Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Low-coverage whole-genome sequencing of plasma cell-free DNA to predict treatment response and progression-free survival in advanced NSCLC receiving EGFR-tyrosine kinase inhibitors or chemotherapy: A pilot prospective study

Pitithat Puranachot
pitithat.pur@cra.ac.th

Chulabhorn Royal Academy

Thivaratana Sinthuwiwat
Chulabhorn Royal Academy

Priyavudh Herabutya
Chulabhorn Royal Academy

Kanthida Jangyubol
Chulabhorn Royal Academy

Bongkod Petcharat
Chulabhorn Royal Academy

Jutamas Yimnoon
Chulabhorn Royal Academy

Jitsopin Phadungsut
Chulabhorn Hospital

Chayanee Samdaengpan
Chulabhorn Hospital

Chumut Phanthunane
Chulabhorn Hospital

Research Article

Keywords: Liquid biopsy monitoring, Cell-free DNA, Non-small cell lung cancer, Progression-free survival, Copy-number alterations, cfDNA fragmentation, Low-coverage whole-genome sequencing

Posted Date: April 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9258115/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Low-coverage whole-genome sequencing of plasma cell-free DNA to predict treatment response and progression-free survival in advanced NSCLC receiving EGFR-tyrosine kinase inhibitors or chemotherapy: A pilot prospective study

Pitithat Puranachot[1,2*], Thivaratana Sinthuwiwat[2,3], Priyavudh Herabutya[3], Kanthida Jangyubol[2,3], Bongkod Petcharat[2,3], Jutamas Yimnoon[3], Jitsopin Phadungsut[4], Chayanee Samdaengpan[5], Chumut Phanthunane[5,6*]

[1*]Princess Srisavangavadhana Faculty of Medicine, Chulabhorn Royal Academy, Bangkok, 10210, Thailand.
[2*]Center for Cancer Discovery and Genomics, Bureau for Research and Innovation Management, Chulabhorn Royal Academy, Bangkok, 10210, Thailand.
[3]Cytogenetics Research Laboratory, Bureau for Research and Innovation Management, Chulabhorn Royal Academy, Bangkok, 10210, Thailand.
[4]Service System Development Unit, Oncology Center, Chulabhorn Hospital, Bangkok, 10210, Thailand.
[5*]Division of Medical Oncology, Chulabhorn Hospital, Bangkok, 10210, Thailand.
[6*]Bureau for Research and Innovation Management, Chulabhorn Royal Academy, Bangkok, 10210, Thailand.

*Corresponding author(s). E-mail(s): pitithat.pur@cra.ac.th; chumut.pha@cra.ac.th;

Contributing authors: thivaratana.sin@cra.ac.th;
priyavudh.her@cra.ac.th; kanthida.jan@cra.ac.th;
bongkod.pet@cra.ac.th; jutamas.yim@cra.ac.th; jitsopin.pha@cra.ac.th;
chayanee.sam@cra.ac.th;

Abstract

Background: Accurately predicting treatment outcomes in patients with advanced non-small cell lung cancer (NSCLC) remains a significant clinical challenge. While circulating tumor DNA (ctDNA) within plasma cell-free DNA (cfDNA) has emerged as a promising non-invasive biomarker, conventional targeted sequencing approaches require prior tumor genetic profiling and may miss the full spectrum of tumor-associated alterations. Low-coverage whole-genome sequencing (lcWGS) offers a cost-effective and unbiased alternative that enables genome-wide detection of copy number alterations (CNAs) and fragmentation-based features without prior tumor genotyping. In this pilot prospective study, plasma cfDNA was collected from 33 patients with advanced NSCLC receiving either EGFR tyrosine kinase inhibitors (EGFR-TKIs) or chemotherapy at diagnosis, and during treatment at week 4, and week 12, as well as from 10 healthy controls. This cfDNA was subjected to lcWGS to explore the potential of ctDNA fraction, CNAs, short-to-long fragment ratio, and ctDNA estimation score (CES) as candidate predictors of treatment response and progression-free survival (PFS).

Results: In this small cohort, patients with progressive disease tended to maintain persistently elevated cfDNA concentrations and ctDNA fractions (¿3%) throughout treatment, with extensive CNA at weeks 4 and 12, which appeared to be associated with shorter PFS in patients receiving chemotherapy ($p = 0.011$). Patients with partial response or stable disease generally demonstrated early clearance or reduction of ctDNA and CNAs. The ctDNA fraction showed promising discriminatory performance for disease progression, with area under the curve (AUC) values of 0.93 and 0.92 at weeks 4 and 12, respectively, although these estimates should be interpreted with caution given the limited sample size. The fragmentation-based analysis showed potentially comparable predictive utility, with AUC values of 0.86 and 0.91 at weeks 4 and 12, respectively.

Conclusion: These preliminary findings suggest that lcWGS of plasma cfDNA may be a feasible and informative approach for monitoring treatment response and exploring factors associated with PFS in patients with advanced NSCLC. The patterns observed across both EGFR-TKI and chemotherapy contexts are encouraging, although the small sample size limits generalizability. Larger prospective validation studies are needed to confirm these findings and establish lcWGS-based cfDNA analysis as a reliable tool in precision oncology.

Keywords: Liquid biopsy monitoring, Cell-free DNA, Non-small cell lung cancer, Progression-free survival, Copy-number alterations, cfDNA fragmentation, Low-coverage whole-genome sequencing

1 Background

Lung cancer is one of the most lethal malignancies worldwide and the second leading cause of cancer-related deaths in Thailand, following liver cancer [1]. Indeed, according to GLOBOCAN 2022, lung cancer accounts for 16.7% of all cancer deaths in Thailand [2, 3]. Most patients (˜70%) present with advanced or metastatic disease, for which systemic chemotherapy has long been the standard of care. However, the advent of genomic profiling has enabled the identification of actionable mutations and the use of targeted therapies, offering superior outcomes compared to chemotherapy.

Among these genetic alterations, mutations in the gene that encodes epidermal growth factor receptor (EGFR) are the most prevalent in Asian lung cancer populations and were the first to be effectively targeted in NSCLC [4, 5] . EIt has been reported that EGFR tyrosine kinase inhibitors (TKIs), such as erlotinib and gefitinib, improve treatment response rates (80% vs. 20% with chemotherapy) and reduce toxicity [6, 7]. Despite these advances, resistance to treatment inevitably emerges through the development of secondary mutations (e.g., EGFR T790M) or the activation of alternative pathways, resulting in residual disease, relapse, and increased treatment burden [8, 9]. Continuous genomic monitoring is therefore critical for selecting the most appropriate sequential therapy. However, the procedures used to obtain tissue samples for genomic monitoring are invasive and costly, and accurate monitoring is often limited by tumor heterogeneity [10].

Cell-free DNA (cfDNA) has emerged as a specimen type that reflects the genomic landscape of tumors and can be obtained in a minimally invasive manner. In patients with advanced non-small cell lung cancer (NSCLC), the circulating tumor DNA (ctDNA) within cfDNA samples can be analyzed to detect mutations, monitor resistance mechanisms, and guide treatment decisions [11]. In 2016, the US Food and Drug Administration approved testing cfDNA for *EGFR* mutations as a non-invasive alternative to tissue biopsy [12]. Most current cfDNA assays target a narrow set of

mutations, which restricts their ability to capture the full spectrum of tumor evolution. Low-coverage whole-genome sequencing (lcWGS) of cfDNA is a promising solution because it enables genome-wide detection of copy-number alterations (CNAs) and chromosomal instability, which is particularly useful in patients with a high tumor burden [13]. Fragmentomic features of cfDNA, such as short-fragment enrichment, also correlate with tumor mutation burden and may improve the prediction of treatment response and disease progression [14].

Nevertheless, important gaps remain in our understanding of the longitudinal dynamics and predictive value of lcWGS-derived cfDNA features in NSCLC patients receiving EGFR-TKIs or chemotherapy. For example, few studies have used serial sampling at multiple time points to monitor cfDNA dynamics focusing on patients receiving immunotherapy, chemotherapy, or radiotherapy [15–17]. In addition, the value of lcWGS-derived cfDNA features for predicting progression-free survival (PFS) is underexplored in patients receiving EGFR-TKIs and chemotherapy. Against this background, we pursued three objectives in this study: (1) to comparatively analyze a set of lcWGS features of cfDNA obtained from advanced NSCLC patients treated with EGFR-TKIs and chemotherapy; (2) to use serial sampling at diagnosis, and week 4 and week 12 from treatment initiation to track changes in the cfDNA concentration, CNAs, ctDNA fraction, and fragmentation features; and (3) to evaluate whether these features predict early PFS outcomes. We further assessed the feasibility and utility of lcWGS cfDNA monitoring in these therapeutic contexts.

2 Methods

2.1 Patients

A total of 37 patients who were diagnosed with NSCLC were recruited at Chulabhorn Hospital between June 2023 and November 2024. Three patients were lost to follow-up by week 12, and one patient died during the study. Hence, 33 patients with NSCLC

who received either EGFR-TKI (n = 18) or conventional chemotherapy (n = 15) were included in this study. The median age of the patients was 63.5 years (range: 32–78 years). Ten healthy, age-unmatched (range: 26–42 years), individuals without known tumor disease (controls) were included as a reference cohort.

Patients with NSCLC, either as locally advanced disease with progression after prior radiotherapy or as metastatic disease, were included. Eligible patients were required to have an Eastern Cooperative Oncology Group performance status [18] of 0–2, have indications for systemic therapy, have had no prior systemic treatment for advanced disease, and have radiologically measurable disease. Clinical data were obtained by reviewing the patients' records (cut-off date: July 18, 2025). Data on the following parameters were collected: tumor histology, disease stage, systemic treatment, treatment duration, actionable genetic alteration (AGA), and tumor response after 12 weeks of treatment. The tumor response was evaluated using the Response Evaluation Criteria in Solid Tumors (version 1.1) and radiologic images, specifically chest/abdominal computed tomography images. PFS was assessed from therapy initiation to disease progression or death. Patients without progression or death at the data cut-off were censored at the last contact date.

2.2 Sample preparation and sequencing

Blood was collected from each patient at three time points: diagnosis (baseline), week 4, and week 12 after treatment initiation. Blood was collected once from each healthy control. At each sample collection, 8 mL of whole blood was drawn directly into a cell-free DNA BCT tube (Streck, La Vista, NE, USA) and stored at room temperature for no more than 7 days before cfDNA extraction and analysis. Each sample was centrifuged at 1,900 × g for 10 min. The plasma supernatant was aliquoted into a 2-mL microtube and stored at -80 °C until cell-free DNA extraction. Plasma isolation

was performed using the QIAamp Circulating Nucleic Acid Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The cfDNA was eluted in 30µL. One microliter of cfDNA was quantified using the Qubit dsDNA HS Assay Kit and Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). The lcWGS of the cfDNA specimens was outsourced to Macrogen Inc. (Seoul, South Korea). The laboratory verified the genomic DNA integrity using TapeStation cfDNA Screen Tape. A TruSeq Nano DNA Library (insert size: 350 bp) was prepared on 7G using the NovaSeq 6000 platform with 100 paired-end flow cells. The raw sequencing data were received from the laboratory online.

2.3 Whole-genome sequencing data pre-processing

The raw sequencing data (FASTQ format) were processed using a custom bioinformatics pipeline that included sequence quality control with FASTQC [19], and read alignment to the human reference genome (GRCh38/hg38) with bwa-mem [20]. Alignment sorting and PCR mark duplication were performed using Picard [21]. Patient genotypes were extracted and compared for relatedness using Somalier [22] to rule out the possibility of samples being mixed up.

2.4 Copy-number variant calling and ctDNA fraction

The ichorCNA was used to determine the CNAs and estimate the ctDNA fraction [23]. All samples that met the cut-off criteria of a genome coverage value ¿ 1× and an ichorCNA median absolute deviation score ¡ 0.15 were selected for further analysis. A Panel of Normals (PoN) was constructed separately from the 10 cfDNA samples from the controls. The ichorCNA parameters were modified as follows: –includeHOMD False, –chrs "c(1:22)", –chrTrain "c(c(1:18),c(20:22))", –estimateNormal True, –estimatePloidy True, –estimateScPrevalence FALSE, –scStates "", –txnE 0.9999, –txnStrength 10000, –ploidy "c(2)", –normal "c(0.8, 0.9, 0.95, 0.99, 0.995, 0.999)", and –maxCN 4. For the

ctDNA fraction, 3% was used as the threshold for distinguishing the "ctDNA+" samples from "ctDNA-" samples. The CNA profiles were visualized using ggplot2 [24] in R (v4.2.0).

2.5 Fragment-length analysis and ctDNA estimation score

The fragment-length analysis was performed using the cfdnakit bioconductor package [25] which extracts fragment-length data from the insert-size information in the alignment file and helps to visualize fragment lengths. Briefly, the short-to-long fragment ratio of a sample ($SL\,Ratio_{sample}$) was calculated by dividing the count of short fragments (100–150 bp) by the count of long fragments (151–250 bp) as follows:

$$SL\,Ratio_{sample} = \frac{N_{F.short}}{N_{F.long}}$$

where $N_{F.short}$ is number of short fragments; $N_{F.long}$ is number of long fragments;

To analyze genome-wide short-fragment aberrations, SL Ratio were calculated within w non-overlapping 1 Mb genomic windows ($SL\,Ratio_w$) as follows

$$SL\,Ratio_w = \frac{N_{F.short_w}}{N_{F.long_w}}$$

$w = \{1, 2, 3, ..., n\}$; where n is number of bins;

$N_{F.short_w}$is number of short fragments in bin w; $N_{F.long_w}$is number of long fragments in bin w.

Each $SL\,Ratio_w$ were adjusted for GC-content and sequence mappability using LOESS regression to ensure data quality using the information provided by the

QDNAseq package [26]. From these corrected values, the SL Ratio was determined for each window and converted into a z-score relative to the genome-wide median as follows.

$$SL\,norm_w = \frac{SL\,Ratio_w - median(\{SL\,Ratio_1,...,SL\,Ratio_n\} - \{SL\,Ratio_w\})}{mad(\{SL\,Ratio_1,...,SL\,Ratio_n\} - \{SL\,Ratio_w\})}$$

$$w = \{1, 2, 3, ..., n\};\ \text{where } n \text{ is number of bins}$$

The MAD is a term representing the median of the absolute deviation from the median. MAD was used as an alternative to the mean ± standard deviation, given its status as a robust measure of the variability of data and resistance to outliers. We calculated MAD from a sample as follows:

$$mad\,(SL\,Ratio_{1..n}) = median\,(|SL\,Ratio_i - median\,(SL\,Ratio_{1..n})|)$$

To isolate true signals, these normalized SL ratios were normalized against the normalized SL Ratio of samples in PoN.

$$zscore_w = \frac{SL\,norm_w - median(\{SL\,norm_{w,1},...,SL\,norm_{w,p}\})}{mad(\{SL\,norm_{w,1},...,SL\,norm_{w,p}\})}$$

$$w = \{1, 2, 3, ..., n\};\ \text{where } n \text{ is number of bins}$$

$$\text{and } p = \{1, 2, 3, ..., m\};\ \text{where } m \text{ is number of samples in PoN}$$

Finally, bins were segmented via circular binary segmentation (CBS) [27], and the median of $zscore$ was calculated per segment ($Z_{segment}$). The ctDNA estimation score (CES), adapted from Raman et al. [28], was determined by the cfdnakit bioconductor

package to comprehensively estimate the ctDNA based on the enrichment of short-fragment cfDNA and CNAs.

$$CES = \left(\sum_{i=1}^{N_{segment}} \left(\left| Z_{segment_i} \times l_{segment_i} \right| \right) / N_{segment} \right)$$

where $Z_{segment_i}$ is the z-score of segment$_i$;

$l_{segment_i}$ is the number of bins in segment$_i$

and $N_{segment}$ is the total number of segments

2.6 Statistical analysis and data visualization

The coefficient of determination R^2 was calculated as the square of Pearson's correlation coefficients. Statistical significance was set at p ¡ 0.05. All statistical analyses were conducted using R (v.4.5.1; R Foundation for Statistical Computing, Vienna, Austria), and all plots were generated using the ggplot2 package.

3 Results

3.1 The patients' demographic, clinical, and diagnostic characteristics

The patients' demographic, clinical, and diagnostic characteristics are recorded (Supplementary Table 1) and summarized in Table 1. The median age was 63.5 years (range: 32–78 years), and there was an almost equal distribution of males and females in the patient cohort (48.5% male, 51.5% female). AGAs were detected in 75.8% of the patients, with sensitizing EGFR mutations being the most prevalent variant, present

in 51.5% of the cohort. The patients received either chemotherapy (45.5%) with/without immunotherapy or targeted therapy (54.5%). The majority of the cohort (97%) received first-line therapy, while 3% received second-line therapy. The median follow-up period was 214 days. In terms of treatment efficacy, 57.6% of the patients achieved a partial response (PR), 27.3% had stable disease (SD), and 15.2% had progressive disease (PD).

Table 1: Patients' Demographic, Clinical, and Diagnostic Characteristics

Demographic		
	N	**%**
Sex		
Male	16	48.5
Female	17	51.5
Actionable genetic alterations		
ALK fusion	1	3.0
BRAF V600E	1	3.0
Sensitizing *EGFR* mutation[a]	17	51.5
EGFR Exon20ins	2	6.1
HER2 mutation	1	3.0
KRAS mutation	2	6.1
ROS1 fusion	1	3.0
None found	8	24.2
Treatment agents		
Chemotherapy	15	45.5
Targeted therapy	18	54.5
Line of treatment		
First line	32	97.0
Second line	1	3.0
Objective response rate[b]		
Partial response	19	57.6
Stable disease	9	27.3
Progression disease	5	15.2
[a] Defined as *EGFR* Exon 19 deletions and the L858R point mutation		
[b] Evaluated 12 weeks after treatment initiation; no complete responses were observed		

3.2 cfDNA concentration was higher in patients with progressed disease

The plasma cfDNA concentration in the 99 samples obtained from the 33 NSCLC patients and in the 10 samples obtained from the controls were measured (Supplementary Table 2). The overall cfDNA concentrations were broadly comparable between the controls (median: 1.5 ng/µL, range: 1.2–7.2) and the patients (median: 1.7 ng/µL, range: 0.3–89.2) (Figure 1A). There was no difference in the cfDNA concentration between the patients with EGFR mutations (median: 1.5 ng/µL, range: 0.35–89.2) and those without EGFR mutations (median: 2.00 ng/µL, range: 0.4–16.2) (Supplementary Figure **??**A). The highest cfDNA concentrations were detected at diagnosis (median: 3.1 ng/µL, range: 0.6–89.2). The cfDNA concentrations were lower than the baseline level at 4 weeks (median: 1.7 ng/µL, range: 0.5–11.4) and had significantly decreased by 12 weeks (median: 1.3 ng/µL, range: 0.4–16.2, $p = 0.001$).

When the clinical outcomes were taken into account, the patients with PD exhibited significantly higher ctDNA concentrations (median: 3.5 ng/µL, range: 0.6–16.2) than those with SD (median: 1.4 ng/µL, range: 0.4–9.8, $p = 0.003$) and PR (median: 1.5 ng/µL, range: 0.4–89.2, $p = 0.011$) (Figure 1B). Meanwhile, the cfDNA concentration did not significantly differ between the patients with PR or SD. Interestingly, the cfDNA dynamics showed a distinct pattern (Figure 1C). In the majority of patients with PR, the cfDNA concentration dropped significantly compared to the baseline at week 4 (median: 1.4 ng/µL, range: 0.5–11.4, $p = 0.034$) and became steady at week 12 (median: 1.3 ng/µL, range: 0.4–3.6). The cfDNA concentration in patients with SD did not change at week 4 (median: 1.7 ng/µL, range: 1.0–9.8) compared to the baseline but dropped significantly at week 12 (median: 0.8 ng/µL, range: 0.4–1.6, $p = 0.008$). In contrast, the cfDNA concentration in most patients with PD did not change dramatically at week 4 (median: 3.0 ng/µL, range: 1.5–8.0) and 12 (median: 3.9 ng/µL, range: 0.7–16.2) compared to the baseline.

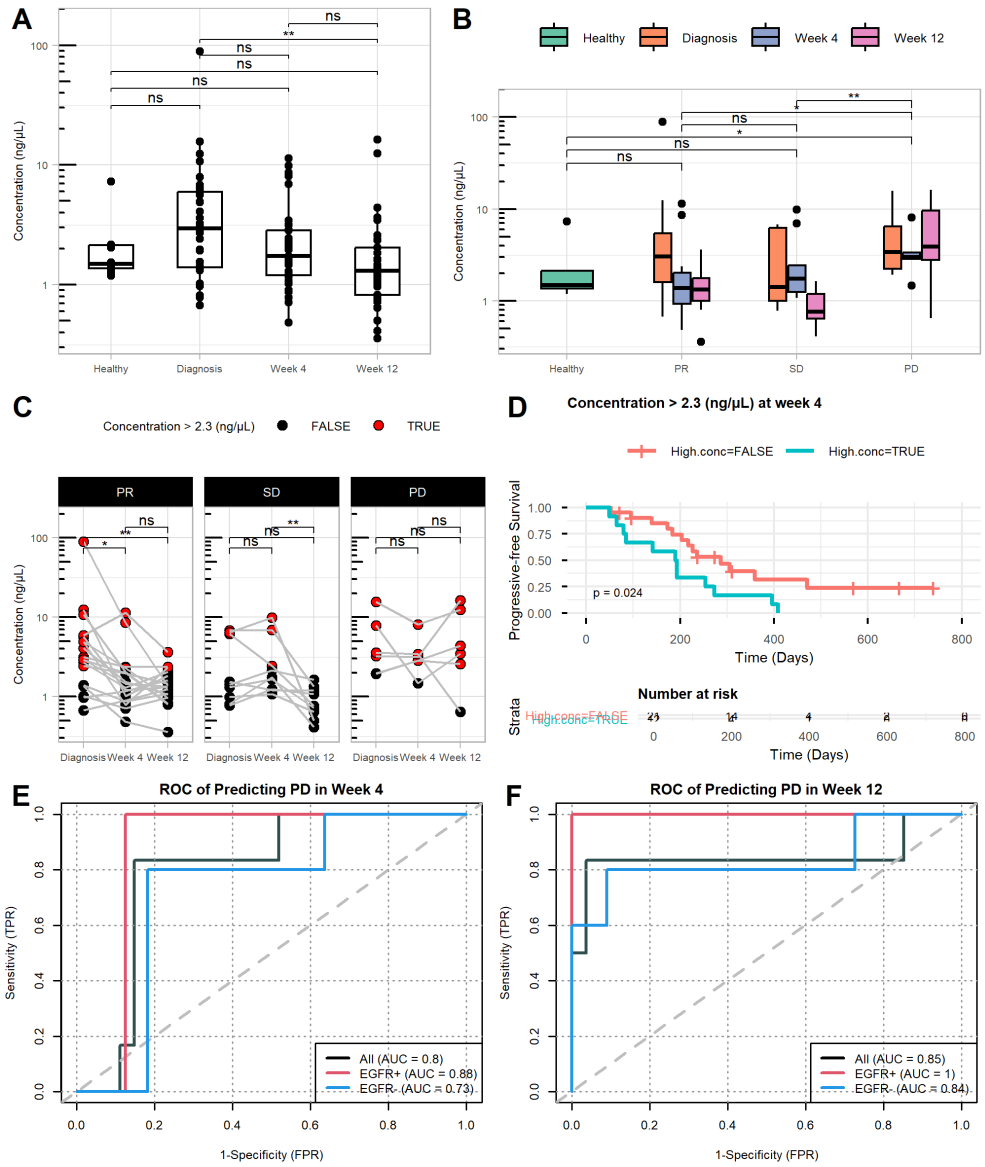


Fig. 1: The cfDNA concentration aligned with the clinical outcome and the ctDNA fraction status. (A) The cfDNA concentration significantly decreased from diagnosis to week 12. (B) When the clinical outcome was considered, patients with PD had significantly higher cfDNA concentrations through all clinical timepoints. (C) The cfDNA concentration decreased in patients with PR or SD from diagnosis to week 12, whereas it remained high and stable in patients with PD. (D) Patients with a high cfDNA concentration at week 4 showed a shorter PFS. The cfDNA concentration at weeks 4 (E) and 12 (F) could potentially predict which patients experience an unfavorable outcome.

Because the cfDNA concentration was significantly higher in patients with PD throughout the course of treatment, we examined the association between the cfDNA concentration and PFS. Interestingly, the cfDNA concentration at diagnosis was not significantly associated with PFS (Supplementary Figure **??**B). A low concentration of cfDNA (¡ 2.3 ng/µL) at week 4 correlated most significantly with PFS ($p = 0.024$) (Figure 1D), followed by the concentration at week 12 ($p = 0.041$,Supplementary Figure **??**C). We then examined the performance of the cfDNA concentration as a predictive marker for disease progression via receiver operating characteristic (ROC) curve analysis. The results showed that the concentration at week 12 (area under the curve [AUC] = 0.85, Figure 1F) outperformed the concentration at week 4 (AUC = 0.8, Figure 1E). Moreover, it demonstrated better performance in patients with EGFR mutations (EGFR+) than in those without EGFR mutations (EGFR-). Overall, the cfDNA concentration was highest in the patients at diagnosis, significantly decreased during and after treatment, and remained comparable between the controls and patients with PR or SD across the time points analyzed in this study. Persistently higher cfDNA levels and no decrease in the level after treatment initiation were associated with a poor therapeutic response and disease progression.

3.3 The ctDNA fraction reflected the treatment response and disease progression

The fraction of ctDNA was determined for each cfDNA sample. In general, the ctDNA fraction in the patients (median: 1.3%, range: 0.3–44) was significantly higher than that in the controls (median: 0.5%, range: 0.3–1, $p = 0.0491$) (Supplementary Figure **??**A). In the patients, there were no significant differences between the ctDNA fraction values at different time points (Supplementary Figure **??**B) or according to the EGFR mutation status (Supplementary Figure **??**C). When we took the clinical outcomes into account, patients with PR or SD were found to have lower and stable ctDNA fractions

(¡ 3%) at weeks 4 and 12, whereas patients with PD had significantly higher ctDNA fractions (median: 5%, range: 0.5–36.8) than patients with PR (median: 1.2%, range: 0.1–44.0, $p = 0.005$) and SD (median: 1.0%, range: 0.3–34.3, $p = 0.006$) throughout the course of treatment (Figure 2A). Next, we classified the samples as ctDNA+ or ctDNA- using a threshold of 3% and evaluated the predictive capacity of this ctDNA fraction status. As shown in Figure 2B, the ctDNA fraction status correlated with the clinical outcome. The cfDNA concentration in the ctDNA+ samples (median: 3.5 ng/µL, range: 0.5–36.2) was higher than that in the ctDNA- samples (median: 1.5 ng/µL, range: 0.4–89.2, $p = 0.003$), particularly in patients with SD and PD (Supplementary Figure **??**D). At diagnosis, almost half of the patients were ctDNA+ (14/33, 42%). The number of ctDNA+ patients was lower at weeks 4 (7/33, 21%) and 12 (9/33, 27%). At 4 weeks, 11% of patients with PR (2/18) and SD (1/9) were ctDNA+, compared to 67% of the patients with PD (4/6). Similarly, at week 12, a small number of patients with PR (3/18, 17%) or SD (1/9, 11%) were ctDNA+, while the majority of patients with PD (5/6, 83%) were ctDNA+. Similar to the cfDNA concentration, the ctDNA fraction showed distinct patterns across the clinical outcomes (Figure 2C). Patients with PR or SD who were ctDNA+ at diagnosis had decreasing ctDNA fractions at week 4, and in most cases, the low ctDNA fractions were maintained at week 12. In contrast, patients with PD had persistent or increasing ctDNA fractions from diagnosis to week 12.

In patients with EGFR+ tumors who were treated with EGFR-TKIs (Supplementary Figure **??**A), all those with PR and detectable ctDNA at diagnosis had ctDNA- at weeks 4 and 12. Patients with SD (3/3, 100%) had ctDNA- throughout treatment. In contrast, a patient (2LB-015) with PD exhibited a continuous rise in the ctDNA fraction from week 4 through week 12, although ctDNA was not detected at diagnosis. In patients with EGFR- tumors who were treated with chemotherapy (Supplementary Figure **??**B), reductions in the ctDNA fraction were also observed among patients with

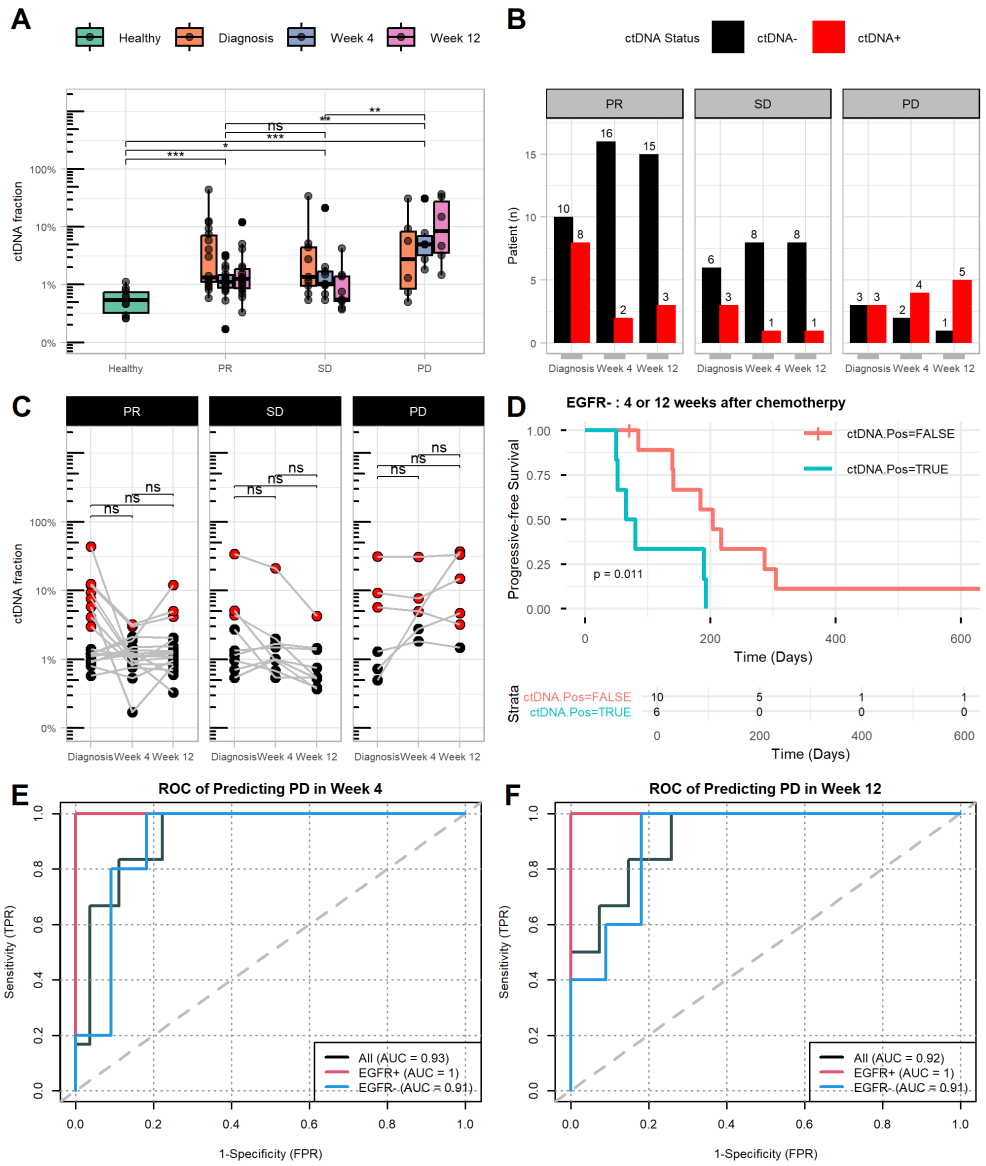


Fig. 2: The ctDNA fraction aligned with the clinical outcome and was associated with PFS. (A) The patients had larger ctDNA fractions than the controls. Patients with PD had larger ctDNA fractions than those with PR or SD. (B) Few patients with PR or SD were ctDNA+ at weeks 4 and 12, whereas more patients with PD were ctDNA+. (C) The ctDNA fraction was larger at diagnosis than at weeks 4 and 12 in patients with PR or SD, whereas the level persisted or increased in patients with PD. (D) In the EGFR- patients, those who were ctDNA+ at weeks 4 and 12 had a shorter PFS period. The ctDNA fraction at weeks 4 (E) and 12 (F) was a strong predictor of PD.

PR (4/5, 80%); however, they were less pronounced compared to those in patients treated with EGFR-TKIs. Similar to the patients with EGFR+ tumors, patients with EGFR- tumors with SD generally had low ctDNA fractions (5/6, 83%), while those with PD tended to have persistently large or increasing ctDNA fractions across all time points (4/5, 80%).

PFS was associated with the ctDNA fraction at various time points. The ctDNA fraction at diagnosis was not significantly associated with PFS (Supplementary Figure **??**A,D). However, in patients with EGFR- tumors, a high ctDNA fraction at week 4 (p = 0.016, Supplementary Figure **??**B) and week 12 (p = 0.011, Supplementary Figure **??**C) strongly correlated with a shorter PFS period. Hence, our results suggested that having a high ctDNA fraction (¿ 3%) at either week 4 or week 12 could be indicative of a shorter PFS period (p = 0.011, Figure2D) in EGFR- patients. However, the PFS of patients with EGFR+ tumors was not significantly associated with having a large ctDNA fraction at any time point (Supplementary Figure **??**D-F). Compared with the cfDNA concentration, the ctDNA fraction was a better predictor of PD at both week 4 (AUC = 0.93, Figure2E) and week 12 (AUC = 0.92, Figure2F). These results indicate that dynamic changes in the ctDNA level during the early phase of treatment reflect the therapeutic response and disease progression.

3.4 Longitudinal CNA dynamics differed by treatment response and therapeutic context

When we analyzed the CNA profiles of the samples, distinct patterns were evident in the samples from the patients with EGFR+ tumors who were treated with EGFR-TKIs and the patients with EGFR- tumors who were treated with chemotherapy. In several of the patients with EGFR+ tumors, widespread CNAs were detectable at diagnosis. CNAs were detected in 47% (8/17) of the patients with EGFR+ tumors at diagnosis (Supplementary Figure **??**). At weeks 4 and 12, most of these patients

had almost no detectable CNAs (17% [3/17]); Supplementary Figures **??** and **??**, respectively). In contrast, a patient with PD exhibited persistent CNAs at weeks 4 and 12. Patients with SD generally had low levels of CNAs at all time points.

In patients with EGFR- tumors, CNAs were also observed at diagnosis across multiple chromosomes. CNAs were detected in 37.5% (6/16) of these patients at diagnosis (Supplementary Figure **??**). At week 4, patients with PR showed some reduction in their levels of CNAs and no ctDNA+. Similarly, most patients with SD (5/6) had undetectable levels of ctDNA (Supplementary Figure **??**). At week 12, 80% (4/5) of the patients with PD continued to show extensive CNA profiles, indicating persistent genomic instability despite undergoing chemotherapy (Supplementary Figure **??**).

3.5 The proportion of short cfDNA fragments aligned with the size of the ctDNA fraction and correlated with clinical outcome

We explored the fragment characteristics of the cfDNA samples, with a focus on the proportion of short cfDNA fragments present at the different time points and its association with the clinical outcomes (Supplementary Figure **??**A-C). The cfDNA fragment size was inferred from the insert size of the mapped paired-end reads. Overall, the fragment-length distribution showed a unique characteristic of cfDNA. Most cfDNA fragments are reportedly 167 bp in length, which corresponds to the size of a piece of DNA wrapped around a histone octamer plus a linker fragment[29]. There are periodic peaks in size every 10 bp below 143 bp, which align with the helical turns of DNA. In this study, the patients with PR (Supplementary Figure **??**A) or SD (Supplementary Figure **??**B) exhibited a decreasing proportion of short cfDNA (¡150 bps) across the three time points, while this proportion remained stable over time in the patients with PD (Supplementary Figure **??**C). In week 12, patients with PD also showed the highest proportion of short cfDNA in the cohort (Figure 3A).

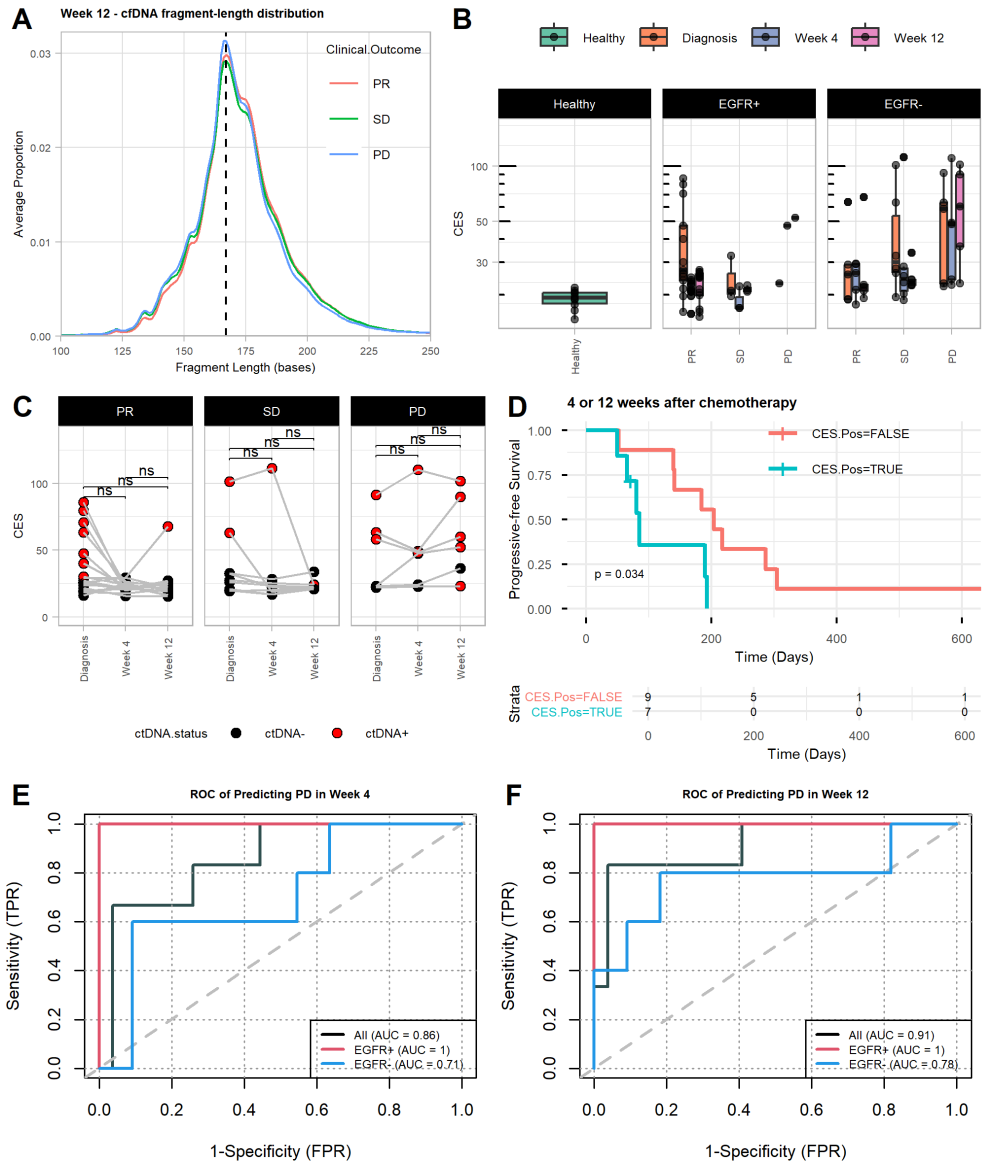


Fig. 3: Fragment-length feature analysis in cfDNA. (A) Fragment-length distribution of cfDNA obtained at week 12. The proportion of short-fragments (¡150 bps) was higher in patients with PD than in SD and PR. (B) CES distribution according to clinical outcome. Patients with PR or SD had mostly low CESs at weeks 4 and 12, while the majority of PD patients showed rising or persistently elevated CES (C). The high CES was strongly associated with a shorter PFS period (D) and showed good performance in predicting PD from cfDNA in week 4 (E) and week 12 (F)

The SL ratio did not significantly differ across the EGFR statuses (Supplementary Figure **??**A), time points (Supplementary Figure **??**B), or clinical outcomes (Supplementary Figure **??**C). However, it only showed a slightly positive correlation ($R^2 = 0.23$) with the size of the ctDNA fraction (Supplementary Figure **??**A). It was higher at every time point in the ctDNA+ samples compared to the ctDNA- samples, and the difference was particularly evident at diagnosis. The SL ratio in the ctDNA+ samples was lower at 12 weeks compared to that at diagnosis. The SL ratio at week 4 did not predict the probability of PFS (Supplementary Figure **??**D).

Compared to the SL ratio, the CES showed a stronger positive correlation ($R^2 = 0.65$) with the ctDNA fraction (Supplementary Figure **??**B). Thus, the CES should be a better indicator of the size of the ctDNA fraction. Patients with PR or SD exhibited decreasing or stable low CESs from diagnosis to week 12, while patients with PD, especially those with EGFR+ tumors, exhibited consistently high CESs (Figure 3B). When we analyzed the CESs of the individual patients, we found that their patterns resembled those of the ctDNA fractions (Figure 3C). The majority of patients with PR or SD had stable low CESs, whereas those with PD exhibited rising or high CESs at weeks 4 and 12. PFS was strongly associated with a high CES at week 4 or week 12 ($p = 0.034$, Figure 3D). The CES at week 4 (AUC = 0.86, Figure 3E) and week 12 (AUC = 0.91, Figure 3F) showed comparable performance in predicting which patients would have an unfavorable outcome.

4 Discussion

In this pilot study, we demonstrated that lcWGS of cfDNA is a feasible and informative approach for monitoring the treatment response in patients with advanced NSCLC. By analyzing longitudinal plasma samples from patients treated with either EGFR-TKIs or chemotherapy, we showed that the cfDNA concentration, CNA profile, ctDNA fraction size, and fragment-related features correlate with clinical outcomes

and can stratify PFS. Previous investigations of cfDNA from patients with NSCLC primarily relied on targeted mutation assays or digital PCR to track EGFR mutations or acquired resistance variants [30–32]. While these approaches have clinical utility, they can only be used to analyze predefined loci and thus cannot capture the broader genomic dynamics that reflect tumor heterogeneity and clonal evolution. In contrast, lcWGS enables genome-wide profiling of CNAs and fragmentomic analysis without prior knowledge of the tumor genotype. In prior studies that involved lcWGS of cfDNA from patients with metastatic squamous NSCLC or patients receiving immunotherapy, associations between tumor burden and survival outcomes were observed [15, 17]. However, to our knowledge, no previous study has examined serial samples obtained at multiple time points from patients receiving either EGFR-TKI or chemotherapy. Therefore, our findings provide novel evidence that cfDNA features derived via lcWGS have predictive value in two distinct therapeutic contexts for NSCLC. First, the cfDNA concentration was high at diagnosis and lower following the initiation of therapy; however, it was consistently high in patients with early progression. When the cfDNA concentration remained ¿ 2.3 ng/µL at week 4, the PFS period was significantly shorter. These findings suggest that a low or decreasing cfDNA concentration in the early stages of treatment may serve as a robust indicator for an early response to treatment. Second, the CNA and ctDNA fraction dynamics of the responders clearly differed from those of the non-responders. The lack of detectable CNAs at 4 and 12 weeks in the patients treated with EGFR-TKIs correlated with the radiographic findings and durable disease control, whereas the patients who received chemotherapy and had consistently high levels of CNAs were more likely to have PD. Hence, longitudinal monitoring of CNAs may help guide treatment decisions by providing timely information on therapeutic efficacy that potentially precedes information on disease progression gained from radiologic modalities. Third, the analyzed fragmentomic features, particularly the CES, strongly correlated with the

ctDNA fraction size, and their levels were consistently higher in patients with PD. While the SL ratio showed modest value, the CES was found to be a more robust metric of tumor-derived cfDNA. These findings suggest that fragmentation signatures complement CNA data and may improve risk stratification in cfDNA-based monitoring of disease progression. Overall, our results highlight the potential of lcWGS-based cfDNA analysis as a non-invasive, broadly applicable biomarker platform. Unlike targeted genotyping, this type of analysis does not require prior tumor profiling, making it suitable for routine monitoring in heterogeneous clinical settings. The ability to identify patients at risk of early progression—regardless of whether they are receiving targeted therapy or chemotherapy—has direct implications for stratification and tailoring therapeutic strategies. Importantly, early detection of progression in patients with EGFR+ tumors could result in timely consideration of next-line therapies, such as third-generation TKIs or combination regimens. This study is limited by its modest sample size and single-center design, which may constrain the generalizability of the results. Additionally, although we included patients treated with EGFR-TKIs or chemotherapy, each treatment subgroup contained relatively few patients. The healthy controls samples used to construct the Panel of Normals (PoN) were collected from age-unmatched healthy individuals rather than from an age- and comorbidity-matched community cohort. Although sex distribution was balanced, the controls were younger than the patient cohort and subclinical comorbidities were not assessed, both of which can independently influence cfDNA release. This mismatch may introduce systematic bias in ichorCNA-derived ctDNA fraction estimates, particularly near the 3% detection threshold. Future studies should construct the PoN from age- and sex-matched healthy individuals with a larger reference set (n ¿ 20) to improve background noise correction. Looking forward, larger prospective studies are required to validate the

cfDNA-derived thresholds and establish standardized workflows for lcWGS in clinical practice. Integrating additional multi-omic cfDNA features, such as methylation patterns, could further enhance the predictive accuracy of cfDNA-derived parameters.

5 Conclusion

In this pilot study, we demonstrated that lcWGS of cfDNA is a promising approach for predicting treatment outcomes and PFS in patients with advanced NSCLC. The observed clinical utility of the cfDNA-derived features across patients treated with EGFR-TKIs or chemotherapy supports the broad applicability of this approach and provides a foundation for larger prospective validation studies in precision oncology.

6 List of abbreviations

AGA	Actionable genetic alteration
AUC	Area under the curve
bwa-mem	Burrows-Wheeler Aligner — Maximal Exact Matches
CBS	Circular binary segmentation
CES	ctDNA estimation score
cfDNA	Cell-free DNA
CNA	Copy-number alteration
CT	Computed tomography
ctDNA	Circulating tumor DNA
ECOG	Eastern Cooperative Oncology Group
EGFR	Epidermal growth factor receptor
EGFR-TKI	EGFR-tyrosine kinase inhibitor
lcWGS	Low-coverage whole-genome sequencing
LOESS	Locally estimated scatterplot smoothing
MAD	Median absolute deviation
NSCLC	Non-small cell lung cancer
PCR	Polymerase chain reaction
PD	Progressive disease
PFS	Progression-free survival
PoN	Panel of Normals
PR	Partial response
RECIST	Response Evaluation Criteria in Solid Tumors
ROC	Receiver operating characteristic
SD	Stable disease
SL ratio	Short-to-long fragment ratio
WGS	Whole-genome sequencing

7 Declarations

Acknowledgments. First of all, we would like to thank all the subjects who participated in this study. We thank Dr. Kristen Sadler and Tom Buckle from Scribendi (www.scribendi.com) for editing a draft of this manuscript.

Author contributions. P.P., T.S., and C.P. conceived and designed the study. P.P. and T.S. wrote the manuscript. J.P., K.J., B.P., and T.S. collected and processed the subject samples. T.S. performed the experiments. P.P. performed bioinformatics data analysis. P.P. drafted, reviewed, edited, and revised the data and manuscript. C.P. is responsible for finalization of the manuscript. All authors have read and approved the final manuscript.

Funding. This study was financially supported by the Chulabhorn Royal Academy (grant number E2566/103). The funder had no role in the study design, data collection, analysis, interpretation, or preparation of the manuscript.

Data availability. The genomics datasets generated and/or analysed during the current study are available in the NCBI Sequence Read Archive (SRA) under BioProject PRJNA1447685.

Ethics approval and consent to participate. The study was conducted in accordance with the Declaration of Helsinki and was approved by the Chulabhorn Royal Academy Institutional Review Board (ethics approval ID: EC 043/2565). Written informed consent to participate has been obtained from all the participants.

Consent for publication. Written informed consent for publication of their personal or clinical details along with any identifying images has been obtained from all the participants and/or their legal representatives.

Competing interests. The authors declare no competing interests.

References

[1] Reungwetwattana T, Oranratnachai S, Puataweepong P, Tangsujaritvijit V, Cherntanomwong P. Lung cancer in Thailand. J Thorac Oncol. 2020;15(11):1714–1721.

[2] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–249.

[3] Ferlay J, Ervik M, Lam F, Laversanne M, Colombet M, Mery L, et al.: Global Cancer Observatory: Cancer Today. Available from: https://gco.iarc.who.int/today.

[4] Shi Y, Au JSK, Thongprasert S, Srinivasan S, Tsai CM, Khoa MT, et al. A prospective, molecular epidemiology study of EGFR mutations in Asian patients with advanced non–small-cell lung cancer of adenocarcinoma histology (PIONEER). J Thorac Oncol. 2014;9(2):154–162.

[5] Midha A, Dearden S, McCormack R. EGFR mutation incidence in non-small-cell lung cancer of adenocarcinoma histology: a systematic review and global map by ethnicity (mutMapII). Am J Cancer Res. 2015;5(9):2892.

[6] Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al. Gefitinib or carboplatin–paclitaxel in pulmonary adenocarcinoma. N Engl J Med. 2009;361(10):947–957.

[7] Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Isobe H, et al. Gefitinib or chemotherapy for non–small-cell lung cancer with mutated EGFR. N Engl J Med. 2010;362(25):2380–2388.

[8] Yu HA, Arcila ME, Rekhtman N, Sima CS, Zakowski MF, Pao W, et al. Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. Clin Cancer Res. 2013;19(8):2240–2247.

[9] Sequist LV, Waltman BA, Dias-Santagata D, Digumarthy S, Turke AB, Fidias P, et al. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. Sci Transl Med. 2011;3(75):75ra26–75ra26.

[10] Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366(10):883–892.

[11] Diaz Jr LA, Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. J Clin Oncol. 2014;32(6):579–586.

[12] Kwapisz D. The first liquid biopsy test approved. Is it a new era of mutation testing for non-small cell lung cancer? Ann Transl Med. 2017;5(3):46.

[13] Heitzer E, Haque IS, Roberts CE, Speicher MR. Current and future perspectives of liquid biopsies in genomics-driven oncology. Nat Rev Genet. 2019;20(2):71–88.

[14] Mouliere F, Chandrananda D, Piskorz AM, Moore EK, Morris J, Ahlborn LB, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. Sci Transl Med. 2018;10(466):eaat4921.

[15] Carbonell C, Frigola J, Pardo N, Callejo A, Iranzo P, Valdivia A, et al. Dynamic changes in circulating tumor DNA assessed by shallow whole-genome sequencing associate with clinical efficacy of checkpoint inhibitors in NSCLC. Mol Oncol. 2023;17(5):779–791.

[16] Balázs Z, Balermpas P, Ivanković I, Willmann J, Gitchev T, Bryant A, et al. Longitudinal cell-free DNA characterization by low-coverage whole-genome sequencing in patients undergoing high-dose radiotherapy. Radiother Oncol. 2024;197:110364.

[17] Janke F, Gasser M, Angeles AK, Riediger AL, Görtz M, Appenheimer L, et al. Low-coverage whole genome sequencing of cell-free DNA to predict and track immunotherapy response in advanced non-small cell lung cancer. J Exp Clin Cancer Res. 2025;44(1):87.

[18] Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol. 1982;5(6):649–656.

[19] Andrews S.: FastQC: A Quality Control Tool for High Throughput Sequence Data. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

[20] Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997. 2013;.

[21] Broad Institute.: Picard Tools. Broad Institute. https://broadinstitute.github.io/picard/.

[22] Pedersen BS, Bhetariya PJ, Brown J, Kravitz SN, Marth G, Jensen RL, et al. Somalier: rapid relatedness estimation for cancer and germline studies using efficient genome sketches. Genome Med. 2020;12(1):62.

[23] Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. Nat Commun. 2017;8(1):1324.

[24] Wickham H. ggplot2. Wiley Interdiscip Rev Comput Stat. 2011;3(2):180–185.

[25] Puranachot P.: cfdnakit: Fragment-length analysis package from high-throughput sequencing of cell-free DNA (cfDNA). Available from: https://bioconductor.org/packages/cfdnakit.

[26] Sie D, Scheinin I, Lieshout Sv, Cordes M, Pinkel D, Albertson DG, et al. QDNAseq: a bioinformatics pipeline for DNA copy number analysis from shallow whole genome sequencing with noise levels near the probabilistic lower limit imposed by read counting. Clin Cancer Res. 2016;22(1_Supplement):52.

[27] Olshen AB, Bengtsson H, Neuvial P, Spellman PT, Olshen RA, Seshan VE. Parent-specific copy number in paired tumor–normal studies using circular binary segmentation. Bioinformatics. 2011;27(15):2038–2046.

[28] Raman L, Van der Linden M, Van der Eecken K, Vermaelen K, Demedts I, Surmont V, et al. Shallow whole-genome sequencing of plasma cell-free DNA accurately differentiates small from non-small cell lung carcinoma. Genome Med. 2020;12(1):35.

[29] Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. Cell. 2016;164(1):57–68.

[30] Chan DLH, Toh GLX, Goh LL. Clinical implementation of plasma EGFR T790M testing using droplet digital PCR in TKI-resistant NSCLC patients. Exp Mol Pathol. 2020;116:104515.

[31] Ntzifa A, Kotsakis A, Georgoulias V, Lianidou E. Detection of EGFR mutations in plasma cfDNA and paired CTCs of NSCLC patients before and after osimertinib therapy using crystal digital PCR. Cancer Res.

2022;82(12_Supplement):5158–5158.

[32] Huang CT, Lin CA, Su TJ, Yang CY, Tsai TH, Hsu CL, et al. Monitoring of T790M in plasma ctDNA of advanced EGFR-mutant NSCLC patients on first-or second-generation tyrosine kinase inhibitors. BMC Cancer. 2023;23(1):234.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1SupplementaryTable1.xlsx
- Additionalfile2SupplementaryTable2.xlsx
- Additionalfile3Supplementaryfigures.pdf