Abstract. Personalized medicine is revolutionizing the diagnosis and treatment of cancer; however, for personalized medicine to be used accurately, patient information is essential to determine the appropriate diagnosis, prognosis and treatment. The detection of genomic mutations in liquid biopsy samples is a non-invasive method of characterizing the genotype of a tumor. However, next generation sequencing-based plasma genotyping only has a sensitivity of ~70%. Identifying potential indicators that may reflect the sensitivity of a liquid biopsy analysis could offer important information for its clinical application. In the present study, 47 pairs of patient-matched plasma and tumor tissue samples obtained from patients with advanced lung cancer were sequenced using a panel of 56 cancer-associated genes. The plasma maximum allele frequency (Max AF) was identified as a novel biomarker to indicate the sensitivity of plasma genotyping. Using the identified somatic mutations in patient tissue biopsy samples as a reference, the sensitivity of the corresponding patient plasma test was investigated. The by-variant sensitivity of the plasma test was 68.1%, with 79 matched and 37 missed genetic aberrances. The by-patient sensitivity was calculated as 83%. Patients with a high plasma Max AF value (>2.2%) demonstrated a higher concordance with the range of mutations identified in the patient-matched tissue samples. The Max AF observed in patient plasma samples was positively correlated with liquid biopsy sensitivity and could be used as a potential indicator of liquid biopsy sensitivity. Therefore, patients with a low plasma Max AF (≤2.2%) may need to undergo further tissue biopsy to allow personalized oncology treatment. In summary, the present study may offer a non-invasive testing method for a sub-group of patients with advanced lung cancer.

Introduction

As the development of sequencing technologies advances, personalized medicine has become increasingly available, particularly for the treatment of cancer (1,2). This offers customized treatment for patients based upon the molecular analysis of tumor-associated biomarkers (3). The identification of driver genes and genomic aberrations, including KRAS proto-oncogene GTPase (KRAS) and epidermal growth factor receptor (EGFR) (4-8), as oncogenes for lung cancer has driven personalized medicine. Based on the number and specific somatic mutations detected in a patient, one could be diagnosed more precisely and treated with targeted drugs (9).

Tumor tissue biopsies offer the ideal sample for molecular analysis to target personalized cancer medicine (10). However, limitations exist due to the level of invasiveness involved in obtaining a biopsy for certain tumor types, the limited feasibility of this technique and tumor heterogeneity, which results in a number of potential false negatives (11). Circulating tumor DNA (ctDNA), which is comprised of small fragments of DNA, has been identified in the blood of patients. Typically, tumor cells undergoing apoptosis or necrosis release ctDNA into the circulatory system (12-14). Indeed, previous studies have detected mutations in driver genes via examination of liquid biopsies (15,16). A previous study of patients with non-small cell lung cancer (NSCLC), which analyzed blood samples for genetic alterations in EGFR, demonstrated...
adequate sensitivity and specificity for the approval of liquid biopsy testing in the clinic (17).

Previous studies have reported that the average concordance rate between tissue and plasma genotyping is ~70% (range, 48-98%) (18-23). A clinical trial used Therascreen EGFR detection kit, approved for diagnostic use in Europe, to detect somatic mutations in tissue and liquid biopsies, demonstrated a sensitivity of 65.7% for plasma genotyping and a concordance of 94.3% among plasma and tumor tissue samples (24). However, implementation of plasma genotyping within the clinic is currently hindered by the number of inconsistencies observed in results compared with tumor biopsies, which is thought to be due to tumor heterogeneity (25). If this is the case, theoretically, using tumor biopsies as a reference, more mutations are likely to be identified in ctDNA compared with tissue DNA (11,26). However, in practice, plasma tests often fail to detect numerous mutations, including genomic aberrations in the driver genes (27). Consequently, within clinical practice, despite the ability to sensitively detect EGFR mutations within ctDNA samples, tissue biopsy is recommended for all patients with negative results regarding the T790M mutation in the gene, in order to eliminate the potential for false negative results from plasma genotyping (26). Furthermore, it is unclear which patient sub-populations may be more likely to carry specific mutations. This means that currently all patients need to undergo costly and invasive procedures to determine diagnosis, prognosis and optimal treatment strategies.

In the present study, targeted sequencing was performed using 47 pairs of blood and tumor tissue biopsies from patients with advanced lung cancer. By comparing the paired genetic profiles of tumors and plasma samples, a parameter was derived to indicate the sensitivity of the plasma test, termed the plasma maximum allelic fraction (Max AF). Max AF is defined as the maximum mutant allele fraction in a plasma sample. Patients with a lower plasma Max AF (≤2.2%) were more likely to carry tissue-specific mutations that are not identifiable through plasma genotyping; Therefore, tumor tissue biopsy may be necessary for these patients for a complete knowledge of their tumor genotype. Such a parameter may prove useful for the accurate diagnosis and personalized medical treatment of patients with lung cancer.

Materials and methods

Patient and sample collection. Patients ≥18-years-old diagnosed with any type of lung cancer at the General Hospital of Southern Theater Command in Guangzhou (China) between January 2015 and December 2016 were eligible for the present study. The inclusion criteria for the study were as follows: i) Sufficient tissue and plasma samples were available for targeted sequencing; and ii) intervals between the tissue biopsy and blood sample collection were ≤14 days. Blood (~10 ml) was obtained and stored in an EDTA-coated tube at 4°C until DNA extraction. In total, 47 patients with advanced lung cancer were enrolled in the current study and paired biopsy and blood samples were obtained from each patient. Patients with no genetic mutations observed in tumor tissues were excluded from further analysis. The maximum allelic fraction (Max AF) was defined as the highest mutant allele fraction detected in a particular sample, regardless of gene or mutation site. The present study was approved by the Ethical Committee at the General Hospital of Southern Theater Command, PLA (Guangzhou, China). All patients provided written informed consent for their participation in the study.

DNA extraction for tissue and plasma samples. Tissue DNA was extracted using the commercially available QIAamp DNA formalin-fixed paraffin-embedded tissue kit (Qiagen, Inc.), according to the manufacturer's protocol. Similarly, blood DNA was extracted using a QIAamp Circulating Nucleic Acid kit (Qiagen, Inc.), as previously described (28). Briefly, plasma were removed from whole blood samples and transferred to separate tubes prior to centrifugation at 16,000 x g for 10 min at 4°C to remove debris. Circulating DNA was extracted from the plasma, according to the manufacturer's protocol. Quantification of both tissue DNA and plasma DNA were performed using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Inc.).

Next generation sequencing (NGS) library preparation and sequencing. The NGS library was prepared as described previously (29). Briefly, nucleotide fragments of 200-400 base pairs were selected using Agencourt AMPure beads (Beckman Coulter, Inc.). The quality of the DNA fragments was evaluated using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Inc.), following hybridization and amplification. Paired samples of tissue and plasma samples were sequenced using a capture-based targeted sequencing panel (Burning Rock Biotech). A sequencing panel consisting of 56 lung cancer-associated genes was used to detect and quantify genomic aberrations.

Statistical analysis. Statistical analysis was performed using SPSS 20.0 software (IBM Corp., Armonk, NY, USA). Analysis was performed with unpaired t-test. Box and whisker plots were generated to present the median values and 95% confidence intervals. Correlation between plasma Max AF and the number of tissue mutations was analyzed by Spearman's correlation. Correlation between the tissue Max AF and the plasma Max AF was analyzed using a linear regression model. Receiver operating characteristic (ROC) analysis was performed to quantify the extent to which Max AF can discriminate patients based on their likelihood to harbor tissue-specific mutations. P<0.05 was considered to indicate a statistically significant difference.

Results

Patient characterization. Patients diagnosed with advanced lung cancer at the General Hospital of Guangzhou Military Command of PLA (Guangzhou, China) between 2015 and 2016 were enrolled in the present study (n=47) with a median age of 59. All 47 patients underwent a biopsy examination and blood test. Both a tumor tissue sample and plasma sample were collected from each patient. Of the 29 patients with known sex information, 13 were male and 16 were female. A total of 45 patients had been diagnosed with NSCLC and two with small cell lung cancer. All patients had been diagnosed with stage IV disease.

Genomic mutations in tissue and plasma samples. Genomic alterations detected in tumor tissues were considered a
A sequencing panel consisting of 56 lung cancer-associated genes was used to detect and quantify genomic aberrations. The panel included driver genes and genes to which targeted therapies exist, both in development and in the clinic. Sequencing was performed on all 47 tumor biopsy samples and their patient-matched plasma samples. Of the 47 tissue biopsy samples analyzed, 40 patients (87%) presented with mutations within the sequencing panel used in the present study. Collectively, 116 variants were identified, spanning 33 genes (Fig. 1A). EGFR was the most frequently mutated gene, with mutations observed in 57% of patients, followed by tumor protein p53 (TP53) and RB transcriptional corepressor 1, with mutations in 56 and 13% of patients, respectively. In addition to carrying EGFR as a driver mutation, two patients presented with ALK receptor tyrosine kinase (ALK) rearrangements, two presented with ERBB2 receptor tyrosine kinase 2 mutations and one with a MET proto-oncogene receptor tyrosine kinase (MET) mutation. One patient with an EGFR mutation also tested positive for a KRAS mutation. In addition, 10 patients carried the EGFR T790M mutation (Fig. 1A).

Subsequently, the spectrum of mutations across the tissue and plasma samples was compared. Collectively, 106 genomic alterations spanning 37 genes within the plasma samples were identified. Using the mutations detected within the patient-matched tumor samples as a reference, a by-variant association of 68.1% was achieved between plasma and tumor samples, with 79 matched, 37 missed and 27 new genomic aberrations that were only present in plasma samples (Fig. 1B), while the by-patient sensitivity was calculated as 83%. Among the 37 mutations uniquely detected within plasma samples, a number of them were classic NSCLC drivers, including EGFR (L858R, 19 del and T790M) and ALK rearrangements. Additionally, four patients carried the EGFR T790M mutation that was only detectable in the tissue biopsy samples, whilst EGFR L858R appeared to be tissue-specific for three patients. The tissue-specific mutations of 19 del and a rearrangement in ALK were also observed only in a single patient for each mutation. Notably, the present data revealed 27 mutations that were only present in plasma samples. This could potentially be due to tumor heterogeneity. The majority of the plasma-specific mutations were either in driver genes or TP53 mutations, including eight EGFR mutations (including five EGFR T790M), one ALK, one KRAS, one MET and eight TP53 mutations. The Venn diagram presented in Fig. 1C summarizes all of the genomic alterations detected across all analyzed samples. In summary, the present data demonstrate a concordance between tissue and plasma samples.

Patients with tissue-specific mutations exhibit a low plasma Max AF. The plasma Max AF of patients with and without tissue-specific mutations was compared to determine whether there was a difference in these patient sub-groups. The plasma Max AF of patients without tissue-specific mutations were observed to be significantly higher compared with patients with tissue specific mutations (P=0.003; Fig. 2A). A receiver

Figure 1. Genomic mutations detected in tumor biopsy tissues and plasma samples. (A) Somatic mutations in 33 tumor-associated genes were detected in biopsy samples. Each column represents one patient. The total number of mutations detected in each patient are presented at the top of the graph. The occurrence rate of each genomic mutation is presented on the left of the graph. Different colors indicate different types of alterations. (B) A comparison of mutations between biopsy tissues and plasma samples. Different colors represent matched, missed or new mutations detected in plasma DNA as compared with tissue samples. Each column represents one patient. (C) Bivariant comparison of somatic alterations detected in biopsy samples and plasma samples. Pla, plasma; Tis, tissue.
operating characteristic analysis was subsequently performed to derive plasma Max AF as a percentage that could be used to differentiate patients on the likelihood of them having tissue-specific mutations. The analysis revealed that patients with a plasma Max AF ≤2.2% were more likely to have tissue-specific mutations, achieving an area under curve of 78% (specificity, 89.5%; sensitivity, 61.9%; Fig. 2B). Utilizing 2.2% as a cutoff, 24 patients were classified as having a high plasma Max AF (>2.2%) and the remaining 16 patients were classified as having a low Max AF (≤2.2%). Among the 24 patients with a high Max AF, 16 possessed no tissue-specific mutations, while 8 exhibited tissue-specific mutations. Among the 16 patients with a low Max AF, 13 possessed tissue-specific mutations and 3 demonstrated no tissue-specific mutations (Fig. 2C). Further analysis revealed that the detection rate of tissue-specific mutations was 81.3% (13/16) in patients with a plasma Max AF <2.2%, while the rate was 33.3% (8/24) in patients with a Max AF ≥2.2 (Fig. 2C). In summary, these results demonstrate that the tumor mutation profile is more likely to be comprehensively reflected by plasma in patients with a high plasma Max AF (>2.2%). In such patients, liquid biopsy may have the potential to replace tissue biopsy.

To further confirm the current finding, the number of tissue-specific mutations was plotted against plasma Max AF for each patient. Patients with more than one tissue-specific mutation consistently demonstrated a lower plasma Max AF (P=0.0003; Fig. 3A). The correlation between the tissue Max AF and the plasma Max AF was also analyzed for each patient using a linear regression model, which revealed a significant positive correlation between these factors (r²=0.285; P<0.001; Fig. 3B). Collectively, these data confirm an inverse association between plasma Max AF and the likelihood of harboring tissue-specific mutations. Furthermore, a positive correlation between tissue Max AF and plasma Max AF was noted in this patient population.

Discussion

The present study derived a parameter, termed Max AF, to differentiate patients with lung cancer with tissue-specific mutations from those without tissue-specific mutations. The sensitivity of plasma genotyping, which is reported to be ~70% accurate compared with tissue biopsy analysis (18-23,25), remains a challenge for the use of this method in clinical settings. The current study revealed an inverse correlation between a patient's plasma Max AF and the likelihood of that patient possessing tissue-specific mutations. Furthermore, a plasma Max AF of 2.2% was derived as a binary classifier to differentiate between patient with and without tissue-specific mutations. Plasma genotyping results of patients with plasma Max AF >2.2% were more likely to reflect the mutation profile in tumor tissue samples.

One possible explanation for missing mutations not detected plasma genotyping is the low levels of ctDNA (30). The ctDNA concentration is significantly diluted by DNA released from normal cells. Another possibility is tumor heterogeneity, where very few clones with certain mutations exist, resulting in the low levels of this genomic aberration present in the ctDNA sample (31-33). In any case, a high sequencing depth is required for the detection of mutations in
plasma samples. In the present study, the average sequencing depths were ~1,000X for tissue samples and ~10,000X for the ctDNA of plasma. However, the sensitivity of the plasma test is still at an unsatisfactory level for accurate prognostic and treatment use within the clinic. A recent study reported that although the detection sensitivity of EGFR mutations in plasma samples was comparable to that of tumor tissue, the detection rate of KRAS mutations was much lower, and this may be due to the limited number of tumor sub-clones in plasma samples (34). The low concordance between tissue and plasma for the additional coexisting mutations in EGFR-mutant NSCLCs is clinically relevant. It has been demonstrated that patients with an EGFR mutant that carry additional mutations in other genes, such as KRAS, are more likely to exhibit worse outcomes when treated with EGFR tyrosine kinase inhibitor drugs (35-37). Consequently, it is important to identify new methods of analysis for plasma samples that reduce the incidence of false negative results.

The present study derived plasma Max AF as a biomarker to reflect plasma test sensitivity. The majority of patients (81%) with a plasma Max AF <2.2% presented with tissue-specific mutations, suggesting that they may benefit from additional tissue biopsy due to a limited amount of ctDNA present in plasma samples. Abundant ctDNA is the key characteristic that determines high sensitivity in liquid biopsy examination (27,38). The concentration of ctDNA can be assessed through measurement of the cell free DNA concentration; however, this further increases the cost of molecular analysis (38). Using the cutoff value of Max AF, patients that are more likely to have false negative results in plasma tests can be easily identified. Therefore, a parameter indicating liquid biopsy sensitivity could contribute to improved clinical practice in the future.

The newly identified mutations in plasma samples not present in tumor samples are largely due to tumor heterogeneity (39,40). In an individual tumor, cells from different regions could have different genetic characteristics largely due to the presence of sub-clones (41). A single biopsy of tumor tissue cannot represent the comprehensive features of the tumor (42). As such, it is possible that more mutations are identified in plasma-derived ctDNA compared with alterations in tissue DNA. As it is difficult to obtain a complete analysis of the molecular features of a tumor through a single biopsy test, it can be recommended that a more comprehensive examination of tissue biopsy samples could potentially extend these findings in the future.

Acknowledgements
Not applicable.

Funding
This study was supported by the Science and Technology Program of Guangzhou, China (grant no. 201607010117).

Availability of data and materials
The data and materials are available upon reasonable request from the corresponding author.

Authors' contributions
YT, XL and GQ worked on the conception and design of the study. YT, XL, ZO, ZH, QZ, YW and MY collected the data. JY assisted with the statistical analysis. YT, XL, ZO, ZH, QZ,
YW, MY, JY, HHZ and GQ analyzed and interpreted the data. YT, XL and HHZ wrote the manuscript in consultation with GQ. All the authors approved the manuscript and are accountable for the content of the manuscript.

Ethics approval and consent to participate
The present study was approved by the Ethical Committee at the General Hospital of Southern Theater Command of PLA (Guangzhou, China). All patients provided written informed consent for their participation in the study.

Patient consent for publication
Not applicable.

Competing interests
All the authors declare no conflict of interest.

References