Neoplasma 2019; 66(4): 652–660

doi:10.4149/neo_2018_181130N910

# Efficacy comparison of targeted next-generation sequencing in the identification of somatic mutations in circulating tumor DNA from different stages of lung cancer

Y. CHEN[1], T. HAN[2], Y. ZHOU[1], B. MAO[2], W. ZHUANG[1,*]

[1]*Department of Thoracic Surgery, Xiangya Hospital of Central South University, Changsha, Hunan, China;* [2]*Beijing Genecast Biotechnology Co., Beijing, China*

**Correspondence: zhuangwei@csu.edu.cn*

**Received November 30, 2018 / Accepted February 20, 2019**

This study aims to assess the potential clinical application of targeted next generation sequencing (NGS)-based deep sequencing for the detection of clinically relevant mutations in circulating tumor DNA (ctDNA) obtained from non-small cell lung cancer (NSCLC) patients. Targeted deep sequencing was performed to identify High Confidence Somatic Variants (HCSVs) in matched tumor tissue DNA (tDNA) and ctDNA in 50 NSCLC patients. Our results demonstrated that NSCLC patients with Stage IV (61.5%) exhibited a higher concordance rate at the mutation level between plasma ctDNA and tDNA samples than patients with Stage I–III (14.5%). Moreover, it is noteworthy that the allele frequency of these detected HCSVs in ctDNA increased with the advance in tumor stage. Besides, using tDNA as a reference, the sensitivity of plasma ctDNA analyzed by deep NGS for actionable EGFR was much higher in patients with Stage IV (66.6%) than in patients with Stage I–III (7.7%). In conclusion, it appears that ctDNA NGS-based deep sequencing is a feasible approach to identify mutations in patients with Stage IV NSCLC. However, additional methods with higher sensitivity and specificity are needed to improve the successful application of this platform in the earlier stages of NSCLC.

*Key words: circulating tumor DNA; lung cancer; next-generation sequencing; target sequencing*

Lung cancer has been the leading cause of cancer-related deaths [1, 2]. Among its common subtypes, non-small cell lung cancer (NSCLC) has often been characterized by unique mutation patterns [3–5]. In order to assess these somatic mutations, circulating tumor DNA (ctDNA) analysis appears to be a promising method [6–10]. The ctDNA analysis can be used to assess single mutations or whole-genomes [11]. Since 2000, allele-specific PCR methods have been applied for the detection of hot-spot mutations in serum and plasma. Furthermore, appropriately designed assays for individual mutations have been able to achieve high sensitivity [12, 13]. In this context, digital PCR assays are quantitative and highly sensitive, and have thereby been used extensively to quantify ctDNA levels [14–16]. However, these methods are typically suitable for investigating a small number of mutations and are often applied to analyze cancer hot-spot mutations. In order to expand the region of sequencing, targeted sequencing using PCR amplicons or hybrid capture has been previously used [6, 8, 17].

Multiple studies have demonstrated that the detection of mutations in plasma ctDNA using targeted deep sequencing shows high concordance with mutations detected in tDNA from advanced and late stage cancers [18–20]. Thus, it makes more sense to monitor advanced cancers, in which the multiple or serial biopsy of tissues is impractical. The study conducted by Yang et al. [21] observed that ctDNA mutation patterns are independent of tumor stage, thereby suggesting that circulating tumor cells enter circulation during the early stage of cancer progression, and tumor DNA dissemination is an early event. This is clinically significant and indicates that mutation detection in ctDNA can serve as a non-invasive and early detection approach. However, the low tumor burden in the early stage of cancer has made mutation detection in ctDNA challenging [11].

Many methods have been developed for the analysis of ctDNA in late-stage cancer patients. However, no specific method has been systematically used for the analysis of early-stage diseases. Recently, several studies have explored the

application of next generation sequencing (NGS)-based deep sequencing in early stage cancer [21–23]. All these studies usually apply small scale gene panels or personalized panels for the analysis. Since each target gene provides an independent opportunity to detect ctDNA variants, a larger panel of genes would thereby increase the probability of detecting at least one gene alteration in plasma obtained from cancer patients.

In the present study, a panel covering 1 407 cancer related genes and spanning 1.15Mb (1.15 million base pairs) of human genome, was used to explore the clinical utility of targeted NGS deep sequencing by evaluating the concordance between ctDNA and tDNA in non-metastatic stages (I–III) and the metastatic stage (IV) of NSCLC.

## Materials and methods

**Patient and tumor tissue collection.** A total of 50 patients with different stages of NSCLC were included in the present study after providing a written informed consent. Leukocytes and plasma samples were collected from each patient along with matched fresh frozen or formalin-fixed paraffin embedded (FFPE) tumor tissue sections. The histopathologic assessment confirmed that all collected tumor tissue samples had ≥10% viable tumor cell content. The study was approved by the Ethics Committee of Xiangya Hospital of Central South University (Changsha, China).

**Sample preparation.** Genomic DNA from fresh frozen tissues was extracted using a Tiangen genomic DNA kit (Tiangen, Beijing, PRC), while a blackPREP FFPE DNA kit (Analytik Jena, GER) was used to extract genomic DNA from FFPE tumor tissue specimen slides. Then, these DNA samples were quantified using a Qubit dsDNA HS Assay kit (Life Technologies, USA).

In addition, peripheral blood lymphocytes and plasma were initially separated by centrifugation at 1 600 × g for 10 minutes. Subsequently, the supernatant plasma was centrifuged again at 16 000 × g for another 10 minutes in a new 2-ml centrifuge tube. Then, DNA was extracted from peripheral blood lymphocytes using a Tiangen whole blood DNA kit (Tiangen, Beijing, China). Similarly, DNA from the centrifuged plasma was extracted using a MagMAX™ Cell-Free DNA isolation kit (Life Technologies, California, USA). Concentrations of both lymphocyte and plasma DNA were assessed using a Qubit dsDNA HS Assay kit or Qubit dsDNA BR Assay kit (Life Technologies, California, USA).

**Targeted sequencing.** For sequencing, genomic DNA was first sheared into 150–200 bp fragments using a Covaris M220 Focused-ultrasonicator™ Instrument (Covaris, Massachusetts, USA). Next, fragmented DNA and cell-free DNA (cfDNA) libraries were constructed using a KAPA HTP Library Preparation Kit (Illumina platforms) (KAPA Biosystems, Massachusetts, USA), according to manufacturer's instructions. Subsequently, the captured DNA fragments were sequenced using the Illumina Novaseq 6000 sequencing system, and 150 bp paired-end reads were identified using a custom capture panel (Genecast, Beijing, China), which covered 1 407 genes associated with cancer diagnosis and prognosis. Raw sequencing data was de-multiplexed using BCL2FASTQ software into individual paired-reads FASTQ files. The sequencing reads were eventually aligned against the hg19 reference genome using BWA MEM software through default parameters and sorted and indexed with Samtools. The duplicated reads from tumor tissues and blood cell samples were marked using the Picard's MarkDuplicate tool and were excluded from subsequent analyses. Further reads and mapping qualities were evaluated using FastQC and Samtools flagstat. The average unique sequence coverage was at least 900× for lymphocytes and 2 400× for tumor tissue samples. The plasma ctDNA samples were sequenced with a high average coverage of at least 4 600×.

**Single nucleotide variant (SNV) and insertions or deletions (INDEL) calling.** Paired-sample variant callings on tumor and plasma were performed using matched blood cell samples as a baseline reference. Varscan 2 somatic calling was initially employed to detect somatic SNPs and Indels using loose parameters, and the calling results were subsequently filtered with custom and more stringent criteria. Especially in tumor tissue samples, only bases with a depth of over 40× were considered, while for plasma samples, the minimum depth was 400×. Both tumor and plasma somatic variants (SVs) were required to have at least one support read for each strand with a background p-value under 0.05. In addition, the minor allele frequency for tumor and plasma SVs were not less than 2% and 0.2%, respectively. Among these SVs, those with a p-value of <0.05 along with a deeper depth of 100× and 1 000× for tumor and plasma, respectively, and a larger allele frequency of over 5% and 1% for tumor and plasma, respectively, were termed as high-confidence somatic variants (HCSVs) in the present study. All others were categorized as low-confidence somatic variants (LCSVs). Moreover, the bases that met the minimum depth requirement (100× for tumors and 1 000× for plasma), but had a minor allele frequency of less than 1% (for tumors) or 0.1% (for plasma), were defined as wild types (WTs), while the other bases were considered to be in the gray zone, that is, no-calls (NCs) due to insufficient information.

Furthermore, common variants and potential background noises introduced by the platform were also excluded to ensure the reliability of the present data. In particular, SVs with a population frequency of >0.5%, according to the ExAC database or 1000 Genomes database, were excluded. In addition to these two databases, a sample pool based on 30 healthy Chinese donors was also used as the standard normal database to identify either prevalent germline mutations in the Chinese population, or technical artifacts prevailing in sequencing platform. Variants with a prevalence of no less than 20% in the standard normal database were blacklisted in the present study and excluded from the downstream analysis. In addition, based on a similar concept

of variants with no less than 20% prevalence, the larger real-life patient sample population, which contained more than 5 000 samples of different cancer types, as tested by Genecast, Beijing, China on a Novaseq 6000 sequencer, were considered as sequencing-specific errors, and were also excluded from further analysis.

Finally, in order to compare the calling results between paired tumor-plasma samples and avoid uncertainty with no-call (NC) data points, merely SVs and WTs were taken into account. In other words, when a specific locus was observed to be NC in one sample, the concordance of this locus was not counted, even if the SV or WT was at the same position in the paired sample.

**Statistical and concordance analysis.** All patients with matching tDNA and ctDNA sequencing results were included in the final analysis, and all statistical tests were performed using the R v3.4.1 software. The concordance of HCSVs between tumor tissues and matched plasma samples was defined as the ratio of two times the number of concordant variants over the total number of HCSVs from either tumor tissues or plasma samples, with NC genomic positions excluded. The Wilcoxon rank-sum test was used to compare two independent continuous variables. When comparing more than two independent samples, the Kruskal-Wallis test followed by Dunn-Bonferroni post hoc test was used. The statistical analyses were performed using SPSS version 20.0 (IBM, New York, NY, USA) software. A p-value of <0.05 was regarded as significant.

$$\text{Concordance}=\frac{2\times\text{Count}((HCSV_{tumor}-NC_{plasma})\cap(HCSV_{plasma}-NC_{tumor}))}{\text{Count}(HCSV_{tumor}-NC_{plasma})+\text{Count}(HCSV_{plasma}-NC_{tumor})}$$

## Results

**Patient characteristics.** The baseline characteristics of patients (n=50) are summarized in Table 1. In the present study, 35 patients were male, and 15 patients were female. At the time of biopsy, the age of these patients ranged from 34 to76 years, with a median age of 57.8 years. In total, 19 patients were categorized as stage I, 11 patients were categorized as stage II, 13 patients were categorized as stage III, and seven were categorized as stage IV NSCLC. In addition, 30 patients had adenocarcinoma, 16 patients had squamous cell carcinoma (SCC), two patients had large cell lung cancer, and two patients had lymphoepithelial carcinoma.

**Somatic variant landscape.** Based on the filtering criteria, a total of 211 HCSVs were identified in tDNA, including 181 SNVs and 30 INDELs. In contrast, 253 HCSVs were identified in plasma ctDNA, including 233 SNVs and 20 INDELs. Among both plasma ctDNA and tDNA, 42 overlapped HCSVs were identified, including 40 SNVs and 2 INDELs (Figure 1A).

In addition, tDNA revealed at least one mutation in 42/50 patients (84%) (7/7 patients in stage IV vs. 35/43 patients in stage I–III). Importantly, the most frequently mutated genes observed in tDNA (N≥3) were *TP53* (48% of patients), *EGFR* (34%), *PIK3CA* (10%), *KRAS* (8%), *GNAS* (6%), *NFE2L2* (6%), *CYP2D6* (6%) and *PTEN* (6%) (Figure 2A). Similarly, for ctDNA, at least one mutation was detected in 45/50 patients (90%) (7/7 patients in stage IV vs. 38/43 patients in

**Table 1. Patient baseline characteristics.**

| Characteristics | Number |
|---|---|
| Age, year | |
| ≥60 | 25 |
| <60 | 25 |
| Gender | |
| Male | 35 |
| Female | 15 |
| Smoking History | |
| Smoking | 21 |
| Nonsmoking | 29 |
| Histology | |
| Adenocarcinoma | 30 |
| Squamous | 16 |
| large cell lung cancer | 2 |
| Lymphoepithelial carcoma | 2 |
| Stage | |
| I | 19 |
| II | 11 |
| III | 13 |
| IV | 7 |

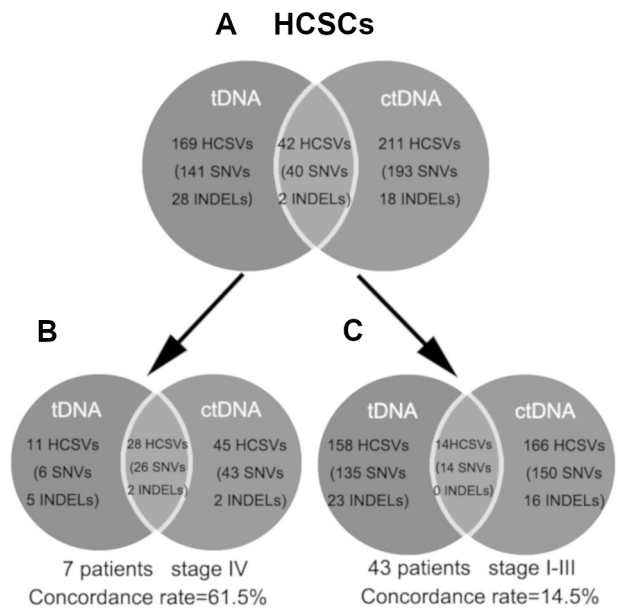


**Figure 1. A) Number of HCSVs identified in tDNA, plasma ctDNA, and both tDNA and plasma ctDNA. B) Concordance rate of HCSVs between plasma ctDNA and tDNA at mutation level in seven patients with late metastatic stage (stage IV) NSCLC. C) Concordance rate of HCSVs between plasma ctDNA and tDNA at mutation level in 43 patients with non-metastatic stage (stage I–III) NSCLC.**

stage I–III), and the most frequently mutated genes (N≥3) observed were *EGFR* (14%), *TP53* (14%), *BRCA2* (10%), *ALK* (8%) and *PIK3CA* (6%) (Figure 2B).

The Kruskal-Wallis test indicated a p-value <0.001, upon comparing the allele frequency of somatic mutations in different stage patients. Furthermore, Dunn-Bonferroni post hoc test also indicated that allele frequency of somatic mutations in stage IV group patients was higher than stage I (p<0.001), II (p<0.001) and III (p=0.067) patients. Similarly, the allele frequency of somatic mutations in stage III group patients was higher than stage II (p=0.009) and I (p=0.054) patients. Collectively, these results indicated that allele frequency of detected HCSVs in ctDNA, increased with the advanced tumor stage (Figure 3A).

Furthermore, 16 of 50 patients (32%) had overlapping HCSVs between both tDNA and ctDNA. Interestingly, all seven stage IV NSCLC patients and nine stage I–III NSCLC patients had overlapping HCSVs. The most frequent overlapping mutations (n≥2) were in the *TP53* (12%), *EGFR* (8%), *PIK3CA* (4%) and *HGF* (4%) genes (Figure 2C). Consistent with the variant frequency of individual HCSVs, the overlapped HCSVs also exhibited a significantly greater ctDNA variant frequency in patients with metastatic stage, when compared to those with non-metastatic stage (p<0.001) (Figure 3B).

**Mutation concordance in tDNA and plasma ctDNA.** Higher variant concordance rates were observed between plasma ctDNA and tDNA in patients with metastatic stage, in comparison to those with earlier and non-metastatic stages (stage IV patients 61.5% vs. stage I–III patients 14.5%). A total of 39 and 73 HCSVs were identified in tDNA and ctDNA of seven patients with late stage NSCLC, respectively. Among these HCSVs, 28 HCSVs were overlapped variants, resulting in a 61.5% concordance rate at the mutation level (Figure 1B). Meanwhile, 43 non-metastatic stage NSCLC patients had 172 HCSVs in the tDNA and 180 HCSVs in the ctDNA. Among these, only 14 variants overlapped, leading to a 14.5% estimated concordance rate (Figure 1C). More specifically, the concordance rates in stage I, II and III NSCLC patients were 2.2%, 31.3% and 15.8%, respectively.

**Analysis of the specificity and sensitivity of the ctDNA NGS-based test for detecting KRAS and actionable EGFR variants.** Using tDNA as a reference, it was identified that the ctDNA NGS-based test exhibited high specificity for

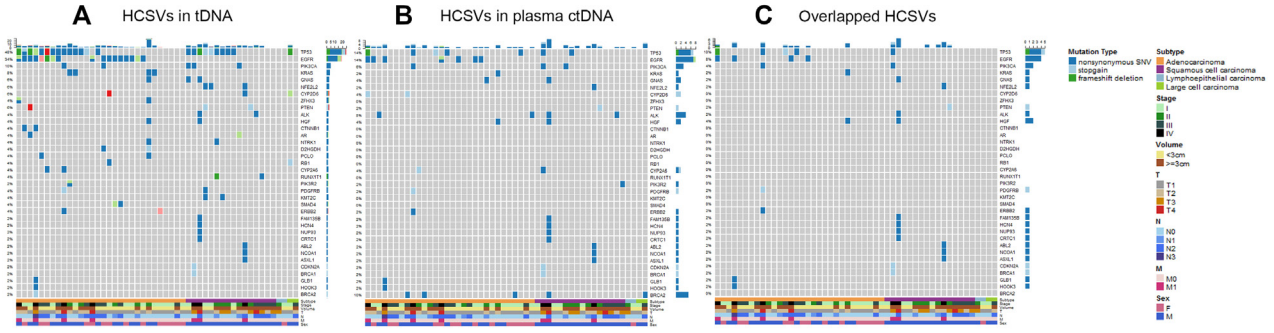


**Figure 2. Mutation type (SNV, stopgain, frameshift deletion, non-frameshift deletion, frameshift insertion, and non-frameshift insertion) identified in tDNA is compared with that found in plasma ctDNA. A) High Confidence Somatic Variants detected in tDNA. B) High Confidence Somatic Variants detected in plasma ctDNA. C) Overlapped High Confidence Somatic Variants in plasma ctDNA and tDNA. Patients were categorized based on histology type, size, stage, gender and age (bottom).**

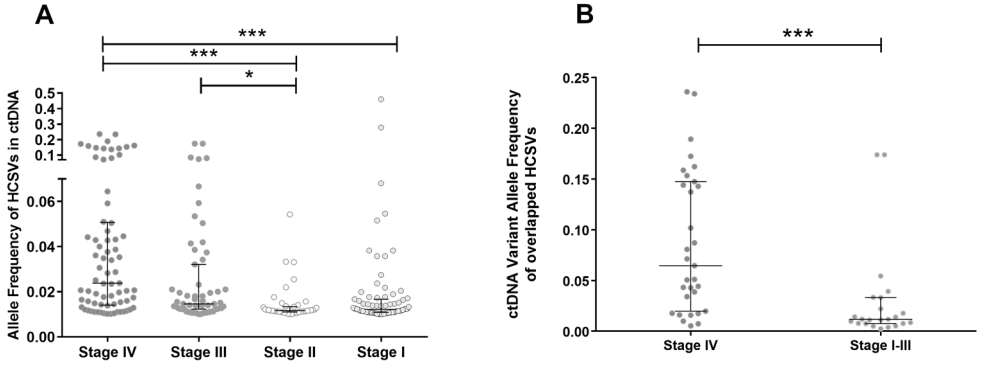


**Figure 3. A) Allele frequency of detected HCSVs in ctDNA increased with the advanced tumor stage. B) The overlapped HCSVs also exhibited a significantly greater ctDNA variant allele frequency in patients with metastatic stage (stage IV), when compared to those with non-metastatic stage (stage I–III). ***p<0.001, *p<0.05. Lines indicate the median value and interquartile range.**

detecting actionable EGFR variants (L858R, G719X, T790M, exon 19 del and exon 20 ins) (100%; 31/31) and KRAS G12X (100%; 39/39) in 43 stage I–III NSCLC patients. However, this assay revealed a low sensitivity of 7.7% (1/13) for EGFR alterations (L858R, G719X, exon 19 del and exon 20 ins) and 25.0% (1/4) for KRAS mutations in stage I–III NSCLC patients. Interestingly, in metastatic NSCLC patients, no KRAS mutation was detected in both tDNA and ctDNA, but the ctDNA NGS-based test revealed a high sensitivity of 66.6% (2/3) and a high specificity of 100% (4/4) for EGFR alterations (Table 2).

**Clinical characteristic and mutation detection in the non-metastatic stage.** Among non-metastatic stage NSCLC patients, the concordance of all variants was much higher in stage II and III (25.5%), than in stage I (2.2%). Furthermore, the concordance of all variants was also much higher in patients who specifically have a tumor size of ≥3 cm (26.1%), when compared to patients with a tumor size of <3 cm (4.0%) (Figure 4B). However, the concordance rate in patients with lymph node involvement (16.7%) was similar to those without lymph node involvement (13.5%) (Figure 5B).

The variant allele frequency of ctDNA in non-metastatic stage NSCLC patients with a tumor size of ≥3 cm was higher, when compared to patients with a tumor size of <3 cm, which had a marginal significance (Figure 4A, p=0.052). However, no significant difference was observed in variant

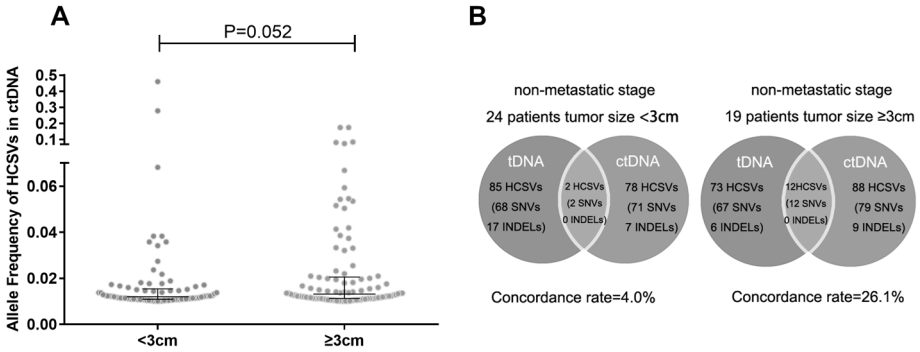


**Figure 4. A) Allele frequency of detected HCSVs in ctDNA in non-metastatic stage NSCLC patients with a tumor size of ≥3 cm was higher, when compared to patients with a tumor size of <3 cm (p=0.052). B) In non-metastatic stage NSCLC, the concordance of HCSVs between plasma ctDNA and tDNA was much higher in patients who have a tumor size of ≥3 cm (26.1%), when compared to patients with a tumor size of <3 cm (4.0%). Lines indicate the median value and interquartile range.**

**Table 2. Summary of KRAS and actionable EGFR variants detection in tDNA and plasma ctDNA using the NGS-based test.**

| Patient ID | Stage | Mutations detected in tDNA (Variant Allele Frequency) | Mutations detected in plasma ctDNA (Variant Allele Frequency) |
|---|---|---|---|
| 2 | I | EGFR exon21: L858R (12.1%) | None |
| 3 | I | EGFR exon21: L858R (27.4%) | None |
| 4 | I | EGFR exon19 del (35.7%) | None |
| 8 | I | EGFR exon21: L858R (78.5%) | EGFR exon21: L858R (1.4%) |
| 12 | I | EGFR exon21: L858R (8.1%) | None |
| 15 | I | EGFR exon21: L858R (16.3%) | None |
| 19 | I | EGFR exon18: G719S (15.2%) | None |
| 33 | II | EGFR exon21: L858R (8.2%) | None |
| 34 | II | EGFR exon19 del (19.2%) | None |
| 32 | III | EGFR exon19 del (39.2%) | None |
| 30 | III | EGFR exon20 ins (48.5%) | None |
| 37 | III | EGFR exon18: G719A (21.5%)<br>EGFR exon20: T790M (21.9%) | None |
| 41 | III | EGFR exon21: L858R (18.4%) | None |
| 45 | IV | EGFR exon19 del (44.4%) | None |
| 48 | IV | EGFR exon19 del (7.8%) | EGFR exon19 del (8.7%) |
| 50 | IV | EGFR exon21: L858R (60.%) | EGFR exon21: L858R (5.1%) |
| 6 | I | KRAS exon2: G12C (14.4%) | None |
| 18 | I | KRAS exon2: G12C (18.3%) | None |
| 20 | II | KRAS exon2: G12A (37.5%) | KRAS exon2: G12A (5.4%) |
| 21 | II | KRAS exon2:.G12A (37.2%) | None |

allele frequency between patients with and without lymph node involvement (Figure 5A, p=0.76). Similarly, no significant difference was observed in the mutation detection between plasma ctDNA and tDNA in patient subtypes based on different clinical characteristics, including gender, age, smoking history and histopathological type.

## Discussion

Numerous studies have demonstrated the high concordance between mutation profiles obtained through ctDNA and tumor biopsy using the NGS test in late-stage NSCLC [24]. Similarly, the present results also revealed, with high confidence, the concordance (61.5%) among all variants between tumor tissues and matched plasma samples from the metastatic stage. Although tissue biopsy is the gold standard for molecular testing, there are several advantages of analyzing ctDNA over a tissue biopsy. First, the ctDNA NGS test makes the detection of actionable mutations possible when a tissue is hard to obtain or the quality of biopsy tissue is poor. Second, the biopsy is stressful for patients, and ctDNA is easily accessible during the entire disease course. In addition, the ctDNA test need less turnaround times, and is sometimes cost effective, considering the full cost of obtaining tissue samples, and the potentially extra costs involved due to biopsy complications [25–27].

However, there is very little evidence to establish the clinical utility of NGS-based ctDNA test for early stage cancer [22, 28]. Targeted sequencing using off-the-shelf panels, which does not require patient-specific optimization, appears to be a more cost-effective method to facilitate the clinical utility of liquid biopsy [3, 22, 29]. Thus, in the present study, a 1.15M cancer panel was used, which included 1 407 genes, to assess whether ctDNA analyses can be useful for identifying important mutations in patients with stage I–III NSCLC.

Based on earlier studies, the concordance rate at the patient level was defined as (patients with true positive + patients with true negative) / the number of patients [22, 28]. It was observed in a study conducted by Chen et al. that based on the screening of 50 genes, 31 of 76 patients with stage I–III NSCLC exhibited concordance mutations in both tDNA and ctDNA samples, while 21 patients had no mutations in both samples. The overall concordance rate was 68.4% [22]. Similarly, a study conducted by Xu et al. [28] revealed a concordant mutation in both tDNA and plasma ctDNA in 20 of 42 enrolled patients with stage III–IV NSCLC. Twelve samples did not have mutations in any of the 50 screened genes, thereby indicating that patient concordance for multiple genes in matched tDNA and ctDNA was 76.2% [28]. Interestingly, the present results are consistent with the results of a previous study on patients with stage II–IV NSCLC. Specifically, in the present study, 100% (7/7) of stage IV NSCLC patients harbored concordance mutations. However, among these stage II–III NSCLC patients, there were only seven of 23 harbor concordant mutations, while among these 20 stage I NSCLC patients, merely one patient had a concordant mutation.

Importantly, the concordance rate at the patient level was not sufficient to evaluate the efficiency of target sequencing. Thus, in the present study, a new concept of the mutation site concordance rate was defined to evaluate the concordance between mutations found in ctDNA and tDNA. As a result, the concordance of all variants with high confidence between tumor tissues and matched plasma samples were observed to be 61.5% at the metastatic stage. In contrast, the concordance of all variants with high confidence between tumor tissues and matched plasma samples was only 14.5% at the non-metastatic stage. In particular, the concordance rate at stage I, II and III was 2.2%, 31.3%, and 15.8%, respectively. This poor concordance rate at localized stage NSCLC, especially at stage I, can be attributed to the fact that the release of ctDNA is low when the tumor burden is low during early stage NSCLC. Thus, its detection is hard [11].

In addition, tDNA was used as a reference to calculate the sensitivity and specificity of detecting EGFR and

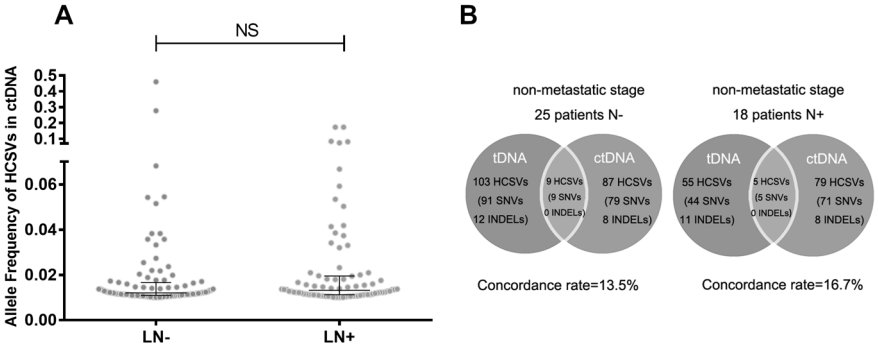


**Figure 5. A) No significant difference was observed in allele frequency of detected HCSVs in ctDNA between non-metastatic stage NSCLC patients with and without lymph node involvement. B) In non-metastatic stage NSCLC, the concordance rate of HCSVs between plasma ctDNA and tDNA in patients with lymph node involvement (16.7%) was similar to those without lymph node involvement (13.5%). NS, not significant. N-, without lymph node involvement. N+, with lymph node involvement. Lines indicate the median value and interquartile range.**

KRAS variants in plasma ctDNA through NGS-based deep sequencing. The present results revealed that the specificity of the plasma ctDNA test is high in both the metastatic and non-metastatic stages. However, the sensitivity of the plasma ctDNA test was lower in the non-metastatic stage (7.7% for EGFR) than in the metastatic stage (66.6% for EGFR).

As mentioned earlier, the present study used a 1.15M panel, which covered 1 407 cancer-related genes, for assessing the importance of ctDNA. This has its own advantage, since a larger scale panel with a larger scale of sequence would provide an additional opportunity to detect rare mutations in a limited amount of plasma. However, as a downside, there is always a risk of high false positives with big size panels. Thus, specific filters need to be applied to increase the specificity. In order to achieve the balance between specificity and sensitivity in the present panel, the investigators considered variants with a calling of more than 1% allele frequency as HCSVs. It has also been observed that ctDNA concentration levels were <0.5% in the majority of lung cancer patients in the early stage, and this number appeared to be a considerably lower than the detection thresholds of the sequencing methods used in the present study [24, 30, 31]. This would definitely pose a big challenge for the detection of low frequency mutations in early stage NSCLC. However, previous studies have also indicated that the concentration of ctDNA in plasma directly correlates with tumor size [32] and stage [24]. In this context, the study conducted by Bettegowda et al. [24] revealed a 100-fold increase in median ctDNA concentration in patients with stage IV disease, when compared to patients with stage I disease. Consistent with this observation, the present study also revealed that the allele frequency of HCSVs in ctDNA was much higher in patients with late stage NSCLC. Furthermore, the allele frequency of overlapped HCSVs was also much higher in patients with stage IV, when compared to patients with stage I–III. It is also possible that at low ctDNA concentrations in the early stages of cancer, merely some mutations could be detected, while others can be missed [11]. Another recent report compared ctDNA levels with tumor volume by imaging in patients with relapsed high-grade serous ovarian cancer, and it was established that ctDNA levels and disease volume were significantly correlated [33]. In this report, the mutant alleles observed in plasma were elevated in fraction by approximately 0.08% and in concentration by six mutant copies per milliliter of plasma for every cubic centimeter of disease [33]. In the present study, it was also observed that the allele frequency of the detected HCSVs was larger in patients with tumors ≥3 cm, when compared to patients with a tumor size of <3 cm. Similarly, another independent study conducted by Chen et al. [22] also confirmed that patients with tumors ≥3 cm had a significantly high ctDNA frequency. All these studies along with the present data demonstrate that liquid biopsy is a feasible option for patients with late stage lung cancer. However, more sensitive methods are required for patients with early stage cancers.

In order to understand the reasons of discordance in plasma ctDNA and tDNA data, intratumor heterogeneity can be an important factor. A recent study conducted by Jamal-Hanjani et al. [29], performed multi-region whole-exon sequencing on 100 early-stage NSCLC tumors resected before systemic therapy. The results demonstrated that 30% of SNVs were identified as subclonal. Another recent study conducted by Xie et al. [34] analyzed 35 pairs of matched primary tumor tissues, metastatic lymph nodes and plasma from treatment-naïve patients with advanced NSCLC, to interrogate heterogeneity and similarity among these three sites, using a panel of 56 NSCLC-related genes. Their results showed 62.0% by-variant concordance rate among primary tumors, metastatic lymph nodes and plasma, and 76.4% by-variant concordance rate between primary tumors and metastatic lymph nodes. It is important to note that since biopsy samples may not include all mutations found within the tumor, the analysis of an individual biopsy might not accurately reflect the genomic architecture of a patient's cancer and could introduce bias to the selection and efficacy of personalized medicine. In study conducted by Xu et al. [28], mutations identified in plasma ctDNA, but not in tDNA, were found in actionable genes, such as EGFR (p.G719X, p.T790M and 19 exon del) and BRAF (p.V600E). Thus, in this context, ctDNA may offer additional information, when compared to tDNA alone [28]. The combination of tDNA and ctDNA for molecular testing can provide a more thorough inspection of genetic variants.

In conclusion, the present study demonstrates that the concordance rate of variants in both ctDNA and tDNA generally trends towards a higher number in advanced stage NSCLC patients. This can be attributed to the increasing and higher ctDNA concentrations in the later stages of cancer patients. Finally, it appears that ctDNA NGS-based deep sequencing is a feasible approach for identifying mutations in patients with late stage (IV) NSCLC patients. However, additional methods with higher sensitivity and specificity would definitely be beneficial to improve the successful application of this platform in the earlier stages of NSCLC.

Acknowledgements: The present study was funded by the Human Provincial Science and Technology Department (2017JJ2345) and the National Natural Science Foundation of China (No. 81673516).

## References

[1] SIEGEL RL, MILLER KD, JEMAL A. Cancer Statistics, 2017. CA Cancer J Clin 2017; 67: 7–30. https://doi.org/10.3322/caac.21387

[2] CHEN Y, PENG X, ZHOU Y, XIA K, ZHUANG W. Comparing the benefits of chemoradiotherapy and chemotherapy for resectable stage III A/N2 non-small cell lung cancer: a meta-analysis. World J Surg Oncol 2018; 16: 8. https://doi.org/10.1186/s12957-018-1313-x

[3] ABBOSH C, BIRKBAK NJ, WILSON GA, JAMAL-HANJANI M, CONSTANTIN T et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature 2017; 545: 446–451. https://doi.org/10.1038/nature22364

[4] IZUMCHENKO E, CHANG X, BRAIT M, FERTIG E, KAGOHARA LT et al. Targeted sequencing reveals clonal genetic changes in the progression of early lung neoplasms and paired circulating DNA. Nat Commun 2015; 6: 8258. https://doi.org/10.1038/ncomms9258

[5] RIMKUNAS VM, CROSBY KE, LI D, HU Y, KELLY ME et al. Analysis of receptor tyrosine kinase ROS1-positive tumors in non-small cell lung cancer: identification of a FIG-ROS1 fusion. Clin Cancer Res 2012; 18: 4449–4457. https://doi.org/10.1158/1078-0432.CCR-11-3351

[6] FORSHEW T, MURTAZA M, PARKINSON C, GALE D, TSUI DW et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med 2012; 4: 136ra68. https://doi.org/10.1126/scitranslmed.3003726

[7] GEVENSLEBEN H, GARCIA-MURILLAS I, GRAESER MK, SCHIAVON G, OSIN P et al. Noninvasive detection of HER2 amplification with plasma DNA digital PCR. Clin Cancer Res 2013; 19: 3276–3284. https://doi.org/10.1158/1078-0432.CCR-12-3768

[8] MURTAZA M, DAWSON SJ, TSUI DW, GALE D, FORSHEW T et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature 2013; 497: 108–112. https://doi.org/10.1038/nature12065

[9] PHALLEN J, SAUSEN M, ADLEFF V, LEAL A, HRUBAN C et al. Direct detection of early-stage cancers using circulating tumor DNA. Sci Transl Med 2017; 9. https://doi.org/10.1126/scitranslmed.aan2415

[10] REINERT T, SCHOLER LV, THOMSEN R, TOBIASEN H, VANG S et al. Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery. Gut 2016; 65: 625–634. https://doi.org/10.1136/gutjnl-2014-308859

[11] WAN JCM, MASSIE C, GARCIA-CORBACHO J, MOULIERE F, BRENTON JD et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nat Rev Cancer 2017; 17: 223–238. https://doi.org/10.1038/nrc.2017.7

[12] THIERRY AR, MOULIERE F, EL MESSAOUDI S, MOLLEVI C, LOPEZ-CRAPEZ E et al. Clinical validation of the detection of KRAS and BRAF mutations from circulating tumor DNA. Nat Med 2014; 20: 430–435. https://doi.org/10.1038/nm.3511

[13] ZONTA E, GARLAN F, PECUCHET N, PEREZ-TORALLA K, CAEN O et al. Multiplex Detection of Rare Mutations by Picoliter Droplet Based Digital PCR: Sensitivity and Specificity Considerations. PLoS One 2016; 11: e0159094. https://doi.org/10.1371/journal.pone.0159094

[14] SACHER AG, PAWELETZ C, DAHLBERG SE, ALDEN RS, O'CONNELL A et al. Prospective Validation of Rapid Plasma Genotyping for the Detection of EGFR and KRAS Mutations in Advanced Lung Cancer. JAMA Oncol 2016; 2: 1014–1022. https://doi.org/10.1001/jamaoncol.2016.0173

[15] SCHIAVON G, HREBIEN S, GARCIA-MURILLAS I, CUTTS RJ, PEARSON A et al. Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer. Sci Transl Med 2015; 7: 313ra182. https://doi.org/10.1126/scitranslmed.aac7551

[16] YUNG TK, CHAN KC, MOK TS, TONG J, TO KF et al. Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non-small cell lung cancer patients. Clin Cancer Res 2009; 15: 2076–2084. https://doi.org/10.1158/1078-0432.CCR-08-2622

[17] NEWMAN AM, BRATMAN SV, TO J, WYNNE JF, ECLOV NC et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med 2014; 20: 548–554. https://doi.org/10.1038/nm.3519

[18] LEBOFSKY R, DECRAENE C, BERNARD V, KAMAL M, BLIN A et al. Circulating tumor DNA as a non-invasive substitute to metastasis biopsy for tumor genotyping and personalized medicine in a prospective trial across all tumor types. Mol Oncol 2015; 9: 783–790. https://doi.org/10.1016/j.molonc.2014.12.003

[19] BEIJE N, HELMIJR JC, WEERTS MJA, BEAUFORT CM, WIGGIN M et al. Somatic mutation detection using various targeted detection assays in paired samples of circulating tumor DNA, primary tumor and metastases from patients undergoing resection of colorectal liver metastases. Mol Oncol 2016; 10: 1575–1584. https://doi.org/10.1016/j.molonc.2016.10.001

[20] WYATT AW, ANNALA M, AGGARWAL R, BEJA K, FENG F et al. Concordance of Circulating Tumor DNA and Matched Metastatic Tissue Biopsy in Prostate Cancer. J Natl Cancer Inst 2017; 109. https://doi.org/10.1093/jnci/djx118

[21] YANG M, TOPALOGLU U, PETTY WJ, PAGNI M, FOLEY KL et al. Circulating mutational portrait of cancer: manifestation of aggressive clonal events in both early and late stages. J Hematol Oncol 2017; 10: 100. https://doi.org/10.1186/s13045-017-0468-1

[22] CHEN K, ZHANG J, GUAN T, YANG F, LOU F et al. Comparison of plasma to tissue DNA mutations in surgical patients with non-small cell lung cancer. J Thorac Cardiovasc Surg 2017; 154: 1123–1131.e2. https://doi.org/10.1016/j.jtcvs.2017.04.073

[23] SAUSEN M, PHALLEN J, ADLEFF V, JONES S, LEARY RJ et al. Clinical implications of genomic alterations in the tumour and circulation of pancreatic cancer patients. Nat Commun 2015; 6: 7686. https://doi.org/10.1038/ncomms8686

[24] BETTEGOWDA C, SAUSEN M, LEARY RJ, KINDE I, WANG Y et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med 2014; 6: 224ra24. https://doi.org/10.1126/scitranslmed.3007094

[25] ZHOU C, WU YL, CHEN G, FENG J, LIU XQ et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. Lancet Oncol 2011; 12: 735-742. https://doi.org/10.1016/S1470-2045(11)70184-X

[26] WU YL, ZHOU C, HU CP, FENG J, LU S et al. Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): an open-label, randomised phase 3 trial. Lancet Oncol 2014; 15: 213–222. https://doi.org/10.1016/S1470-2045(13)70604-1

[27] SHI Y, ZHANG L, LIU X, ZHOU C, ZHANG L et al. Icotinib versus gefitinib in previously treated advanced non-small-cell lung cancer (ICOGEN): a randomised, double-blind phase 3 non-inferiority trial. Lancet Oncol 2013; 14: 953–961. https://doi.org/10.1016/S1470-2045(13)70355-3

[28] XU S, LOU F, WU Y, SUN DQ, ZHANG JB et al. Circulating tumor DNA identified by targeted sequencing in advanced-stage non-small cell lung cancer patients. Cancer Lett 2016; 370: 324–331. https://doi.org/10.1016/j.canlet.2015.11.005

[29] JAMAL-HANJANI M, WILSON GA, MCGRANAHAN N, BIRKBAK NJ, WATKINS TBK et al. Tracking the Evolution of Non-Small-Cell Lung Cancer. N Engl J Med 2017; 376: 2109–2121. https://doi.org/10.1056/NEJMoa1616288

[30] TANIGUCHI K, UCHIDA J, NISHINO K, KUMAGAI T, OKUYAMA T et al. Quantitative detection of EGFR mutations in circulating tumor DNA derived from lung adenocarcinomas. Clin Cancer Res 2011; 17: 7808–7815. https://doi.org/10.1158/1078-0432.CCR-11-1712

[31] CROWLEY E, DI NICOLANTONIO F, LOUPAKIS F, BARDELLI A. Liquid biopsy: monitoring cancer-genetics in the blood. Nat Rev Clin Oncol 2013; 10: 472–484. https://doi.org/10.1038/nrclinonc.2013.110

[32] THIERRY AR, MOULIERE F, GONGORA C, OLLIER J, ROBERT B et al. Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. Nucleic Acids Res 2010; 38: 6159–6175. https://doi.org/10.1093/nar/gkq421

[33] PARKINSON CA, GALE D, PISKORZ AM, BIGGS H, HODGKIN C et al. 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 2016; 13: e1002198. https://doi.org/10.1371/journal.pmed.1002198

[34] XIE F, ZHANG Y, MAO X, ZHENG X, HAN-ZHANG H et al. Comparison of genetic profiles among primary lung tumor, metastatic lymph nodes and circulating tumor DNA in treatment-naive advanced non-squamous non-small cell lung cancer patients. Lung Cancer 2018; 121: 54–60. https://doi.org/10.1016/j.lungcan.2018.05.002