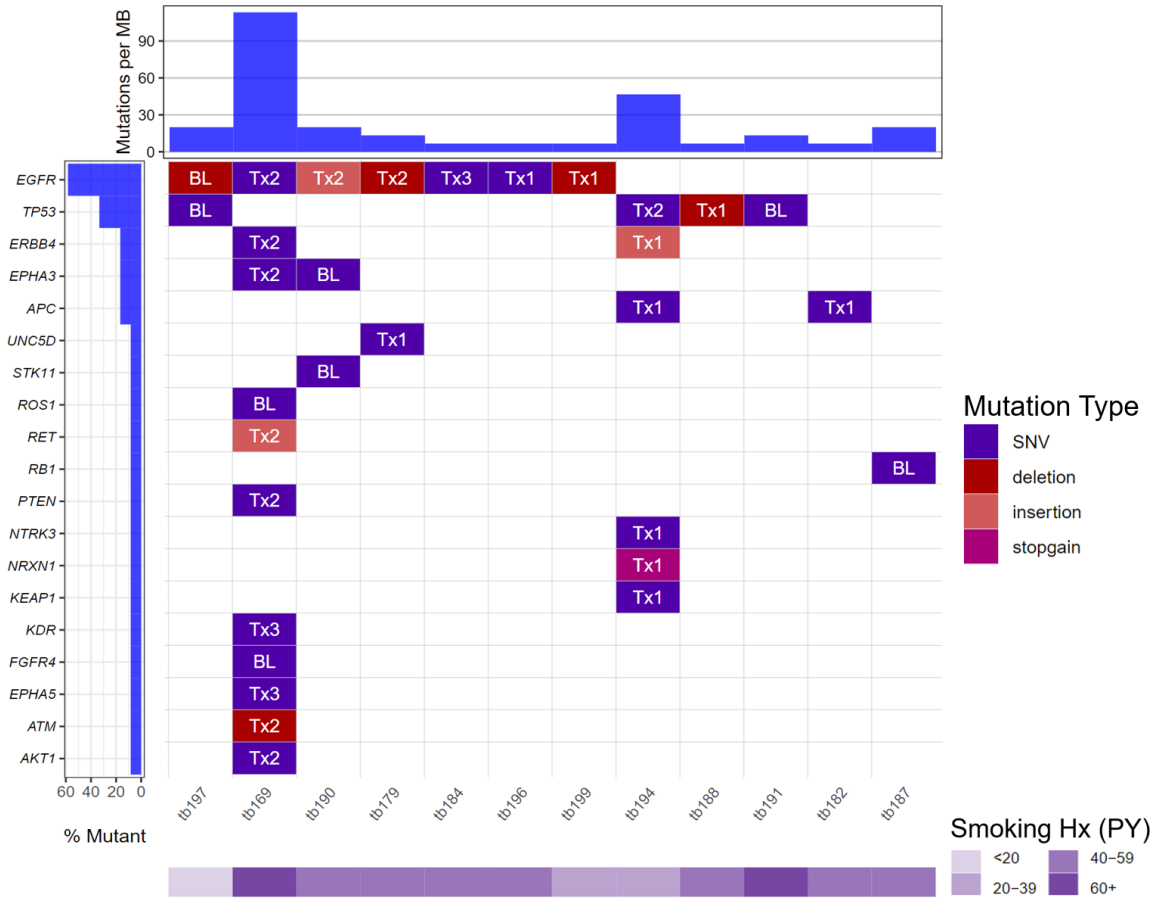


Figure 4.7. Non-synonymous mutations detected in cell-free DNA libraries collected at baseline and during radiation treatment of NSCLC patients.
A waterfall plot showing single nucleotide variants (SNVs) and insertion/deletion events detected in cell-free DNA collected at baseline (BL) or on-treatment (Tx) as shown in Fig. 4.1B. The first draw in which the mutation was called is indicated. Mutation burden was calculated for the target enrichment capture space defined in the text. Smoking history is shown as pack-years (PY). Mutations appearing in this plot were called from full-sized bams and filtered as described in the text using probabilistic error rate thresholds calculated from library mean depth.

To assess the contribution of clonal hematopoiesis (CH)-derived variants to cell-free DNA, we compared the MAFs of WBC calls to the MAFs of ctDNA calls from all cell-free DNA libraries Fig. 4.8. We reasoned that CH-derived cell-free DNA would be present in plasma at similar fractions to WBC DNA, however, tumor-derived ctDNA

would not be detectable in the WBC compartment. We found 97 mutation calls from cell-free DNA and WBC DNA (43 and 38, respectively) passing filters. Only 9 calls had supporting reads in both compartments and 3 of those were called in both compartments. Mutations with supporting reads in both compartments had similar MAFs suggesting that they were derived from CH (slope=0.8, R^2=0.41). The majority of calls did not have supporting reads in their corresponding tissue compartment. Shared mutations occurred in TP53 (n=4) and in genes not typically associated with CH (CSMD1, ERBB4, OR6F1, and NF1, n=4, see Appendix B Table 6.6). This suggests that filtering ctDNA call sets by simply omitting mutations found in genes commonly associated with CH may not be sufficient to remove false positives arising from CH. These results highlight the value of sequencing patient-matched WBC DNA when calling low-MAF ctDNA mutations.

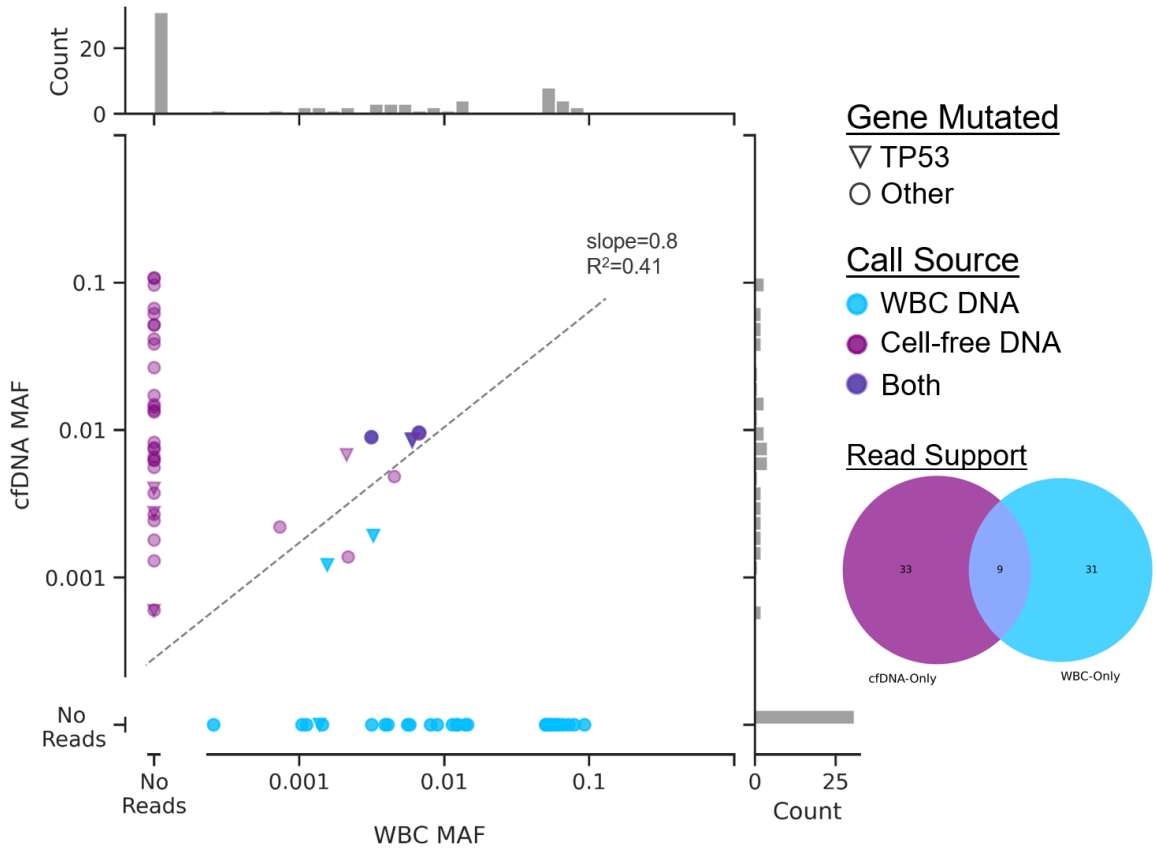


Figure 4.8. Prevalence of mutant reads in cell-free DNA and white blood cell (WBC)-derived DNA.

Mutant allele fractions (MAF) are shown for mutation calls passing filters (without filtering for matched normal) in all cell-free DNA (lilac) and matched WBC DNA (blue) libraries. Mutations in TP53 are marked (▼). Mutations called in both tissue compartments are in dark blue (n=3, TP53, ERBB4, and NF1). Inset Venn diagram shows the distributions of read support in each tissue compartment for all mutations called (n=73 total, n=9 with reads in both tissue compartments).

4.3.4 Detection rates of ctDNA decreased after EBRT

To test our hypothesis that sampling cell-free DNA during EBRT improves ctDNA detection rates, we considered the number of mutation calls passing filters at baseline (BL) and at the first on-treatment blood draw (Tx1). We accounted for the variability in depth between BL and Tx1 libraries in each patient by randomly subsampling either

library to a mean depth equal to the lower of the two (Fig. 4.9, Table 4.2). We subsampled the healthy control libraries to these same depths and determined the parameters such that FPR=0 in those libraries (see Appendix B Table 6.2). We found that the number of total calls passing filters across all patients decrease from BL in the Tx1 blood draw ($p = 0.048$; Fig. 4.10A). The proportion of patients with detectable ctDNA (>0 calls passing filters) also decreased significantly from 43% at BL to 7% at Tx1 ($p=0.023$).

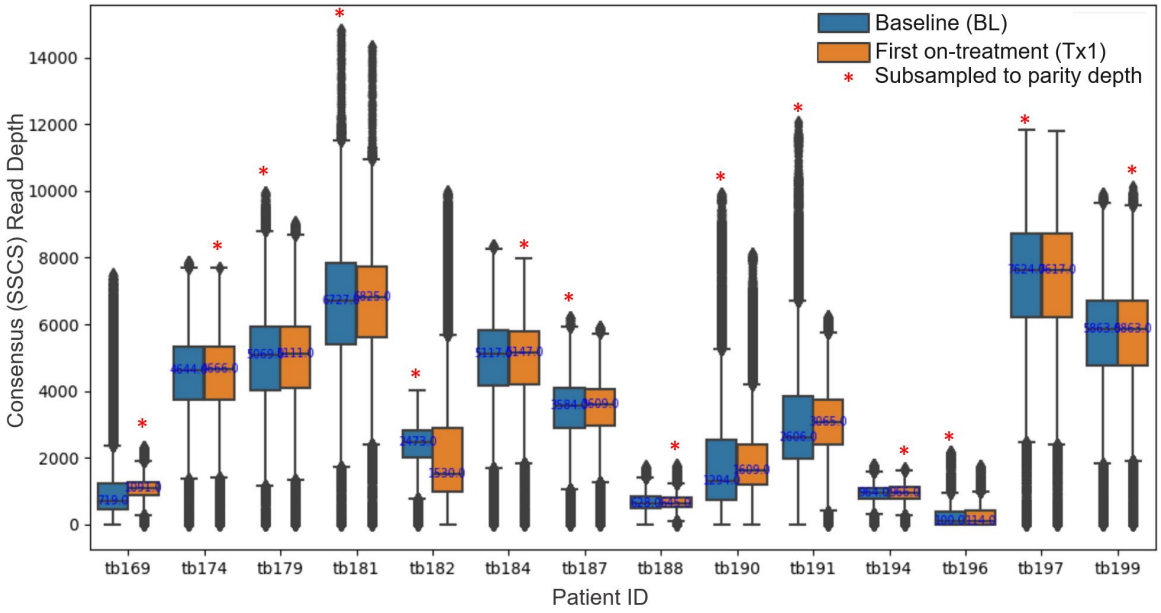


Figure 4.9. Post-parity subsampling depth distributions for BL and Tx1 libraries pairs.

For each patient, the depth distributions for the parity subsampled bams from baseline (BL) and first on-treatment (Tx1) blood draws are shown. Of the BL/Tx1 pairs, the library that was subsampled to achieve parity mean depth is indicated with a red asterisk (see Appendix B Tables 6.1 and 6.2) for full-sized and subsampled mean depths for all libraries).

When considering the cumulative ctDNA detection rate after multiple blood draws, despite decreased ctDNA detection rates overall, we found that the percentage of

ctDNA positive patients increased from 43% (6 out of 14) at BL, to 71% (10 out of 14) when including all draws from BL to Tx3. These results underscore the value of multiple serial collections when assessing patients with very low ctDNA levels. Interestingly, we found that EGFR mutations were detectable in patients during EBRT but not at baseline (tb169, tb184, tb190, tb196, tb197, and tb199, see Fig. 4.7). Without multiple draws for each healthy control, it is unclear if these results are simply due to multiple sampling and unrelated to EBRT. We did not detect gene fusions in any patient at any time point using multiple software tools nor was the number of calls passing filters associated with mean library depth (see Appendix B Table 6.5).

Patient ID	NSCLC Stage	BL[a]	Tx1[a]	Tx2[a]	Tx3[a]	Tx4[b]	Tx5[b]	Tx6[b]	Tx7[b]
tb169	IA2	2	0	12	3	-	-	-	-
tb174	IA2	0	0	-	-	-	-	-	-
tb179	IA2	0	0	1	-	-	-	-	-
tb181	IA1	0	0	0	-	-	-	-	-
tb182	IA3	1	0	0	-	-	-	-	-
tb184	IA1	0	0	0	1	0	0	0	-
tb187	IA2	1	0	0	0	1	1	0	0
tb188	IA2	0	0	0	-	-	-	-	-
tb190	IA2	1	0	1	-	-	-	-	-
tb191	IA2	2	0	0	-	-	-	-	-
tb194	IIB	1	0	0	-	-	-	-	-
tb196	IIIA	0	1	0	0	0	-	-	-
tb197	IA1	0	0	0	1	-	-	-	-
tb199	IA1	0	0	0	-	-	-	-	-
ctDNA Detection Rate		**43%**	**7%**	**23%**	**60%**				
Cumulative Detection Rate		**43%**	**50%**	**57%**	**71%**				

Table 4.2: Number of mutation calls passing filters for each blood draw[a]

[a] BL or Tx1 libraries subsampled to achieve parity mean depth and using empirical error threshold such that the FPR of subsampled healthy controls was 0; [b] Libraries analyzed at full depth using probabilistic error threshold (resulting FPR of full-depth healthy controls of 10%); FPR = false positive rate; BL = baseline; Tx = on- or post-treatment.

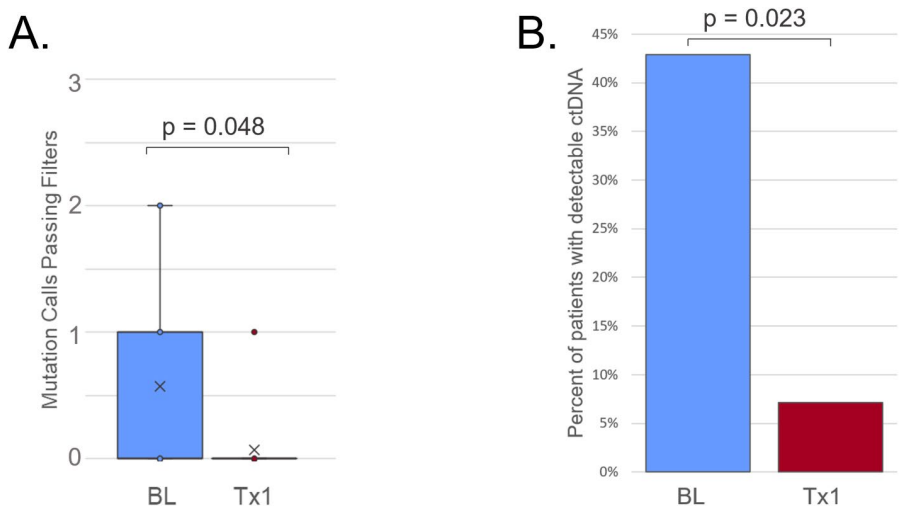


Figure 4.10. Comparison of final mutation call counts and detectable ctDNA between baseline (BL) and first on-treatment (Tx1) blood draws.
The number of ctDNA calls passing filters in BL and Tx1 draws, as reported in Table 4.2, was compared using a paired, two-tailed Student's t-test (significance = 0.05) (**A**). The proportion of patients in Table 4.2 found to be ctDNA-positive between BL and Tx1 was compared using a two-tailed proportions z-score test (significance = 0.05) (**B**).

4.3.5 ctDNA dynamics during EBRT were mixed among NSCLC patients

To assess ctDNA dynamics during EBRT, for each patient we estimated the total number of mutant genomes per mL of plasma at every time point for all calls made at any time point using non-subsampled bam files. We set the error rate threshold for mutation call filtering to the value predicted by our probabilistic model using a given library's mean depth. Notably, this approach yielded one false positive in our healthy controls (FPR = 10%) and a slightly lower BL detection rate (36%, 5/14), which increased to 93% after 3 draws were considered (see Appendix B Table 6.4). We found that 5 patients had increased total ctDNA abundance at Tx1, while 6 patients had decreased levels. Eight patients had at least one post-BL draw with increased ctDNA

levels (Fig. 4.11). Notably, 2 of the patients with samples collected following EBRT had peak ctDNA abundance between 300-600 hours (tb184 and tb187, Fig. 4.11B). When we considered only BL and Tx1 draws for all patients, we saw no significant difference in mean mutant hGE/ml plasma ($p=0.65$; Fig. 4.11C and Table 4.3). However, both patients with stage II and III disease were among those with increased ctDNA at Tx1 vs BL, which is consistent with previous work in more advanced NSCLC and metastatic breast cancer [32, 125].

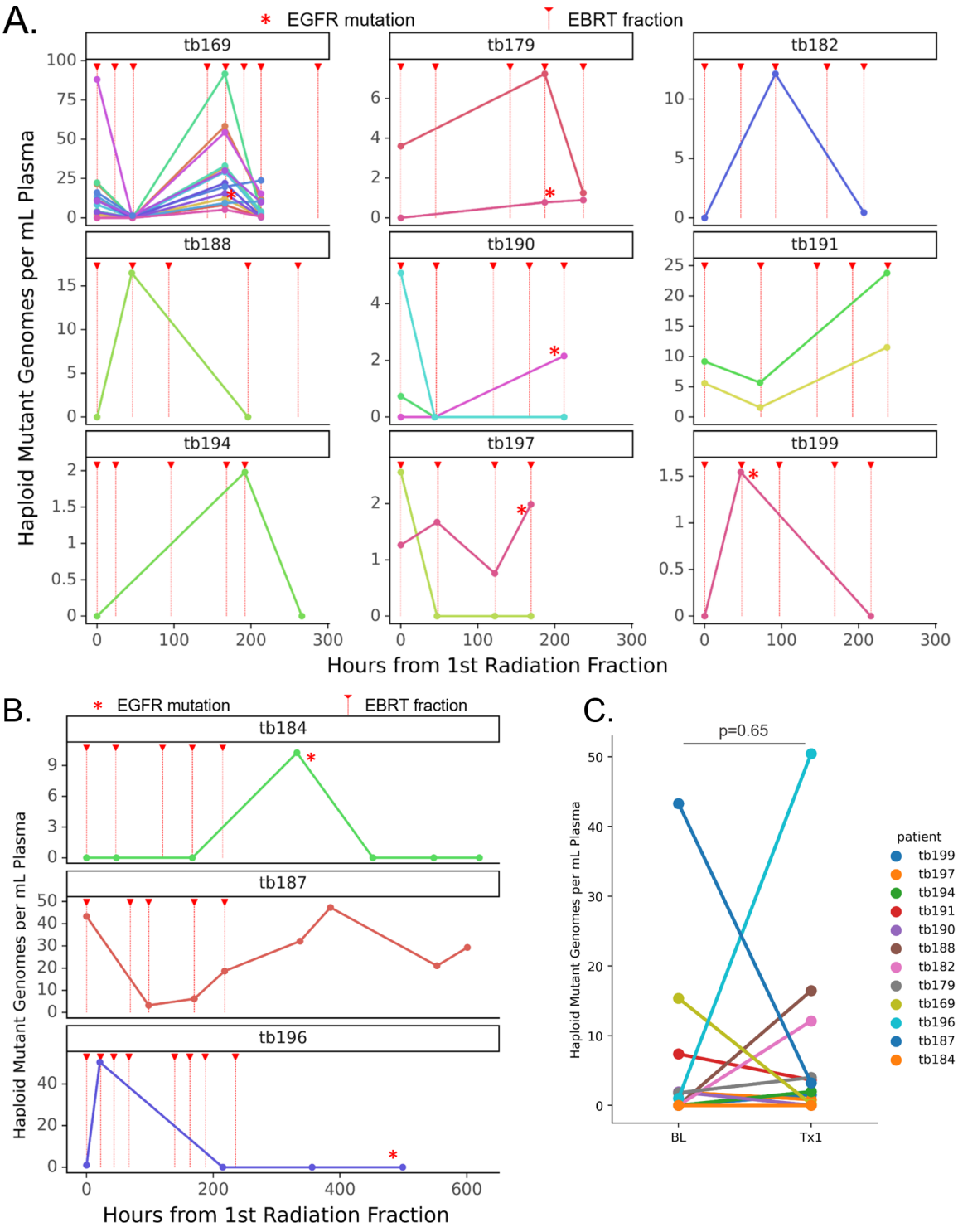


Figure 4.11. ctDNA dynamics during fractionated external-beam radiation treatment (EBRT) and comparison of total ctDNA between baseline (BL) and the first on-treatment blood draw (Tx1).

(caption continued on next page)

(continued from Fig. 4.11 on previous page)

ctDNA abundance is shown for each patient during EBRT as mutant haploid genome equivalents per ml of plasma (hGE/ml plasma) for individual mutation calls passing filters at any time point in full-sized bam files. Red triangles and dashed lines indicate radiation treatment (see Table 4.1 and Fig 4.1B). Red asterisks indicate mutations in EGFR which were observed in patients tb169, tb184, and tb196. All blood draws that occurred on the same day as EBRT were taken ~30-60 minutes prior to treatment. Nine patients had blood collected at BL (0 hours) and throughout treatment (**A**) and 3 patients also had draws collected after treatment completion (**B**). For each patient, read counts were summed for all mutation calls passing filters at BL and Tx1 (see Table 4.3) and used to calculate total hGE/ml plasma (**C**) for statistical comparison between those time points (paired, two-tailed Student's t-test).

	Baseline Draw (BL)*			1st On-Treatment Draw (Tx1)*			
Patient	Mut Rd Count	Ref Rd Count	hGE/ml plas*	Mut Rd Count	Ref Rd Count	hGE/ml plas*	Delta hGE/ml plas
tb169	96	18696	15.4	2	71226	0.2	-15.2
tb179	6	15525	1.8	11	11376	4.0	2.2
tb182	0	6061	0.0	8	1292	12.1	12.1
tb184	0	5082	0.0	0	6235	0.0	0.0
tb187	14	2269	43.3	1	1742	3.2	-40.1
tb188	0	1300	0.0	13	4778	16.5	16.5
tb190	22	14774	1.9	0	10515	0.0	-1.9
tb191	36	10708	7.4	16	15610	3.6	-3.7
tb194†	0	1074	0.0	6	12195	2.0	2.0
tb196†	1	2095	1.0	32	803	50.5	49.5
tb197	6	18332	1.9	1	16470	0.8	-1.1
tb199	0	6439	0.0	2	8548	1.5	1.5

Table 4.3. Aggregate mutant and reference read counts for each patient at baseline (BL) and first on-treatment (Tx1) blood collections.

* BL vs. Tx1 hGE/ml plasma, p = 0.51 (paired, two-tailed Student's t-test, sig. = 0.05); † stage II and III disease (all other patients were stage I, see Table 4.1); see Fig. 4.11C for associated plots.

4.3.6 ctDNA detection and abundance were not significantly associated with clinical observations

We analyzed multiple clinical parameters against ctDNA measurements at baseline and the first time point including: smoking history, NSCLC stage at diagnosis, longest tumor dimension at diagnosis and at 3-month follow-up, treatment response (RECIST 1.1 criteria at follow-up), and outcome at study completion (progression vs. non-progression at any point within ~2 years). We found no significant differences among our comparisons.

4.4 Discussion

Disease stage at diagnosis is a critical factor in overall patient outcome and survival in NSCLC. However, as early-stage NSCLC is typically discovered, diagnosed, and treated from imaging alone, the incidence of overdiagnosis and treatment of non-cancerous nodules and other masses is likely underestimated [183, 198, 199]. Although liquid biopsy seems ideally suited as a companion diagnostic, low ctDNA detection rates and highly-variable ctDNA levels in early-stage NSCLC still pose a significant technical challenge.

There is evidence that radiation treatment of solid tumors can potentiate ctDNA abundance in human subjects and animal models, but not without significant variability among subjects, and it is typically more pronounced with more advanced disease [32, 67, 125, 129]. Yet, it is unclear what factors, both biological and physiological, might drive this variability. In this study, we explored how ctDNA detection and characterization in early-stage NSCLC might improve during radiation treatment by expanding on this previous work to characterize ctDNA before, during, and after radiation treatment.

Contrary to previous findings, ctDNA detection rates in our largely stage I NSCLC cohort decreased between baseline and on-treatment draws. However, ctDNA abundance varied between patients during treatment. Some of this variability was likely due to stochastic mechanisms considering the low number of mutant reads detected in many of the libraries, but we cannot rule out potential biological sources either. Our probabilistic error thresholding approach yield 93% ctDNA detection rates after 4 draws in NSCLC patients, but multiple draws from the same healthy control were not available. Further studies using multiple serial sampling of healthy individuals would help clarify this result.

This study was also hindered by asynchronous blood collection timing between patients, particularly for the first on-treatment draw. Asynchronous EBRT fractions might also impact ctDNA dynamics depending on the influence of cumulative radiation dosage on ctDNA shedding. Based on previous work, we believe earlier and more frequent sample collection, at 3-, 6-, or 12-hour intervals, would be more informative of the effects of EBRT on ctDNA abundance. In locally advanced NSCLC patients, Breadner and colleagues [125] showed increased ctDNA abundance over BL with peaks at 3 or 24 hours in patients receiving chemotherapy after ICI (no RT) or chemoRT, respectively. They also observed increased abundance in patients undergoing palliative fractionated RT alone at 36 and 60 hours after the initial fraction. Their cohorts received fractional doses ranging from 2 to 9 Gy per fraction whereas our cohort received less fractions at higher doses (mean = 11.3 Gy per fraction, range = 4-12 Gy), which may have impacted our findings given our first post-baseline sampling was at a median of 47 hours. It may be that at higher fractional doses, cell death is more acute and ctDNA shedding more immediate. However, it is possible that chemotherapy is a more effective means to induce ctDNA shedding than RT or that early-stage patients do not consistently shed ctDNA in the same way as locally advanced patients do; that ctDNA detection in patients with smaller tumors depends on much more than assay sensitivity. It has also been

observed that changes in ctDNA abundance during therapy may be predictive of treatment efficacy, where temporary increases in ctDNA during treatment followed by ctDNA clearance is associated with improved patient outcomes (see our review, [38]). Additional studies should be done to clarify these questions.

Our study was also unable to validate our mutation calls with matched tumor biopsies due to availability. This would help reduce uncertainty around mutation calling and filtering. Future studies in which solid tumor genotyping is done in conjunction with on-treatment ctDNA analysis could establish the ground truth of ctDNA shedding and dynamics during therapy. Moreover, any such studies in early-stage NSCLC patients should include larger, more gender diverse cohorts to improve statistical analysis. Doing more studies that utilize deeper sequencing, more patients, and earlier blood collection, might yet reveal some utility in sample collection during treatment as a companion diagnostic. As sequencing cost continuing to plummet, multiple draws over the course of a patient's care will become more feasible and routine, thereby improving the diagnostic confidence of liquid biopsy for guiding treatment.

4.5 Funding source

This project was supported by funding (FULL5080319) from the Cancer Early Detection Advanced Research Center at Oregon Health & Science University, Knight Cancer Institute.

Conclusion

The work presented in this dissertation explores the biological factors that influence ctDNA abundance, particularly during treatment. It assesses the utility of patient-specific ctDNA assays to detect subclinical disease and radiation treatment to improve ctDNA detection rates of patient-agnostic assays in lung cancer patients. Although there are many clinical and biological correlates to ctDNA abundance, it remains unclear why ctDNA levels are high in some patients and not others, even within the same disease and stage. Despite this state of affairs, our data and the vast majority of existing studies that use patient-specific ctDNA assays suggest that this approach is largely robust in detecting minimal residual disease and recurrence with clinically-significant lead time. Unfortunately, our attempt to leverage radiation treatment for patient-agnostic detection and characterization of ctDNA in lung cancer patients was less promising.

In order to better understand the biological determinants of ctDNA detection and dynamics, it is clear that a significant effort must be made toward deconvoluting the myriad of variables that seem to be implicated. In Chapter II I outlined a range of these determinants, from the overt, such as tumor burden and aggressiveness, to the nuanced, such as macrophage presence and abscopal effects. Carefully designed experiments and studies will need to be done to determine the degree that each factor influences the variability of ctDNA abundance among patients and cancer types. New models may need to be

developed as we reach the limits to which our current models are able to isolate and control for these variables.

Regardless of what efforts are made to improve our understanding of the factors contributing to ctDNA abundance, the implementation of ctDNA assays in all stages of cancer care will march on. Fortunately, we will continue to reap the benefits of such assays regardless of whether or not the underlying mechanisms of analyte generation and dynamics are truly understood. This fact is particularly true for tumor types and stages with higher levels of ctDNA, such as colorectal cancer and metastatic disease.

Finally, although the methods utilized in my work relied on the detection of tumor-derived mutations, such as SNVs and indels, it is likely that multiomic approaches will be necessary to achieve the sensitivity and specificity required for cancer screening, early-detection, and monitoring for subclinical disease. Base-substitutions and indels are possibly the simplest way to distinguish ctDNA from healthy cell-free DNA in sequencing data. The use of more novel features for ctDNA enrichment and detection, such as methylation patterns, fragment lengths, and fragment boundary locations, illustrates the breadth of information available in these molecules; we are only just beginning to understand these features and how they relate to the tissue of origin and host biology. As our ability to collect and analyze large amounts of data improves, we may also discover new modalities of distinction. Furthermore, integrating over all of these features, as well as the large number of other analytes available in liquid biopsy, in

combination with machine learning approaches, will dramatically improve the sensitivity and specificity of ctDNA assays.

As we continue to push the boundaries of assay sensitivity and detection of early or pre-cancerous lesions and micro-metastasis, we will be faced with more abstract questions about what constitutes cancerous disease. It is likely that the stochastic genetic drift found in hematopoietic stem cells, for example, also occurs in other tissues and will eventually be readily detectable via liquid biopsy. The challenge will then be to identify those occasions and individuals for which intervention is appropriate and necessary. Once again, in the effort to effectively detect and treat disease, we may again redefine what it means to be healthy.

Appendix A: Additional Tables from Chapter III

WES Sequencing Library	Mean Depth	On-target Capture Effiency	Total Calls Passing Filters	Total Coding Calls
Pt1 Buffy Coat	120	0.75	-	-
Pt1 Primary Bx	364	0.75	256	81
Pt2 Buffy Coat	122	0.75	-	-
Pt2 Primary Bx	342	0.73	358	106
Pt3 Buffy Coat	119	0.73	-	-
Pt3 Primary Bx	302	0.74	1351	456
Pt3 Metastasis Bx	304	0.74	1735	604
Pt4 Buffy Coat	119	0.75	-	-
Pt4 Primary Bx	346	0.74	402	135
Pt5 Buffy Coat	124	0.75	-	-
Pt5 Primary Bx	313	0.73	5922	585

Appendix A Table 5.1: Chapter III WES metrics and mutation counts

Patient	Mutation Call Source	Number of sites	total neg cntl wt rd count	total neg cntl mut rd count	Total neg ctrl depth	Min error rate (%)	Max error rate (%)	Aggregate error rate (%)	1 error in N reads
1	Primary Biopsy	28	825316	53	825369	0.0020	0.0375	0.0064	15573
2	Primary Biopsy	35	1147864	92	1147956	0.0025	0.0368	0.0080	12478
3	Primary Private	17	412565	18	412583	0.0032	0.0071	0.0044	22921
3	Primary & Met Shared	14	292340	14	292354	0.0017	0.0091	0.0048	20882
3	(Patient 3 combined)	31	704905	32	704937	0.0017	0.0091	0.0045	22029
4	Primary Biopsy	39	3521623	235	3521858	0.0008	0.0301	0.0067	14987
5	Primary Biopsy	40	1342296	64	1342360	0.0014	0.0258	0.0048	20974

Appendix A Table 5.2: Chapter III Patient-specific hybrid capture panel evaluation in negative controls.

Patient	Month	Mutation Call Source	Neg. Control Mut. Read Count	Neg. Control WT Read Count	Sample Mut. Read Count	Sample WT Read Count	Negative Control VAF (%)	Sample VAF (%)	Overlap p-value	Min. Required Mut. Reads (binomial test, alpha=0.9)	ng cfDNA/ml Plasma	Mutant hGE/mL Plasma	Lowest Significant VAF (%)*
1	0.0	Primary Biopsy	53	825316	1187	70084	0.0064	1.67	< 0.0001	7	61.1	339.3	
1	1.4	Primary Biopsy	53	825316	1128	127435	0.0064	0.88	< 0.0001	12	137.5	402.1	
1	3.6	Primary Biopsy	53	825316	1135	130097	0.0064	0.86	< 0.0001	12	130.8	377.2	
1	5.0	Primary Biopsy	53	825316	1374	118034	0.0064	1.15	< 0.0001	11	24.0	92.1	
1	6.9	Primary Biopsy	53	825316	1088	97684	0.0064	1.10	< 0.0001	10	12.4	45.5	
1	14.3	Primary Biopsy	53	825316	1375	111417	0.0064	1.22	< 0.0001	11	11.1	45.3	
2	0.0	Primary Biopsy	92	1147864	9	201752	0.0080	0.005	0.007	21	47.5	0.7	
2	0.5	Primary Biopsy	92	1147864	4	162942	0.0080	0.003	0.17	18	61.2	0.5	
2	4.2	Primary Biopsy	92	1147864	9	173283	0.0080	0.005	0.18	19	15.1	0.3	
2	10.7	Primary Biopsy	92	1147864	23	184383	0.0080	0.013	0.005	20	14.8	0.6	0.013
2	13.5	Primary Biopsy	92	1147864	21	21006	0.0080	0.100	< 0.0001	3	5.8	1.9	
2	15.6	Primary Biopsy	92	1147864	10	213458	0.0080	0.005	0.068	22	14.6	0.2	
2	18.0	Primary Biopsy	92	1147864	20	279028	0.0080	0.007	0.1757	28	12.7	0.3	
3	0.0	Primary Private	18	412565	30	90514	0.0044	0.033	< 0.0001	7	8.8	96.7	
3	1.8	Primary Private	18	412565	27	121504	0.0044	0.022	0.0001	8	32.4	240.0	
3	3.5	Primary Private	18	412565	111	127103	0.0044	0.087	< 0.0001	9	24.4	709.8	
3	4.2	Primary Private	18	412565	3	144844	0.0044	0.002	0.22	10	24.4	17.1	
3	10.1	Primary Private	18	412565	31	96929	0.0044	0.032	< 0.0001	7	12.4	131.9	0.032
3	0.0	Primary & Met Shared	14	292340	160	62213	0.0048	0.257	< 0.0001	5	8.8	749.6	
3	1.8	Primary & Met Shared	14	292340	125	81831	0.0048	0.153	< 0.0001	7	32.4	1648.5	
3	3.5	Primary & Met Shared	14	292340	1412	86893	0.0048	1.599	< 0.0001	7	24.4	13001.4	
3	4.2	Primary & Met Shared	14	292340	3	101515	0.0048	0.003	0.27	8	5.2	5.2	
3	10.1	Primary & Met Shared	14	292340	216	66731	0.0048	0.323	< 0.0001	6	12.4	1329.5	
4	0.0	Primary Biopsy	235	3521623	172	160456	0.0067	0.107	< 0.0001	15	3	115.5	
4	0.4	Primary Biopsy	235	3521623	494	203600	0.0067	0.242	< 0.0001	18	8	677.6	
4	1.4	Primary Biopsy	235	3521623	29	266181	0.0067	0.011	0.047	23	160	582.5	0.011
4	4.6	Primary Biopsy	235	3521623	5	114164	0.0067	0.004	0.1055	11	7	10.4	
4	4.6	Primary Biopsy	235	3521623	29	134977	0.0067	0.022	< 0.0001	13	0	1.6	
4	6.1	Primary Biopsy	235	3521623	5	208223	0.0067	0.002	0.01	19	11	9.2	
5	0.0	Primary Biopsy	65	1358753	54	405053	0.0048	0.0133	0.0001	25	74.0	328.1	
5	1.3	Primary Biopsy	65	1358753	13	164500	0.0048	0.0079	0.16	12	19.0	50.1	0.0079
5	3.0	Primary Biopsy	65	1358753	5	180035	0.0048	0.0028	0.16	12	37.7	35.2	
5	4.9	Primary Biopsy	65	1358753	8	177422	0.0048	0.0045	0.066	12	26.6	39.9	
5	5.1	Primary Biopsy	65	1358753	5	215050	0.0048	0.0023	0.10	15	50.5	38.7	
5	7.4	Primary Biopsy	65	1358753	1	206177	0.0048	0.0005	0.002	14	12.2	2.0	

Appendix A Table 5.3: Chapter III DIDA-Seq cfDNA read counts and statistical data.

* Used to calculate average limit of detection which did not include Patient 1 as all time-points had significant VAFs

Appendix B: Additional Tables from Chapter IV

Patient	Cohort	BL	Tx1	Tx2	Tx3	Tx4	Tx5	Tx6	Tx7	Final_BL	Final_Tx1	subsampled_draw	subsampled_readcount
tb169	NSCLC	1090	4483	1163	1795	-	-	-	-	1090	1090	Tx1b	2467021
tb174	NSCLC	4495	4638	-	-	-	-	-	-	4495	4495	Tx1	10599660
tb179	NSCLC	6695	4986	6438	-	-	-	-	-	4983	4986	BL	11926402
tb181	NSCLC	7360	6583	7805	-	-	-	-	-	6579	6583	BL	15326037
tb182	NSCLC	5128	2400	1176	-	-	-	-	-	2399	2400	BL	5588961
tb184	NSCLC	4932	5895	4531	7591	8092	5888	3304	-	4932	4932	Tx1	11466119
tb187	NSCLC	3706	3465	1758	4917	4240	5411	4706	4810	3464	3465	BL	8030188
tb188	NSCLC	689	2849	2772	-	-	-	-	-	689	684	Tx1b	1419106
tb190	NSCLC	2174	1945	2695	-	-	-	-	-	1944	1945	BL	4075216
tb191	NSCLC	3106	6080	7656	-	-	-	-	-	3106	3096	Tx1	6893247
tb194	NSCLC	943	9273	2416	-	-	-	-	-	943	942	Tx2	2143537
tb196	NSCLC	618	250	477	459	439	-	-	-	250	250	BL	530272
tb197	NSCLC	8259	7376	8108	7268	-	-	-	-	7376	7376	BL	16824183
tb199	NSCLC	5669	7379	417	-	-	-	-	-	5669	5668	Tx1	13206473
tb2	Healthy	7976	-	-	-	-	-	-	-	See Appendix B Table 6.1 for mean depths of subsampled healthy controls			
tb3	Healthy	4607	-	-	-	-	-	-	-				
tb4	Healthy	7159	-	-	-	-	-	-	-				
tb5	Healthy	8054	-	-	-	-	-	-	-				
tb6	Healthy	10458	-	-	-	-	-	-	-				
tb11	Healthy	5795	-	-	-	-	-	-	-				
tb12	Healthy	6379	-	-	-	-	-	-	-				
tb67	Healthy	10175	-	-	-	-	-	-	-				
tb68	Healthy	6873	-	-	-	-	-	-	-				
tb71	Healthy	9703	-	-	-	-	-	-	-				

Appendix B Table 6.1: Chapter IV On-target mean consensus sequence depth of sequencing libraries.

On-target mean consensus sequence depth of full-sized (not subsampled) bams and subsampled pairs (BL/Tx1), final parity depth, and read count (paired end).

		NSCLC Patient ID																	
		tb169	tb174	tb179	tb181	tb182	tb184	tb187	tb188	tb190	tb191	tb194	tb196	tb197	tb199
Target DP (mean DP of NSCLC Patient):		1090	4495	4983	6579	2399	4932	3464	684	1944	3096	942	250	7376	5668
Healthy Patient ID	Healthy Patient Mean DP	Subsamped Depth													
tb2	7976	1091	4479	5085	6600	2452	4916	3426	625	1911	3303	933	252	7410	5691
tb3	4607	1078	4426	4607	4607	2421	4608	3384	618	1887	3262	921	248	4608	4608
tb4	7159	1118	4591	5212	6775	2512	5039	3511	642	1957	3384	956	258	7159	5839
tb5	8054	1103	4522	5133	6670	2474	4962	3458	632	1928	3333	943	255	7486	5750
tb6	10458	1117	4581	5201	6757	2506	5029	3502	641	1953	3376	955	257	7584	5824
tb11	5795	1075	4405	5000	5795	2411	4834	3367	617	1881	3245	919	248	5796	5601
tb12	6379	1087	4459	5063	6397	2439	4894	3408	623	1900	3285	929	250	6379	5669
tb67	10175	1083	4444	5046	6555	2432	4878	3399	621	1896	3276	925	250	7356	5650
tb68	6873	1096	4485	5092	6614	2455	4923	3434	628	1915	3309	938	252	6874	5704
tb71	9703	1104	4527	5138	6674	2478	4968	3460	633	1932	3335	944	254	7489	5752
Average DS Depth		1095.2	4491.9	5057.7	6344.4	2458.0	4905.1	3435.0	628.1	1916.1	3310.6	936.3	252.4	6814.1	5608.8
Delta		5.2	-3.1	74.7	-234.6	59.0	-26.9	-29.0	-55.9	-27.9	214.6	-5.7	2.4	-561.9	-59.2
Avg w/out mins		1095.2	4491.9	5107.8	6663.6	2458.0	4938.1	3435.0	628.1	1916.1	3310.6	936.3	252.4	7465.1	5720.1
Delta w/out mins		5.2	-3.1	124.8	84.6	59.0	6.1	-29.0	-55.9	-27.9	214.6	-5.7	2.4	89.1	52.1
Prior target read count		2589579	10636350	12075512	15112074	5816443	11674079	8129727	1483999	4532358	7835183	2157891	596237	17612077	13524254
Average Reads per 1x		2364	2368	2364	2268	2366	2364	2367	2363	2365	2367	2305	2362	2359	2364
Empirical Error Threshold Parameters where FPR=0															
--err_thr		3E-08	3E-09	3E-10	3E-09	2E-09	5E-10	6E-09	3E-08	4E-08	5E-09	4E-08	5E-07	3E-09	1E-10
--hs_err_thr		1E-06	1E-07	1E-07	1E-07	1E-07	1E-07	1E-07	1E-06	1E-07	1E-07	1E-06	1E-06	1E-07	1E-07
--tp53		1E-06	1E-07	1E-08	1E-07	1E-07	5E-08	1E-07	1E-06	1E-07	1E-07	1E-06	1E-06	1E-07	1E-07
--exp_offset		0	0	0	0	0	0	0	2	0	0	0	0	0	0

Appendix B Table 6.2: Chapter IV Library depth and calculated error rate thresholds and filtering parameters.

Mean depth targets for healthy controls to achieve parity depth of patient's BL/Tx1 pair. Empirically determined error rate thresholds are also shown for each patient target depth (these parameters were provided to filtering algorithm in Fig. 7.1, bottom).

patient	chr	position (hg38)	ref base	alt base	alt cnt	DP	MAF	healthy plasma bg dp	healthy plasma bg maf	CalledIn	hotspots*	min sscs req	BC alt cnt	BC ref cnt	aa ch	mut type	gene
tb169	chr5	177095368	G	C	20	3088	0.00648	199902	3.00E-05	BL	None	8	0	13325	E480Q	SNV	FGFR4
tb169	chr6	117326279	A	T	23	449	0.05122	131862	0	BL	None	10	0	12861	N1834K	SNV	ROS1
tb169	chr7	55191821	C	G	4	1715	0.00233	171030	0	Tx2	['C>A','C>G','C>T']	6	0	12625	L591V	SNV	EGFR
tb169	chr10	43112867	T	+A	11	3022	0.00364	218052	0	Tx2	None	10	0	17363	F301fs	Ins	RET
tb169	chr14	104780190	G	C	9	1642	0.00548	189162	0	Tx2	None	8	0	12576	R25G	SNV	AKT1
tb169	chr5	177095372	T	G	21	2872	0.00731	202342	0	Tx2	None	10	0	13285	M481R	SNV	FGFR4
tb169	chr5	177095368	G	C	28	2818	0.00994	199902	3.00E-05	Tx2	None	8	0	13325	E480Q	SNV	FGFR4
tb169	chr3	89450279	A	C	12	913	0.01314	171910	0	Tx2	None	8	0	15363	N867H	SNV	EPHA3
tb169	chr6	117319892	A	T	12	895	0.01341	159556	0	Tx2	None	9	0	15019	S1972R	SNV	ROS1
tb169	chr11	108329141	T	-T	12	843	0.01424	182330	0	Tx2	None	10	0	14018	Y2404fs	del	ATM
tb169	chr2	211725102	A	G	11	744	0.01478	148576	0	Tx2	None	9	0	12886	S239P	SNV	ERBB4
tb169	chr6	117326279	A	T	18	725	0.02483	131862	0	Tx2	None	9	0	12861	N1834K	SNV	ROS1
tb169	chr10	87957859	A	T	17	644	0.02640	172972	0	Tx2	None	9	0	13067	Q214L	SNV	PTEN
tb169	chr2	211722503	C	G	19	459	0.04139	131900	0	Tx2	None	8	0	13404	C258S	SNV	ERBB4
tb169	chr7	55191821	C	G	2	3537	0.00057	171030	0	Tx3	['C>A','C>G','C>T']	5	0	12625	L591V	SNV	EGFR
tb169	chr4	55110472	A	C	8	1069	0.00748	185358	0	Tx3	None	7	0	15720	Y396D	SNV	KDR
tb169	chr4	65367386	T	G	17	997	0.01705	153606	0	Tx3	None	8	0	11340	E611A	SNV	EPHA5
tb179	chr8	35748598	C	T	10	5891	0.00170	168038	0	Tx1	None	7	0	6790	T608I	SNV	UNC5D
tb179	chr7	55191823	G	-G	2	7189	0.00028	170218	1.17E-05	Tx2	None	12	0	5344	L591fs	del	EGFR
tb182	chr5	112837901	A	C	8	1300	0.00615	178128	0	Tx1	None	6	0	5773	L751F	SNV	APC
tb184	chr7	55174773	G	A	4	8008	0.00050	164632	0	Tx3	['GAATTAAGAGAAGCA>G', 'GAATTAAGA>G']	2	0	1155	E479K	SNV	EGFR
tb187	chr13	48303986	C	G	14	2283	0.00613	107798	0	BL	None	2	0	648	P25R	SNV	RB1
tb187	chr13	48303986	C	G	8	2262	0.00354	107798	0	Tx4	None	2	0	648	P25R	SNV	RB1
tb187	chr13	48303986	C	G	12	2697	0.00445	107798	0	Tx5	None	4	0	648	P25R	SNV	RB1
tb188	chr17	7675088	C	-C	13	4922	0.00264	206494	0	Tx1b	['C>T','C>A']	10	0	8740	R43fs	del	TP53
tb190	chr19	1220687	A	G	10	8352	0.00120	166730	0	BL	None	8	0	4984	K235R	SNV	STK11
tb190	chr3	89210218	T	G	12	1473	0.00815	161170	0	BL	None	11	0	5787	V171G	SNV	EPHA3
tb190	chr7	55181321	A	+C	14	5410	0.00259	218204	6.42E-05	Tx2	None	13	0	6035	N504fs	Ins	EGFR
tb191	chr17	7674872	T	C	12	4612	0.00260	175640	0	BL	['T>C']	9	0	4883	Y88C	SNV	TP53
tb191	chr17	7675224	G	A	24	6132	0.00391	190264	0	BL	None	20	0	5504	L91F	SNV	TP53
tb194	chr15	87880401	G	A	11	212	0.05189	184104	0	Tx1	None	10	0	9167	R713C	SNV	NTRK3
tb194	chr2	211387120	C	+A	15	244	0.06148	152652	1.31E-05	Tx1	None	7	0	9734	A1056fs	Ins	ERBB4
tb194	chr19	10499829	C	A	11	165	0.06667	176078	0	Tx1	None	4	0	8790	E69X	stopgain	KEAP1
tb194	chr19	10489800	C	A	15	156	0.09615	185662	0	Tx1	None	4	0	8728	R460M	SNV	KEAP1
tb194	chr2	50236893	G	A	14	131	0.10687	167310	0	Tx1	None	10	0	11413	R113X	stopgain	NRXN1
tb194	chr5	112840275	A	G	8	74	0.10811	183282	0	Tx1	None	7	0	11234	K1543E	SNV	APC
tb194	chr17	7675189	G	C	6	12201	0.00049	201622	0	Tx2	None	2	0	10675	C9W	SNV	TP53
tb196	chr7	55181378	C	T	32	835	0.03832	208326	0	Tx1	['C>T']	7	0	4481	T523M	SNV	EGFR
tb197	chr7	55191823	G	-G	2	9311	0.00022	170218	1.17E-05	BL	None	18	0	4982	L591fs	del	EGFR
tb197	chr17	7674890	T	C	4	9047	0.00044	171764	0	BL	['T>C']	3	0	3797	H82R	SNV	TP53
tb197	chr7	55191823	G	-G	3	8368	0.00036	170218	1.17E-05	Tx3	None	26	0	4982	L591fs	del	EGFR
tb199	chr7	55191823	G	-G	2	8578	0.00023	170218	1.17E-05	Tx1	None	12	0	7408	L591fs	del	EGFR

Appendix B Table 6.3: Chapter IV Final mutation calls passing filter from all time points.

* Hotspots previously defined by Newman et al 2014 [24]

		Blood Draw Library[a]							
patient	cohort	BL	Tx1	Tx2	Tx3	Tx4	Tx5	Tx6	Tx7
tb169	NSCLC	2	0	12	3	-	-	-	-
tb174	NSCLC	0	0	-	-	-	-	-	-
tb179	NSCLC	0	1	1	-	-	-	-	-
tb181	NSCLC	0	0	0	-	-	-	-	-
tb182	NSCLC	0	1	0	-	-	-	-	-
tb184	NSCLC	0	0	0	1	0	0	0	-
tb187	NSCLC	1	0	0	0	1	1	0	0
tb188	NSCLC	0	1	0	-	-	-	-	-
tb190	NSCLC	2	0	1	-	-	-	-	-
tb191	NSCLC	2	0	0	-	-	-	-	-
tb194	NSCLC	0	6	1	-	-	-	-	-
tb196	NSCLC	0	1	0	0	0	-	-	-
tb197	NSCLC	2	0	0	1	-	-	-	-
tb199	NSCLC	0	1	0	-	-	-	-	-
	(n)	14	14	13	5	3	2	2	1
	ctDNA+	5	6	4	3	1	1	0	0
	Rate	0.36	0.43	0.31	0.60	0.33	0.50	0.00	0.00
	Cumul. ctDNA+	5	11	11	13	13	13	13	13
	Cumul. Rate	0.36	0.79	0.79	0.93	0.93	0.93	0.93	0.93
tb2	Healthy	0	-	-	-	-	-	-	-
tb3	Healthy	0	-	-	-	-	-	-	-
tb4	Healthy	0	-	-	-	-	-	-	-
tb5	Healthy	0	-	-	-	-	-	-	-
tb6	Healthy	7	-	-	-	-	-	-	-
tb11	Healthy	0	-	-	-	-	-	-	-
tb12	Healthy	0	-	-	-	-	-	-	-
tb67	Healthy	0	-	-	-	-	-	-	-
tb68	Healthy	0	-	-	-	-	-	-	-
tb71	Healthy	0	-	-	-	-	-	-	-
	(n)	10	-	-	-	-	-	-	-
	ctDNA+	1	-	-	-	-	-	-	-
	Rate	0.10	-	-	-	-	-	-	-

Appendix B Table 6.4: Chapter IV Number of mutation calls passing filters for each blood draw

[a] All libraries were assessed at full-depth using error rate thresholds determined using our probabilistic approach and the library's mean depth

Patient	library	mean depth	call count[a]
tb11	BL	5795	0
tb12	BL	6379	0
tb169	BL	1090	2
tb169	Tx2	1795.1	3
tb169	Tx4	1163	12
tb169	Tx6	4483	0
tb174	Tx1	4495	0
tb174	Tx7	4638	0
tb179	BL	4986	1
tb179	Tx2	6695	0
tb179	Tx6	6438	1
tb181	Tx1	6583	0
tb181	Tx3	7360	0
tb181	Tx7	7805	0
tb182	BL	1176	0
tb182	Tx2	2400	1
tb182	Tx4	5128	0
tb184	Tx1	3304	0
tb184	Tx1	4531	0
tb184	Tx3	5887.5	0
tb184	Tx3	5895	0
tb184	Tx5	4932	0
tb184	Tx5	8092.1	0
tb184	Tx7	7591.2	1
tb187	BL	4810	0
tb187	BL	4917.2	0
tb187	Tx2	1758	0
tb187	Tx2	4706	0
tb187	Tx4	3465	0
tb187	Tx4	5410.6	1
tb187	Tx6	3706	1
tb187	Tx6	4239.9	1
tb188	Tx3	2772	0
tb188	Tx5	2849	1
tb188	Tx7	689	0
tb190	BL	2174	2
tb190	Tx4	2695	1
tb190	Tx6	1945	0
tb191	Tx1	3106	2
tb191	Tx5	7656	0
tb191	Tx7	6080	0
tb194	BL	9273	6
tb194	Tx2	943	0
tb194	Tx6	2416.4	1
tb196	Tx1	250	1
tb196	Tx3	439.3	0
tb196	Tx3	618	0
tb196	Tx5	459.2	0
tb196	Tx7	477	0
tb197	BL	8108	0
tb197	Tx2	7376	0
tb197	Tx4	8259	2
tb197	Tx6	7268.1	1
tb199	Tx1	417	0
tb199	Tx3	7379	1
tb199	Tx5	5669	0
tb2	BL	7976	0
tb3	BL	4607	0
tb4	BL	7159	0
tb5	BL	8054	0
tb6	BL	10458	7
tb67	BL	10175	0
tb68	BL	6873	0
tb71	BL	9703	0

Appendix B Table 6.5: Chapter IV Library mean consensus sequencing read depth (SSCS) and mutation calls passing filters

[a] All libraries were assessed at full-depth using error rate thresholds determined using our probabilistic approach and the library's mean depth

							cell-free DNA**			WBC DNA***		
patient	mutation	blood draw(s) called in*	hotspot?	type	gene	amino acid change	supporting reads	total reads	MAF	supporting reads	total reads	MAF
tb169	chr10:43112867:T>+A	Tx2	FALSE	insertion	RET	F301fs	11	3022	0.0036	0	17363	0
tb169	chr10:87957859:A>T	Tx2	FALSE	SNV	PTEN	Q214L	17	644	0.0264	0	13067	0
tb169	chr11:108329141:T>-T	Tx2	FALSE	deletion	ATM	Y2404fs	12	843	0.0142	0	14018	0
tb169	chr14:104780190:G>C	Tx2	FALSE	SNV	AKT1	R25G	9	1642	0.0055	0	12576	0
tb169	chr2:211722503:C>G	Tx2	FALSE	SNV	ERBB4	C258S	19	459	0.0414	0	13404	0
tb169	chr2:211725102:A>G	Tx2	FALSE	SNV	ERBB4	S239P	11	744	0.0148	0	12886	0
tb169	chr3:89450279:A>C	Tx2	FALSE	SNV	EPHA3	N867H	12	913	0.0131	0	15363	0
tb169	chr4:55110472:A>C	Tx3	FALSE	SNV	KDR	Y396D	8	1069	0.0075	0	15720	0
tb169	chr4:65367386:T>G	Tx3	FALSE	SNV	EPHA5	E611A	17	997	0.0171	0	11340	0
tb169	chr5:177095368:G>C	BL, **Tx2**	FALSE	SNV	FGFR4	E480Q	20	3088	0.0065	0	13325	0
tb169	chr5:177095372:T>G	Tx2	FALSE	SNV	FGFR4	M481R	21	2872	0.0073	0	13285	0
tb169	chr6:117319892:A>T	Tx2	FALSE	SNV	ROS1	S1972R	12	895	0.0134	0	15019	0
tb169	chr6:117326279:A>T	**BL**, Tx2	FALSE	SNV	ROS1	N1834K	23	449	0.0512	0	12861	0
tb169	chr7:55191821:C>G	**Tx2**, Tx3	TRUE	SNV	EGFR	L591V	4	1715	0.0023	0	12625	0
tb174	chr3:89341914:G>A	WBC	FALSE	SNV	EPHA3	S377N	0	5268	0	8	7818	0.0010
tb174	chr4:65324122:G>A	WBC	FALSE	SNV	EPHA5	P1015S	0	4259	0	7	7386	0.0009
tb179	chr1:247712671:T>+AG	Tx1, WBC	FALSE	insertion	OR6F1	M29fs	13	6197	0.0021	4	6297	0.0006
tb179	chr1:247712724:T>A	Tx2, WBC	FALSE	SNV	OR6F1	D11V	19	4007	0.0047	13	2930	0.0044
tb179	chr8:35748598:C>T	Tx1	FALSE	SNV	UNC5D	T608I	10	5891	0.0017	0	6790	0
tb181	chrX:111952329:G>C	WBC	FALSE	SNV	TRPC5	A31G	0	4277	0	40	2778	0.0144
tb182	chr5:112837901:A>C	Tx1	FALSE	SNV	APC	L751F	8	1300	0.0062	0	5773	0
tb184	chr17:31229289:A>G	Tx4, WBC	FALSE	SNV	NF1	S892G	13	10163	0.0013	3	1442	0.0021
tb184	chr17:31325889:T>A	WBC	FALSE	stopgain	NF1	Y1614X	0	5822	0	7	790	0.0089
tb184	chr17:31336771:A>G	WBC	FALSE	SNV	NF1	D2074G	0	6294	0	10	1255	0.0080
tb184	chr17:7674229:C>T	BL, Tx1, Tx2, Tx3, Tx4, Tx5, **Tx6**, WBC	TRUE	SNV	TP53	G113D	37	3140	0.0118	12	2040	0.006
tb184	chr4:55092634:A>T	WBC	FALSE	SNV	KDR	F1018I	0	5737	0	13	938	0.0139
tb184	chr7:41690315:G>C	WBC	FALSE	SNV	INHBA	L206V	0	6957	0	11	2753	0.0040
tb184	chr7:55174773:G>A	Tx3	TRUE	SNV	EGFR	E479K	4	8008	0.0005	0	1155	0
tb184	chr9:8460418:A>G	WBC	FALSE	SNV	PTPRD	Y869H	0	5528	0	10	826	0.0121
tb187	chr11:108304817:C>T	WBC	FALSE	SNV	ATM	T1880M	0	3367	0	67	722	0.0928
tb187	chr13:48303986:C>G	**BL**, Tx4, Tx5	FALSE	SNV	RB1	P25R	14	2283	0.0061	0	648	0
tb187	chr14:104780146:C>A	WBC	FALSE	SNV	AKT1	K39N	0	4199	0	70	1062	0.0659
tb187	chr15:87877088:C>A	WBC	FALSE	SNV	NTRK3	L767F	0	4300	0	49	789	0.0621
tb187	chr15:88135320:G>T	WBC	FALSE	SNV	NTRK3	P329T	0	5154	0	90	1527	0.0589
tb187	chr17:31327781:C>A	WBC	FALSE	SNV	NF1	P1830T	0	4162	0	40	799	0.0501
tb187	chr17:7675167:A>-AATC	BL, **Tx1**, Tx2, Tx3, Tx4, Tx5, Tx6, WBC	FALSE	deletion	TP53	D16fs	32	3823	0.0084	3	1480	0.002
tb187	chr2:212124847:G>A	WBC	FALSE	stopgain	ERBB4	R47X	0	3945	0	12	1062	0.0113
tb187	chr4:65490541:C>A	WBC	FALSE	SNV	EPHA5	R413M	0	3989	0	76	1303	0.0583
tb187	chr7:41700080:C>T	WBC	FALSE	SNV	INHBA	G99R	0	2493	0	14	1140	0.0123
tb187	chr9:136496739:G>T	WBC	FALSE	SNV	NOTCH1	P2334T	0	4848	0	49	889	0.0551
tb188	chr17:7674872:T>C	WBC	TRUE	SNV	TP53	Y88C	11	7498	0.0015	1	897	0.0011
tb188	chr17:7675088:C>-C	Tx1b	TRUE	deletion	TP53	R43fs	13	4922	0.0026	0	8740	0
tb190	chr19:1220687:A>G	BL	FALSE	SNV	STK11	K235R	10	8352	0.0012	0	4984	0
tb190	chr2:211679135:C>G	BL, WBC	FALSE	SNV	ERBB4	W513C	20	2079	0.0096	25	3791	0.0066
tb190	chr3:89210218:T>G	BL	FALSE	SNV	EPHA3	V171G	12	1473	0.0081	0	5787	0
tb190	chr7:55181321:A>+C	Tx2	FALSE	insertion	EGFR	N504fs	14	5410	0.0026	0	6035	0
tb191	chr17:7674872:T>C	BL	TRUE	SNV	TP53	Y88C	12	4612	0.0026	0	4883	0
tb191	chr17:7675224:G>A	BL	FALSE	SNV	TP53	L91F	24	6132	0.0039	0	5504	0
tb194	chr15:87880401:G>A	Tx1	FALSE	SNV	NTRK3	R713C	11	212	0.0519	0	9167	0
tb194	chr17:7675157:G>-G	WBC	TRUE	deletion	TP53	P20fs	0	999	0	13	10112	0.0013
tb194	chr17:7675189:G>C	Tx2	FALSE	SNV	TP53	C9W	6	12201	0.0005	0	10675	0
tb194	chr19:10489800:C>A	Tx1	FALSE	SNV	KEAP1	R460M	15	156	0.0962	0	8728	0
tb194	chr19:10499829:C>A	Tx1	FALSE	stopgain	KEAP1	E69X	11	165	0.0667	0	8790	0
tb194	chr2:211387120:C>+A	Tx1	FALSE	insertion	ERBB4	A1056fs	15	244	0.0615	0	9734	0
tb194	chr2:50236893:G>A	Tx1	FALSE	stopgain	NRXN1	R113X	14	131	0.1069	0	11413	0
tb194	chr5:112840275:A>G	Tx1	FALSE	SNV	APC	K1543E	8	74	0.1081	0	11234	0
tb194	chr8:3307815:C>T	Tx2, **Tx3**, WBC	FALSE	SNV	CSMD1	C1277Y	56	9414	0.0059	32	10480	0.0031
tb194	chr9:136496759:C>T	WBC	FALSE	SNV	NOTCH1	R2327Q	0	968	0	13	9681	0.0013
tb194	chrX:79363078:C>T	WBC	FALSE	SNV	ITM2A	R58H	0	449	0	16	5208	0.0031
tb196	chr17:7675217:T>C	Tx1, WBC	TRUE	SNV	TP53	K93R	5	608	0.0082	9	2847	0.0032
tb196	chr5:112840518:T>G	WBC	FALSE	SNV	APC	Y1624D	0	581	0	12	2182	0.0055
tb196	chr7:55181378:C>T	Tx1	TRUE	SNV	EGFR	T523M	32	835	0.0383	0	4481	0
tb196	chr9:136497164:T>C	WBC	FALSE	SNV	NOTCH1	D2192G	0	2295	0	10	2617	0.0038
tb196	chr9:136502441:C>A	WBC	FALSE	SNV	NOTCH1	V1739L	0	3247	0	25	4393	0.0057
tb199	chr15:87929380:C>A	WBC	FALSE	SNV	NTRK3	K550N	0	7258	0	252	5028	0.0501
tb199	chr15:88136563:C>A	WBC	FALSE	SNV	NTRK3	E223D	0	6299	0	325	6215	0.0523
tb199	chr17:31229154:C>A	WBC	FALSE	SNV	NF1	L847I	0	3687	0	139	2549	0.0545
tb199	chr4:55106747:C>A	WBC	FALSE	SNV	KDR	Q492H	0	5385	0	265	5082	0.0521
tb199	chr5:177093285:C>A	WBC	FALSE	SNV	FGFR4	P402Q	0	2564	0	149	2062	0.0723
tb199	chr7:55191823:G>-G	Tx1, **Tx2**, WBC	FALSE	deletion	EGFR	L591fs	2	8578	0.0002	0	7408	0
tb199	chr9:136496137:C>A	WBC	FALSE	SNV	NOTCH1	E2534D	0	7119	0	510	6459	0.0790
tb199	chr9:136497526:C>A	WBC	FALSE	SNV	NOTCH1	E2071D	0	4368	0	196	3886	0.0504

(continued on next page)

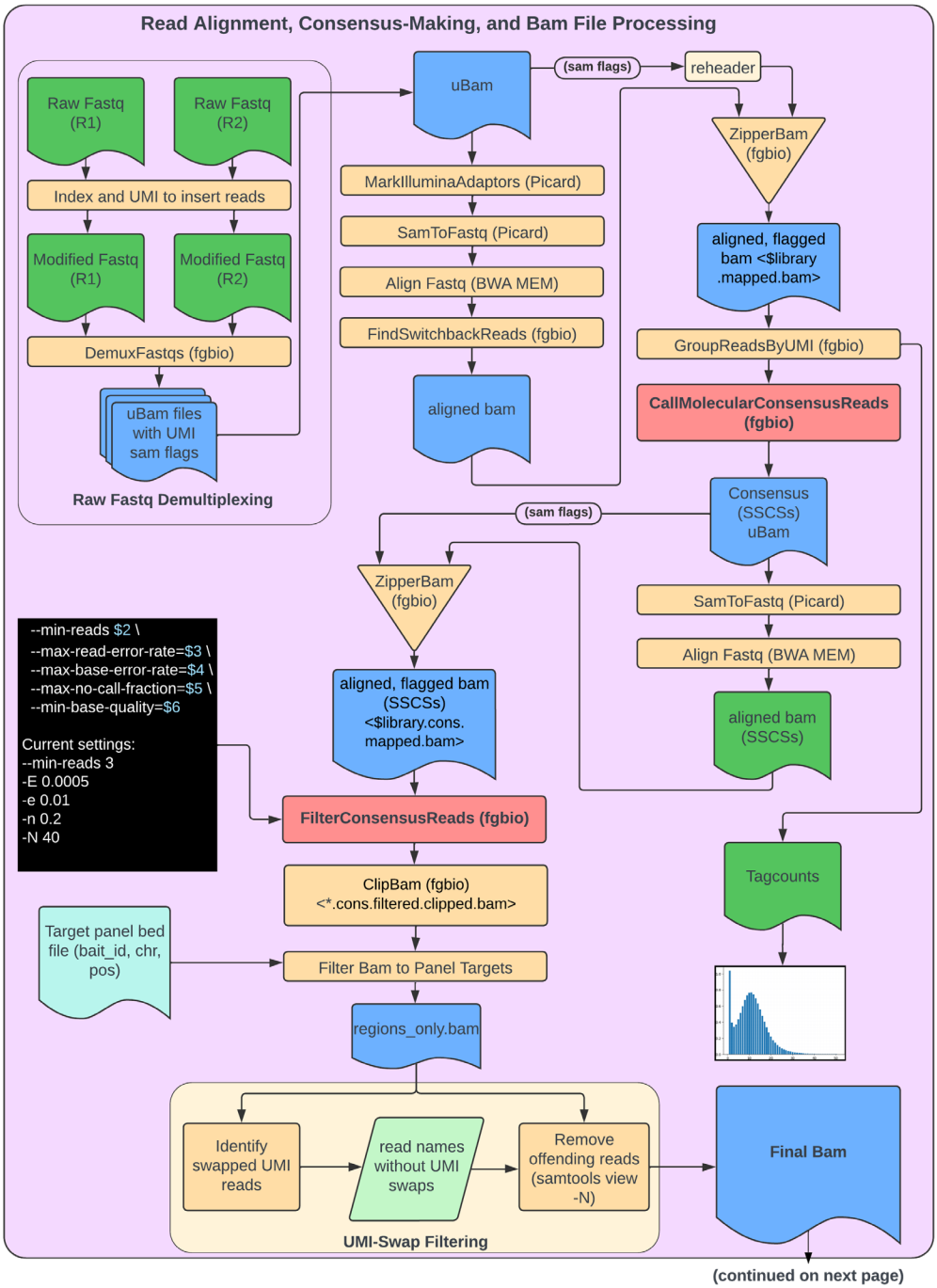
(continued from page 136)

Appendix B Table 6.6: Chapter IV Mutation call passing filters from full-sized cell-free DNA and WBC DNA libraries

Note: cell-free DNA calls were note filtered for WBC presence (as done for detection and dynamics analysis)

* Cell-free DNA library with maximum mutant allele frequency (MAF) is bolded
** If mutation was called in multiple blood draws, read counts are shown for the cell-free DNA library with the maximum MAF
*** WBC DNA from baseline (BL) blood draw was used to construct sequencing libraries

Appendix C: Computational workflow schematic for low-MAF mutation detection and characterization from Chapter IV



Appendix C Figure 7.1. Complete schematic of low-MAF *de novo* mutation calling computational workflow.

The schematic shown is an expanded version of that shown in Fig. 4.2. Note that color coding is not applicable in alignment workflow on previous page.

References

[1] J. Moss, J. Magenheim, D. Neiman, H. Zemmour, N. Loyfer, A. Korach, Y. Samet, M. Maoz, H. Druid, P. Arner, K.Y. Fu, E. Kiss, K.L. Spalding, G. Landesberg, A. Zick, A. Grinshpun, A.M.J. Shapiro, M. Grompe, A.D. Wittenberg, B. Glaser, R. Shemer, T. Kaplan, Y. Dor, Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease, Nat Commun 9(1) (2018) 5068.

[2] P. Mandel, P. Metais, [Nuclear Acids In Human Blood Plasma], C R Seances Soc Biol Fil 142(3-4) (1948) 241-3.

[3] J. Aucamp, A.J. Bronkhorst, C.P.S. Badenhorst, P.J. Pretorius, The diverse origins of circulating cell-free DNA in the human body: a critical re-evaluation of the literature, Biol Rev Camb Philos Soc 93(3) (2018) 1649-1683.

[4] T. Watanabe, S. Takada, R. Mizuta, Cell-free DNA in blood circulation is generated by DNase1L3 and caspase-activated DNase, Biochem Biophys Res Commun 516(3) (2019) 790-795.

[5] F. Stephan, G. Marsman, L.M. Bakker, I. Bulder, F. Stavenuiter, L.A. Aarden, S. Zeerleder, Cooperation of factor VII-activating protease and serum DNase I in the release of nucleosomes from necrotic cells, Arthritis Rheumatol 66(3) (2014) 686-93.

[6] S.N. Tamkovich, A.V. Cherepanova, E.V. Kolesnikova, E.Y. Rykova, D.V. Pyshnyi, V.V. Vlassov, P.P. Laktionov, Circulating DNA and DNase activity in human blood, Ann N Y Acad Sci 1075 (2006) 191-6.

[7] D.S.C. Han, Y.M.D. Lo, The Nexus of cfDNA and Nuclease Biology, Trends Genet 37(8) (2021) 758-770.

[8] K. Chen, H. Zhao, Y. Shi, F. Yang, L.T. Wang, G. Kang, Y. Nie, J. Wang, Perioperative Dynamic Changes in Circulating Tumor DNA in Patients with Lung Cancer (DYNAMIC), Clin Cancer Res 25(23) (2019) 7058-7067.

[9] X.Y. Zhong, M.R. Burk, C. Troeger, A. Kang, W. Holzgreve, S. Hahn, Fluctuation of maternal and fetal free extracellular circulatory DNA in maternal plasma, Obstet Gynecol 96(6) (2000) 991-6.

[10] K. Brodbeck, S. Schick, B. Bayer, K. Anslinger, K. Kruger, Z. Mayer, S. Holdenrieder, S. Peldschus, Biological variability of cell-free DNA in healthy females at rest within a short time course, Int J Legal Med 134(3) (2020) 911-919.

[11] A.T. Madsen, J.A. Hojbjerg, B.S. Sorensen, A. Winther-Larsen, Day-to-day and within-day biological variation of cell-free DNA, EBioMedicine 49 (2019) 284-290.

[12] J.A. Hojbjerg, A.T. Madsen, H.H. Schmidt, S.F. Sorensen, M. Stougaard, P. Meldgaard, B.S. Sorensen, Intra-individual variation of circulating tumour DNA in lung cancer patients, Mol Oncol 13(10) (2019) 2098-2106.

[13] J.T. Wagner, H.J. Kim, K.C. Johnson-Camacho, T. Kelley, L.F. Newell, P.T. Spellman, T.T.M. Ngo, Diurnal stability of cell-free DNA and cell-free RNA in human plasma samples, Sci Rep 10(1) (2020) 16456.

[14] L. Alghofaili, H. Almubarak, K. Gassem, S.S. Islam, S. Coskun, N. Kaya, B. Karakas, Cell-free DNA levels of twins and sibling pairs indicate individuality and possible use as a personalized biomarker, PLoS One 14(10) (2019) e0223470.

[15] A.J. Bronkhorst, V. Ungerer, S. Holdenrieder, Early detection of cancer using circulating tumor DNA: biological, physiological and analytical considerations, Crit Rev Clin Lab Sci (2019) 1-17.

[16] E.A. Klein, D. Richards, A. Cohn, M. Tummala, R. Lapham, D. Cosgrove, G. Chung, J. Clement, J. Gao, N. Hunkapiller, A. Jamshidi, K.N. Kurtzman, M.V. Seiden, C. Swanton, M.C. Liu, Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set, Ann Oncol 32(9) (2021) 1167-1177.

[17] T.M. Butler, C.T. Boniface, K. Johnson-Camacho, S. Tabatabaei, D. Melendez, T. Kelley, J. Gray, C.L. Corless, P.T. Spellman, Circulating tumor DNA dynamics using patient-customized assays are associated with outcome in neoadjuvantly treated breast cancer, Cold Spring Harb Mol Case Stud 5(2) (2019).

[18] S. Avanzini, D.M. Kurtz, J.J. Chabon, E.J. Moding, S.S. Hori, S.S. Gambhir, A.A. Alizadeh, M. Diehn, J.G. Reiter, A mathematical model of ctDNA shedding predicts tumor detection size, Sci Adv 6(50) (2020).

[19] P. Razavi, B.T. Li, D.N. Brown, B. Jung, E. Hubbell, R. Shen, W. Abida, K. Juluru, I. De Bruijn, C. Hou, O. Venn, R. Lim, A. Anand, T. Maddala, S. Gnerre, R. Vijaya Satya, Q. Liu, L. Shen, N. Eattock, J. Yue, A.W. Blocker, M. Lee, A. Sehnert, H. Xu, M.P. Hall, A. Santiago-Zayas, W.F. Novotny, J.M. Isbell, V.W. Rusch, G. Plitas, A.S. Heerdt, M. Ladanyi, D.M. Hyman, D.R. Jones, M. Morrow, G.J. Riely, H.I. Scher, C.M. Rudin, M.E. Robson, L.A. Diaz, Jr., D.B. Solit, A.M. Aravanis, J.S. Reis-Filho, High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants, Nat Med 25(12) (2019) 1928-1937.

[20] D. Ferreira, J. Miranda, P. Martins-Lopes, F. Adega, R. Chaves, Future Perspectives in Detecting EGFR and ALK Gene Alterations in Liquid Biopsies of Patients with NSCLC, Int J Mol Sci 22(8) (2021).

[21] V.A. Adalsteinsson, G. Ha, S.S. Freeman, A.D. Choudhury, D.G. Stover, H.A. Parsons, G. Gydush, S.C. Reed, D. Rotem, J. Rhoades, D. Loginov, D. Livitz, D. Rosebrock, I. Leshchiner, J. Kim, C. Stewart, M. Rosenberg, J.M. Francis, C.Z. Zhang, O. Cohen, C. Oh, H. Ding, P. Polak, M. Lloyd, S. Mahmud, K. Helvie, M.S. Merrill, R.A. Santiago, E.P. O'Connor, S.H.

Jeong, R. Leeson, R.M. Barry, J.F. Kramkowski, Z. Zhang, L. Polacek, J.G. Lohr, M. Schleicher, E. Lipscomb, A. Saltzman, N.M. Oliver, L. Marini, A.G. Waks, L.C. Harshman, S.M. Tolaney, E.M. Van Allen, E.P. Winer, N.U. Lin, M. Nakabayashi, M.E. Taplin, C.M. Johannessen, L.A. Garraway, T.R. Golub, J.S. Boehm, N. Wagle, G. Getz, J.C. Love, M. Meyerson, Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors, Nat Commun 8(1) (2017) 1324.

[22] M.W. Snyder, M. Kircher, A.J. Hill, R.M. Daza, J. Shendure, Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin, Cell 164(1-2) (2016) 57-68.

[23] E. Heitzer, I.S. Haque, C.E.S. Roberts, M.R. Speicher, Current and future perspectives of liquid biopsies in genomics-driven oncology, Nat Rev Genet 20(2) (2019) 71-88.

[24] A.M. Newman, S.V. Bratman, J. To, J.F. Wynne, N.C. Eclov, L.A. Modlin, C.L. Liu, J.W. Neal, H.A. Wakelee, R.E. Merritt, J.B. Shrager, B.W. Loo, Jr., A.A. Alizadeh, M. Diehn, An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage, Nat Med 20(5) (2014) 548-54.

[25] A.M. Newman, A.F. Lovejoy, D.M. Klass, D.M. Kurtz, J.J. Chabon, F. Scherer, H. Stehr, C.L. Liu, S.V. Bratman, C. Say, L. Zhou, J.N. Carter, R.B. West, G.W. Sledge, J.B. Shrager, B.W. Loo, Jr., J.W. Neal, H.A. Wakelee, M. Diehn, A.A. Alizadeh, Integrated digital error suppression for improved detection of circulating tumor DNA, Nat Biotechnol 34(5) (2016) 547-555.

[26] J.J. Chabon, E.G. Hamilton, D.M. Kurtz, M.S. Esfahani, E.J. Moding, H. Stehr, J. Schroers-Martin, B.Y. Nabet, B. Chen, A.A. Chaudhuri, C.L. Liu, A.B. Hui, M.C. Jin, T.D. Azad, D. Almanza, Y.J. Jeon, M.C. Nesselbush, L. Co Ting Keh, R.F. Bonilla, C.H. Yoo, R.B. Ko, E.L. Chen, D.J. Merriott, P.P. Massion, A.S. Mansfield, J. Jen, H.Z. Ren, S.H. Lin, C.L. Costantino, R. Burr, R. Tibshirani, S.S. Gambhir, G.J. Berry, K.C. Jensen, R.B. West, J.W. Neal, H.A. Wakelee, B.W. Loo, Jr., C.A. Kunder, A.N. Leung, N.S. Lui, M.F. Berry, J.B. Shrager, V.S. Nair, D.A. Haber, L.V. Sequist, A.A. Alizadeh, M. Diehn, Integrating genomic features for non-invasive early lung cancer detection, Nature 580(7802) (2020) 245-251.

[27] C. Abbosh, N.J. Birkbak, G.A. Wilson, M. Jamal-Hanjani, T. Constantin, R. Salari, J. Le Quesne, D.A. Moore, S. Veeriah, R. Rosenthal, T. Marafioti, E. Kirkizlar, T.B.K. Watkins, N. McGranahan, S. Ward, L. Martinson, J. Riley, F. Fraioli, M. Al Bakir, E. Gronroos, F. Zambrana, R. Endozo, W.L. Bi, F.M. Fennessy, N. Sponer, D. Johnson, J. Laycock, S. Shafi, J. Czyzewska-Khan, A. Rowan, T. Chambers, N. Matthews, S. Turajlic, C. Hiley, S.M. Lee, M.D. Forster, T. Ahmad, M. Falzon, E. Borg, D. Lawrence, M. Hayward, S. Kolvekar, N. Panagiotopoulos, S.M. Janes, R. Thakrar, A. Ahmed, F. Blackhall, Y. Summers, D. Hafez, A. Naik, A. Ganguly, S. Kareht, R. Shah, L. Joseph, A. Marie Quinn, P.A. Crosbie, B.

Naidu, G. Middleton, G. Langman, S. Trotter, M. Nicolson, H. Remmen, K. Kerr, M. Chetty, L. Gomersall, D.A. Fennell, A. Nakas, S. Rathinam, G. Anand, S. Khan, P. Russell, V. Ezhil, B. Ismail, M. Irvin-Sellers, V. Prakash, J.F. Lester, M. Kornaszewska, R. Attanoos, H. Adams, H. Davies, D. Oukrif, A.U. Akarca, J.A. Hartley, H.L. Lowe, S. Lock, N. Iles, H. Bell, Y. Ngai, G. Elgar, Z. Szallasi, R.F. Schwarz, J. Herrero, A. Stewart, S.A. Quezada, K.S. Peggs, P. Van Loo, C. Dive, C.J. Lin, M. Rabinowitz, H. Aerts, A. Hackshaw, J.A. Shaw, B.G. Zimmermann, T.R. consortium, P. consortium, C. Swanton, Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution, Nature 545(7655) (2017) 446-451.

[28] V. Potapov, J.L. Ong, Examining Sources of Error in PCR by Single-Molecule Sequencing, PLoS One 12(1) (2017) e0169774.

[29] N. Stoler, A. Nekrutenko, Sequencing error profiles of Illumina sequencing instruments, NAR Genom Bioinform 3(1) (2021) lqab019.

[30] S.R. Kennedy, M.W. Schmitt, E.J. Fox, B.F. Kohrn, J.J. Salk, E.H. Ahn, M.J. Prindle, K.J. Kuong, J.C. Shen, R.A. Risques, L.A. Loeb, Detecting ultralow-frequency mutations by Duplex Sequencing, Nat Protoc 9(11) (2014) 2586-606.

[31] C. Boniface, C. Deig, C. Halsey, T. Kelley, M.B. Heskett, C.R. Thomas, Jr., P.T. Spellman, N. Nabavizadeh, The Feasibility of Patient-Specific Circulating Tumor DNA Monitoring throughout Multi-Modality Therapy for Locally Advanced Esophageal and Rectal Cancer: A Potential Biomarker for Early Detection of Subclinical Disease, Diagnostics (Basel) 11(1) (2021).

[32] B.E. Johnson, A.L. Creason, J.M. Stommel, J.M. Keck, S. Parmar, C.B. Betts, A. Blucher, C. Boniface, E. Bucher, E. Burlingame, T. Camp, K. Chin, J. Eng, J. Estabrook, H.S. Feiler, M.B. Heskett, Z. Hu, A. Kolodzie, B.L. Kong, M. Labrie, J. Lee, P. Leyshock, S. Mitri, J. Patterson, J.L. Riesterer, S. Sivagnanam, J. Somers, D. Sudar, G. Thibault, B.R. Weeder, C. Zheng, X. Nan, R.F. Thompson, L.M. Heiser, P.T. Spellman, G. Thomas, E. Demir, Y.H. Chang, L.M. Coussens, A.R. Guimaraes, C. Corless, J. Goecks, R. Bergan, Z. Mitri, G.B. Mills, J.W. Gray, An omic and multidimensional spatial atlas from serial biopsies of an evolving metastatic breast cancer, Cell Rep Med 3(2) (2022) 100525.

[33] M. Labrie, A. Li, A. Creason, C. Betts, J. Keck, B. Johnson, S. Sivagnanam, C. Boniface, H. Ma, A. Blucher, Y.H. Chang, K. Chin, J. Vuky, A.R. Guimaraes, M. Downey, J.Y. Lim, L. Gao, K. Siex, S. Parmar, A. Kolodzie, P.T. Spellman, J. Goecks, L.M. Coussens, C.L. Corless, R. Bergan, J.W. Gray, G.B. Mills, Z.I. Mitri, Multiomics analysis of serial PARP inhibitor treated metastatic TNBC inform on rational combination therapies, NPJ Precis Oncol 5(1) (2021) 92.

[34] C. Boniface, K. Baker, C. Dieg, C. Halsey, R. Rahmani, M. Deffenbach, C. Thomas, G. Tormoen, N. Nabavizadeh, P. Spellman, Abstract PR05: Radiation-assisted Amplification Sequencing (RAMP-Seq): Evaluating the use of stereotactic body radiation therapy (SBRT) for enriching circulating tumor DNA in liquid biopsies, Clinical Cancer Research 26(11 Supplement) (2020) PR05-PR05.

[35] C. Bettegowda, M. Sausen, R.J. Leary, I. Kinde, Y. Wang, N. Agrawal, B.R. Bartlett, H. Wang, B. Luber, R.M. Alani, E.S. Antonarakis, N.S. Azad, A. Bardelli, H. Brem, J.L. Cameron, C.C. Lee, L.A. Fecher, G.L. Gallia, P. Gibbs, D. Le, R.L. Giuntoli, M. Goggins, M.D. Hogarty, M. Holdhoff, S.M. Hong, Y. Jiao, H.H. Juhl, J.J. Kim, G. Siravegna, D.A. Laheru, C. Lauricella, M. Lim, E.J. Lipson, S.K. Marie, G.J. Netto, K.S. Oliner, A. Olivi, L. Olsson, G.J. Riggins, A. Sartore-Bianchi, K. Schmidt, M. Shih I, S.M. Oba-Shinjo, S. Siena, D. Theodorescu, J. Tie, T.T. Harkins, S. Veronese, T.L. Wang, J.D. Weingart, C.L. Wolfgang, L.D. Wood, D. Xing, R.H. Hruban, J. Wu, P.J. Allen, C.M. Schmidt, M.A. Choti, V.E. Velculescu, K.W. Kinzler, B. Vogelstein, N. Papadopoulos, L.A. Diaz, Jr., Detection of circulating tumor DNA in early- and late-stage human malignancies, Sci Transl Med 6(224) (2014) 224ra24.

[36] C.A. Parkinson, D. Gale, A.M. Piskorz, H. Biggs, C. Hodgkin, H. Addley, S. Freeman, P. Moyle, E. Sala, K. Sayal, K. Hosking, I. Gounaris, M. Jimenez-Linan, H.M. Earl, W. Qian, N. Rosenfeld, J.D. Brenton, Exploratory Analysis of TP53 Mutations in Circulating Tumour DNA as Biomarkers of Treatment Response for Patients with Relapsed High-Grade Serous Ovarian Carcinoma: A Retrospective Study, PLoS Med 13(12) (2016) e1002198.

[37] M. Heidary, M. Auer, P. Ulz, E. Heitzer, E. Petru, C. Gasch, S. Riethdorf, O. Mauermann, I. Lafer, G. Pristauz, S. Lax, K. Pantel, J.B. Geigl, M.R. Speicher, The dynamic range of circulating tumor DNA in metastatic breast cancer, Breast Cancer Res 16(4) (2014) 421.

[38] C.T. Boniface, P.T. Spellman, Blood, Toil, and Taxoteres: Biological Determinates of Treatment-Induce ctDNA Dynamics for Interpreting Tumor Response, Pathol Oncol Res 28 (2022) 1610103.

[39] J.C.M. Wan, C. Massie, J. Garcia-Corbacho, F. Mouliere, J.D. Brenton, C. Caldas, S. Pacey, R. Baird, N. Rosenfeld, Liquid biopsies come of age: towards implementation of circulating tumour DNA, Nat Rev Cancer 17(4) (2017) 223-238.

[40] E. Heitzer, L. Auinger, M.R. Speicher, Cell-Free DNA and Apoptosis: How Dead Cells Inform About the Living, Trends Mol Med 26(5) (2020) 519-528.

[41] J.D. Merker, G.R. Oxnard, C. Compton, M. Diehn, P. Hurley, A.J. Lazar, N. Lindeman, C.M. Lockwood, A.J. Rai, R.L. Schilsky, A.M. Tsimberidou, P. Vasalos, B.L. Billman, T.K. Oliver, S.S. Bruinooge, D.F. Hayes, N.C.

Turner, Circulating Tumor DNA Analysis in Patients With Cancer: American Society of Clinical Oncology and College of American Pathologists Joint Review, J Clin Oncol 36(16) (2018) 1631-1641.

[42] L. Keller, Y. Belloum, H. Wikman, K. Pantel, Clinical relevance of blood-based ctDNA analysis: mutation detection and beyond, Br J Cancer 124(2) (2021) 345-358.

[43] J.J. Adashek, F. Janku, R. Kurzrock, Signed in Blood: Circulating Tumor DNA in Cancer Diagnosis, Treatment and Screening, Cancers (Basel) 13(14) (2021).

[44] S. De Michino, M. Aparnathi, A. Rostami, B.H. Lok, S.V. Bratman, The Utility of Liquid Biopsies in Radiation Oncology, Int J Radiat Oncol Biol Phys 107(5) (2020) 873-886.

[45] S. Khier, L. Lohan, Kinetics of circulating cell-free DNA for biomedical applications: critical appraisal of the literature, Future Sci OA 4(4) (2018) FSO295.

[46] M.J.M. Magbanua, L.B. Swigart, H.T. Wu, G.L. Hirst, C. Yau, D.M. Wolf, A. Tin, R. Salari, S. Shchegrova, H. Pawar, A.L. Delson, A. DeMichele, M.C. Liu, A.J. Chien, D. Tripathy, S. Asare, C.J. Lin, P. Billings, A. Aleshin, H. Sethi, M. Louie, B. Zimmermann, L.J. Esserman, L.J. van 't Veer, Circulating tumor DNA in neoadjuvant-treated breast cancer reflects response and survival, Ann Oncol 32(2) (2021) 229-239.

[47] S. Morbelli, A. Alama, G. Ferrarazzo, S. Coco, C. Genova, E. Rijavec, F. Bongioanni, F. Biello, M.G. Dal Bello, G. Barletta, M. Massollo, I. Vanni, R. Piva, A. Nieri, M. Bauckneht, G. Sambuceti, F. Grossi, 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 58(11) (2017) 1764-1769.

[48] D.C. Garcia-Olmo, J. Samos, M.G. Picazo, A.I. Asensio, I. Toboso, D. Garcia-Olmo, Release of cell-free DNA into the bloodstream leads to high levels of non-tumor plasma DNA during tumor progression in rats, Cancer Lett 272(1) (2008) 133-40.

[49] L. Xi, T.H. Pham, E.C. Payabyab, R.M. Sherry, S.A. Rosenberg, M. Raffeld, Circulating Tumor DNA as an Early Indicator of Response to T-cell Transfer Immunotherapy in Metastatic Melanoma, Clin Cancer Res 22(22) (2016) 5480-5486.

[50] Y.M. Lo, S.F. Leung, L.Y. Chan, A.T. Chan, K.W. Lo, P.J. Johnson, D.P. Huang, Kinetics of plasma Epstein-Barr virus DNA during radiation therapy for nasopharyngeal carcinoma, Cancer Res 60(9) (2000) 2351-5.

[51] F.J. Hilke, F. Muyas, J. Admard, B. Kootz, D. Nann, S. Welz, O. Riess, D. Zips, S. Ossowski, C. Schroeder, K. Clasen, Dynamics of cell-free tumour DNA correlate with treatment response of head and neck cancer patients receiving radiochemotherapy, Radiother Oncol 151 (2020) 182-189.

[52] Q. Zhang, J. Luo, S. Wu, H. Si, C. Gao, W. Xu, S.E. Abdullah, B.W. Higgs, P.A. Dennis, M.S. van der Heijden, N.H. Segal, J.E. Chaft, T. Hembrough, J.C. Barrett, M.D. Hellmann, Prognostic and Predictive Impact of Circulating Tumor DNA in Patients with Advanced Cancers Treated with Immune Checkpoint Blockade, Cancer Discov 10(12) (2020) 1842-1853.

[53] Y. Song, C. Hu, Z. Xie, L. Wu, Z. Zhu, C. Rao, L. Liu, Y. Chen, N. Liang, J. Chen, C. Hu, N. Yang, J. Hu, W. Zhao, G. Tong, X. Dong, D. Zheng, M. Jin, J. Chen, M. Huang, Y. He, R. Rosell, G. Lippi, M. Mino-Kenudson, H. Han-Zhang, X. Mao, L. Zhang, H. Liu, J.K. Field, S. Chuai, J. Ye, Y. Han, S. Lu, A.M.E.L.C.C.G. Written on behalf of, Circulating tumor DNA clearance predicts prognosis across treatment regimen in a large real-world longitudinally monitored advanced non-small cell lung cancer cohort, Transl Lung Cancer Res 9(2) (2020) 269-279.

[54] Y. Wang, L. Yang, H. Bao, X. Fan, F. Xia, J. Wan, L. Shen, Y. Guan, H. Bao, X. Wu, Y. Xu, Y. Shao, Y. Sun, T. Tong, X. Li, Y. Xu, S. Cai, J. Zhu, Z. Zhang, Utility of ctDNA in predicting response to neoadjuvant chemoradiotherapy and prognosis assessment in locally advanced rectal cancer: A prospective cohort study, PLoS Med 18(8) (2021) e1003741.

[55] V. Anagnostou, P.M. Forde, J.R. White, N. Niknafs, C. Hruban, J. Naidoo, K. Marrone, I.K.A. Sivakumar, D.C. Bruhm, S. Rosner, J. Phallen, A. Leal, V. Adleff, K.N. Smith, T.R. Cottrell, L. Rhymee, D.N. Palsgrove, C.L. Hann, B. Levy, J. Feliciano, C. Georgiades, F. Verde, P. Illei, Q.K. Li, E. Gabrielson, M.V. Brock, J.M. Isbell, J.L. Sauter, J. Taube, R.B. Scharpf, R. Karchin, D.M. Pardoll, J.E. Chaft, M.D. Hellmann, J.R. Brahmer, V.E. Velculescu, Dynamics of Tumor and Immune Responses during Immune Checkpoint Blockade in Non-Small Cell Lung Cancer, Cancer Res 79(6) (2019) 1214-1225.

[56] S. Xia, J. Ye, Y. Chen, A. Lizaso, L. Huang, L. Shi, J. Su, H. Han-Zhang, S. Chuai, L. Li, Y. Chen, Parallel serial assessment of somatic mutation and methylation profile from circulating tumor DNA predicts treatment response and impending disease progression in osimertinib-treated lung adenocarcinoma patients, Transl Lung Cancer Res 8(6) (2019) 1016-1028.

[57] R. Silva, B. Moran, A.M. Baird, C.J. O'Rourke, S.P. Finn, R. McDermott, W. Watson, W.M. Gallagher, D.J. Brennan, A.S. Perry, Longitudinal analysis of individual cfDNA methylome patterns in metastatic prostate cancer, Clin Epigenetics 13(1) (2021) 168.

[58] B. Gold, M. Cankovic, L.V. Furtado, F. Meier, C.D. Gocke, Do circulating tumor cells, exosomes, and circulating tumor nucleic acids have clinical utility? A report of the association for molecular pathology, J Mol Diagn 17(3) (2015) 209-24.

[59] S. Grabuschnig, A.J. Bronkhorst, S. Holdenrieder, I. Rosales Rodriguez, K.P. Schliep, D. Schwendenwein, V. Ungerer, C.W. Sensen, Putative

Origins of Cell-Free DNA in Humans: A Review of Active and Passive Nucleic Acid Release Mechanisms, Int J Mol Sci 21(21) (2020).

[60] A. Kustanovich, R. Schwartz, T. Peretz, A. Grinshpun, Life and death of circulating cell-free DNA, Cancer Biol Ther 20(8) (2019) 1057-1067.

[61] F.S. de Miranda, V.G. Barauna, L. Dos Santos, G. Costa, P.F. Vassallo, L.C.G. Campos, Properties and Application of Cell-Free DNA as a Clinical Biomarker, Int J Mol Sci 22(17) (2021).

[62] J. Phallen, M. Sausen, V. Adleff, A. Leal, C. Hruban, J. White, V. Anagnostou, J. Fiksel, S. Cristiano, E. Papp, S. Speir, T. Reinert, M.W. Orntoft, B.D. Woodward, D. Murphy, S. Parpart-Li, D. Riley, M. Nesselbush, N. Sengamalay, A. Georgiadis, Q.K. Li, M.R. Madsen, F.V. Mortensen, J. Huiskens, C. Punt, N. van Grieken, R. Fijneman, G. Meijer, H. Husain, R.B. Scharpf, L.A. Diaz, Jr., S. Jones, S. Angiuoli, T. Orntoft, H.J. Nielsen, C.L. Andersen, V.E. Velculescu, Direct detection of early-stage cancers using circulating tumor DNA, Sci Transl Med 9(403) (2017).

[63] F. Diehl, K. Schmidt, M.A. Choti, K. Romans, S. Goodman, M. Li, K. Thornton, N. Agrawal, L. Sokoll, S.A. Szabo, K.W. Kinzler, B. Vogelstein, L.A. Diaz, Jr., Circulating mutant DNA to assess tumor dynamics, Nat Med 14(9) (2008) 985-90.

[64] M.L. Cheng, E. Pectasides, G.J. Hanna, H.A. Parsons, A.D. Choudhury, G.R. Oxnard, Circulating tumor DNA in advanced solid tumors: Clinical relevance and future directions, CA Cancer J Clin 71(2) (2021) 176-190.

[65] J. Aucamp, A.J. Bronkhorst, D.L. Peters, H.C. Van Dyk, F.H. Van der Westhuizen, P.J. Pretorius, Kinetic analysis, size profiling, and bioenergetic association of DNA released by selected cell lines in vitro, Cell Mol Life Sci 74(14) (2017) 2689-2707.

[66] W. Wang, P. Kong, G. Ma, L. Li, J. Zhu, T. Xia, H. Xie, W. Zhou, S. Wang, Characterization of the release and biological significance of cell-free DNA from breast cancer cell lines, Oncotarget 8(26) (2017) 43180-43191.

[67] A. Rostami, M. Lambie, C.W. Yu, V. Stambolic, J.N. Waldron, S.V. Bratman, Senescence, Necrosis, and Apoptosis Govern Circulating Cell-free DNA Release Kinetics, Cell Rep 31(13) (2020) 107830.

[68] A. Otandault, J.D. Abraham, Z. Al Amir Dache, A. Khalyfa, I. Jariel-Encontre, T. Forne, C. Prevostel, S. Chouaib, D. Gozal, A.R. Thierry, Hypoxia differently modulates the release of mitochondrial and nuclear DNA, Br J Cancer 122(5) (2020) 715-725.

[69] A.J. Bronkhorst, J.F. Wentzel, J. Aucamp, E. van Dyk, L. du Plessis, P.J. Pretorius, Characterization of the cell-free DNA released by cultured cancer cells, Biochim Biophys Acta 1863(1) (2016) 157-65.

[70] A.J. Bronkhorst, J.F. Wentzel, V. Ungerer, D.L. Peters, J. Aucamp, E.P. de Villiers, S. Holdenrieder, P.J. Pretorius, Sequence analysis of cell-free

DNA derived from cultured human bone osteosarcoma (143B) cells, Tumour Biol 40(9) (2018) 1010428318801190.

[71] Z.Y. Hu, N. Xie, C. Tian, X. Yang, L. Liu, J. Li, H. Xiao, H. Wu, J. Lu, J. Gao, X. Hu, M. Cao, Z. Shui, M. Xiao, Y. Tang, Q. He, L. Chang, X. Xia, X. Yi, Q. Liao, Q. Ouyang, Identifying Circulating Tumor DNA Mutation Profiles in Metastatic Breast Cancer Patients with Multiline Resistance, EBioMedicine 32 (2018) 111-118.

[72] Y. Zhou, Y. Xu, Y. Gong, Y. Zhang, Y. Lu, C. Wang, R. Yao, P. Li, Y. Guan, J. Wang, X. Xia, L. Yang, X. Yi, Q. Sun, Clinical factors associated with circulating tumor DNA (ctDNA) in primary breast cancer, Mol Oncol 13(5) (2019) 1033-1046.

[73] E. Ortolan, V. Appierto, M. Silvestri, R. Miceli, S. Veneroni, S. Folli, G. Pruneri, A. Vingiani, A. Belfiore, V. Cappelletti, M. Vismara, F. Dell'Angelo, L. De Cecco, G.V. Bianchi, F.G. de Braud, M.G. Daidone, S. Di Cosimo, Blood-based genomics of triple-negative breast cancer progression in patients treated with neoadjuvant chemotherapy, ESMO Open 6(2) (2021) 100086.

[74] C. Abbosh, N.J. Birkbak, G.A. Wilson, M. Jamal-Hanjani, T. Constantin, R. Salari, J. Le Quesne, D.A. Moore, S. Veeriah, R. Rosenthal, T. Marafioti, E. Kirkizlar, T.B.K. Watkins, N. McGranahan, S. Ward, L. Martinson, J. Riley, F. Fraioli, M. Al Bakir, E. Gronroos, F. Zambrana, R. Endozo, W.L. Bi, F.M. Fennessy, N. Sponer, D. Johnson, J. Laycock, S. Shafi, J. Czyzewska-Khan, A. Rowan, T. Chambers, N. Matthews, S. Turajlic, C. Hiley, S.M. Lee, M.D. Forster, T. Ahmad, M. Falzon, E. Borg, D. Lawrence, M. Hayward, S. Kolvekar, N. Panagiotopoulos, S.M. Janes, R. Thakrar, A. Ahmed, F. Blackhall, Y. Summers, D. Hafez, A. Naik, A. Ganguly, S. Kareht, R. Shah, L. Joseph, A.M. Quinn, P.A. Crosbie, B. Naidu, G. Middleton, G. Langman, S. Trotter, M. Nicolson, H. Remmen, K. Kerr, M. Chetty, L. Gomersall, D.A. Fennell, A. Nakas, S. Rathinam, G. Anand, S. Khan, P. Russell, V. Ezhil, B. Ismail, M. Irvin-Sellers, V. Prakash, J.F. Lester, M. Kornaszewska, R. Attanoos, H. Adams, H. Davies, D. Oukrif, A.U. Akarca, J.A. Hartley, H.L. Lowe, S. Lock, N. Iles, H. Bell, Y. Ngai, G. Elgar, Z. Szallasi, R.F. Schwarz, J. Herrero, A. Stewart, S.A. Quezada, K.S. Peggs, P. Van Loo, C. Dive, C.J. Lin, M. Rabinowitz, H. Aerts, A. Hackshaw, J.A. Shaw, B.G. Zimmermann, C. Swanton, Corrigendum: Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution, Nature 554(7691) (2018) 264.

[75] D. Fornvik, K.E. Aaltonen, Y. Chen, A.M. George, C. Brueffer, R. Rigo, N. Loman, L.H. Saal, L. Ryden, Detection of circulating tumor cells and circulating tumor DNA before and after mammographic breast compression in a cohort of breast cancer patients scheduled for neoadjuvant treatment, Breast Cancer Res Treat 177(2) (2019) 447-455.

[76] K.Z. Chen, F. Lou, F. Yang, J.B. Zhang, H. Ye, W. Chen, T. Guan, M.Y. Zhao, X.X. Su, R. Shi, L. Jones, X.F. Huang, S.Y. Chen, J. Wang,

Circulating Tumor DNA Detection in Early-Stage Non-Small Cell Lung Cancer Patients by Targeted Sequencing, Sci Rep 6 (2016) 31985.

[77] M.S. Cho, C.H. Park, S. Lee, H.S. Park, Clinicopathological parameters for circulating tumor DNA shedding in surgically resected non-small cell lung cancer with EGFR or KRAS mutation, PLoS One 15(3) (2020) e0230622.

[78] A.C. McEvoy, L. Warburton, Z. Al-Ogaili, L. Celliers, L. Calapre, M.R. Pereira, M.A. Khattak, T.M. Meniawy, M. Millward, M. Ziman, E.S. Gray, Correlation between circulating tumour DNA and metabolic tumour burden in metastatic melanoma patients, BMC Cancer 18(1) (2018) 726.

[79] Z. Chen, A. Fadiel, F. Naftolin, K.D. Eichenbaum, Y. Xia, Circulation DNA: biological implications for cancer metastasis and immunology, Med Hypotheses 65(5) (2005) 956-61.

[80] V. Ungerer, A.J. Bronkhorst, P. Van den Ackerveken, M. Herzog, S. Holdenrieder, Serial profiling of cell-free DNA and nucleosome histone modifications in cell cultures, Sci Rep 11(1) (2021) 9460.

[81] M. Fleischhacker, B. Schmidt, Circulating nucleic acids (CNAs) and cancer--a survey, Biochim Biophys Acta 1775(1) (2007) 181-232.

[82] J.J. Choi, C.F. Reich, 3rd, D.S. Pisetsky, Release of DNA from dead and dying lymphocyte and monocyte cell lines in vitro, Scand J Immunol 60(1-2) (2004) 159-66.

[83] K. Utani, A. Okamoto, N. Shimizu, Generation of micronuclei during interphase by coupling between cytoplasmic membrane blebbing and nuclear budding, PLoS One 6(11) (2011) e27233.

[84] O.P. Kisurina-Evgenieva, O.I. Sutiagina, G.E. Onishchenko, Biogenesis of Micronuclei, Biochemistry (Mosc) 81(5) (2016) 453-64.

[85] H.D. Ryoo, A. Bergmann, The role of apoptosis-induced proliferation for regeneration and cancer, Cold Spring Harb Perspect Biol 4(8) (2012) a008797.

[86] R. Lugano, M. Ramachandran, A. Dimberg, Tumor angiogenesis: causes, consequences, challenges and opportunities, Cell Mol Life Sci 77(9) (2020) 1745-1770.

[87] J. Wang, A. Huang, Y.P. Wang, Y. Yin, P.Y. Fu, X. Zhang, J. Zhou, Circulating tumor DNA correlates with microvascular invasion and predicts tumor recurrence of hepatocellular carcinoma, Ann Transl Med 8(5) (2020) 237.

[88] H. Shirasu, H. Taniguchi, J. Watanabe, M. Kotaka, K. Yamazaki, K. Hirata, M. Yokota, Y. Emi, M. Ikenaga, K. Kato, N. Akazawa, T. Yamaguchi, M. Ikeda, A. Aleshin, D. Kotani, S. Mishima, H. Yukami, E. Oki, I. Takemasa, T. Kato, Y. Nakamura, T. Yoshino, O-11 Monitoring molecular residual disease by circulating tumor DNA in resectable colorectal cancer: Molecular subgroup analyses of a prospective observational study

GALAXY in CIRCULATE-Japan, Annals of Oncology 32 (2021) S222-S223.

[89] S.A. Nabavizadeh, J.B. Ware, S. Guiry, M.P. Nasrallah, J.J. Mays, J.E. Till, J. Hussain, A. Abdalla, S.S. Yee, Z.A. Binder, D.M. O'Rourke, S. Brem, A.S. Desai, R. Wolf, E.L. Carpenter, S.J. Bagley, Imaging and histopathologic correlates of plasma cell-free DNA concentration and circulating tumor DNA in adult patients with newly diagnosed glioblastoma, Neurooncol Adv 2(1) (2020) vdaa016.

[90] S.J. Bagley, S.A. Nabavizadeh, J.J. Mays, J.E. Till, J.B. Ware, S. Levy, W. Sarchiapone, J. Hussain, T. Prior, S. Guiry, T. Christensen, S.S. Yee, M.P. Nasrallah, J.J.D. Morrissette, Z.A. Binder, D.M. O'Rourke, A.J. Cucchiara, S. Brem, A.S. Desai, E.L. Carpenter, Clinical Utility of Plasma Cell-Free DNA in Adult Patients with Newly Diagnosed Glioblastoma: A Pilot Prospective Study, Clin Cancer Res 26(2) (2020) 397-407.

[91] R. Mair, F. Mouliere, C.G. Smith, D. Chandrananda, D. Gale, F. Marass, D.W.Y. Tsui, C.E. Massie, A.J. Wright, C. Watts, N. Rosenfeld, K.M. Brindle, Measurement of Plasma Cell-Free Mitochondrial Tumor DNA Improves Detection of Glioblastoma in Patient-Derived Orthotopic Xenograft Models, Cancer Res 79(1) (2019) 220-230.

[92] S.R. McKeown, Defining normoxia, physoxia and hypoxia in tumours-implications for treatment response, Br J Radiol 87(1035) (2014) 20130676.

[93] A.L. Harris, Hypoxia--a key regulatory factor in tumour growth, Nat Rev Cancer 2(1) (2002) 38-47.

[94] M.B. Giacona, G.C. Ruben, K.A. Iczkowski, T.B. Roos, D.M. Porter, G.D. Sorenson, Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls, Pancreas 17(1) (1998) 89-97.

[95] S. Jahr, H. Hentze, S. Englisch, D. Hardt, F.O. Fackelmayer, R.D. Hesch, R. Knippers, DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells, Cancer Res 61(4) (2001) 1659-65.

[96] R. Cortese, I. Almendros, Y. Wang, D. Gozal, Tumor circulating DNA profiling in xenografted mice exposed to intermittent hypoxia, Oncotarget 6(1) (2015) 556-69.

[97] H.K. Lindberg, X. Wang, H. Jarventaus, G.C. Falck, H. Norppa, M. Fenech, Origin of nuclear buds and micronuclei in normal and folate-deprived human lymphocytes, Mutat Res 617(1-2) (2007) 33-45.

[98] M. Al-Asmakh, L. Hedin, Microbiota and the control of blood-tissue barriers, Tissue Barriers 3(3) (2015) e1039691.

[99] Y.M. Lo, M.S. Tein, T.K. Lau, C.J. Haines, T.N. Leung, P.M. Poon, J.S. Wainscoat, P.J. Johnson, A.M. Chang, N.M. Hjelm, Quantitative analysis

of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis, Am J Hum Genet 62(4) (1998) 768-75.

[100] L. Zhu, A. Nazeri, C.P. Pacia, Y. Yue, H. Chen, Focused ultrasound for safe and effective release of brain tumor biomarkers into the peripheral circulation, PLoS One 15(6) (2020) e0234182.

[101] C.P. Pacia, L. Zhu, Y. Yang, Y. Yue, A. Nazeri, H. Michael Gach, M.R. Talcott, E.C. Leuthardt, H. Chen, Feasibility and safety of focused ultrasound-enabled liquid biopsy in the brain of a porcine model, Sci Rep 10(1) (2020) 7449.

[102] J.J. Choi, C.F. Reich, 3rd, D.S. Pisetsky, The role of macrophages in the in vitro generation of extracellular DNA from apoptotic and necrotic cells, Immunology 115(1) (2005) 55-62.

[103] N. Jiang, C.F. Reich, 3rd, D.S. Pisetsky, Role of macrophages in the generation of circulating blood nucleosomes from dead and dying cells, Blood 102(6) (2003) 2243-50.

[104] G. Herbreteau, A. Vallee, A.C. Knol, S. Theoleyre, G. Quereux, E. Varey, A. Khammari, B. Dreno, M.G. Denis, Circulating Tumor DNA Early Kinetics Predict Response of Metastatic Melanoma to Anti-PD1 Immunotherapy: Validation Study, Cancers (Basel) 13(8) (2021).

[105] J. Westman, S. Grinstein, P.E. Marques, Phagocytosis of Necrotic Debris at Sites of Injury and Inflammation, Front Immunol 10 (2019) 3030.

[106] Y.M. Lo, J. Zhang, T.N. Leung, T.K. Lau, A.M. Chang, N.M. Hjelm, Rapid clearance of fetal DNA from maternal plasma, Am J Hum Genet 64(1) (1999) 218-24.

[107] P.S. Patel, B.P. Patel, R.M. Rawal, G.N. Raval, M.M. Patel, J.B. Patel, F.P. Jha, D.D. Patel, Evaluation of serum alkaline DNase activity in treatment monitoring of head and neck cancer patients, Tumour Biol 21(2) (2000) 82-9.

[108] M. Martin, J. Leffler, K.I. Smolag, J. Mytych, A. Bjork, L.D. Chaves, J.J. Alexander, R.J. Quigg, A.M. Blom, Factor H uptake regulates intracellular C3 activation during apoptosis and decreases the inflammatory potential of nucleosomes, Cell Death Differ 23(5) (2016) 903-11.

[109] C. Kirsch, S. Weickmann, B. Schmidt, M. Fleischhacker, An improved method for the isolation of free-circulating plasma DNA and cell-free DNA from other body fluids, Ann N Y Acad Sci 1137 (2008) 135-9.

[110] V.J. Gauthier, L.N. Tyler, M. Mannik, Blood clearance kinetics and liver uptake of mononucleosomes in mice, J Immunol 156(3) (1996) 1151-6.

[111] M. Korabecna, S. Opatrna, J. Wirth, K. Rulcova, J. Eiselt, F. Sefrna, A. Horinek, Cell-free plasma DNA during peritoneal dialysis and hemodialysis and in patients with chronic kidney disease, Ann N Y Acad Sci 1137 (2008) 296-301.

[112] T.H.T. Cheng, P. Jiang, J.C.W. Tam, X. Sun, W.S. Lee, S.C.Y. Yu, J.Y.C. Teoh, P.K.F. Chiu, C.F. Ng, K.M. Chow, C.C. Szeto, K.C.A. Chan, R.W.K. Chiu, Y.M.D. Lo, Genomewide bisulfite sequencing reveals the origin and time-dependent fragmentation of urinary cfDNA, Clin Biochem 50(9) (2017) 496-501.

[113] E.C. Hung, T.K. Shing, S.S. Chim, P.C. Yeung, R.W. Chan, K.W. Chik, V. Lee, N.B. Tsui, C.K. Li, C.S. Wong, R.W. Chiu, Y.M. Lo, Presence of donor-derived DNA and cells in the urine of sex-mismatched hematopoietic stem cell transplant recipients: implication for the transrenal hypothesis, Clin Chem 55(4) (2009) 715-22.

[114] D.B. Costa, B. Halmos, A. Kumar, S.T. Schumer, M.S. Huberman, T.J. Boggon, D.G. Tenen, S. Kobayashi, BIM mediates EGFR tyrosine kinase inhibitor-induced apoptosis in lung cancers with oncogenic EGFR mutations, PLoS Med 4(10) (2007) 1669-79; discussion 1680.

[115] J. Phallen, A. Leal, B.D. Woodward, P.M. Forde, J. Naidoo, K.A. Marrone, J.R. Brahmer, J. Fiksel, J.E. Medina, S. Cristiano, D.N. Palsgrove, C.D. Gocke, D.C. Bruhm, P. Keshavarzian, V. Adleff, E. Weihe, V. Anagnostou, R.B. Scharpf, V.E. Velculescu, H. Husain, Early Noninvasive Detection of Response to Targeted Therapy in Non-Small Cell Lung Cancer, Cancer Res 79(6) (2019) 1204-1213.

[116] F.G. Blankenberg, In vivo imaging of apoptosis, Cancer Biol Ther 7(10) (2008) 1525-32.

[117] P. Maier, L. Hartmann, F. Wenz, C. Herskind, Cellular Pathways in Response to Ionizing Radiation and Their Targetability for Tumor Radiosensitization, Int J Mol Sci 17(1) (2016).

[118] K. Patsch, N. Matasci, A. Soundararajan, P. Diaz, D.B. Agus, D. Ruderman, M.E. Gross, Monitoring dynamic cytotoxic chemotherapy response in castration-resistant prostate cancer using plasma cell-free DNA (cfDNA), BMC Res Notes 12(1) (2019) 275.

[119] T. Moser, J. Waldispuehl-Geigl, J. Belic, S. Weber, Q. Zhou, S.O. Hasenleithner, R. Graf, J.A. Terzic, F. Posch, H. Sill, S. Lax, K. Kashofer, G. Hoefler, H. Schoellnast, E. Heitzer, J.B. Geigl, T. Bauernhofer, M.R. Speicher, On-treatment measurements of circulating tumor DNA during FOLFOX therapy in patients with colorectal cancer, NPJ Precis Oncol 4(1) (2020) 30.

[120] I. Lyskjaer, C.S. Kronborg, M.H. Rasmussen, B.S. Sorensen, C. Demuth, M. Rosenkilde, A.F.B. Johansen, M. Knudsen, S. Vang, S.R.P. Krag, K.G. Spindler, C.L. Andersen, Correlation between early dynamics in circulating tumour DNA and outcome from FOLFIRI treatment in metastatic colorectal cancer, Sci Rep 9(1) (2019) 11542.

[121] H. Wang, B. Zhang, D. Chen, W. Xia, J. Zhang, F. Wang, J. Xu, Y. Zhang, M. Zhang, L. Zhang, Y. Lu, Y. Geng, P. Huang, P. Huang, H. Wang, S.

Pan, Real-time monitoring efficiency and toxicity of chemotherapy in patients with advanced lung cancer, Clin Epigenetics 7 (2015) 119.

[122] S. Kruger, V. Heinemann, C. Ross, F. Diehl, D. Nagel, S. Ormanns, S. Liebmann, I. Prinz-Bravin, C.B. Westphalen, M. Haas, A. Jung, T. Kirchner, M. von Bergwelt-Baildon, S. Boeck, S. Holdenrieder, Repeated mutKRAS ctDNA measurements represent a novel and promising tool for early response prediction and therapy monitoring in advanced pancreatic cancer, Ann Oncol 29(12) (2018) 2348-2355.

[123] B.S. Chera, S. Kumar, C. Shen, R. Amdur, R. Dagan, R. Green, E. Goldman, J. Weiss, J. Grilley-Olson, S. Patel, A. Zanation, T. Hackman, J. Blumberg, S. Patel, B. Thorp, M. Weissler, W. Yarbrough, N. Sheets, W. Mendenhall, X.M. Tan, G.P. Gupta, Plasma Circulating Tumor HPV DNA for the Surveillance of Cancer Recurrence in HPV-Associated Oropharyngeal Cancer, J Clin Oncol 38(10) (2020) 1050-1058.

[124] E.J. Moding, Y. Liu, A.B. Hui, J. He, Y. Qiao, T. Xu, L. Yao, S. Gandhi, Z. Liao, M. Das, K.J. Ramchandran, S.K. Padda, J.W. Neal, H.A. Wakelee, B.W. Loo, S.H. Lin, A.A. Alizadeh, M. Diehn, Abstract PO-069: Circulating tumor DNA kinetics to identify genomic predictors of rapid response to chemoradiation in non-small cell lung cancer, Clinical Cancer Research 27(8 Supplement) (2021) PO-069.

[125] D.A. Breadner, M.D. Vincent, R. Correa, M. Black, A. Warner, M. Sanatani, V. Bhat, C. Morris, G. Jones, A. Allan, D.A. Palma, J. Raphael, Exploitation of treatment induced tumor lysis to enhance the sensitivity of ctDNA analysis: A first-in-human pilot study, Lung Cancer 165 (2022) 145-151.

[126] S.A. Leon, B. Shapiro, D.M. Sklaroff, M.J. Yaros, Free DNA in the serum of cancer patients and the effect of therapy, Cancer Res 37(3) (1977) 646-50.

[127] N.G. Medvedeva, I.V. Panyutin, I.G. Panyutin, R.D. Neumann, Phosphorylation of histone H2AX in radiation-induced micronuclei, Radiat Res 168(4) (2007) 493-8.

[128] M. Durante, S.C. Formenti, Radiation-Induced Chromosomal Aberrations and Immunotherapy: Micronuclei, Cytosolic DNA, and Interferon-Production Pathway, Front Oncol 8 (2018) 192.

[129] G.M. Walls, L. McConnell, J. McAleese, P. Murray, T.B. Lynch, K. Savage, G.G. Hanna, D.G. de Castro, Early circulating tumour DNA kinetics measured by ultra-deep next-generation sequencing during radical radiotherapy for non-small cell lung cancer: a feasibility study, Radiat Oncol 15(1) (2020) 132.

[130] E.L. Chen, A.A. Chaudhuri, B.Y. Nabet, J.J. Chabon, D.J. Merriott, B.W. Loo, A.A. Alizadeh, M. Diehn, Analysis of Circulating Tumor DNA Kinetics during Stereotactic Ablative Radiation Therapy for Non-Small Cell Lung

Cancer, International Journal of Radiation Oncology*Biology*Physics 102(3, Supplement) (2018) e676.

[131] A.A. Chaudhuri, A.F. Lovejoy, J.J. Chabon, A. Newman, H. Stehr, D.J. Merriott, J.N. Carter, T.D. Azad, S. Padda, M.F. Gensheimer, H.A. Wakelee, J.W. Neal, B.W. Loo, A.A. Alizadeh, M. Diehn, Circulating Tumor DNA Analysis during Radiation Therapy for Localized Lung Cancer Predicts Treatment Outcome, International Journal of Radiation Oncology*Biology*Physics 99(2, Supplement) (2017) S1-S2.

[132] R.H. Mole, Whole body irradiation; radiobiology or medicine?, Br J Radiol 26(305) (1953) 234-41.

[133] M.H.D. Neumann, S. Bender, T. Krahn, T. Schlange, ctDNA and CTCs in Liquid Biopsy - Current Status and Where We Need to Progress, Comput Struct Biotechnol J 16 (2018) 190-195.

[134] S.P. Ng, S. Meas, H. Bahig, H.D. Skinner, A.S. Garden, J. Phan, D.I. Rosenthal, G.B. Gunn, S.J. Frank, W.H. Morrison, E.M. Sturgis, R. Ferrarotto, C.D. Fuller, A. Lucci, C.S. Hall, Detection of Circulating Tumor Cells During Radiation Therapy in Patients with Head and Neck Cancer, International Journal of Radiation Oncology*Biology*Physics 102(3, Supplement) (2018) e177.

[135] J.H. Lee, G.V. Long, A.M. Menzies, S. Lo, A. Guminski, K. Whitbourne, M. Peranec, R. Scolyer, R.F. Kefford, H. Rizos, M.S. Carlino, Association Between Circulating Tumor DNA and Pseudoprogression in Patients With Metastatic Melanoma Treated With Anti-Programmed Cell Death 1 Antibodies, JAMA Oncol 4(5) (2018) 717-721.

[136] Y. Ma, Q. Wang, Q. Dong, L. Zhan, J. Zhang, How to differentiate pseudoprogression from true progression in cancer patients treated with immunotherapy, Am J Cancer Res 9(8) (2019) 1546-1553.

[137] R. Varaljai, S. Elouali, S.S. Lueong, K. Wistuba-Hamprecht, T. Seremet, J.T. Siveke, J.C. Becker, A. Sucker, A. Paschen, P.A. Horn, B. Neyns, B. Weide, D. Schadendorf, A. Roesch, The predictive and prognostic significance of cell-free DNA concentration in melanoma, J Eur Acad Dermatol Venereol 35(2) (2021) 387-395.

[138] S.B. Goldberg, A. Narayan, A.J. Kole, R.H. Decker, J. Teysir, N.J. Carriero, A. Lee, R. Nemati, S.K. Nath, S.M. Mane, Y. Deng, N. Sukumar, D. Zelterman, D.J. Boffa, K. Politi, S.N. Gettinger, L.D. Wilson, R.S. Herbst, A.A. Patel, Early Assessment of Lung Cancer Immunotherapy Response via Circulating Tumor DNA, Clin Cancer Res 24(8) (2018) 1872-1880.

[139] K. Kato, J. Uchida, Y. Kukita, T. Kumagai, K. Nishino, T. Inoue, M. Kimura, F. Imamura, Transient appearance of circulating tumor DNA associated with de novo treatment, Sci Rep 6 (2016) 38639.

[140] A.L. Riediger, S. Dietz, U. Schirmer, M. Meister, I. Heinzmann-Groth, M. Schneider, T. Muley, M. Thomas, H. Sultmann, Mutation analysis of

circulating plasma DNA to determine response to EGFR tyrosine kinase inhibitor therapy of lung adenocarcinoma patients, Sci Rep 6 (2016) 33505.

[141] L. Horn, J.G. Whisenant, H. Wakelee, K.L. Reckamp, H. Qiao, T.A. Leal, L. Du, J. Hernandez, V. Huang, G.R. Blumenschein, S.N. Waqar, S.P. Patel, J. Nieva, G.R. Oxnard, R.E. Sanborn, T. Shaffer, K. Garg, A. Holzhausen, K. Harrow, C. Liang, L.P. Lim, M. Li, C.M. Lovly, Monitoring Therapeutic Response and Resistance: Analysis of Circulating Tumor DNA in Patients With ALK+ Lung Cancer, J Thorac Oncol 14(11) (2019) 1901-1911.

[142] A.W. Wyatt, A.A. Azad, S.V. Volik, M. Annala, K. Beja, B. McConeghy, A. Haegert, E.W. Warner, F. Mo, S. Brahmbhatt, R. Shukin, S. Le Bihan, M.E. Gleave, M. Nykter, C.C. Collins, K.N. Chi, Genomic Alterations in Cell-Free DNA and Enzalutamide Resistance in Castration-Resistant Prostate Cancer, JAMA Oncol 2(12) (2016) 1598-1606.

[143] A. Romanel, D. Gasi Tandefelt, V. Conteduca, A. Jayaram, N. Casiraghi, D. Wetterskog, S. Salvi, D. Amadori, Z. Zafeiriou, P. Rescigno, D. Bianchini, G. Gurioli, V. Casadio, S. Carreira, J. Goodall, A. Wingate, R. Ferraldeschi, N. Tunariu, P. Flohr, U. De Giorgi, J.S. de Bono, F. Demichelis, G. Attard, Plasma AR and abiraterone-resistant prostate cancer, Sci Transl Med 7(312) (2015) 312re10.

[144] D.A. Breadner, M.D. Vincent, R.J.M. Correa, M. Black, A. Warner, M. Qu, D. Logan, M.S. Sanatani, J. Younus, B.P. Yaremko, G. Rodrigues, P.S. Blanchette, J. Laba, V. Bhat, C.D. Morris, E. Green, G. Jones, A.L. Allan, D.A. Palma, J. Raphael, Exploitation of treatment induced tumor lysis to enhance sensitivity of ctDNA analysis: A first-in-human pilot study, Journal of Clinical Oncology 38(15_suppl) (2020) 3530-3530.

[145] P.P. van de Donk, L. Kist de Ruijter, M.N. Lub-de Hooge, A.H. Brouwers, A.J. van der Wekken, S.F. Oosting, R.S. Fehrmann, D.J.A. de Groot, E.G. de Vries, Molecular imaging biomarkers for immune checkpoint inhibitor therapy, Theranostics 10(4) (2020) 1708-1718.

[146] B. O'Leary, S. Hrebien, J.P. Morden, M. Beaney, C. Fribbens, X. Huang, Y. Liu, C.H. Bartlett, M. Koehler, M. Cristofanilli, I. Garcia-Murillas, J.M. Bliss, N.C. Turner, Early circulating tumor DNA dynamics and clonal selection with palbociclib and fulvestrant for breast cancer, Nat Commun 9(1) (2018) 896.

[147] B.J. Noordman, B.P.L. Wijnhoven, S.M. Lagarde, J.J. Boonstra, P. Coene, J.W.T. Dekker, M. Doukas, A. van der Gaast, J. Heisterkamp, E.A. Kouwenhoven, G.A.P. Nieuwenhuijzen, J.E.N. Pierie, C. Rosman, J.W. van Sandick, M.J.C. van der Sangen, M.N. Sosef, M.C.W. Spaander, R. Valkema, E.S. van der Zaag, E.W. Steyerberg, J.J.B. van Lanschot, S.A.-s. group, Neoadjuvant chemoradiotherapy plus surgery versus active surveillance for oesophageal cancer: a stepped-wedge cluster randomised trial, BMC Cancer 18(1) (2018) 142.

[148] M.J.M. van der Valk, D.E. Hilling, E. Bastiaannet, E. Meershoek-Klein Kranenbarg, G.L. Beets, N.L. Figueiredo, A. Habr-Gama, R.O. Perez, A.G. Renehan, C.J.H. van de Velde, I. Consortium, Long-term outcomes of clinical complete responders after neoadjuvant treatment for rectal cancer in the International Watch & Wait Database (IWWD): an international multicentre registry study, Lancet 391(10139) (2018) 2537-2545.

[149] S.R. Markar, A. Karthikesalingam, S. Thrumurthy, D.E. Low, Volume-outcome relationship in surgery for esophageal malignancy: systematic review and meta-analysis 2000-2011, J Gastrointest Surg 16(5) (2012) 1055-63.

[150] M. Reece, H. Saluja, P. Hollington, C.S. Karapetis, S. Vatandoust, G.P. Young, E.L. Symonds, The Use of Circulating Tumor DNA to Monitor and Predict Response to Treatment in Colorectal Cancer, Front Genet 10 (2019) 1118.

[151] J.C.M. Wan, K. Heider, D. Gale, S. Murphy, E. Fisher, F. Mouliere, A. Ruiz-Valdepenas, A. Santonja, J. Morris, D. Chandrananda, A. Marshall, A.B. Gill, P.Y. Chan, E. Barker, G. Young, W.N. Cooper, I. Hudecova, F. Marass, R. Mair, K.M. Brindle, G.D. Stewart, J.E. Abraham, C. Caldas, D.M. Rassl, R.C. Rintoul, C. Alifrangis, M.R. Middleton, F.A. Gallagher, C. Parkinson, A. Durrani, U. McDermott, C.G. Smith, C. Massie, P.G. Corrie, N. Rosenfeld, ctDNA monitoring using patient-specific sequencing and integration of variant reads, Sci Transl Med 12(548) (2020).

[152] A. Zviran, R.C. Schulman, M. Shah, S.T.K. Hill, S. Deochand, C.C. Khamnei, D. Maloney, K. Patel, W. Liao, A.J. Widman, P. Wong, M.K. Callahan, G. Ha, S. Reed, D. Rotem, D. Frederick, T. Sharova, B. Miao, T. Kim, G. Gydush, J. Rhoades, K.Y. Huang, N.D. Omans, P.O. Bolan, A.H. Lipsky, C. Ang, M. Malbari, C.F. Spinelli, S. Kazancioglu, A.M. Runnels, S. Fennessey, C. Stolte, F. Gaiti, G.G. Inghirami, V. Adalsteinsson, B. Houck-Loomis, J. Ishii, J.D. Wolchok, G. Boland, N. Robine, N.K. Altorki, D.A. Landau, Genome-wide cell-free DNA mutational integration enables ultra-sensitive cancer monitoring, Nat Med (2020).

[153] K. Cervena, P. Vodicka, V. Vymetalkova, Diagnostic and prognostic impact of cell-free DNA in human cancers: Systematic review, Mutat Res 781 (2019) 100-129.

[154] R.J. Lee, G. Gremel, A. Marshall, K.A. Myers, N. Fisher, J.A. Dunn, N. Dhomen, P.G. Corrie, M.R. Middleton, P. Lorigan, R. Marais, Circulating tumor DNA predicts survival in patients with resected high-risk stage II/III melanoma, Ann Oncol 29(2) (2018) 490-496.

[155] D. Zheng, X. Ye, M.Z. Zhang, Y. Sun, J.Y. Wang, J. Ni, H.P. Zhang, L. Zhang, J. Luo, J. Zhang, L. Tang, B. Su, G. Chen, G. Zhu, Y. Gu, J.F. Xu, Plasma EGFR T790M ctDNA status is associated with clinical outcome in advanced NSCLC patients with acquired EGFR-TKI resistance, Sci Rep 6 (2016) 20913.

[156] K. Cibulskis, M.S. Lawrence, S.L. Carter, A. Sivachenko, D. Jaffe, C. Sougnez, S. Gabriel, M. Meyerson, E.S. Lander, G. Getz, Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples, Nat Biotechnol 31(3) (2013) 213-9.

[157] NIH, Sequence Read Archive (SRA) [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2009 - [cited 2020 May 5]. Available from: https://www.ncbi.nlm.nih.gov/sra/.

[158] A.H. Ramos, L. Lichtenstein, M. Gupta, M.S. Lawrence, T.J. Pugh, G. Saksena, M. Meyerson, G. Getz, Oncotator: cancer variant annotation tool, Hum Mutat 36(4) (2015) E2423-9.

[159] J.T. Robinson, H. Thorvaldsdottir, A.M. Wenger, A. Zehir, J.P. Mesirov, Variant Review with the Integrative Genomics Viewer, Cancer Res 77(21) (2017) e31-e34.

[160] J.A. Montoya, G.F. P, D. González-Sánchez, Statistical inference for the Weitzman overlapping coefficient in a family of distributions, Applied Mathematical Modelling 71 (2019) 558-568.

[161] J. Garcia-Aguilar, S. Patil, J.K. Kim, J.B. Yuval, H. Thompson, F. Verheij, M. Lee, L.B. Saltz, o.b.o.t.O. Consortium, Preliminary results of the organ preservation of rectal adenocarcinoma (OPRA) trial, Journal of Clinical Oncology 38(15_suppl) (2020) 4008-4008.

[162] P.R. Rohatgi, S.G. Swisher, A.M. Correa, T.T. Wu, Z. Liao, R. Komaki, G. Walsh, A. Vaporciyan, P.M. Lynch, D.C. Rice, J.A. Roth, J.A. Ajani, Failure patterns correlate with the proportion of residual carcinoma after preoperative chemoradiotherapy for carcinoma of the esophagus, Cancer 104(7) (2005) 1349-55.

[163] S.G. Urba, M.B. Orringer, A. Turrisi, M. Iannettoni, A. Forastiere, M. Strawderman, Randomized trial of preoperative chemoradiation versus surgery alone in patients with locoregional esophageal carcinoma, J Clin Oncol 19(2) (2001) 305-13.

[164] E. Ancona, A. Ruol, S. Santi, S. Merigliano, V.C. Sileni, H. Koussis, G. Zaninotto, L. Bonavina, A. Peracchia, Only pathologic complete response to neoadjuvant chemotherapy improves significantly the long term survival of patients with resectable esophageal squamous cell carcinoma: final report of a randomized, controlled trial of preoperative chemotherapy versus surgery alone, Cancer 91(11) (2001) 2165-74.

[165] A.C. Berger, J. Farma, W.J. Scott, G. Freedman, L. Weiner, J.D. Cheng, H. Wang, M. Goldberg, Complete response to neoadjuvant chemoradiotherapy in esophageal carcinoma is associated with significantly improved survival, J Clin Oncol 23(19) (2005) 4330-7.

[166] L.R. Chirieac, S.G. Swisher, J.A. Ajani, R.R. Komaki, A.M. Correa, J.S. Morris, J.A. Roth, A. Rashid, S.R. Hamilton, T.T. Wu, Posttherapy

pathologic stage predicts survival in patients with esophageal carcinoma receiving preoperative chemoradiation, Cancer 103(7) (2005) 1347-55.

[167] L. Kleinberg, J.P. Knisely, R. Heitmiller, M. Zahurak, R. Salem, B. Burtness, E.I. Heath, A.A. Forastiere, Mature survival results with preoperative cisplatin, protracted infusion 5-fluorouracil, and 44-Gy radiotherapy for esophageal cancer, Int J Radiat Oncol Biol Phys 56(2) (2003) 328-34.

[168] R.V. Scheer, A.J. Fakiris, P.A. Johnstone, Quantifying the benefit of a pathologic complete response after neoadjuvant chemoradiotherapy in the treatment of esophageal cancer, Int J Radiat Oncol Biol Phys 80(4) (2011) 996-1001.

[169] S.H. Bailey, D.A. Bull, D.H. Harpole, J.J. Rentz, L.A. Neumayer, T.N. Pappas, J. Daley, W.G. Henderson, B. Krasnicka, S.F. Khuri, Outcomes after esophagectomy: a ten-year prospective cohort, Ann Thorac Surg 75(1) (2003) 217-22; discussion 222.

[170] S.S. Biere, K.W. Maas, M.A. Cuesta, D.L. van der Peet, Cervical or thoracic anastomosis after esophagectomy for cancer: a systematic review and meta-analysis, Dig Surg 28(1) (2011) 29-35.

[171] E.Y. Kuo, Y. Chang, C.D. Wright, Impact of hospital volume on clinical and economic outcomes for esophagectomy, Ann Thorac Surg 72(4) (2001) 1118-24.

[172] A. Habr-Gama, J. Gama-Rodrigues, G.P. Sao Juliao, I. Proscurshim, C. Sabbagh, P.B. Lynn, R.O. Perez, Local recurrence after complete clinical response and watch and wait in rectal cancer after neoadjuvant chemoradiation: impact of salvage therapy on local disease control, Int J Radiat Oncol Biol Phys 88(4) (2014) 822-8.

[173] M. Maas, R.G. Beets-Tan, D.M. Lambregts, G. Lammering, P.J. Nelemans, S.M. Engelen, R.M. van Dam, R.L. Jansen, M. Sosef, J.W. Leijtens, K.W. Hulsewe, J. Buijsen, G.L. Beets, Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer, J Clin Oncol 29(35) (2011) 4633-40.

[174] J.D. Smith, J.A. Ruby, K.A. Goodman, L.B. Saltz, J.G. Guillem, M.R. Weiser, L.K. Temple, G.M. Nash, P.B. Paty, Nonoperative management of rectal cancer with complete clinical response after neoadjuvant therapy, Ann Surg 256(6) (2012) 965-72.

[175] M.J. Gollub, D.H. Gultekin, O. Akin, R.K. Do, J.L. Fuqua, 3rd, M. Gonen, D. Kuk, M. Weiser, L. Saltz, D. Schrag, K. Goodman, P. Paty, J. Guillem, G.M. Nash, L. Temple, J. Shia, L.H. Schwartz, Dynamic contrast enhanced-MRI for the detection of pathological complete response to neoadjuvant chemotherapy for locally advanced rectal cancer, Eur Radiol 22(4) (2012) 821-31.

[176] N.K. Cheedella, A. Suzuki, L. Xiao, W.L. Hofstetter, D.M. Maru, T. Taketa, K. Sudo, M.A. Blum, S.H. Lin, J. Welch, J.H. Lee, M.S. Bhutani, D.C. Rice,

A.A. Vaporciyan, S.G. Swisher, J.A. Ajani, Association between clinical complete response and pathological complete response after preoperative chemoradiation in patients with gastroesophageal cancer: analysis in a large cohort, Ann Oncol 24(5) (2013) 1262-6.

[177] H. Tranchart, J.H. Lefevre, M. Svrcek, J.F. Flejou, E. Tiret, Y. Parc, What is the incidence of metastatic lymph node involvement after significant pathologic response of primary tumor following neoadjuvant treatment for locally advanced rectal cancer?, Ann Surg Oncol 20(5) (2013) 1551-9.

[178] I. Garcia-Murillas, G. Schiavon, B. Weigelt, C. Ng, S. Hrebien, R.J. Cutts, M. Cheang, P. Osin, A. Nerurkar, I. Kozarewa, J.A. Garrido, M. Dowsett, J.S. Reis-Filho, I.E. Smith, N.C. Turner, Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer, Sci Transl Med 7(302) (2015) 302ra133.

[179] J. Tie, J.D. Cohen, Y. Wang, L. Li, M. Christie, K. Simons, H. Elsaleh, S. Kosmider, R. Wong, D. Yip, M. Lee, B. Tran, D. Rangiah, M. Burge, D. Goldstein, M. Singh, I. Skinner, I. Faragher, M. Croxford, C. Bampton, A. Haydon, I.T. Jones, S.K. C, T. Price, M.J. Schaefer, J. Ptak, L. Dobbyn, N. Silliman, I. Kinde, C. Tomasetti, N. Papadopoulos, K. Kinzler, B. Volgestein, P. Gibbs, Serial circulating tumour DNA analysis during multimodality treatment of locally advanced rectal cancer: a prospective biomarker study, Gut 68(4) (2019) 663-671.

[180] T.D. Azad, A.A. Chaudhuri, P. Fang, Y. Qiao, M.S. Esfahani, J.J. Chabon, E.G. Hamilton, Y.D. Yang, A. Lovejoy, A.M. Newman, D.M. Kurtz, M. Jin, J. Schroers-Martin, H. Stehr, C.L. Liu, A.B. Hui, V. Patel, D. Maru, S.H. Lin, A.A. Alizadeh, M. Diehn, Circulating Tumor DNA Analysis for Detection of Minimal Residual Disease After Chemoradiotherapy for Localized Esophageal Cancer, Gastroenterology 158(3) (2020) 494-505 e6.

[181] NIH, Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER Research Data, 8 Registries, Nov 2021 Sub (1975-2019) - Linked To County Attributes - Time Dependent (1990-2019) Income/Rurality, 1969-2020 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2022, based on the November 2021 submission.

[182] H.J. de Koning, C.M. van der Aalst, P.A. de Jong, E.T. Scholten, K. Nackaerts, M.A. Heuvelmans, J.J. Lammers, C. Weenink, U. Yousaf-Khan, N. Horeweg, S. van 't Westeinde, M. Prokop, W.P. Mali, F.A.A. Mohamed Hoesein, P.M.A. van Ooijen, J. Aerts, M.A. den Bakker, E. Thunnissen, J. Verschakelen, R. Vliegenthart, J.E. Walter, K. Ten Haaf, H.J.M. Groen, M. Oudkerk, Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial, N Engl J Med 382(6) (2020) 503-513.

[183] T. National Lung Screening Trial Research, T.R. Church, W.C. Black, D.R. Aberle, C.D. Berg, K.L. Clingan, F. Duan, R.M. Fagerstrom, I.F. Gareen, D.S. Gierada, G.C. Jones, I. Mahon, P.M. Marcus, J.D. Sicks, A. Jain, S. Baum, Results of initial low-dose computed tomographic screening for lung cancer, N Engl J Med 368(21) (2013) 1980-91.

[184] M. Jamal-Hanjani, G.A. Wilson, N. McGranahan, N.J. Birkbak, T.B.K. Watkins, S. Veeriah, S. Shafi, D.H. Johnson, R. Mitter, R. Rosenthal, M. Salm, S. Horswell, M. Escudero, N. Matthews, A. Rowan, T. Chambers, D.A. Moore, S. Turajlic, H. Xu, S.M. Lee, M.D. Forster, T. Ahmad, C.T. Hiley, C. Abbosh, M. Falzon, E. Borg, T. Marafioti, D. Lawrence, M. Hayward, S. Kolvekar, N. Panagiotopoulos, S.M. Janes, R. Thakrar, A. Ahmed, F. Blackhall, Y. Summers, R. Shah, L. Joseph, A.M. Quinn, P.A. Crosbie, B. Naidu, G. Middleton, G. Langman, S. Trotter, M. Nicolson, H. Remmen, K. Kerr, M. Chetty, L. Gomersall, D.A. Fennell, A. Nakas, S. Rathinam, G. Anand, S. Khan, P. Russell, V. Ezhil, B. Ismail, M. Irvin-Sellers, V. Prakash, J.F. Lester, M. Kornaszewska, R. Attanoos, H. Adams, H. Davies, S. Dentro, P. Taniere, B. O'Sullivan, H.L. Lowe, J.A. Hartley, N. Iles, H. Bell, Y. Ngai, J.A. Shaw, J. Herrero, Z. Szallasi, R.F. Schwarz, A. Stewart, S.A. Quezada, J. Le Quesne, P. Van Loo, C. Dive, A. Hackshaw, C. Swanton, T.R. Consortium, Tracking the Evolution of Non-Small-Cell Lung Cancer, N Engl J Med 376(22) (2017) 2109-2121.

[185] N. Cancer Genome Atlas Research, Comprehensive molecular profiling of lung adenocarcinoma, Nature 511(7511) (2014) 543-50.

[186] S. Waldeck, J. Mitschke, S. Wiesemann, M. Rassner, G. Andrieux, M. Deuter, J. Mutter, A.M. Luchtenborg, D. Kottmann, L. Titze, C. Zeisel, M. Jolic, U. Philipp, S. Lassmann, P. Bronsert, C. Greil, J. Rawluk, H. Becker, L. Isbell, A. Muller, S. Doostkam, B. Passlick, M. Borries, J. Duyster, J. Wehrle, F. Scherer, N. von Bubnoff, Early assessment of circulating tumor DNA after curative-intent resection predicts tumor recurrence in early-stage and locally advanced non-small-cell lung cancer, Mol Oncol 16(2) (2022) 527-537.

[187] P. Goldstraw, K. Chansky, J. Crowley, R. Rami-Porta, H. Asamura, W.E. Eberhardt, A.G. Nicholson, P. Groome, A. Mitchell, V. Bolejack, S. International Association for the Study of Lung Cancer, A.B. Prognostic Factors Committee, I. Participating, S. International Association for the Study of Lung Cancer, B. Prognostic Factors Committee Advisory, I. Participating, 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.

[188] H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform, Bioinformatics 26(5) (2010) 589-95.

[189] P. Danecek, J.K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M.O. Pollard, A. Whitwham, T. Keane, S.A. McCarthy, R.M. Davies, H. Li, Twelve years of SAMtools and BCFtools, Gigascience 10(2) (2021).

[190] Picard toolkit, Broad Institute, GitHub repository (https://broadinstitute.github.io/picard/) (2019).

[191] K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data, Nucleic Acids Res 38(16) (2010) e164.

[192] M.J. Landrum, J.M. Lee, M. Benson, G.R. Brown, C. Chao, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, W. Jang, K. Karapetyan, K. Katz, C. Liu, Z. Maddipatla, A. Malheiro, K. McDaniel, M. Ovetsky, G. Riley, G. Zhou, J.B. Holmes, B.L. Kattman, D.R. Maglott, ClinVar: improving access to variant interpretations and supporting evidence, Nucleic Acids Res 46(D1) (2018) D1062-D1067.

[193] S. Chen, M. Liu, T. Huang, W. Liao, M. Xu, J. Gu, GeneFuse: detection and visualization of target gene fusions from DNA sequencing data, Int J Biol Sci 14(8) (2018) 843-848.

[194] A.M. Newman, S.V. Bratman, H. Stehr, L.J. Lee, C.L. Liu, M. Diehn, A.A. Alizadeh, FACTERA: a practical method for the discovery of genomic rearrangements at breakpoint resolution, Bioinformatics 30(23) (2014) 3390-3.

[195] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, I. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, C. SciPy, SciPy 1.0: fundamental algorithms for scientific computing in Python, Nat Methods 17(3) (2020) 261-272.

[196] Z.L. Skidmore, A.H. Wagner, R. Lesurf, K.M. Campbell, J. Kunisaki, O.L. Griffith, M. Griffith, GenVisR: Genomic Visualizations in R, Bioinformatics 32(19) (2016) 3012-4.

[197] D.C. Koboldt, Best practices for variant calling in clinical sequencing, Genome Med 12(1) (2020) 91.

[198] B. Heleno, V. Siersma, J. Brodersen, Estimation of Overdiagnosis of Lung Cancer in Low-Dose Computed Tomography Screening: A Secondary Analysis of the Danish Lung Cancer Screening Trial, JAMA Intern Med 178(10) (2018) 1420-1422.

[199] J. Brodersen, T. Voss, F. Martiny, V. Siersma, A. Barratt, B. Heleno, Overdiagnosis of lung cancer with low-dose computed tomography screening: meta-analysis of the randomised clinical trials, Breathe (Sheff) 16(1) (2020) 200013.