



OPEN ACCESS

ORIGINAL ARTICLE

Resectable lung lesions malignancy assessment and cancer detection by ultra-deep sequencing of targeted gene mutations in plasma cell-free DNA

Muyun Peng,¹ Yuancai Xie,² Xiaohua Li,³ Youhui Qian,⁴ Xiaonian Tu,³ Xumei Yao,³ Fangsheng Cheng,³ Feiyue Xu,³ Deju Kong,³ Bing He,⁵ Chaoyu Liu,³ Fengjun Cao,⁶ Haoxian Yang,⁷ Fenglei Yu,¹ Chuanbo Xu,³ Geng Tian⁸

For numbered affiliations see end of article.

Correspondence to

Dr Fenglei Yu, Department of Thoracic Surgery, Second Xiangya Hospital, Changsha, China; yufenglei@csu.edu.cn, Dr Chuanbo Xu, Vienomics Corp. Ltd, Shenzhen 95129, China; xcb@vienomics.com and Dr Geng Tian, Department of Medical Oncology, Shenzhen Second People's Hospital, The First Affiliated Hospital of Shenzhen University, Shenzhen, China; tiangeng666@aliyun.com

MP, YX and XL contributed equally.

Received 23 October 2018
Revised 12 February 2019
Accepted 5 March 2019
Published Online First 13 April 2019

ABSTRACT

Background Early detection of lung cancer to allow curative treatment remains challenging. Cell-free circulating tumour (ct) DNA (ctDNA) analysis may aid in malignancy assessment and early cancer diagnosis of lung nodules found in screening imagery.

Methods The multicentre clinical study enrolled 192 patients with operable occupying lung diseases. Plasma ctDNA, white cell count genomic DNA (gDNA) and tumour tissue gDNA of each patient were analysed by ultra-deep sequencing to an average of 35 000× of the coding regions of 65 lung cancer-related genes.

Results The cohort consists of a quarter of benign lung diseases and three quarters of cancer patients with all histopathology subtypes. 64% of the cancer patients are at stage I. Gene mutations detection in tissue gDNA and plasma ctDNA results in a sensitivity of 91% and specificity of 88%. When ctDNA assay was used as the test, the sensitivity was 69% and specificity 96%. As for the lung cancer patients, the assay detected 63%, 83%, 94% and 100%, for stages I, II, III and IV, respectively. In a linear discriminant analysis, combination of ctDNA, patient age and a panel of serum biomarkers boosted the overall sensitivity to 80% at a specificity of 99%. 29 out of the 65 genes harboured mutations in the patients with lung cancer with the largest number found in *TP53* (30% plasma and 62% tumour tissue samples) and *EGFR* (20% and 40%, respectively).

Conclusion Plasma ctDNA was analysed in lung nodule assessment and early cancer detection, while an algorithm combining clinical information enhanced the test performance.

Trial registration number NCT03081741.

INTRODUCTION

Lung cancer is the leading cause of cancer-related deaths, accounting for an estimated 1.6 million deaths each year globally.¹ The prognosis of lung cancer is dependent on the stage of diagnosis, with 5-year overall survival rate decreasing dramatically from stage IA (85%) to stage IV disease (6%).^{2,3} It is clear that to screen and diagnose lung cancer earlier will save lives.³

Current method for lung cancer screening is low-dose computed tomography (LDCT).⁴ The National Lung Screening Trial demonstrated a 20% reduction in lung cancer mortality for LDCT

compared with X-ray and a 6.7% all-cause mortality reduction.⁵ However, the imagery technique often results in indeterminate nodules⁶ and the false positive results lead to unnecessary invasive diagnostic procedures and increased deaths from avoidable surgeries.⁷

Traditionally, biopsy is used to determine malignancy of lung nodules. This approach has significant limitations as being difficult or even impossible. Molecular tests using pervasive biofluid samples, so-called liquid biopsy, are promising and urgently needed in the thoracic clinic, as demonstrated in thyroid nodules assessment where a number of molecular tests are available.⁸ In pulmonary nodule malignancy assessment, there is also the report of blood proteomics biomarkers by the PANOPTIC team.⁹

In theory, circulating tumour (ct) DNA (ctDNA) is exquisitely specific for an individual's tumour as by definition somatic mutations are identified by their presence in tumour DNA and absence in matched normal DNA. This bypasses the issues related to the false-positivity encountered with other biomarkers, such as protein biomarkers. This specific and promising early detection method has garnered tremendous attention for cancer in general and lung in particular.^{10,11} The growing interest has drawn attention from international societies such as the International Association of the Study of Lung Cancer (IASLC) that has issued statement regarding liquid biopsy in the management of non-small cell lung cancer (NSCLC).¹²

Early detection of cancer by blood test was shown possible even with microscopic tumour proceeding radiography.¹³ Most recently, a preliminary analysis of early to mid-stage (stages I–III) patients with lung cancer as part of the Circulating Cancer Genome Atlas (CCGA) pan-cancer study was reported.¹⁴

In addition to mutation analysis, ctDNA epigenetics has also been evaluated and researched for early cancer detection. The studies examined DNA methylation regulation¹⁵ or the hypermethylation of the promoter regions of genes¹⁶ for the potential biomarkers of lung cancer detection.

In this study, we report a multicentre clinical trial result on genetic alterations in patients that undergo surgical resection with either benign nodules or early to midstages lung cancer. By using ultra-deep next generation sequencing (NGS) sequencing to



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Peng M, Xie Y, Li X, et al. *J Med Genet* 2019;**56**:647–653.

detect very low-frequency ctDNA in a background of mostly non-tumour-derived cfDNA and comparing with the tumour tissue, we develop an assay to distinguish benign versus malignant lung lesions and detect the early cancer pathogenesis.

METHODS

Patients and clinical data

The study design of the clinical trial (ClinicalTrials.gov) is prospective, cross-sectional, longitudinal and observational. It is conducted at four tier A hospitals in China to recruit patients with lung occupying diseases identified by imaging evaluation to be treated by surgery. Briefly, consecutive patients with benign lung nodules or cancers of stages I–III planned for surgical resection were assessed for eligibility, and those met the criteria (online supplementary table S1) were approached by the study team for consenting. The ones that signed the informed consent form were enrolled, and biological samples (two tubes of 10 mL peripheral blood prior to surgery collected in a cell-free DNA [cfDNA] BCT blood collection tube [Streck, Omaha, Nebraska, USA] and 10 slides of formalin-fixed paraffin-embedded (FFPE) tissues from surgery) and related clinical data including serum biomarkers were collected.

Basic demographic and clinical data were collected using the IRB-approved clinical protocol and case report form. The initial discovery of the space occupying lesions was done through thoracic CT scan, and the read out was provided by the radiologists at the clinical site of the hospitals. The final diagnosis of the malignancy status of the nodules and cancer TNM staging was established by the pathologists of the hospital using the resected tissues.

Sequencing analysis

The above-collected DNA samples were analysed by our proprietary Sec-Seq technique as described previously.¹⁷ Briefly, blood sample were processed to separate the plasma from blood cells by centrifugation. cfDNA, gDNA of white cell count (WCC) and FFPE were extracted using QIAamp Circulating Nucleic Acid Kit, DNA mini kit, DNA FFPE Tissue kit, respectively (Qiagen, Hilden, Germany). The concentration of extracted DNA was measured using Qubit 3.0 dsDNA high-sensitivity assay (Life Technologies, Carlsbad, California, USA).

Capture probes were designed for 65 cancer-associated genes covering 241 kb genomic regions (online supplementary table S2) and synthesised by IDT (Michigan, USA). Indexed libraries were constructed using KAPA HyperPlus Kit (KK8514). Barcoding was employed to reduce noise. Postcapture multiplexed libraries were amplified with Illumina backbone primers for 16 cycles of PCR using 1 × KAPA HiFi Hot Start Ready Mix and sequenced on Illumina NovaSeq platform (Illumina, California, USA) at Novogen (Nanjing, China).

Horizon's Partners Spike-in control (Horizon, Cambridge, UK) was used in a serial dilution (from 0.0005 to 1) using the wild-type reference genome and the provided reference standard. Reference variants include *EGFR* (L858R and T790M), *KRAS* (G12D), *NRAS* (Q61K and A59T) and *PIK3CA* (E545K) or NA12878 (12 mutations) and NA24385 (29 mutations).

Bioinformatics data analysis

Reads with quality score <30 or having >5% of positions differ from the rest of the reads targeting the same region were removed. The results were then mapped to the human reference genome (hg19) using BWA (V.0.7.15-r1140). We used the start mapping positions, the length and the dual barcode on both side

of merged paired-end fragments to form reads groups amplified from every primary cfDNA molecules and to identify incorrect base produced due to PCR errors.

Variant calling for single nucleotide variation or insertion/deletion was performed using samtools mpileup tool (V.1.3.1). For ctDNA samples, a variant was selected as a candidate somatic mutation when: (1) two distinct paired reads (each redundantly sequenced at least three times) contained the mutation, (2) effective reads depth >500 (captured primary cfDNA molecules >500) and (3) the corresponding allele frequency in WCC is less than 1%.

Mutation annotation and classification

The variants were called by SnpEff (V.4.3o) and annotated by COSMIC (V.85), ExAc, ClinVar and 1000 Genome. The following variants were eliminated: (1) intergenomic or intronic (except for splicing junction); (2) synonymous; and (3) variant allele frequency (VAF) <0.2% in ctDNA or <1% in FFPE samples. Previously reported and confirmed pathogenic mutations in the clinical samples of lung cancer of all human races and ethnicities will be considered as lung cancer related.

Statistical methods

Data were summarised using descriptive statistics. Fisher's exact test was used to compare any two subgroups. Wilcoxon rank-sum test was used to compare median age between any two subgroups (stages I, II and III) or mutant groups (mutant vs wild type).

Linear discriminant analysis (LDA) was performed on improving mutation analysis. The model considers the age of the patients, ctDNA mutations and the serum biomarkers. It was developed using the 10-fold cross-validation by dividing the samples into training and validation subsets. The test sensitivity and specificity was calculated, and the area under the receiver operator curve (AUROC) was plotted.

RESULTS

Patient demographic and clinical analysis

In total, 192 patients with pulmonary space occupying lesions (136 malignant and 56 benign) pathologically diagnosed and surgically treated were included in this analysis. These patients were recruited from four clinical sites: Xiangya No. 2 Hospital in Changsha, Hunan Province, Beijing University Shenzhen Hospital, Shenzhen, Huizhou People's Hospital, Huizhou, and No.2 People's Hospital in Shenzhen, Guangdong Province.

The average age is 56.5 (range 26–79) years, and male proportion is 59%. These numbers are 50.1 (26–73) years and 55% for the benign group, 59.1 (27–79) years and 60% for lung cancer group with a statistically significant difference. No statistically significant difference in terms of smoke status (32% vs 35%) and family history (7% vs 8%) was found between the two groups (table 1).

For the benign lesions, the most common diseases diagnosed are pneumonia (n=14, 25%), tuberculosis (n=12, 21%), pulmonary fibrosis (n=4, 7%) and necrosis granuloma (n=3, 5%).

Lung cancer distribution was 87, 29 and 17 in stages I, II and III, respectively. The average size of the nodules for lung cancer patients is 2.9 (range 0.5–9.0) cm, and for each subgroup: (1) stage I: 2.2 (0.5–4.0) cm; (2) stage II: 3.8 (1.0–7.0) cm; and (3) stage III: 5.0 (1.3–9.0) cm (table 2). The average size of the nodules in the benign group is 2.3 (0.3–6.0) cm, which is statistically smaller than that of the malignant group.

Table 1 Patient demographics

Patients	Non-cancer	Cancer	Total	P value
	n=56	n=136	n=192	
Sex, n (%)				0.53
Male	31 (55.4)	82 (60.3)	113 (58.9)	
Female	25 (44.6)	54 (39.7)	79 (41.1)	
Average age (range)	50.1 (26–73)	59.1 (27–79)	56.5 (26–79)	1.27E-07
Smoking history, n (%)				0.68
Yes	18 (32.1)	48 (35.3)	66 (34.4)	
No	38 (67.9)	88 (64.7)	126 (65.6)	
Family history, n (%)				0.83
Yes	4 (7.1)	11 (8.1)	15 (7.8)	
No	52 (92.9)	125 (91.9)	177 (92.2)	
Lung nodule size (mm, average max dimension)	23.3	29.4	27.6	0.04

Stage IV lung cancer was to be excluded according to the trial design, but due to the fact of late coming of pathological result, we have samples from three patients of this stage collected and analysed. The data for these samples are not statistically meaningful and therefore are included for completeness in the online supplementary tables but not taking into account in the result tables and figures in the main text.

All of the patients are with solitary pulmonary nodules, except for two patients who had two nodules each that are malignant.

Genetic profiling and mutation burden

For each patient, three biospecimen samples including plasma ctDNA, WCC and FFPE tumour tissue were sequenced. For ctDNA, the average sequencing depth is 35 000 with 1350 unique reads after deduplication.

In total, 312 occurrences and 274 unique somatic mutations were found in 29 genes from either plasma ctDNA or tissue DNA in 120 cancer and 5 benign cases (online supplementary table S3).

In the benign lesions, 2 out of 56 patients (3.6%) had four non-driver gene mutations in ctDNA, and three patients (5.4%) had two non-driver mutations in FFPE samples.

Among the lung cancer patients, 88% (120 out of 136 patients) were found to harbour at least one mutation in ctDNA or tumour tissue. When analysed by stage of cancer, the class, that is, whether driver or non-driver mutation, and number of mutations increase as the stage advances (figure 1).

Mutations were found in 9 known lung cancer genes of *ALK*, *BRAF*, *EGFR*, *HER2*, *KRAS*, *MET*, *NRAS*, *PIK3CA* and *ROS1* (figure 2) out of the 12 genes defined as drivers for lung cancer.¹⁸

Table 2 Lung cancer histological diagnosis and stages

	Number of patients (percentage)	Average nodule size (mm, largest dimension)
Pathological diagnosis		
Adenocarcinoma	100 (73.5)	24.3
Squamous carcinoma	28 (20.6)	49.8
Small cell lung cancer	1 (0.7)	18
Others	7 (5.2)	22.3
Clinical stage		
I	87 (64.0)	22
II	29 (21.3)	39
III	17 (12.5)	52.1
IV	3 (2.2)	21

The most commonly mutated genes in the patients with lung cancer are TP53 (44%) and EGFR (35%) (online supplementary figure S1).

The most commonly occurred mutation is *EGFR* L858R (found in 24 samples) followed by *EGFR* exon 19 deletion (in 15 samples). The largest number of mutations was found in a patient who harboured 15 mutations.

Concordance between ctDNA and tumour tissue

Concordance is defined as at least one gene mutation is the same in both the plasma ctDNA and the FFPE tumour tissue gDNA of a patient. When completely no mutations were found in both the blood ctDNA and tissue gDNA, it is also considered as concordant of the two samples.

Among the 136 malignant cases, the overall concordance rate is 27%. Concordance was higher in the driver genes, 46%. The shared mutation rate increases as the stage of the cancer advances: 14%, 48%, 41% and 67% at stage I, II, III, and IV, respectively (online supplementary figure S2).

Comparison with serum biomarkers

A panel of six serum protein tumour biomarkers was also analysed, which has a sensitivity of 51% with a specificity of 83% (table 3). These markers include NSE, CYFRA 21–1, CEA, ProGRP, CA-125 and SCC. When the most sensitive marker of CYFRA 21–1 was considered alone, the sensitivity was merely 25% at a specificity of 95% (table 3).

In comparison, the profiling by ctDNA showed a higher sensitivity in detecting lung cancer. The sensitivity increases as the stage advances and ctDNA outperform the serum biomarkers in all stages (figure 3).

Linear discriminant lung cancer algorithm

The mutation profiling classifies the lung nodules from benign to malignant (figure 1). The benign samples harbour very little to no mutations. The cancer patients have increasing level of mutations both in terms of category and numbers as stage progresses. From stage I to III, there are more mutations found, and they are more often in driver genes.

For lung cancer detection, the overall sensitivity of plasma ctDNA was 69% at the specificity of 96% (table 3). According to cancer stages, the sensitivity rate is 63%, 83% and 94% for stages I, II and III, respectively (figure 3).

We further conducted an LDA where patient age, smoking status and serum protein markers were considered (online

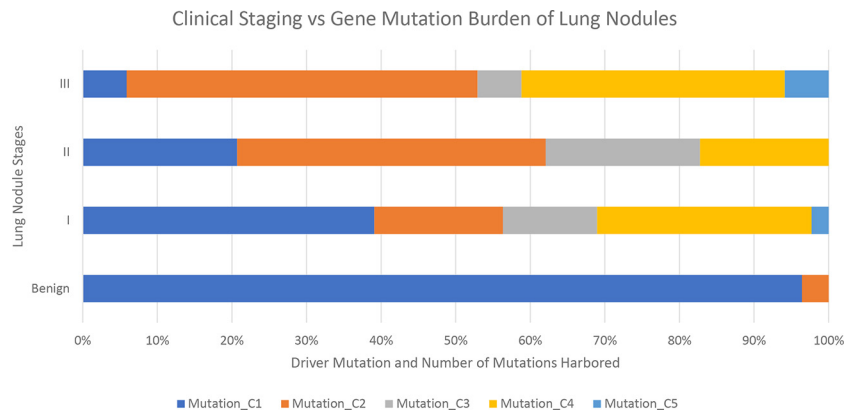


Figure 1 Lung cancer gene mutation burden in benign and various stages of patients. Mutation types: C1: no mutation found; C2: one non-driver mutation found; C3: one driver mutation found; C4: both driver mutation and non-driver mutation found; C5: more than one non-driver mutation found.

supplementary figure S2). The combined model of ctDNA mutations and serum biomarkers improved the sensitivity and specificity to 80% and 99%, respectively (table 3).

DISCUSSION

Lung cancer screening recommended by the guidelines targets populations at high risk of developing lung cancer, such as

patients aged above 55 years, heavy smokers and having chronic obstructive pulmonary disease (COPD) and with family history of lung cancer. In general, by definition of 50–500 mm³,⁶ screening by imagery techniques could result in about 30%–40% indeterminate nodules that need to be further evaluated. In our cohort, we have >60% nodules in this size range making it an urgent need for malignancy assessment.

Our patient population drawn from major hospitals in southern and central China has a median age of 56.5 years, just above the threshold for lung cancer screening. The clinical outcome confirmed the age of above 55 years as a risk factor for lung cancer: the median age of the patients with benign nodules (50.1 years) was about 10 years younger than those having lung cancer (median age of 59.1 years). Although not strictly a screening population, our cross-sectional cohort drawn from the consecutive patients with sign and symptom revealed pulmonary nodules of both benign or malignant nature represents a closer step towards high-risk screening.

Smoking status and family history were not significantly different between the benign and the malignant cases in our cohort. This result is somewhat uncommon given that the role of smoking in lung cancer is considered documented. However, our cohort may not be large enough to observe the effect, especially when considering our cases (early stage cancers) and controls (benign lung lesions) are not the typically studied previously.

Many potential explanations could still be explored. On one hand, our Chinese cohort has 41.1% of women, which has less

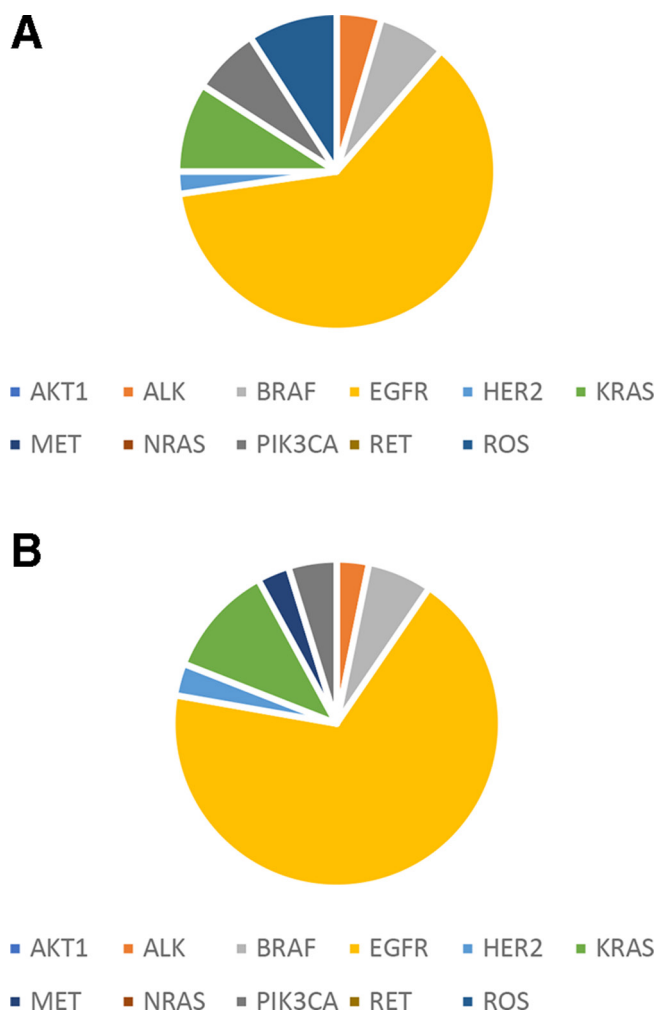


Figure 2 Driver mutation distribution in patients with lung cancer patients. (A) Plasma ctDNA. (B) FFPE gDNA. ctDNA, circulating tumour DNA; gDNA, genomic DNA.

Table 3 Lung nodule malignancy prediction diagnostic test performance

Description	No. of patients	Sensitivity (%)	Specificity (%)
Plasma ctDNA and FFPE	192	91	88
ctDNA mutation (average of all stages)	192	68	96
Stage I lung cancer	87	63	96
Stage II lung cancer	29	83	96
Stage III lung cancer	17	94	96
Panel of six serum biomarkers (BMs)	175	51	83
LDA model of ctDNA and BMs (all stages)	174	80	99

ctDNA, circulating tumour DNA; FFPE, formalin fixed paraffin embedded; LDA, linear discriminant analysis.

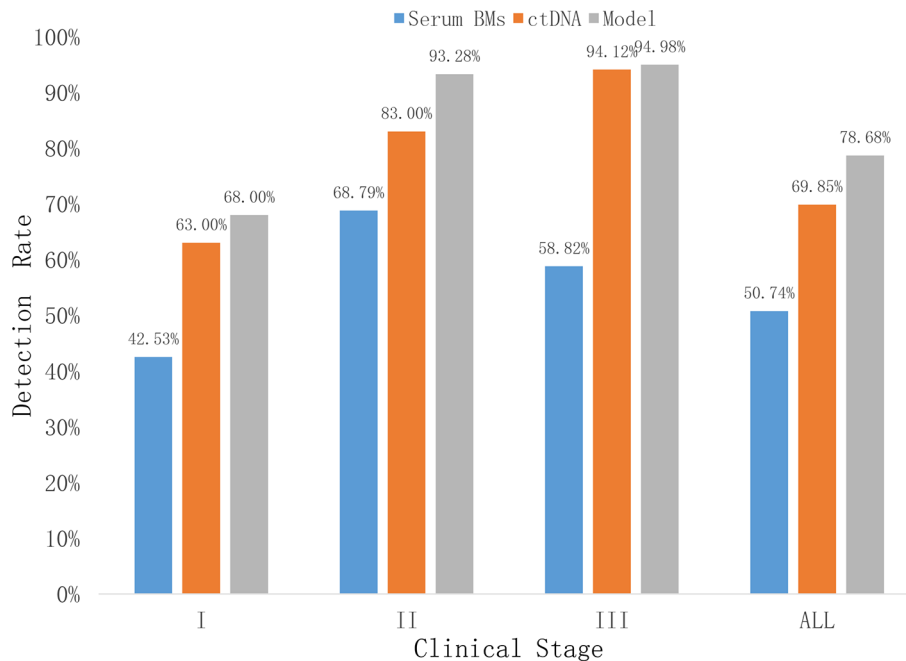


Figure 3 Lung cancer detection rate by ctDNA, serum biomarkers and LDA model per stage. Model=LDA model. ctDNA, circulating tumour DNA; LDA, linear discriminant analysis.

of smokers. Gender distribution also corroborates: women are more in the benign (44.6%) than the cancer group (39.7%). On the other hand, the impact of smoking should be considered more broadly as an environmental factor than tobacco smoking alone. For example, the time spent on Chinese-style cooking could be a potential risk. Unfortunately, no such data were collected. As a proxy, however, the gender distribution difference could be explored since women usually spends more time in the kitchen. While we have more male patients in the cancer group (about 60%), the opposite is true for the benign group, which has less (about 55%).

On the potential interaction of smoking and genes, although there are studies about the association of smoking and tumour driving gene mutations, such as *EGFR*, *ALK*, *KRAS* and *TP53* genes (reviewed in refs 19–22), these are usually studied in patients with late-stage cancer, while ours is heavily centred on the early stage patients (>85% for stages I and II). It is shown that histopathology, gender and ethnicity could also impact the mutation profile of smoking versus non-smoking lung cancers.¹⁹

It is reported that tumour burden of lung cancer corresponds to its size.²³ Our data confirm that the size of the nodule relates to malignancy and progression. The average tumour size increases from 22 mm in stage I to 38 mm in stage II and to 50 mm in stage III (table 2). However, there is no significant difference between a benign nodule and that of the stage I cancer (tables 1 and 2).

The concordance rate between tissue and corresponding plasma ctDNA also reflects the challenge of early cancer liquid biopsy. Our study is heavy in stage I patients (64%, table 2) which has a rate of 32% and causes the overall rate of 36%. The rate increases as stage advances—up to the highest of 78% and the average of stages I–III is 53%. In CCGA, a set of 73 early to midstage (stages I–III) lung cancer samples showed a similar rate of 59% (95% CI 47% to 70%).¹³ In another very small study of 31 paired lung cancer tissues and plasma DNA samples with 10 000-fold ctDNA sequencing depth, the

concordance of mutation between tumour tissue DNA and ctDNA was merely 3.9%.²⁴ Ours is more like CCGA in that we both sequenced ctDNA to the depth >40 000×. Another meta-analysis has also put the pooled sensitivity in the range of 60%.²⁵

The gene mutations shared between the plasma ctDNA and the FFPE tumour tissue increase as the lung cancer stage advances. This is in alignment with the previous report that stage IV tumour has the highest concordance,²⁶ as well as that as the tumour is getting larger, the amount of DNA fragments it sheds into the blood stream will increase.^{23 27} In regards to the cases where there is a discordance between mutations in plasma ctDNA and tissue, many factors could be contributing.

For the mutations found in tissue but not ctDNA samples, it could be due to the challenge of the weak ctDNA signal at early stage^{28 29} like the case of our cohort. The kinetics of ctDNA in the circulation and hence the timing of blood sampling in relation to tissue sampling may have influence that is not yet well understood. The effect of ageing induced clonal haematopoiesis or ageing clonal expansion³⁰ could also play a role. In addition, there is report that leukocytes³¹ harbour and release oncogenic gDNA into the blood and therefore may have impact on the detection of somatic mutations. All these issues will be further studied.

In terms of the mutations found in ctDNA but not in tumour tissue, we believe the most likely explanation is the issue of intertumour and intratumour heterogeneity.^{32 33} Blood sample is more homogeneous and could provide a holistic view of the genetic profile released in the plasma ctDNA by the tumour, while tissue sampling is localised to some specific clone or clones of the cancerous lesions.

Another issue to be watched for in ctDNA analysis is the noisy background of the cfDNA. The majority of variations found are from the WBC, a phenomenon called clonal hematopoiesis.³⁴ For this reason, the ctDNA studies should include the matched WCC sequencing such as ours.

Although somatic mutations showed strong feasibility of detecting malignancy and staging the cancer by plasma ctDNA, there are more factors to be considered. Therefore, combining clinical and genomic features improved the test performance as shown by our LDA modelling (table 3). Integrated classifiers have also been explored in terms of plasma protein biomarkers such as the ones tested by some panel components in our cohort.³⁵ One study involving 60 patients with NSCLC, 40 patients with COPD and 40 healthy controls showed that combining the serum cfDNA concentrations and integrity and CEA improves sensitivity to 93.3%.³⁶

Liquid biopsy starts moving into cancer clinics in therapy selection.^{37–40} A very recent cohort study of over 300 patients with advanced stage lung cancer using ctDNA and/or matched tumour tissue NGS mutation testing guidance for therapy selection showed the utility of liquid biopsy in increasing the positivity of drug selection and treatment outcome.³⁹ Another single-centre study of 102 patients investigated the role of ctDNA in detecting driver gene or other actionable mutations for lung cancer therapy precision and resistance management including serial sampling, in the context where tissue biopsy is limited or could be unavailable.⁴⁰ Another study identified 17 miRNA species in the exosomes of the blood that are differentially expressed in cancer (both NSCLC and SCLC) and controls.⁴¹

Early detection of lung cancer using blood samples is emerging, and similar level of detection performance as ours has been reported.^{13 14} The CCGA study reported a sensitivity of 54% and specificity of 98% for early-stage cancer detection in 127 patients with lung cancer. The potential use of ctDNA for early detection of other cancers has also been reported.^{42 43}

There are, however, a number of limitation and challenges. First of all, the sample sizes of the early detection studies are usually small especially the number of healthy controls. Second, ctDNA amount correlates with cancer stage.⁴⁴ Therefore, the consensus is that ultra-deep sequencing of 40 000× is required to detect the low-frequency mutations in the 10 mL blood. Finally, the less-than-expected driver mutation concordance between ctDNA and tumour DNA may reflect genetic heterogeneity and indicate tumour evolution⁴⁵ suggesting that other types of genes and mutations should be considered as well.

Author affiliations

¹Department of Thoracic Surgery, Second Xiangya Hospital, Changsha, China

²Department of Thoracic Surgery, Peking University Shenzhen Hospital, Shenzhen, China

³Vienomics Corp Ltd, Shenzhen, China

⁴Department of Thoracic Surgery, Shenzhen Second People's Hospital, Shenzhen, China

⁵Department of Clinical Pharmacy, Regents of the University of Michigan, Ann Arbor, Michigan, USA

⁶Oncology Center, Hubei University of Medicine, Shiyan, China

⁷Department of Thoracic Surgery, Sun Yat-sen University Cancer Center, Guangzhou, China

⁸Department of Medical Oncology, Shenzhen Second People's Hospital, The First Affiliated Hospital of Shenzhen University, Shenzhen, China

Acknowledgements The authors would like to thank the physicians and nurses in the Department of Thoracic Surgery, No. 2 Xiangya Hospital, Thoracic Department, Peking University Shenzhen Hospital, Department of Oncology, Shenzhen Second People's Hospital, Oncology Center and Department of Thoracic Surgery, Sun Yat-sen University Cancer Center for their work in enrolling the patients, collecting biospecimen samples and related clinical data. We would like to thank the laboratory staff at Vienomics for processing the samples and conducting the sequencing procedures. Finally, but not the least, we would like to express our highest appreciation of the participation and contribution of our patients in the study.

Contributors MP, YX, YQ, FCh, HY, FY and GT participated in patient recruitment; FCa, CL and CX participated in clinical trial and patient sample management; XL, XY and FX participated in genomic sequencing; XT, DK, BH and CX participated in bioinformatics data analysis; and CX, XL, FY and GT participated in manuscript writing and review.

Funding This work was supported by Innovation Fund of Shenzhen China (Grant No: CKCY2016082916544973 and Grant No: CYZZ20170406170950746); Technological Innovation Research Program of Shenzhen China (Grant No: JSGG20160428090301587 and Grant No: JSGG201704141042216477); State Key Research Program of China (Grant No: 2016YFA0501604); the Young Scientist Innovation Team Project of Hubei Colleges (Grant No: T201510); the Key Project of Health and Family Planning Commission of Hubei Province (Grant No: WJ20172023), Science technology and innovation committee of Shenzhen for research projects (JSGG20160428090301587) and Scientific Research Project of Shenzhen Health and Family Planning System (SZLY2017008/SZLY2018020).

Competing interests None declared.

Patient consent for publication Obtained.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- 1 Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;136:E359–86.
- 2 Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, Nicholson AG, Groome P, Mitchell A, Bolejack V. The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (Eighth) Edition of the TNM classification for lung cancer. *J Thorac Oncol* 2016;11:39–51.
- 3 Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, Miettinen OS. Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med* 2006;355:1763–71.
- 4 Patz EF, Goodman PC, Bepler G. Screening for lung cancer. *N Engl J Med* 2000;343:1627–33.
- 5 Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD; Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395–409.
- 6 Kakinuma R, Muramatsu Y, Yamamichi J, Gomi S, Oubel E, Moriyama N. Evaluation of the 95% limits of agreement of the volumes of 5-year clinically stable solid nodules for the development of a follow-up system for indeterminate solid nodules in CT lung cancer screening. *J Thorac Dis* 2018;10:175–89.
- 7 Tammemägi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, Chaturvedi AK, Silvestri GA, Riley TL, Commins J, Berg CD. Selection criteria for lung-cancer screening. *N Engl J Med* 2013;368:728–36.
- 8 Nicholson KJ, Yip L. An update on the status of molecular testing for the indeterminate thyroid nodule and risk stratification of differentiated thyroid cancer. *Curr Opin Oncol* 2018;30:8–15.
- 9 Silvestri GA, Tanner NT, Kearney P, Vachani A, Massion PP, Porter A, Springmeyer SC, Fang KC, Midthun D, Mazzone PJ, Team PT; Assessment of plasma proteomics biomarker's ability to distinguish benign from malignant lung nodules: results of the PANOPTIC (Pulmonary Nodule plasma proteomic classifier) trial. *Chest* 2018;154:491–500.
- 10 Matikas A, Syrigos KN, Agelaki S. Circulating biomarkers in non-small-cell lung cancer: current status and future challenges. *Clin Lung Cancer* 2016;17:507–16.
- 11 Levy B, Hu ZI, Cordova KN, Close S, Lee K, Becker D. Clinical utility of liquid diagnostic platforms in non-small cell lung cancer. *Oncologist* 2016;21:1121–30.
- 12 Rolfo C, Mack PC, Scagliotti GV, Baas P, Barlesi F, Bivona TG, Herbst RS, Mok TS, Peled N, Pirker R, Raez LE, Reck M, Riess JW, Sequist LV, Shepherd FA, Sholl LM, Tan DSW, Wakelee HA, Wistuba II, Wynes MW, Carbone DP, Hirsch FR, Gandara DR. Liquid Biopsy for Advanced Non-Small Cell Lung Cancer (NSCLC): a statement paper from the IASLC. *Journal of Thoracic Oncology* 2018;13:1248–68.
- 13 Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, Le Quesne J, Moore DA, Veeriah S, Rosenthal R, Marafioti T, Kirkizlar E, Watkins TBK, McGranahan N, Ward S, Martinson L, Riley J, Fraioli F, Al Bakir M, Grönroos E, Zambrana F, Endozo R, Bi WL, Fennessy FM, Sponer N, Johnson D, Laycock J, Shafi S, Czyżewska-Khan J, Rowan A, Chambers T, Matthews N, Turajlic S, Hiley C, Lee SM, Forster MD, Ahmad T, Falzon M, Borg E, Lawrence D, Hayward M, Kolvekar S, Panagiotopoulos N, Jones SM, Thakrar R, Ahmed A, Blackhall F, Summers Y, Hafez D, Naik A, Ganguly A, Kareh S, Shah R, Joseph L, Marie Quinn A, Crosbie PA, Naidu B, Middleton G, Langman G, Trotter S, Nicolson M, Remmen H, Kerr K, Chetty M, Gomersall L, Fennell DA, Nakas A, Rathinam S, Anand G, Khan S, Russell P, Ezhil V, Ismail B, Irvin-Sellers M, Prakash

- V, Lester JF, Kornaszewska M, Attanoos R, Adams H, Davies H, Oukrif D, Akarca AU, Hartley JA, Lowe HL, Lock S, Iles N, Bell H, Ngai Y, Elgar G, Szallasi Z, Schwarz RF, Herrero J, Stewart A, Quezada SA, Peggs KS, Van Loo P, Dive C, Lin CJ, Rabinowitz M, Aerts H, Hackshaw A, Shaw JA, Zimmermann BG, Swanton C. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* 2017;545:446–51.
- 14 Klein EA, Hubbell E, Maddala T, Aravanis A, Beausang JF, Filippova D, Gross S, Jamshidi A, Kurtzman K, Shen L, Valouev A, Venn O, Zhang N, Smith DA, Yeatman TJ, Tibshirani R, Williams RT, Hartman AR, Seiden M, Liu MC. *Development of a comprehensive cell-free DNA (cfDNA) assay for early detection of multiple tumor types: The Circulating Cell-free Genome Atlas (CCGA) study. Abstract#161733.* Chicago, IL, USA: ASCO Conference, 2018.
 - 15 Greenberg AK, Rimal B, Felner K, Zafar S, Hung J, Eylers E, Phalan B, Zhang M, Goldberg JD, Crawford B, Rom WN, Naidich D, Merali S. S-adenosylmethionine as a biomarker for the early detection of lung cancer. *Chest* 2007;132:1247–52.
 - 16 Ostrow KL, Hoque MO, Loyo M, Brait M, Greenberg A, Siegfried JM, Grandis JR, Gaither Davis A, Bigbee WL, Rom W, Sidransky D. Molecular analysis of plasma DNA for the early detection of lung cancer by quantitative methylation-specific PCR. *Clin Cancer Res* 2010;16:3463–72.
 - 17 Xia L, Li Z, Zhou B, Tian G, Zeng L, Dai H, Li X, Liu C, Lu S, Xu F, Tu X, Deng F, Xie Y, Huang W, He J. Statistical analysis of mutant allele frequency level of circulating cell-free DNA and blood cells in healthy individuals. *Sci Rep* 2017;7:7526.
 - 18 Lovly C, Horn L, Pao W. Molecular profiling of lung cancer. My cancer genome. 2018 <https://www.mycancergenome.org/content/disease/lung-cancer/> (Accessed 13th July 2018).
 - 19 Chapman AM, Sun KY, Ruestow P, Cowan DM, Madl AK. Lung cancer mutation profile of EGFR, ALK, and KRAS: Meta-analysis and comparison of never and ever smokers. *Lung Cancer* 2016;102:122–34.
 - 20 Nagahashi M, Sato S, Yuza K, Shimada Y, Ichikawa H, Watanabe S, Takada K, Okamoto T, Okuda S, Lyle S, Takabe K, Tsuchida M, Wakai T. Common driver mutations and smoking history affect tumor mutation burden in lung adenocarcinoma. *J Surg Res* 2018;230:181–5.
 - 21 Desrichard A, Kuo F, Chowell D, Lee KW, Riaz N, Wong RJ, Chan TA, Morris LGT. Tobacco smoking-associated alterations in the immune microenvironment of squamous cell carcinomas. *J Natl Cancer Inst* 2018;110:1386–92.
 - 22 Tseng CH, Chiang CJ, Tseng JS, Yang TY, Hsu KH, Chen KC, Wang CL, Chen CY, Yen SH, Tsai CM, Huang MS, Ho CC, Yu CJ, Tsai YH, Chen JS, Chou TY, Tsai MH, Chen HY, Su KY, Chen JWW, Chen HW, Yu SL, Liu TW, Chang GC. EGFR mutation, smoking, and gender in advanced lung adenocarcinoma. *Oncotarget* 2017;8:98384–93.
 - 23 Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayr FO, Hesch RD, Knippers R. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res* 2001;61:1659–65.
 - 24 Ye M, Li S, Huang W, Wang C, Liu L, Liu J, Liu J, Pan H, Deng Q, Tang H, Jiang L, Huang W, Chen X, Shao D, Peng Z, Wu R, Zhong J, Wang Z, Zhang X, Kristiansen K, Wang J, Yin Y, Mao M, He J, Liang W. Comprehensive targeted super-deep next generation sequencing enhances differential diagnosis of solitary pulmonary nodules. *J Thorac Dis* 2018;10:S820–S829.
 - 25 Luo J, Shen L, Zheng D. Diagnostic value of circulating free DNA for the detection of EGFR mutation status in NSCLC: a systematic review and meta-analysis. *Sci Rep* 2014;4:6269.
 - 26 Douillard JY, Ostoros G, Cobo M, Ciuleanu T, Cole R, McWalter G, Walker J, Dearden S, Webster A, Milenkova T, McCormack R. Gefitinib treatment in EGFR mutated caucasian NSCLC: circulating-free tumor DNA as a surrogate for determination of EGFR status. *J Thorac Oncol* 2014;9:1345–53.
 - 27 Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, Thornton K, Agrawal N, Sokoll L, Szabo SA, Kinzler KW, Vogelstein B, Diaz LA. Circulating mutant DNA to assess tumor dynamics. *Nat Med* 2008;14:985–90.
 - 28 Aravanis AM, Lee M, Klausner RD. Next-generation sequencing of circulating tumor DNA for Early cancer detection. *Cell* 2017;168:571–4.
 - 29 Bettgowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, Bartlett BR, Wang H, Luber B, Alani RM, Antonarakis ES, Azad NS, Bardelli A, Brem H, Cameron JL, Lee CC, Fecher LA, Gallia GL, Gibbs P, Le D, Giuntoli RL, Goggins M, Hogarty MD, Holdhoff M, Hong SM, Jiao Y, Juhl HH, Kim JJ, Siravegna G, Laheru DA, Lauricella C, Lim M, Lipson EJ, Marie SK, Netto GJ, Oliner KS, Olivi A, Olsson L, Riggins GJ, Sartore-Bianchi A, Schmidt K, Shih IM, Oba-Shinjo SM, Siena S, Theodorescu D, Tie J, Harkins TT, Veronesi S, Wang TL, Weingart JD, Wolfgang CL, Wood LD, Xing D, Hruban RH, Wu J, Allen PJ, Schmidt CM, Choti MA, Velculescu VE, Kinzler KW, Vogelstein B, Papadopoulos N, Diaz LA. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* 2014;6:224ra24.
 - 30 Risques RA, Kennedy SR. Aging and the rise of somatic cancer-associated mutations in normal tissues. *PLoS Genet* 2018;14:e1007108.
 - 31 Chennakrishnaiah S, Meehan B, D’Asti E, Montermini L, Lee TH, Karatzas N, Buchanan M, Tawil N, Choi D, Divangahi M, Basik M, Rak J. Leukocytes as a reservoir of circulating oncogenic DNA and regulatory targets of tumor-derived extracellular vesicles. *J Thromb Haemost* 2018;16:1800–13.
 - 32 Lee WC, Diao L, Wang J, Zhang J, Roarty EB, Varghese S, Chow CW, Fujimoto J, Behrens C, Cascone T, Peng W, Kalhor N, Moran CA, Weissferdt A, Johnson FM, William WN, Swisher SG, Lee JJ, Hong WK, Heymach JV, Wistuba II, Futreal PA, Zhang J. Multiregion gene expression profiling reveals heterogeneity in molecular subtypes and immunotherapy response signatures in lung cancer. *Mod Pathol* 2018;31:947–55.
 - 33 Kang F, Wang Z, Li G, Wang S, Liu D, Zhang M, Zhao M, Yang W, Wang J. Inter-heterogeneity and intra-heterogeneity of α ₃ in non-small cell lung cancer and small cell lung cancer patients as revealed by ⁶⁸Ga-RGD₂ PET imaging. *Eur J Nucl Med Mol Imaging* 2017;44:1520–8.
 - 34 Hu Y, Ulrich BC, Supplee J, Kuang Y, Lizotte PH, Feeney NB, Guibert NM, Awad MM, Wong KK, Jänne PA, Paweletz CP, Oxnard GR. False-positive plasma genotyping due to clonal hematopoiesis. *Clin Cancer Res* 2018;24:4437–43.
 - 35 Guida F, Sun N, Bantis LE, Muller DC, Li P, Taguchi A, Dhillon D, Kundnani DL, Patel NJ, Yan Q, Byrnes G, Moons KGM, Tjønneland A, Panico S, Agnoli C, Vineis P, Palli D, Bueno-de-Mesquita B, Peeters PH, Agudo A, Huerta JM, Dorronsoro M, Barranco MR, Ardanaz E, Travis RC, Byrne KS, Boeing H, Steffen A, Kaaks R, Hüsing A, Trichopoulou A, Lagiou P, La Vecchia C, Severi G, Boutron-Ruault MC, Sandanger TM, Weidpass E, Nøst TH, Tsilidis K, Riboli E, Grankvist K, Johansson M, Goodman GE, Feng Z, Brennan P, Johansson M, Hanash SM. Assessment of lung cancer risk on the basis of a biomarker panel of circulating proteins. *JAMA Oncol* 2018;4:e182078.
 - 36 Soliman SE, Alhanafy AM, Habib MSE, Hagag M, Ibrahim RAL. Serum circulating cell free DNA as potential diagnostic and prognostic biomarker in non small cell lung cancer. *Biochem Biophys Rep* 2018;15:45–51.
 - 37 Cabanero M, Tsao M. Circulating tumour DNA in EGFR-mutant non-small-cell lung cancer. *Curr Oncol* 2018;25:38–S44.
 - 38 Hu Y, Alden RS, Odegaard JI, Fairclough SR, Chen R, Heng J, Feeney N, Nagy RJ, Shah J, Ulrich B, Gutierrez M, Lanman RB, Garber JE, Paweletz CP, Oxnard GR. Discrimination of germline EGFR T790M mutations in plasma cell-free DNA allows study of prevalence across 31,414 cancer patients. *Clin Cancer Res* 2017;23:7351–9.
 - 39 Aggarwal C, Thompson JC, Black TA, Katz SI, Fan R, Yee SS, Chien AL, Evans TL, Bauml JM, Alley EW, Ciunci CA, Berman AT, Cohen RB, Lieberman DB, Majmudar KS, Savitch SL, Morrisette JJD, Hwang WT, Elenitoba-Johnson KSJ, Langer CJ, Carpenter EL. Clinical implications of plasma-based genotyping with the delivery of personalized therapy in metastatic non-small cell lung cancer. *JAMA Oncol* 2018.
 - 40 Thompson JC, Yee SS, Troxel AB, Savitch SL, Fan R, Balli D, Lieberman DB, Morrisette JD, Evans TL, Bauml J, Aggarwal C, Kosteva JA, Alley E, Ciunci C, Cohen RB, Bagley S, Stonehouse-Lee S, Sherry VE, Gilbert E, Langer C, Vachani A, Carpenter EL. Detection of therapeutically targetable driver and resistance mutations in lung cancer patients by next-generation sequencing of cell-free circulating tumor DNA. *Clin Cancer Res* 2016;22:5772–82.
 - 41 Poroyko V, Mirzapoirozova T, Nam A, Mambetsariev I, Mambetsariev B, Wu X, Husain A, Vokes EE, Wheeler DL, Salgia R. Exosomal miRNAs species in the blood of small cell and non-small cell lung cancer patients. *Oncotarget* 2018;9:19793–806.
 - 42 Leary RJ, Sausen M, Kinde I, Papadopoulos N, Carpten JD, Craig D, O’Shaughnessy J, Kinzler KW, Parmigiani G, Vogelstein B, Diaz LA, Velculescu VE. Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Sci Transl Med* 2012;4:162ra154.
 - 43 Schwarzenbach H, Müller V, Milde-Langosch K, Steinbach B, Pantel K. Evaluation of cell-free tumour DNA and RNA in patients with breast cancer and benign breast disease. *Mol Biosyst* 2011;7:2848–54.
 - 44 Diehl F, Li M, Dressman D, He Y, Shen D, Szabo S, Diaz LA, Goodman SN, David KA, Juhl H, Kinzler KW, Vogelstein B. Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci U S A* 2005;102:16368–73.
 - 45 Pezzuto A, Manicone M, Scaini MC, Ricci A, Mariotta S, Zamarchi R, Rossi E. What information could the main actors of liquid biopsy provide? - a representative case of non-small cell lung cancer (NSCLC). *J Thorac Dis* 2018;10:E570–6.