Peer Review File

Article Information: https://dx.doi.org/10.21037/tlcr-21-564

Reviewer A

The authors performed a large-scale comparison of ADx-ARMS PCR and digital PCR for the detection of EGFR T790M mutations. Although this is not the first study comparing those platforms, it is the largest study investigating the T790M variant specifically. The moderate sensitivity of assays that are used for detection of EGFR mutations in ctDNA remains an issue and causes the persistent need for tissue biopsy when plasma results are negative. Therefore, Xu and colleagues have made an important effort aiming at increasing the sensitivity of T790M mutation detection with digital PCR. Unfortunately, the authors compared the performance of their digital PCR to the performance of the ADx-ARMS PCR which is known to yield inferior results compared other assays, such as the cobas assay, which is the only FDA-approved assay. This limits the utility of this study. Moreover, testing the performance of the digital PCR assay with tumor tissue as a reference remains of great importance, this analysis was performed for a subset of 45 paired samples. From this analysis, the sensitivity of the digital PCR assay was only 53.85%. This sensitivity is slightly inferior when compared to the reported sensitivity of 61% for the cobas assay for the T790M mutation (PMID 28428148). Therefore, this assay seems to have limited utility and the conclusions of this study are not fully supported by the results.

Major comments:
1. The digital PCR assay was not compared to currently used and approved assays but to the inferior ADx-ARMS assay. The authors should elaborate on this in the discussion.

Reply: Thank you for bringing this point to our attention as we hadn't discussed it in great detail in the paper. As this was a clinical trial, the ultimate goal of the study was to register the dPCR kit with the NMPA. Because we were complying with registration requirements, the ADx-ARMS PCR assay had to be used as the primary assay because it was an NMPA approved assay and had been used routinely in labs. The Cobas® EGFR Mutation Test V2 had not gained NMPA approval at the time which was why it was relegated as a secondary assay. After using the Cobas assay, it became clear that it was a superior assay to ADx-ARMS PCR as it had better consistency, especially at lower allele frequencies. However, we did test ADx-ARMS PCR and dPCR concordant and unconcordant samples with the Cobas PCR assay and found that ADx-ARMS PCR and dPCR concordant samples were also concordant with Cobas PCR. This gave us confidence the results from ADx-ARMS PCR combined with Cobas PCR would be an equivalent to the trials done with Cobas PCR as a comparator assay.

Changes in the text: Added to Methods section, Study design and participants, Lines 201-208:

The samples were tested for EGFR p.T790M by dPCR and ADx-ARMS® EGFR mutation detection kit (ADx-ARMS PCR) as a comparator assay; Cobas® EGFR Mutation Test v2 (Cobas PCR) was utilized for the samples with inconsistent results between ADx-ARMS PCR and dPCR.

ADx-ARMS PCR was chosen as the primary reference assay as it was the only NMPA approved liquid biopsy assay at the time of the study.

Added to Discussion section, Lines 467-483:

The main goal of our study is to evaluate the sensitivity and specificity of the chip-based dPCR assay for clinical use. Because there were no digital PCR assays approved for clinical use on the market that had extensive clinical validation data, a qPCR assay, which was the most similar assay to dPCR, was used. The ADx-ARMS PCR assay was chosen as the comparator assay because it had been extensively validated clinically and had been routinely used in diagnostic labs. For samples that were unconcordant between dPCR and ADx-ARMS PCR, Cobas PCR was used to retest those samples. By testing non-concordant and concordant samples with Cobas PCR, the results showed that it was a better assay than ADx-ARMS PCR but the sensitivity was not as good as dPCR. Using mutant allele frequency determined by dPCR, it also became evident that ADx-ARMS PCR and Cobas PCR had false negative test results due to lack of detection sensitivity at lower allele frequency especially between 0.1% and 1%. This is partly due to the relative quantification of both methods, unlike dPCR which uses absolute quantification, and due to the threshold being set manually for qPCR. For dPCR, positive samples were further confirmed by the mutation status obtained from paired tissue samples, suggesting improved sensitivity did not lead to compromised specificity.

2. The authors used a cut-off at 0.1% allele frequency to reduce false positive results. However, it would be more convenient to use a number of mutation positive droplets since a cut-off based on allele frequency artificially increases the LOD of the digital PCR. The number of mutation positive droplets in the negative controls (in the manuscript referred to as 'special samples') should be used as a cut-off.

Reply: Thank you for the inclusion of 'droplets' in your comment as we hadn't realized the type of dPCR was not explicitly stated in the text. Chip-based dPCR was used in this study, and this discrepancy has been clarified in the text. The number of mutation positive wells could be misleading if the input amount of DNA was not inconsistent. Unlike contrived cell-line samples, the input amount of DNA for clinical samples is more varied. For this reason, we performed analytical validation on the input amount of DNA ranging from 20-80ng prior to the clinical trial with all samples being processed from the universal 20mL of blood taken from all patients. The 20mL of blood was then split into three parts for all three assays involved in the trial. More

details regarding input amount of DNA and respective background level will be further discussed in reply to Comment 7.

In the manuscript, 'special' and 'specific' samples are referring to the same group of patients; going forward they are referred to as the specific samples. This discrepancy has been clarified in the manuscript. The specific samples, which were patients with either lung nodules or small-lung-cell-cancer, cannot act as negative controls because the patients may still be positive for the mutation without having malignant cancer. If the quantity of mutant DNA is not enough to be detected by the assay but is still present in the body, it would bring hesitancy to the validity of the negative control. Additionally, even with healthy volunteers would not be able to act as a negative control for the same reason as there is no guarantee that they don't have the mutation. There are too many confounding variables with these patient samples for them to be accurate negative controls. An appropriate negative control for this experiment would be a cell-line sample known to lack the mutation.

Changes in the text: Chip-based dPCR: Added to Methods section, EGFR p. T790M mutation detection, Lines 246-250:

The reaction mixture was loaded onto a silicon chip using a QuantStudio™ 3D Digital PCR Chip Loader (catalog# 4482592, Thermo Fisher Scientific), and the PCR reaction was performed on a ProFlex™ 2x Flat PCR System (catalog# 4484078, Thermo Fisher Scientific).

0.1% cut-off value: Added to Methods section, EGFR p. T790M mutation detection, Lines: 255-264

The number of target DNA copies was converted from the number of wells using the Poisson distribution. Mutant allele frequency was calculated as:

*Copies of FAM positive DNA/(Copies of FAM positive DNA + Copies of VIC positive DNA) * 100%.*

The cut-off value used was the LoD of EGFR p.T790M digital PCR assay at 0.1% mutant allele frequency 30. The receiver operating characteristic curve (ROC curve) was used to validate the cut-off value in this study.

0.1% cut-off value: Added to Discussion section, Lines 485-495:

The cut-off value of 0.1% that was used was determined by a previously published study. In this study by Zhou et al., they investigated the detection of EGFR p.T790M mutation in NSCLC patients using cell-line samples. One of the criteria for finding a sample mutation positive for dPCR was a mutant allele frequency greater than 0.1%, using the equation cited in the Methods section. This limit of detection was determined using both 20ng and 80ng of input DNA making the criteria applicable for the applications of this paper. To validate the cut-off value, the ROC curve analysis as well as the corresponding Youden Index value was used. While the highest Youden Index value was 0.11%, due to minimal differences between the sensitivity, specificity, PPV, NPV, and concordance between 0.11% and 0.1% cut-off the decided cut-off value for this assay was 0.1% mutant allele frequency.

Specific samples, Added to Methods section, Study design and participants, Lines: 196-197:

Of the specific samples, 4 patients had lung nodules and 12 patients had small-cell-lung-cancer.

Specific samples, Added to Results section, Comparison of EGFR p.T790M detection using dPCR, ADx-ARMS PCR, and Cobas PCR, Lines 309-311:

Among the 1026 samples, there are 4 patients with lung nodules and 12 patients with small-cell-lung-cancer enrolled as specific samples. As expected, all 16 samples were tested to be EGFR p.T790M negative.

3. In addition to point 2, how cut-offs were determined should be described in the method section. Did the authors exclude all samples with a variant allele frequency <0.1% from the analysis?

Reply: We appreciate your help in determining what we are missing in the Methods section. While the cut-off was determined by a previously published paper by Zhou et al. (2018), the cut-off value was validated by the ROC curve.

No, samples with allele frequencies below 0.1% were not excluded. Because the sensitivity of the assays tested did not detect allele frequencies below 0.1%, the samples with allele frequencies below 0.1% were considered mutation negative.

Changes in the text: Added to Methods section, EGFR p. T790M mutation detection, Lines: 255-264

The number of target DNA copies was converted from the number of wells using the Poisson distribution. Mutant allele frequency was calculated as:

*Copies of FAM positive DNA/(Copies of FAM positive DNA + Copies of VIC positive DNA) * 100%.*

The cut-off value used was the LoD of EGFR p.T790M digital PCR assay at 0.1% mutant allele frequency 30. The receiver operating characteristic curve (ROC curve) was used to validate the cut-off value in this study.

4. Methods, line 164-167: It is not clear what is meant by 'specific samples,' this is explained in the results section as patients with SCLC or lung nodules. These were presumably negative controls. Please clarify this in the method section.

Reply: In the manuscript, 'special' and 'specific' samples are referring to the same group of patients; going forward they are referred to as the specific samples. This discrepancy has been clarified in the manuscript. As stated previously in Comment 2, these patients would not be appropriate negative controls because of the potential for them to have the mutation without being detected by the assay. A negative control for this study would be to use known mutation negative contrived samples.

Changes in the text: Added to Methods section, Study design and participants, Lines 196-197:

Of the specific samples, 4 patients had lung nodules and 12 patients had small-cell-lung-cancer.

Specific samples, Added to Results section, Comparison of *EGFR* p.T790M detection using dPCR, ADx-ARMS PCR, and Cobas PCR, Lines 309-311:

Among the 1026 samples, there are 4 patients with lung nodules and 12 patients with small-cell-lung-cancer enrolled as specific samples. As expected, all 16 samples were tested to be EGFR p.T790M negative.

5. In addition to point 4, the authors state that the 16 special samples were determined by statisticians? How? Was a power calculation performed? If so, please elaborate on this. A number of 16 negative controls seems quite low.

Reply: Thank you for your comment as it made us aware of points of confusion in the paper. The 16 samples were not determined by statisticians. The 16 samples were a subpopulation of the overall cohort of 1045 patients. The statisticians determined the minimum number of patients needed to be enrolled in the overall study. They were not used to determine the minimum number of patients for each sub-population. Because the number of enrolled patients far exceeds the minimum requirement, the portion of the sentence regarding statisticians was removed to avoid confusion.

As stated in Comment 1, the ultimate goal of the study was to register the dPCR kit with the NMPA. One of the requirements was to have a subpopulation that encompassed the 'specific samples'; 16 samples was determined to be enough for this study. For this reason, the inclusion of the specific samples did not have anything to with analysis.

Changes in the text: N/A

6. The amount of DNA (input) that was used for the ADx-ARMS and cobas assay was not mentioned. The authors should clarify this in the methods as different inputs could result in lower/higher LODs and an unfair comparison. Did the authors correct for DNA input in their comparison?

Reply: Thank you for bringing this to our attention; the amount of input DNA for all assays tested was the same (20-80ng). As there is no difference in the input amount, there would not be a discrepancy between the different assays tested when LoD is taken into consideration. Additionally, the goal of the study was to compare the assay performance under the same conditions, and not to determine and compare the ultimate LoD for different assays. Therefore, as long as the amount of input DNA was the same for all assays, it would be a fair comparison and there would be no need to correct for DNA input.

Changes in the text: Added to Methods section, EGFR p. T790M mutation detection, Lines: 270-271.

All assays used the same amount input cfDNA.

7. The amount of DNA input for the digital PCR was 20-80 nanograms. This is a large variation that will also affect the LOD. In addition, a digital PCR chip would be overloaded when using more than 50 nanograms of DNA. Did the authors define a maximum of DNA input? Were chips overloaded?

Reply: We appreciate your concern regarding overloaded chips and input DNA. The input amount of DNA varies because of the amount of cfDNA in the collected 20mL of blood varies between patients. The maximum DNA input allowed is 80ng.

The chips were overloaded for the wild-type allele but were not for the *EGFR* p.T790M mutation. Because we were working with dPCR chips, there are 20,000 compartments per chip. 20ng of input DNA would result in 6,000 copies of the wild-type allele and about 6-7 copies of the mutant allele. On the other hand, 80ng of input DNA would result in 24,000 copies of the wild-type allele and around 24 copies of the mutant allele. In both cases, the chip would be overloaded and more than one copy will be in a single well for wildtype allele but not the mutant allele. However, because the Poisson distribution was used to convert the number of wells to number of copies in the reaction, and the calculation of the mutant allele frequency incorporates the wild-type allele, the overloading did not affect the comparison of chips.

The range of 20ng to 80ng was used to make the assay more robust and tolerable for end users. As it is difficult to accurately quantitate cfDNA at low concentration and it is beneficial to avoid false negative results when increased amount of DNA is used, we wanted to incorporate an appropriate range of input DNA so that it can have more usability. The high end of 80ng was used to determine the background. In a previously published paper by Zhou et al. (2018), background signal of 20ng and 80ng of input DNA was analyzed. We found that, for samples that were mutation negative, there was more background signal for 80ng of input DNA (more than half had 1-3 mutation positive wells) when compared to 20ng of input DNA (majority of ~70% had no mutation positive wells). With this information, we used 5 mutation positive wells as the background signal; the criteria of determining a sample as positive would need 6 or more mutation positive wells. In the same study, they also determined that 20ng of input DNA was the minimum amount of input DNA to avoid false negative results; 20ng of input DNA had an LoD of around 0.1% which is why it is part of the criteria for established a sample as positive.

In summary, we determined background using 80ng of input DNA and the LoD using 20ng of input DNA. We allow users to use 20-80ng as a range of sample amount to best avoid false negative results.

Changes in the text: N/A

8. Methods, line 164-167: the authors state that this was a prospective study. When

consultating this trial in the Chinese Clinical Trial Registry it states that an inclusion criteria for this study was: Available homologous tissue samples are the remaining paraffin-embedded tissue samples collected during the same period (within four weeks before and after the blood sampling) of the patient's plasma samples. However, tissue analysis was only reported for 45 cases. Please explain this discrepancy. Were samples collected prospectively?

Reply: We appreciate your effort on evaluating this study. On the Chinese Clinical Trial Registry inclusion criterion, the first point on the list of five, which discusses needing to be tested for the mutation, is what is necessary to be included in the study. The remaining four points, including the criterion pointed out, are optional but not necessary for inclusion.

The tissues samples were collected around the same time as the blood samples. The small number of cases with tissue samples was due to most patients not being willing to provide tissue samples or the location of the tumor made tissue biopsy too risky. The paired tissue samples, while providing insightful information, was not required of all patients because the goal of the study was to compare two liquid biopsy tests.

Changes in the text: N/A

9. Line 473-474: the authors conclude that the newly –developed digital PCR has a high sensitivity. Based on comparison with tumor tissue, the sensitivity was 53.85%. These results do not support the conclusion. Please mitigate.

Reply: We thank you for voicing your concern regarding the sensitivity of dPCR and how we've explained the significance. The wording of 'high sensitivity' is misleading and has been fixed in the manuscript. What we were trying to convey was that the sensitivity of this assay was high compared to the qPCR assays tested in this study. The 61% sensitivity of Cobas® EGFR Mutation Test V2 described (PMID: 28428148) is due to differences in the input amount of DNA (as explained thoroughly in the Discussion section, referenced below).

Changes in the text: Added to Results section, Comparison of *EGFR* p.T790M detection in paired tissue and plasma samples using digital PCR and ARMS-PCR, Lines 386-393:

Taking a closer look at the results, among the 13 out of 45 samples tested positive in tissue biopsy, only 4 were tested positive in plasma samples by ADx-ARMS PCR. 4 were tested positive by Cobas PCR, and 7 were tested positive by dPCR. The additional 3 positive samples determined negative by ADx-ARMS PCR and positive by dPCR had an allele frequency of 0.21%, 0.94% and 1.66%, which is at an allele frequency close to or below the limit of detection of ADx-ARMS PCR. This validates the notion that these positive samples are real; the additional positive samples are confirmed to be true positives by comparing their status to tissue biopsy.

Added to Discussion section, Lines 491-516:

A previously published study cites that the sensitivity of Cobas PCR is 61% vs. 53.85% from our study when results from paired tissue and plasma samples were compared35. However, this is not a fair comparison because the input amount of DNA is different between the published study and ours. 50ng of input DNA Cobas PCR were used in the published study while majority of the samples only had close to 20ng of cfDNA36. This means that the limit of detection according to the user manual would be around 23 copies/reaction. Additionally, ADx-ARMS® EGFR mutation detection kit required 10ng of input DNA per reaction for a limit of detection of 1% mutant allele frequency, or 6-7 copies/reaction37. Another source did site the LoD for ADx-ARMS PCR ranges from 0.5 to 7.02% depending on the mutation, but which mutation corresponded to what LoD was not described and the input amount of DNA to achieve that LoD was not described as well38. On the other hand, the dPCR detection kit used 20ng of input DNA making the limit of detection around 6-7 copies/reaction. Due to the input amount of DNA of the qPCR assays being larger than the dPCR assay, the established limit of detection for the two qPCR assays would not be a fair comparison with dPCR. Additionally, 25mL of blood was required for the qPCR assays while 10mL of blood was used in this study. In this study, we used the same amount of input DNA for all assays and found that, although the Cobas PCR assay had better sensitivity than ADx-ARMS PCR, dPCR had the best sensitivity out of all the assays tested. This is most clearly shown when comparing the results of paired tissue and plasma samples. Of the 13 samples determined to be positive by tissue biopsy, dPCR found 7 samples positive while Cobas PCR found 4 samples positive, and ADx-ARMS PCR found 4 samples positive. The samples were confirmed to be true positive because they were compared against tissue biopsy. When the detection of all assays is compared equally, dPCR still has better sensitivity than both assays although the sensitivity was numerically low, at 53.85%.

Minor comments:
10. Results line 337-338: the authors state that the results further support the necessity of liquid biopsies to avoid false negative results from tissue biopsy. With a sensitivity of 53.85% for the digital PCR assay, this statement could not be supported by the results. Instead, these results support the fact that tissue and plasma testing should be complementary.

Reply: We thank you for bringing this point to our attention. We do think that the results further support the need for liquid biopsy in the realm of companion diagnostics; however, we also agree that tissue and plasma testing should be complementary. With current NCCN Guidelines, liquid and tissue biopsies are used together; however, we are pushing for liquid biopsies to improve as it compensates for the false negatives of tissue biopsy.

When we discuss avoiding false negative results from tissue biopsy, it is not because the sensitivity of tissue biopsy is poor. It is because there are other factors

that influence the mutations that can be detected via tissue biopsy; one of the biggest factors being tumor heterogeneity. For patients with large or multiple tumors, they may have mutations that make them eligible for target therapy; however, the tissue sample may not represent those mutations as the sample is relatively small compared to overall tumor burden. Liquid biopsy would resolve this issue as it analyzed circulating tumor DNA, which would have the mutations present throughout large and/or multiple tumors.

Currently, the National Comprehensive Cancer Network (NCCN Guidelines In Oncology (NCCN Guidelines®): Non-Small Cell Lung Cancer, Metastatic) has established guidelines that liquid biopsy tests can be conducted before tissue biopsy; for patients that are found mutation positive, they can go directly to targeted therapy. For patients that are found mutation negative, they should receive a tissue biopsy test, if possible. These guidelines support the use and betterment of liquid biopsy tests as it is the first test conducted. Additionally, we found that with tissue negative samples, ADx-ARMS PCR found 2 positive samples and dPCR found an additional 1 positive samples. The concordance between ADx-ARMS PCR and dPCR gives support that the samples are true positive. For the additional one sample determined positive by dPCR, it had an allele frequency of 0.29%, this is within the limit of detection for dPCR but not for ADx-ARMS PCR. These results show that improvements in sensitivity of liquid biopsy would help determine more samples positive for mutations that would not otherwise be detected in tissue biopsy samples.

Changes in the text: N/A

11. Line 164 methods section: the trial number that the authors referred to is a number from the Chinese Clinical Trial Registry, not ClinicalTrial.gov. Please adjust this.

Reply: Thank you for bringing this error to our attention; it has been noted and appropriate changes have been made to the text.

Changes in the text: Added to Methods section, Study design and participants; Line: 193-196.

This prospective study (Chinese Clinical Trial Registry identifier: ChiCTR2100043147) included 1045 patients enrolled from 15 medical centers across China, including 1029 patients diagnosed with advanced NSCLC and 16 patients with specific samples (Table 1).

12. This manuscript needs an English writing revision.
a. Line 268: lung modules. Please change to lung nodules
b. Line 343: analysis were. Please change to analyses were
c. Line 321: they. Please change to ‘we’ or ‘in our analysis’, if that is what is meant

Reply: The appropriate edits were made to the text.

Changes in the text: N/A

Reviewer B

The authors demonstrate that digital PCR is a more sensitive approach than ARMS-PCR to detect the EGFR T790M mutation in circulating tumor DNA. This has already been published many times, but in this case, more than 1000 patients/samples have been tested.

More importantly, the authors demonstrate that patients tested positive by dPCR responded to osimertinib, even if they were tested negative with ARMS.

Major comments :

- A positive sample is determined by a VAF greater than a cut-off value, but also a minimal mutant positive micro-reactions. This is clearly stated in the discussion, but in the results section, only the VAF is mentionned. Could the authors provide more details on how they analyze the data (refering to their previously published paper) ?

Reply: Thank you for bringing this point to our attention; for this study, a positive sample is determined by a VAF above the cut-off value, not the minimal mutant positive micro-reactions. This is because the number of wells needs to be normalized to accommodate the range of input DNA to make the results comparable. Because the input amount of DNA affects the amount of background signal, the allele frequency, which adjusts for input DNA by including the number of wild-type gene positive copies, was used as a measure instead of absolute FAM (or mutant positive) copies. Additionally, the criteria mentioned in the Discussion section is for contrived cell-line samples; these samples don't reflect the range of input DNA seen in clinical samples.

In the previous paper by Zhou et al. (2018), the VAF was calculated as: (copies of FAM positive DNA)/(copies of FAM positive DNA + copies of VIC positive DNA)*100%. FAM positive DNA is mutation positive. VIC positive is wild-type gene positive. Using the normalized allele frequency, the cut-off value was set to 0.1% with an input amount of DNA of at least 20ng. Input DNA lower than 20ng resulted in false negative results. It was also determined that samples with less than 5 FAM positive wells was considered negative. The number of mutant positive wells was determined by the number of mutant positive micro-reactions when no mutation is present.

Changes in the text: Added to Discussion section, Lines 485-495:

The cut-off value of 0.1% that was used was determined by a previously published study. In this study by Zhou et al., they investigated the detection of EGFR p.T790M mutation in NSCLC patients using cell-line samples. One of the criteria for finding a sample mutation positive for dPCR was a mutant allele frequency greater than 0.1%, using the equation cited in the Methods section. This limit of detection was determined using both 20ng and 80ng of input DNA making the criteria applicable for

the applications of this paper. To validate the cut-off value, the ROC curve analysis as well as the corresponding Youden Index value was used. While the highest Youden Index value was 0.11%, due to minimal differences between the sensitivity, specificity, PPV, NPV, and concordance between 0.11% and 0.1% cut-off the decided cut-off value for this assay was 0.1% mutant allele frequency.

- The data presented by Fig 2 are very interesting. In light with the previous comment, could the authors also present a similar figure with the number of positive micro-reactions (or the absolute number of T790M copies / mL) ?

Reply: Thank you for your suggestion; however, as discussed in Comment 1, because the allele frequency adjusts for the amount of input DNA and the absolute number of mutant copies does not, it would not make sense to include that figure. The distribution would need to be normalized to the amount of input DNA or the figure would show misleading information about the number of copies present because of the varying degrees of background signal with each sample. If contrived cell-line samples were used, it would be more appropriate to show the absolute number of copies.

Changes in the text: N/A

- The COBAS generates a semi-quantitative index (ISQ), which is supposed to be related to the number of mutated copies. Could the authors present data comparing ISQ to VAF and number of T790M copies ?

Reply: We appreciate you bring SQI to our attention. When looking into the SQI, I was unable to find any reference to it in the user manual for the Cobas ARMS test (versions 3-8, available on diagnostics.roche.com/). The inclusion of SQI would also not be a compatible comparison with VAF because the data is based off of relative quantification, or the trend of the proportion of mutated vs. wild-type copies. Digital PCR, on the other hand, is based off of absolute quantification. The comparison then could only be made with qPCR which may take the reader away from the purpose of the study which is to compare qPCR and digital PCR with the emphasis that digital PCR has improved sensitivity. When researching the index more, I found that the validity of the index results is muddled. While some such as González de Aledo-Castillo et al. (2021) and Marchetti et al (2015) found a significant correlation between SQI, VAF, and copy numbers, others such as Macías et al. (2019) did not. Since the index has not been concluded to be interchangeable with VAF and copy numbers, SQI is interesting but not necessary for this study.

Changes in the text: N/A

Minor comments :
- Lane 446 : NGS does not have
- Legend to Figure 2 : dPCR (not DPCR)

Reply: The appropriate changes were made to the text.

Changes in the text: N/A

Reviewer C
The Authors have compared the performance of different ctDNA detection kit in Lung cancer. The study addresses a pertinent issue in the field, especially when it comes to disease surveillance. There are however somemajor concerns.
Major comments

1) How was the genotype of the patients determined before sampling, because the cause mention that having a pT790M mutation was an inclusion criteria.

Reply: Thank you for your comment. Genotyping is not needed for cancer patients. Because we are interested in somatic mutations and genotyping is typically conducted for germline mutations, it would not make sense in this context to conduct genotyping.

The need of the p.T790M mutation in the inclusion criterion should not have been included and has since been removed. The mutation was screened for before sampling. The PCR assay analysis and tissue biopsy results were analyzed after collection. To be eligible, the patients must have been tested for the mutation as patients that were ineligible were removed due to one of the assay tests failing. Patients were not ineligible because they did not have the mutation.

Changes in the text: N/A

2) at what time point were the samples collected? Before systemic therapy or at disease relapse? This will affect the reported ORR and DCR if sampling was not synchronized. For example, a patient sampled at progression with higher abundance of mutant clones will tend to responfd better than a patient at baseline with few mutant clones.

Reply: We appreciate your concerns regarding sample collection. All samples were collected from patients who were clinically resistant to first generation TKI therapy and had not received third generation TKI therapy. The latter point was not stated explicitly in the manuscript and has since been added for clarification. These patients were selected because the presence of *EGFR* p.T790M in patients with acquired resistance would make them eligible for targeted therapy.

In regards to ORR and DCR, patients did not respond equally to therapy. This

is partly due to the total number of mutations but there are other factors at play. When looking at Table 5, the ORR and DCR of patient samples found positive with ARMS-PCR and digital PCR (which included samples that were dPCR positive and ARMS-PCR positive and samples that were dPCR positive and ARMS-PCR negative) were comparable. This implies that the digital PCR and ARMS-PCR were comparable in detecting patient samples that are truly positive for the mutation. When looking at the ORR and DCR of patient samples that were dPCR positive and ARMS-PCR negative, these samples had comparative results with the dPCR positive and ARMS-PCR positive ORR and DCR. This is notable as this indicates that these samples, with low allele frequencies, were not false positives, since these patients benefitted at a comparable degree from target therapy.

Changes in the text: N/A

3) The authors should include data demonstrating the EGFR mutation profile in the course of treatment and compare the different kits. This will the most important strenght of the study.

Reply 3-5: Comments 3-5 discuss monitoring patients so they will all be addressed in this reply. We appreciate your suggestion; however, for the goals and purposes of this study, continual monitoring of patients is irrelevant. The goal of the study was to compare qPCR and digital PCR with the hypothesis that digital PCR has better sensitivity than qPCR.

In a previously published paper by Zhou et al. (2018), HER2 amplification was monitored during the course of trastuzumab treatment using digital PCR for a patient with stage IV cancer. HER2 amplification correlated with partial response to treatment (sharp decrease in HER2 amplification), stable disease (consistent, low HER2 amplification), and progressive disease (increase in HER2 amplification). These results support the idea that continual collection of plasma throughout treatment would correspond to disease progression. However, for the purposes of this study, continual monitoring is not necessary.

Changes in the text: N/A

4) It will be interesting to compare the different kits at baseline using bóth tumor tissue and corresponding plasma and follow the EGFR mutation profile during treatment.

Reply: See the reply for Comment 3.

Changes in the text: N/A

5) it will be important to compare the EGFR mutation profile with CT-reported response and associated tumor dynamics

Reply: See the reply for Comment 3.

Changes in the text: N/A

Minor
1) There are some typos and language issues to be corrected e.g. lines 122 and 134, just to name a few.

Reply: The appropriate changes were made to the text.

Changes in the text: N/A

2) The Authors concentrated their cfDNA to 6 μl. They however mention that 6,53μ was used for the dPCR. How is this possible?

Reply: We thank you for catching this error; appropriate changes were made to the text.

Changes in the text: Methods section, Sample collection and DNA purification, Lines: 229-231.

Each sample was eluted with 100 µL buffer AVE and concentrated to 20µL using the DNA Clean&Concentrator-5 kit (Zymo Research, Irvine, CA).