The Journal of Liquid Biopsy 4 (2024) 100141

Contents lists available at ScienceDirect

The Journal of Liquid Biopsy

journal homepage: www.sciencedirect.com/journal/the-journal-of-liquid-biopsy

In vitro size-selection of short circulating tumor DNA fragments from late-stage lung cancer patients enhance the detection of mutations and aneuploidies

Christoffer Trier Maansson [a,b,c], Louise Skov Thomsen [a], Laura Stokkebro [a], Julie Gabe Dissing [a], Maiken Parm Ulhoi [a,b,d], Anders Lade Nielsen [c], Peter Meldgaard [d], Boe Sandahl Sorensen [a,b,*]

[a] Department of Clinical Biochemistry, Aarhus University Hospital, Aarhus, Denmark
[b] Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
[c] Department of Biomedicine, Aarhus University, Aarhus, Denmark
[d] Department of Oncology, Aarhus University Hospital, Aarhus, Denmark

ARTICLE INFO

Keywords:
Liquid biopsy
Size-selection
NGS
Lung cancer
Fragmentomics

ABSTRACT

Introduction: Recent studies have demonstrated differences between the fragment length profiles of cell-free DNA (cfDNA) from cancer patients and healthy individuals. This has led to the development of *in vitro* size-selection procedures which can isolate the short fragments that are enriched with mutated circulating tumor DNA (ctDNA). This has yet to be investigated in a large cohort of lung cancer patients.

Materials and methods: We used plasma samples from 35 stage III and IV lung cancer patients and performed targeted next-generation sequencing (NGS) and variant calling from cfDNA with and without size-selection of short fragments. We identified clonal hematopoiesis (CH) and germline mutations using targeted NGS on paired buffy coat (BC) samples. In addition, we performed a genome-wide copy-number alteration analysis on the cfDNA samples with and without size-selection.

Results: ctDNA containing tumor mutations had a different fragment length profile compared to cfDNA fragments with CH or germline mutations. *In vitro* size-selection resulted in a median 1.36-fold (interquartile range (IQR): 0.63 to 2.48) mutational allele fraction (MAF) enrichment of tumor mutations whereas CH/germline mutations had a median 0.95-fold (IQR: 0.62 to 1.05) MAF enrichment. Key oncogenic drivers, including *KRAS* and *EGFR* were more likely to have a MAF increase with size-selection. Size-selection also increased the number plasma aneuploidy positive samples from 8 of 35 to 20 of 35.

Conclusion: This study expands the knowledge regarding ctDNA fragmentation in lung cancer patients and we demonstrate that *in vitro* size-selection can increase MAF of tumor mutations and plasma aneuploidy calls. Size-selection could lead to increased sensitivity of ctDNA detection, which is crucial for clinical implementation of liquid biopsies. This study is the largest of its kind studying cfDNA samples from 35 lung cancer patients containing 109 mutations in total.

1. Introduction

Lung cancer remains one of the most frequent types of cancer worldwide and it is the leading cause of cancer death [1,2]. Liquid biopsies, containing cell-free DNA (cfDNA), have been studied extensively in recent years to detect somatic variants in the plasma from cancer patients [3–7]. Plasma cfDNA originating from the tumor is known as circulating tumor DNA (ctDNA), and mutations in ctDNA is an evolving biomarker in lung cancer [7–9]. However, detecting very low amounts of ctDNA among high levels of cfDNA of non-tumor origin, remains a technical challenge for wide implementation of ctDNA in clinical practice. CtDNA is generally detected using PCR-based methods, including droplet digital PCR (ddPCR) [10] or next-generation sequencing (NGS) [11]. While ddPCR provides very high sensitivity for detecting a limited number of tumor mutations [12], NGS enables the possibility of monitoring a range of mutations and thereby detecting e.g. the emergence of resistance mutations [13,14]. Yet, the positive percent agreement

* Corresponding author. Department of Clinical Biochemistry, Faculty of Health, Aarhus University Hospital, Palle Juul-Jensens Boulevard 69, 8200, Aarhus N, Denmark.
E-mail address: boesoere@rm.dk (B.S. Sorensen).

https://doi.org/10.1016/j.jlb.2024.100141
Received 29 December 2023; Received in revised form 17 January 2024; Accepted 17 January 2024

Abbreviations

BC	Buffy coat
CAPP-seq	Cancer personalized profiling by deep sequencing
cfDNA	cell-free DNA
CH	Clonal hematopoiesis
CI	Confidence interval
CNA	Copy-number alteration
ctDNA	circulating tumor DNA
ddPCR	droplet digital PCR
HPC	high-performance computer
indel	insertion and deletion
IQR	Interquartile range
MAF	Mutational allele fraction
NGS	next-generation sequencing
NOS	Not otherwise specified
NSCLC	Non-small cell lung cancer
PA	Plasma aneuploidy
SCLC	Small cell lung cancer
SNV	Single nucleotide variant
sWGS	shallow whole-genome sequencing
tMAD	trimmed median absolute deviation from copy number neutrality
WES	Whole-exome sequencing
WGS	Whole-genome sequencing
WT	Wild-type

between oncogenic mutations detected in the lung tumor tissue and in plasma with targeted sequencing is about 75 %, but markedly lower for oncogenic gene-fusions [15,16]. Therefore, methodological approaches enhancing the preferential detection of ctDNA relative to the cfDNA of non-tumor origin cell have pertinence and this has led to the discovery of different biological features of ctDNA different from that of non-tumor cfDNA [4,17–20].

Previous studies have shown that short cfDNA fragments (<150bp) contain a larger fraction of mutated fragments compared to the entire cfDNA pool [4,17,18,21,22]. Based on this, *in vitro* size-selection methods have been developed to isolate the short cfDNA fragments, thereby enhancing the ratio of ctDNA relative to non-tumor cfDNA [4,21,23,24]. Generally, the mutational allele fraction (MAF) measured with either NGS or ddPCR is increased with *in vitro* size-selection of short fragments and this has led to the discovery of somatic mutations not detected in unselected cfDNA. In addition, *in-silico* size-selection has been used in error reduction strategies in order to differentiate between signal and noise for ctDNA detection with NGS [17,25].

Studies evaluating the effect of *in vitro* size-selection of cfDNA from lung cancer patients remain limited [26]. In the presented study, we analyzed plasma samples with and without size-selection using targeted sequencing to observe the effect of size-selection on the MAF and genome-wide copy number alterations (CNA). In addition, the presence of clonal hematopoiesis (CH) and germline mutations in cfDNA can affect the interpretation of liquid biopsy analyses, resulting in false positive results [8,27]. While sequencing of peripheral blood cells in the buffy coat (BC) samples is the golden standard for filtering out CH/germline mutations, we hypothesize that CH/germline mutations could have different fragmentation profiles compared to mutations of known tumor origin.

2. Materials and methods

2.1. Patients

This study was performed in accordance with the declaration of Helsinki and accepted by the Central Denmark Region Committee on Biomedical Research Ethics (No. 1-10-72-83-14). Written informed consent was obtained from all individuals. The original study included 36 patients diagnosed with stage III and IV lung cancer but one patient (pt02) had insufficient cfDNA yield following purification and was therefore excluded from the analysis. The demographics of the remaining patients (n = 35) are displayed in Table 1. Plasma samples from five *EGFR*-mutated lung cancer patients obtained during treatment were used as validation. These patients were part of a previous study (ID: NCT02284633) and are described in Supplementary Table 1. Plasma from these patients was subjected to cfDNA purification with or without size-selection and was applied to whole-genome sequencing (WGS).

Table 1
Demographics of patients included in the study. Non-small cell lung cancer (NSCLC), SCLC (Small cell lung cancer), NOS (Not otherwise specified).

All patients, n (%)	35 (100)
Age in years, mean (range)	69 (45–86)
Sex, n (%)	
Male	12 (34)
Female	23 (66)
Smoking status, n (%)	
Never	4 (11)
Former	21 (60)
Current	10 (29)
Lung cancer, n (%)	
NSCLC	31 (89)
SCLC	4 (11)
Cancer subtype, n (%)	
Adenocarcinoma	16 (46)
Squamous cell carcinoma	12 (34)
Small cell carcinoma	4 (11)
NOS	2 (6)
Neuroendocrine carcinoma	1 (3)
Stage, n (%)	
III	4 (11)
IV	31 (89)
Metastasis, n (%)	
M0	4 (11)
M1a	8 (23)
M1b	15 (43)
M1c	8 (23)
Prior treatment, n (%)	
Yes	24 (69)
No	11 (31)

2.2. cfDNA purification and quality control

Peripheral blood was drawn in 10 mL EDTA tubes and centrifuged at room temperature within 2 h at 1400×*g* for 15 min. Plasma and peripheral blood cells were aliquoted separately and stored at −80 °C. The plasma was thawed and split into two samples which were subjected to cfDNA purification or cfDNA purification followed by *in vitro* size-selection. The cfDNA was purified using the AVENIO cfDNA Isolation Kit (Roche Sequencing, Mannheim, Germany) according to the manufacturer's instructions. Following purification, quality control of the cfDNA was established; cfDNA content was measured using Qubit dsDNA

HS assay kit (Thermo Fischer Scientific, Waltham, MA, USA), and fragment sizes were estimated using the Agilent 2100 Bioanalyzer System (Agilent, Santa Clara, CA, USA) to exclude contamination of genomic DNA from peripheral blood cells.

2.3. Cancer personalized profiling by deep sequencing (CAPP-seq)

cfDNA was sequenced using hybridization-capture CAPP-seq [5] via the AVENIO ctDNA surveillance kit (Roche Sequencing, Mannheim, Germany). The surveillance panel targets 197 hypermutated genes in lung cancer, covering approximately 198 kb. The enriched gene fragments were sequenced using the NextSeq 500 (Illumina, San Diego, CA, USA) and the data were analyzed using the AVENIO Oncology Analysis Software. The resulting position deduplicated BAM files were used for further analysis. The amount of plasma and cfDNA used for each sample is displayed in Supplementary Table 2 as well as the resulting number of deduplicated reads.

2.4. Whole-genome-sequencing

WGS was performed at MOMA NGS Core Center at Aarhus University Hospital. Libraries were prepared using the Twist Library Preparation EF kit 2.0 (Twist Bioscience, San Francisco, CA, USA) and the cfDNA was sequenced using the NovaSeq 6000 (Illumina, San Diego, CA, USA). The bioinformatic analysis was performed using the high-performance computer (HPC) facility, GenomeDK, at Aarhus University. Reads were trimmed using fastp [28] and were aligned to hg38 using bwa-mem [29]. Aligned reads were sorted and duplicate reads were marked using picard tools MarkDuplicates [30].

2.5. Identification of circulating tumor DNA and clonal hematopoiesis/ germline mutations

Mutations were identified using the AVENIO Oncology Analysis Software. For each patient, every mutation identified by the AVENIO software was manually verified in BAM files. Subsequently, each mutation was searched for in both input and size-selection BAM files from that patient as previously described [31]. The identified mutated fragments and wild-type (WT) fragments for the position of the mutation were used to calculate the MAF for the mutation. For all patients, genomic DNA from peripheral blood cells in BC samples was extracted and sequenced with CAPP-seq to discover germline or CH mutations. Mutations identified with the AVENIO software in BC samples were classified as CH/germline mutations and CH mutations specifically were classified as CH/germline mutations with a MAF below 10 % in plasma samples [32].

2.6. In vitro size-selection

In vitro size-selection of cfDNA was performed using PippinHT (Sage Science, Beverly, MA, USA) according to the manufacturer's instructions. For each sample, 20 µl of purified cfDNA was added to a 3 % agarose gel cassette. The range mode was used to collect fragments with 95–230 bp with a pause at 152 bp. This resulted in a collection of two 30 µl cfDNA fractions and the fraction containing the short fragments (95–152 bp) was subjected to quality control and CAPP-seq.

2.7. Plasma aneuploidy analysis for targeted sequencing

Genome-wide plasma aneuploidies were identified in plasma samples as previously described [27,33]. In brief, input cfDNA or size-selection cfDNA samples were compared to BC samples from the same individual using CNVkit [34]. Reads in the NGS target regions and resulting background reads were used to detect genome-wide CNA. Sequencing coverages were corrected for GC-biases, sizes of target regions and sequence mappability. Circular binary segmentation [35] was used to estimate genomic segments with constant copy numbers relative to BC sample. Plasma aneuploidy (PA) scores were calculated for each sample to provide a collective estimate representing the genome-wide aneuploidy for the sample [27]. For each sample, the log 2 copy ratio of each segment was centered to have a mean value of 0 and transformed to linear values. For each chromosomal arm the arm-level copy-ratio was estimated by calculating weighted averages based on the fraction of the chromosomal arm overlapped by the segment. Arm-level copy-ratios were transformed into z-scores based on the mean and standard deviation of the arm-level copy-ratios in the paired BC reference sample. The five arm-level z-scores with the absolute largest deviation from 0 were compared to the z-scores from the BC samples and used to calculate a per-sample PA score as described in Ref. [33]. PA scores for BC samples were calculated using a leave one-out approach where all of the BC samples except the sample of interest were used to create a collective reference sample, which was compared to the BC sample of interest. The BC sample with the highest PA score was used to set the threshold for a CNA-positive sample.

2.8. Plasma aneuploidy analysis for WGS

BAM files from WGS were converted to Wig files using HMMCopy readCounter [36] where the genome is segmented into 500 kb bins and the reads are counted in each bin. Genome-wide copy-ratios for each bin as well as the tumor fraction of each sample were estimated using ichorCNA [37]. The copy-ratios were corrected for any mappability- or GC-bias and regions with low mappability, including centromeres, were excluded from the analysis. Segments representing genomic regions with different copy-numbers were estimated with hidden Markov model using IchorCNA.

2.9. Software and statistics

Mutated fragments were identified as described above in R version 4.3.1. BAM files were loaded into R using the *chromstaR* [38] package and individual reads were analyzed using the *GenomicRanges* [39] and *Rsamtools* [40] packages. Plots and graphs were created using the *dplyr* [41], *ggpubr* [42], and *ggplot2* [43] packages. Continuous paired data were compared using a Wilcoxon signed rank test and continuous unpaired data were compared using a Mann-Whitney *U* test. Categorical data were compared using a Fisher's exact test. A *P*-value <0.05 was considered statistically significant.

3. Results

3.1. Fragmentomic characterization of tumor and CH/germline mutations

The demographics of the patients included in the study are presented in Table 1. The study had an overweight of women (66 %) and most patients had a smoking history (89 %). All patients had late-stage lung cancer (stage III-IV) and 69 % had received treatment prior to blood sampling.

Sequencing of the input cfDNA revealed the expected peak in fragment length at 165 bp reflecting that the fragments correspond to 145 bp DNA wrapped around a histone octamer to form a nucleosome core particle plus 20 bp linker DNA (Fig. 1A). Using CAPP-seq we identified mutations in 33/35 (94 %) patients (Supplementary Fig. 1).

We isolated fragments containing tumor mutations *in silico* and compared the length of the fragments with non-mutated WT fragments from the same genomic region. This revealed a difference in the fragment length profile between the two types of fragments (Fig. 1B). By sequencing the BC samples with CAPP-seq we identified 20 germline and CH mutations in 11/35 patients. CH mutations specifically were classified as mutations detected in buffy coat samples with a MAF below 10 % in plasma samples [32]. No association between age and CH detection was observed in this study (Supplementary Fig. 2). Comparing the length of fragments containing CH/germline mutations with the corresponding WT cfDNA did not reveal a similar shift in the fragment length profile as

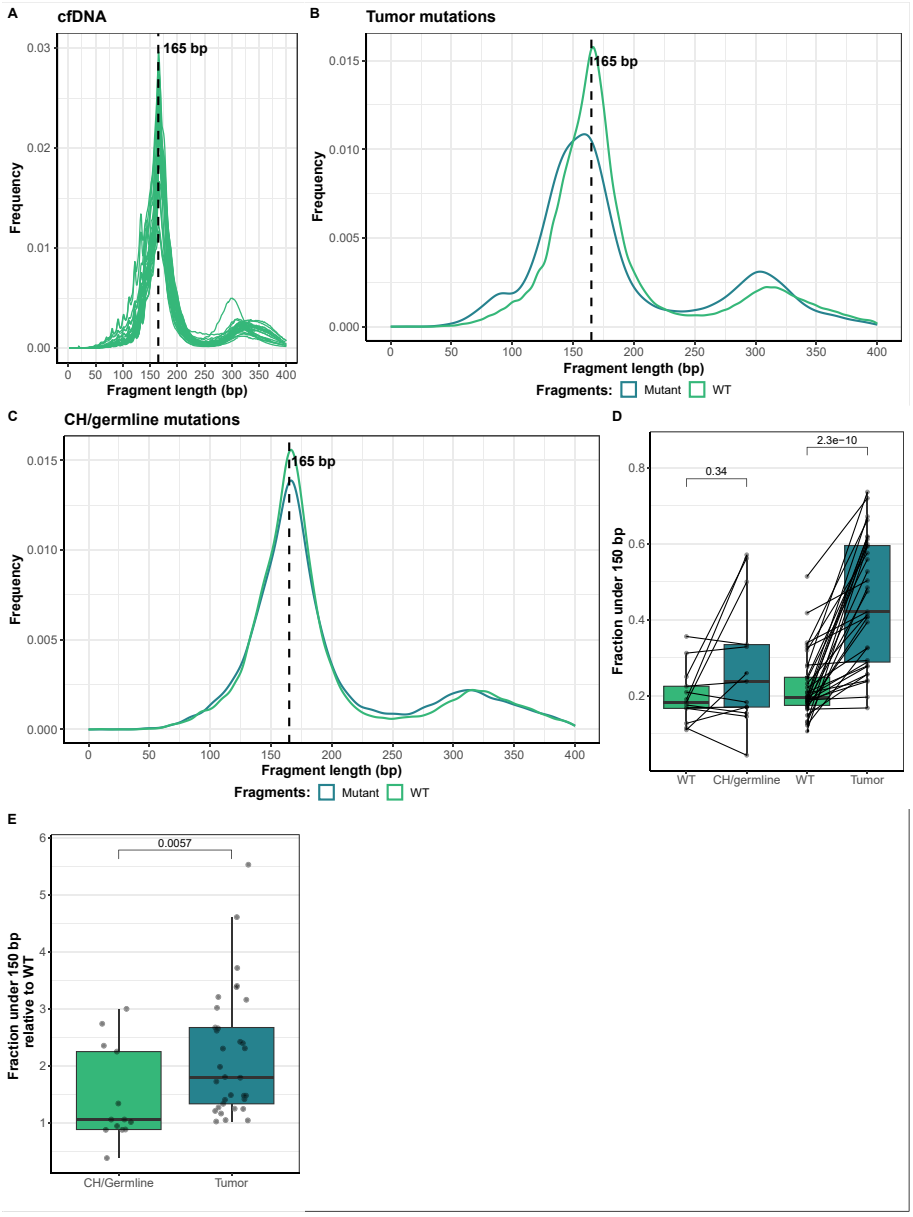

Fig. 1. Fragmentomic profiling of tumor and CH/germline mutations. (**A**) Fragment length observed in input cfDNA samples. (**B**–**C**) Fragment length of cfDNA fragments containing tumor mutations (**B**) and CH/germline mutations (**C**) compared to the corresponding wild-type (WT) cfDNA. (**D**) Paired comparison of the fraction of short fragments (<150 bp) for fragments with CH/germline or tumor mutations and the corresponding WT cfDNA. *P*-values were calculated using a Wilcoxon signed rank test. (**E**) The relative difference in the fraction of short fragments between mutated and WT cfDNA for CH/germline and tumor mutations. *P*-value was calculated using a Mann-Whitney *U* test.

was observed for the tumor mutations (Fig. 1C). Quantifying the fraction of fragments shorter than 150 bp revealed a significant increase of fragments with tumor mutations compared to WT fragments, whereas no significant difference was observed for the CH/germline mutations (Fig. 1D). Next, we calculated the relative difference between the fraction of short fragments for mutated fragments and corresponding WT fragments (Fig. 1E). This revealed that fragments containing tumor mutations had significantly larger relative fraction of short fragments (median = 1.80, IQR: 1.33 to 2.67) than CH/germline mutations (median = 1.05, IQR: 0.88 to 2.25). Collectively, the results in Fig. 1 illustrate the difference in cfDNA fragmentation between mutations from the tumor and mutations of germline or CH origin.

3.2. Size-selection of cfDNA increase MAF

Next, we analyzed the effect of *in vitro* size-selection on the mutational allele fraction for CH/germline and tumor mutations. *In vitro* size-selection resulted in a massive increase of short fragments (<150 bp) from median = 17.4 % (IQR: 13.9 %–22.4 %) in input cfDNA samples to median = 88.0 % (IQR: 80.9 %–91.8 %) in size-selected samples (Supplementary Fig. 3). For tumor mutations, the input cfDNA MAF (median = 1.45 %, IQR: 0.59 %–4.84 %) increased with size-selection (median = 2.37 %, IQR: 0.60 %–11.54 %, Fig. 2A and Supplementary Fig. 4). In contrast, the CH/germline MAF in input samples (median = 2.30 %, IQR: 0.78 %–48.50 %) revealed a slight decrease with size-selection (median = 2.05 %, IQR: 0.74 %–46.07 %). Next, we calculated the relative difference between the MAF observed in size-selection and input samples (Fig. 2B). Tumor mutations had significantly larger fold enrichment in MAF (median = 1.36, IQR: 0.63 to 2.48) compared to CH/germline mutations (median = 0.95, IQR: 0.62–1.05 %), which had a MAF decrease. In addition to the relative enrichment, we also calculated the absolute change in MAF between size-selection and input samples (Fig. 2C). This demonstrated that tumor mutations also had a significant larger absolute change in MAF (median = 0.86 %, IQR: −0.33 %–3.21 %) compared to the CH/germline mutations (median = −0.71 %, IQR: −1.61 %–0.21 %). We hypothesized that CH/germline mutations could be classified based on the absolute change in MAF between input and size-selection samples. To investigate this, we performed receiver operating characteristics analysis on the differences in MAF (Fig. 2D). This revealed an area under the curve,

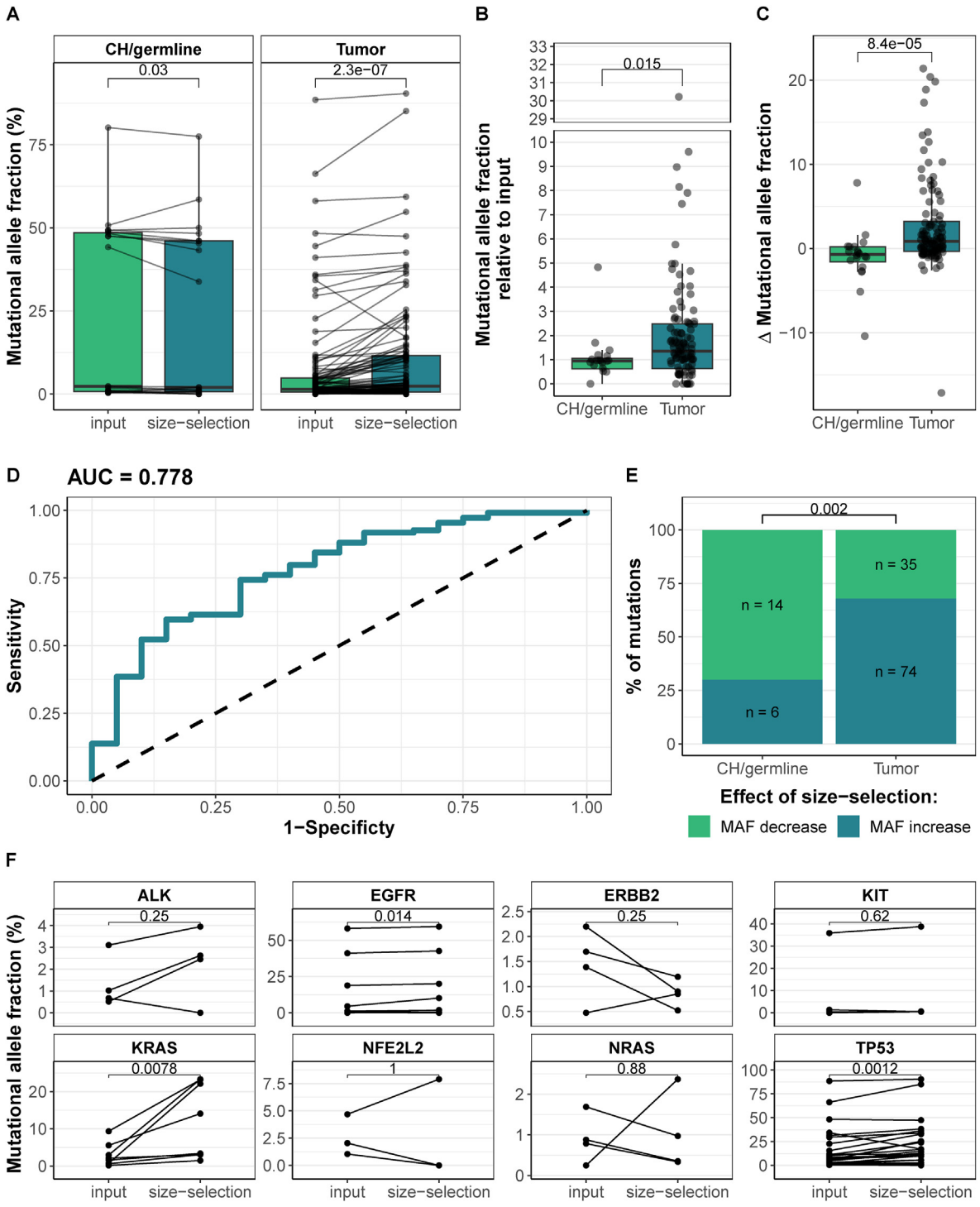

Fig. 2. Effect of size-selection on the MAF of tumor and CH/germline mutations. (**A**) The MAF for CH/germline and tumor mutations in input cfDNA and size-selection cfDNA. (**B**) The relative difference in MAF for CH/germline and tumor mutations between size-selection cfDNA and input cfDNA. (**C**) The absolute change in MAF for CH/germline and tumor mutations between size-selection cfDNA and input cfDNA. (**D**) ROC curve illustrating the ability to classify CH/germline mutations based on the absolute change in MAF between size-selection and input cfDNA. (**E**) Comparison of the effect of size-selection on tumor and CH/germline mutations. (**F**) The MAF of individual genes observed in input cfDNA and size-selection cfDNA. *P*-values were calculated using a Wilcoxon signed rank test (**A** and **F**), Mann-Whitney *U* test (**B** and **C**), or a Fisher's exact test (**E**).

AUC = 0.778 (95 % confidence interval (CI): 0.666 to 0.889), with a sensitivity of 0.67 at a specificity of 0.70. Next, we investigated whether the MAF decreased or increased with size-selection for each individual tumor and CH/germline mutation (Fig. 2E). A significant difference was observed between CH/germline mutations (6/20, 30 % had MAF increase) and tumor mutations (74/109, 68 % had MAF increase). The effect of size-selection on CH/germline and tumor mutations was not affected by whether the patient had received prior treatment or was treatment naïve (Supplementary Fig. 5).

While most tumor mutations had MAF increase with size-selection (Fig. 2E), some mutations also decreased with size-selection. We hypothesized that mutations in different genes would be affected differently

with size-selection. We investigated mutations observed in a minimum of three individuals to see whether size-selection affected the mutations differently (Fig. 2F). A significant increase in MAF was observed for *EGFR, KRAS* and *TP53*, whereas the other five genes demonstrated varying effect of size-selection. For *KRAS* 8/8 (100 %) mutations increase with size-selection. Similar results were observed for *EGFR* where 8/10 (80 %) mutations increased with size-selection.

3.3. Size-selection results in increased CNA compared to input cfDNA

While ctDNA often is detected using somatic variant calling of single nucleotide variants (SNV) and small insertion and deletions (indels), larger structural rearrangements can also be used to detect ctDNA [4,27]. Using cfDNA and BC samples analyzed with targeted sequencing it has been demonstrated that it is possible to detect genome-wide CNA from liquid biopsies [27,44]. We hypothesized that size-selection could increase the amount of CNA detected in the plasma sample compared to the input sample (Supplementary Fig. 6). To compare samples, we estimated a per-sample plasma aneuploidy score (called PA score) to give a single estimate of the combined amount of aneuploidy in a given sample. In addition, we estimated the PA scores in BC samples, where the maximum PA score detected was used as a threshold for positive samples (Fig. 3A). As a result, the maximum PA score detected was 7.38 (patient 33), which was then used to classify cfDNA samples as CNA positive (PA>7.38) or negative (PA≤7.38). PA scores in paired input and size-selection samples are presented in Fig. 3B. The PA scores measured in size-selection samples (median = 10.08, IQR: 4.71 to 16.49) were significantly higher than in input samples (median = −1.27, IQR: −3.78 to 6.67). In addition, significantly more samples were classified as CNA positive for size-selection (20/35, 57 %) compared to input samples (8/35, 23 %) (Fig. 3C). These results were not dependent on whether the patients had received prior treatment or not (Supplementary Fig. 7). Fig. 3D highlights three examples: Patient 20 which has a low PA score in both input (PA score = −3.21) and size-selection (PA score = −4.65), patient27 which has a high PA score in both input (PA score = 24.0) and size-selection (PA score = 32.7), and patient07 which has a low PA score in input (PA score = 5.18), but high in size-selection (PA score = 15.8). It is evident that the aneuploidies detected in the input sample are more pronounced in the size-selection sample for patient27 and patient07. To validate these findings, we sequenced cfDNA from five patients (Supplementary Table 1) using WGS with and without size-selection and quantified aneuploidies using IchorCNA (Fig. 3E). Size-selection also resulted in a large enrichment of short cfDNA fragments for these patients (Supplementary Fig. 8). CNA was only detected in patient 02 (1/5, 20 %), but the IchorCNA-estimated tumor fraction increased from 11.15 % to 13.27 % following size-selection. In addition, the aneuploidies detected in the input sample were also detected in the size-selection sample and demonstrated a larger difference from 0. In conclusion, the results in Fig. 3 show that size-selection can increase the level of aneuploidies detected in liquid biopsies and it classifies more samples as CNA positive.

4. Discussion

To our knowledge, the cohort described in the presented study is the largest cohort of lung cancer patients (n = 35) investigated with *in vitro* size-selection. In the study, we demonstrate that the ctDNA fragments containing tumor mutations are enriched in the short cfDNA fraction and that *in vitro* size-selection of short cfDNA fragments can increase the MAF of somatic mutations. Importantly, we demonstrate that the cfDNA fragmentation is different for mutated fragments of tumor origin compared to germline or CH mutated fragments. This discrimination is important because it can help understand the clinical relevance of mutations detected in liquid biopsies. While our ROC analysis of whether the absolute change in MAF between samples with and without size-selection could classify germline/CH mutations with an AUC = 0.778 (95 % CI: 0.666 to 0.889), sequencing of BC samples is still the most consistent method of detecting these mutations. BC sequencing also enables CNA analysis of the liquid biopsy sample and therefore serves a dual purpose [27,45]. Here, we demonstrate that size-selection increases the number of CNA positive samples and the PA score is higher in size-selection samples compared to the input sample.

This study is in alignment with previous studies analyzing the effect of *in vitro* size-selection of cfDNA from different types of cancer. Underhill et al. described how polyacrylamide gel electrophoresis could be used to extract short cfDNA fragments. This resulted in a 1.1–9.1-fold enrichment of the *EGFR-T790M* mutation measured with ddPCR in 4 lung cancer cases [24]. Mouliere and colleagues made a comprehensive study of *in vitro* size-selection of cfDNA from different cancer types [4]. They apply PippinHT for *in vitro* size-selection, which has also been used in our study. The study primarily focused on shallow WGS (sWGS) which they used to calculate a trimmed median absolute deviation from copy number neutrality (tMAD) score. They demonstrated how both *in vitro* and *in silico* size-selection increased the tMAD score about 2-fold for pre-treatment and post-treatment samples. In addition, they performed whole exome sequencing (WES) for somatic variant calling on 6 pre- and post-treatment samples and observe a MAF median increase of 4.27 and 1.5, respectively. Hellwig et al. performed *in vitro* size-selection after the NGS library preparation which resulted in an about 2-fold MAF enrichment [23]. Performing the size-selection after library preparation could be beneficial for the detection of rare variants, because the PCR amplification could decrease the risk of losing mutated fragments at low frequencies. Ishida and colleagues used a different approach to size-selection and applied solid-phase reversible immobilization beads which isolated cfDNA sizes between 100 and 400 bp [21]. This resulted in a modest increase of fragments between 90 and 150 bp from 3.0.% to 25.9 % and the mean MAF of driver genes increased from 6.8 % to 10.7 % measured with digital PCR.

Using targeted NGS on plasma samples we observe a median MAF-fold increase of 1.36, IQR: 0.63 to 2.48. This was measured in 35 samples containing a total of 109 mutations. The enrichment is slightly lower compared to some of the studies described above, however, for the key oncogenic driver mutations, *KRAS* and *EGFR*, we observe a large increase in MAF with size-selection. In addition to the specific mutations, multiple factors could explain the lower enrichment, including time of blood sampling (pre- or post-treatment), cancer types, cancer stage, and whether size-selection was performed before or after library preparation. Future studies will need to investigate in greater detail at which step it is most beneficial to perform size-selection during the NGS protocol. In addition, while our current study focuses on late-stage lung cancer patients, it would be of great interest to investigate longitudinal ctDNA monitoring with size-selection in early-stage cancers treated with curative intend. In early-stage cancers it is essential to use ctDNA assays with high sensitivity to detect minimal residual disease following e.g. resection [46] or definitive radiotherapy [47]. In this setting, it is crucial to filter out CH variants to avoid false positive results in post-treatment plasma samples, which otherwise could result in unnecessary adjuvant treatment [48,49]. The combination of both somatic variant calling and CNA analysis in size-selection samples, as described in this study, could increase the sensitivity of detecting minimal residual disease in these patients.

5. Conclusion

This study expands the knowledge regarding ctDNA fragmentation and we describe the differences between mutated fragments of tumor and healthy cell origin. We demonstrate that *in vitro* size-selection increases the MAF of tumor mutations suggesting that the ctDNA material is enriched in the short cfDNA fragments, whereas germline/CH mutations are not. In addition, we demonstrate the targeted ctDNA analysis can be expanded to copy-number alternations and demonstrate that size-selection increases CNAs detected in liquid biopsies.

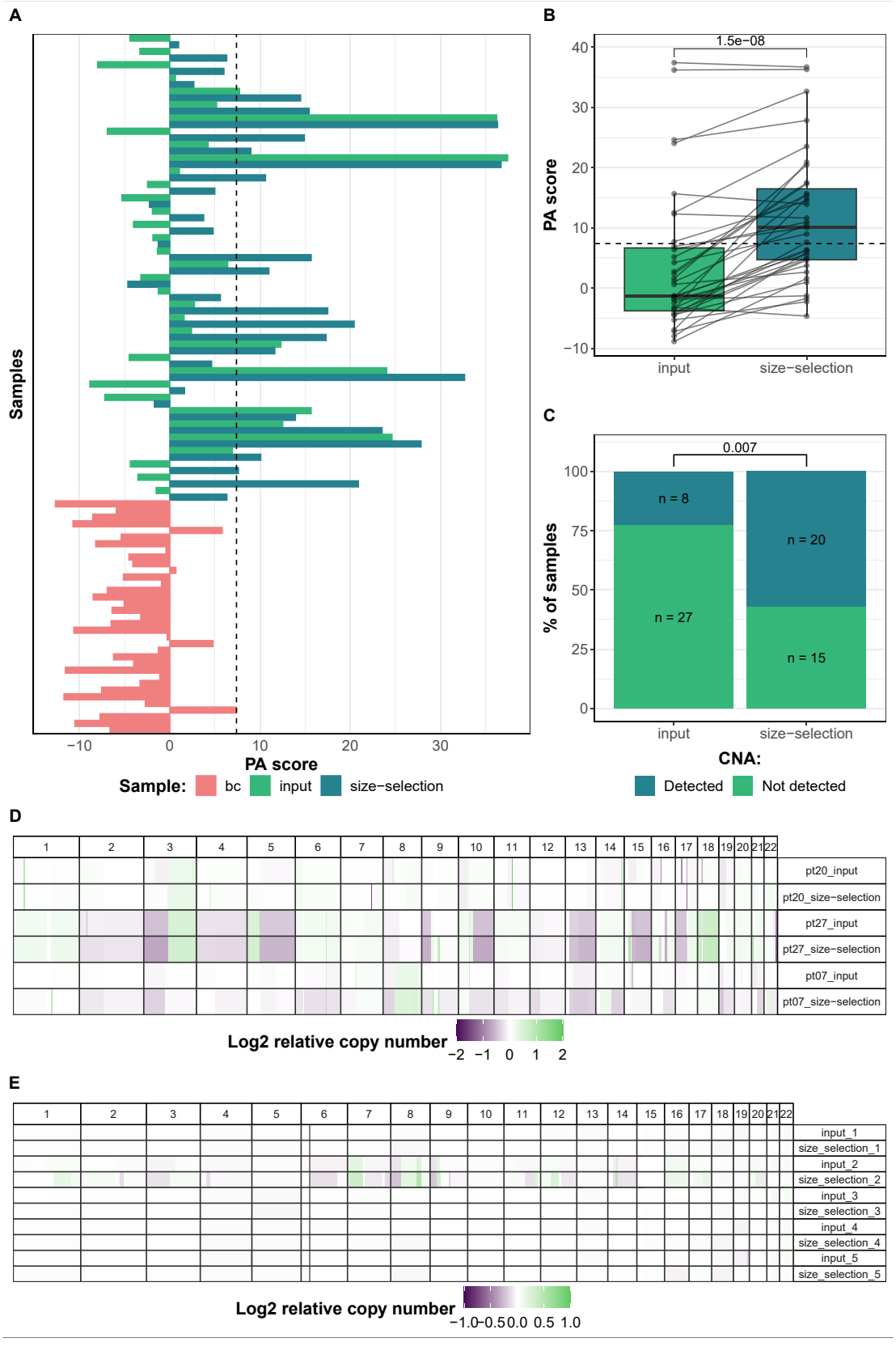

Fig. 3. The effect of size-selection on CNA analysis. (**A**) Plasma aneuploidy (PA) scores in paired input/size-selection samples and in BC samples. The threshold of a CNA positive sample is represented as a dashed line. (**B**) Pairwise comparison of PA scores observed in input and size-selection samples. The threshold of a CNA positive sample is represented as a dashed line. *P*-value was calculated using a Wilcoxon signed rank test. (**C**) Classification of whether CNA was detected in input and size-selection cfDNA samples. *P*-value was calculated using a Fisher's exact test. (**D**) Three examples (pt20, pt27, and pt06) of genome-wide CNA analysis with and without size-selection using CAPP-seq. (**E**) Genome-wide CNA levels in five patients analyzed with whole-genome sequencing with and without size-selection.

Author contributions

CTM: Conceptualization, Methodology, Software, Investigation, Formal analysis, Data curation, Validation, Visualization, Writing – original draft, Writing – review & editing. LST: Methodology, Investigation, Data curation, Writing – review & editing. LS: Methodology, Software, Writing – review & editing. JGD: Data curation, Resources, Writing – review & editing. MPU: Data curation, Resources, Writing – review & editing. ALN: Conceptualization, Project administration, Funding acquisition, Writing – review & editing. PM: Data curation, Resources, Conceptualization, Project administration, Funding acquisition, Writing – review & editing. BSS: Data curation, Resources, Conceptualization, Project administration, Supervision, Funding acquisition, Writing – review & editing.

Data availability

The data associated with this study is available on reasonable request from the corresponding author.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Birgit Westh Mortensen for excellent technical assistance and the MOMA NGS core center for assistance with WGS. We thank GenomeDK for giving access to their HPC facility supported by Aarhus University, Denmark.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jlb.2024.100141.

References

[1] Globocan TGCo-. Cancer fact sheet - tracea, broncus and lung: international agency gor Research on cancer by WHO. 2020 [Available from: http://gco.iarc.fr/today/data/factsheets/cancers/15-Lung-fact-sheet.pdf.
[2] Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature 2018;553(7689):446–54.
[3] Diehl F, Schmidt K, Choti MA, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med 2008;14(9):985–90.
[4] Mouliere F, Chandrananda D, Piskorz AM, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. Sci Transl Med 2018;10(466).
[5] Newman AM, Bratman SV, To J, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med 2014;20(5):548–54.
[6] Oxnard GR, Paweletz CP, Kuang Y, et al. Noninvasive detection of response and resistance in EGFR-mutant lung cancer using quantitative next-generation genotyping of cell-free plasma DNA. Clin Cancer Res 2014;20(6):1698–705.
[7] Jee J, Lebow ES, Yeh R, et al. Overall survival with circulating tumor DNA-guided therapy in advanced non-small-cell lung cancer. Nat Med 2022;28(11):2353–63.
[8] Weber S, van der Leest P, Donker HC, et al. Dynamic changes of circulating tumor DNA predict clinical outcome in patients with advanced non-small-cell lung cancer treated with immune checkpoint inhibitors. JCO Precis Oncol 2021;5:1540–53.
[9] Iams WT, Kopparapu PR, Yan Y, et al. Blood-based surveillance monitoring of circulating tumor DNA from patients with SCLC detects disease relapse and predicts death in patients with limited-stage disease. JTO Clin Res Rep 2020;1(2):100024.
[10] Henriksen TV, Tarazona N, Frydendahl A, et al. Circulating tumor DNA in stage III colorectal cancer, beyond minimal residual disease detection, toward assessment of adjuvant therapy efficacy and clinical behavior of recurrences. Clin Cancer Res 2022;28(3):507–17.
[11] Stensgaard S, Thomsen A, Helstrup S, et al. Blood tumor mutational burden and dynamic changes in circulating tumor DNA predict response to pembrolizumab treatment in advanced non-small cell lung cancer. Transl Lung Cancer Res 2023;12(5):971–84.
[12] Palacín-Aliana I, García-Romero N, Asensi-Puig A, et al. Clinical utility of liquid biopsy-based actionable mutations detected via ddPCR. Biomedicines 2021;9(8).
[13] Chabon JJ, Simmons AD, Lovejoy AF, et al. Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients. Nat Commun 2016;7:11815.
[14] Blakely CM, Watkins TBK, Wu W, et al. Evolution and clinical impact of co-occurring genetic alterations in advanced-stage EGFR-mutant lung cancers. Nat Genet 2017;49(12):1693–704.
[15] Sugimoto A, Matsumoto S, Udagawa H, et al. A large-scale prospective concordance study of plasma- and tissue-based next-generation targeted sequencing for advanced non-small cell lung cancer (LC-SCRUM-Liquid). Clin Cancer Res 2023;29(8):1506–14.
[16] Maansson CT, Andersen ER, Ulhoi MP, et al. DNAfusion: an R/Bioconductor package for increased sensitivity of detecting gene fusions in liquid biopsies. BMC Bioinf 2023;24(1):131.
[17] Chabon JJ, Hamilton EG, Kurtz DM, et al. Integrating genomic features for non-invasive early lung cancer detection. Nature 2020;580(7802):245–51.
[18] Markus H, Chandrananda D, Moore E, et al. Refined characterization of circulating tumor DNA through biological feature integration. Sci Rep 2022;12(1):1928.
[19] Budhraja KK, McDonald BR, Stephens MD, et al. Genome-wide analysis of aberrant position and sequence of plasma DNA fragment ends in patients with cancer. Sci Transl Med 2023;15(678):eabm6863.
[20] Jiang P, Sun K, Peng W, et al. Plasma DNA end-motif profiling as a fragmentomic marker in cancer, pregnancy, and transplantation. Cancer Discov 2020;10(5):664–73.
[21] Ishida Y, Takano S, Maekawa S, et al. Fractionated small cell-free DNA increases possibility to detect cancer-related gene mutations in advanced colorectal cancer. JGH Open 2020;4(5):978–86.
[22] Jiang P, Chan CW, Chan KC, et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. Proc Natl Acad Sci USA 2015;112(11):E1317–25.
[23] Hellwig S, Nix DA, Gligorich KM, et al. Automated size selection for short cell-free DNA fragments enriches for circulating tumor DNA and improves error correction during next generation sequencing. PLoS One 2018;13(7):e0197333.
[24] Underhill HR, Kitzman JO, Hellwig S, et al. Fragment length of circulating tumor DNA. PLoS Genet 2016;12(7):e1006162.
[25] Wan JCM, Heider K, Gale D, et al. ctDNA monitoring using patient-specific sequencing and integration of variant reads. Sci Transl Med 2020;12(548).
[26] Underhill HR. Leveraging the fragment length of circulating tumour DNA to improve molecular profiling of solid tumour malignancies with next-generation sequencing: a pathway to advanced non-invasive diagnostics in precision Oncology? Mol Diagn Ther 2021;25(4):389–408.
[27] Sivapalan L, Iams WT, Belcaid Z, et al. Dynamics of sequence and structural cell-free DNA landscapes in small-cell lung cancer. Clin Cancer Res 2023;29(12):2310–23.
[28] Chen S, Zhou Y, Chen Y, et al. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34(17):i884–90.
[29] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754–60.
[30] Broad institute. Picard Tools. 3.1.0 ed, https://broadinstitute.github.io/picard/.
[31] Vessies DCL, Schuurbiers MMF, van der Noort V, et al. Combining variant detection and fragment length analysis improves detection of minimal residual disease in postsurgery circulating tumour DNA of stage II-IIIA NSCLC patients. Mol Oncol 2022;16(14):2719–32.
[32] Yaung SJ, Fuhlbrück F, Peterson M, et al. Clonal hematopoiesis in late-stage non-small-cell lung cancer and its impact on targeted panel next-generation sequencing. JCO Precis Oncol 2020;4:1271–9.
[33] Leary RJ, Sausen M, Kinde I, et al. Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. Sci Transl Med 2012;4(162):162ra54.
[34] Talevich E, Shain AH, Botton T, et al. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput Biol 2016;12(4):e1004873.
[35] Olshen AB, Venkatraman ES, Lucito R, et al. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 2004;5(4):557–72.
[36] Ha G, Roth A, Lai D, et al. Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. Genome Res 2012;22(10):1995–2007.
[37] Adalsteinsson VA, Ha G, Freeman SS, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. Nat Commun 2017;8(1):1324.
[38] Taudt A, Nguyen MA, Heinig M, et al. chromstaR: tracking combinatorial chromatin state dynamics in space and time. bioRxiv 2016:038612.
[39] Lawrence M, Huber W, Pagès H, et al. Software for computing and annotating genomic ranges. PLoS Comput Biol 2013;9(8):e1003118.
[40] Morgan M, Pagès H, Obenchain V, Hayden N. Rsamtools: binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. Bioconductor; 2023. R package version 2.16.0 ed.
[41] Wickham H, François R, Henry L, Müller K, Vaughan D. Dplyr: a grammar of data manipulation. R package version 1.1.2 ed2023.
[42] Kassambara A. Ggpubr: 'ggplot2' based publication ready plots. R package version 0.6.0 ed2023.
[43] Wickham H. ggplot2: elegant graphics for data analysis. R package version 3.4.2 ed2023.
[44] Zhang J, Zhou N, Deng H, et al. Prognostic value of circulating tumor DNA using target next-generation sequencing in extensive-stage small-cell lung cancer. Lung Cancer 2023;178:11–9.

[45] Maansson CT, Meldgaard P, Sorensen BS. Can liquid biopsy dynamics stratify patients with small cell lung cancer? Transl Lung Cancer Res 2023;12(12):2565–9.
[46] Waldeck S, Mitschke J, Wiesemann S, et al. 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 2022;16(2):527–37.
[47] Lebow ES, Shaverdian N, Eichholz JE, et al. ctDNA-based detection of molecular residual disease in stage I-III non-small cell lung cancer patients treated with definitive radiotherapy. Front Oncol 2023;13:1253629.
[48] Rolfo C, Cardona AF, Cristofanilli M, et al. Challenges and opportunities of cfDNA analysis implementation in clinical practice: perspective of the International Society of Liquid Biopsy (ISLB). Crit Rev Oncol Hematol 2020;151:102978.
[49] Denis MG, Herbreteau G, Pons-Tostivint E. Molecular minimal residual disease in resected non-small cell lung cancer (NSCLC): results of specifically designed interventional clinical trials eagerly awaited. Transl Lung Cancer Res 2023;12(2):200–3.