Liquid Biopsies in Oncology Drug and Device Development Part 2

Transcript: Session IV, Liquid Biopsy Test Development

Reena Philip: ... The test development. So last year when we had the liquid biopsy workshop, we presented a cast study and we had a very, very great session actually last year and I'm hoping this will be similar to last year. But it is slightly different this year. First, we have the regulatory landscape, current landscape. That will be

[00:00:30] presented by Dr. Eunice Lee. She is the brand chief in CDRH, Office of In Vitro Diagnostics.

Then we have Dr. Carl Barrett from AstraZeneca. He is the Vice-President of Translational Science in the Oncology Innovative Medicines Division at AstraZeneca. He's going to talk about examination of analytical impacting

[00:01:00] concordance of plasma versus tumor testing by next generation sequencing. When can the liquid biopsy replace the tumor biopsy? That's going to be a really exciting talk as you will see from his slides.

In a continuation of that, whether we can replace a tumor biopsy, we need to have reference samples and that's presented by Dr. Mickey Williams from NCI. He directs the Molecular Characterization Lab, Division of Cancer Treatment and

[00:01:30] Diagnostics at NCI. Mickey is going to talk about the pre-competitive effort to deliver reference control materials for CT DNA assays.

And then the last talk is by Dr. Meijuan Li. When we were in the planning sessions of this workshop, we were thinking about we should talk about early detection using CT DNA assays, so I'm happy to have Dr. Meijuan Li. She is going to present in a statistical study the signs for early detection.

[00:02:00] As we did in the last two sessions, please hold off your questions until the last after all the four speakers are done. We will have the panel discussion as well as the floor will be open for Q&A. With that, I invite Dr. Eunice Lee.

Reena Philip: ... landscape for really good biopsy diagnostics for oncology applications.

Eunice Lee: Thank you.

So today I'm going to provide a regulatory update for liquid biopsy tests in oncology.

[00:00:30] So as we all know, biomarkers and the tests that are used to detect them, can play a significant role in the management of patients in oncology. Biomarker testing can be used at many different stages for many different reason as shown here, and as was discussed in the last session. Also, as we've heard throughout the day, liquid biopsies are really an attractive alternative to tumor tissue testing,

[00:01:00] for some of the reasons listed here, which I won't go into, as it was discussed in the last session.

In the first session today, it was discussed about the different types of analytes that can be detected in liquid biopsy samples which include circulating, sulfur nucleic acid, circulating tumor cells, and exosomes. And for the purposes of my talk today, I'll really focus on regulatory considerations for test validation, and for
[00:01:30] tests that are intended to detect circulating sulfur nucleic acids for treatment selection purposes, and so their use is really companion diagnostics.

So, just a little bit of background. Companion diagnostic was formerly defined in The Final Guidance Document that was issued in 2014. It's defined as a test that provides information that's essential for the safe and effective use of a corresponding therapeutic. To date, there are over 40 in vitro companion
[00:02:00] diagnostic approvals that are paired with a corresponding therapeutic and among those, there are about 30 unique tests that were approved, that have been approved. And one of those tests is an approved liquid biopsy companion diagnostic.

So that approved test is the Cobas EGFR Mutation Test V2, which is used with plasma specimens. Last year, there were two approvals for this test. The first
[00:02:30] approval was in June, for the detection of EGFR XM19 deletions and L58R mutations for the selection of patients for treatment with TARCEVA or ERLOTNIB. A few months later, the test was approved for a second indication which was the detection of EGFR T790M Mutation for selection of patients for treatment with OSIMERTINIB.

So, the Cobas approval was discussed last year as a case study. So I won't discuss
[00:03:00] the details about the approvals and rather I will touch on some practical considerations for liquid biopsy tests, and then touch on how some of these were addressed in the Cobas test situation.

So, as with all molecular tests, really, test validation is considered for the entire test system. It's from specimen collection all the way through to result reporting.
[00:03:30] This includes the pre analytical steps. In the case of the Cobas test, indeed, specimen collection and specimen handling, validation studies, as well as specimen stability studies were provided.

Another consideration is that in the analytical validation studies, often times procured clinical specimens are utilized. It's important that the clinical specimens
[00:04:00] should represent the intended use, population for the test, and in the case with liquid biopsy tests, often clinical samples may be limiting. This is especially true for rare mutations.

In certain cases, and for certain types of studies, contrived samples may be used. And one example is with the Cobas EGFR test. Contrived samples were used in
[00:04:30] the study that was used to establish the limit of detection or the LOD of the test. Once the LOD was established, then clinical specimens were used in a confirmatory study to confirm the limit of detection. In this way, clinical specimens could still be used in the LOD study, but using fewer replicates as well as fewer dilution levels.

[00:05:00] Another consideration, is that currently there is a lack of reference methods and materials for these types of tests. This makes analytical accuracy very challenging. For the Cobas test, the sponsor used a validated and GS comparator method in the accuracy studies, and it was important that the sponsor provided some limited amount of validation for that NGS comparator method in order to support its use as a comparator.

[00:05:30] With respect to clinical performance, relevance to clinical outcomes should be demonstrated. For the Cobas test, the device was actually studied in the same pivotal clinical trial that was used to support the primary efficacy population that supported drug approval. In this case, patients were enrolled into the clinical trials using tumor testing. A type of bridging study had to be performed with the final plasma assay. In the case for T790M, it turned out that there was very

[00:06:00] limited clinical data. For patients who had a positive T790M result based on the plasma test, but who had a negative or unknown tumor status. So given this limited amount of data, this lead to a post market commitment where the sponsor is currently collecting additional clinical outcome data for this population.

[00:06:30] I've listed here a couple of ongoing challenges as they relate to liquid biopsy tests. First, to what extent can contrived samples be used in validation studies? Next, is it possible for liquid biopsy tests to obtain follow on claims to approved companion diagnostic tissue tests? And then finally, are there additional considerations for validation of liquid biopsy and GS panels?

[00:07:00] So, I'll try and touch on each one of these to share our current thinking. Since time is very limited, perhaps if there is additional interest in time, then we can explore these questions a little further during the panel discussion.

[00:07:30] First, with respect to contrived samples. Our general recommendation is that clinical specimens are used in the validation studies as much as possible. But really, at least they should be used in the key analytical studies to the extent possible. And what we consider as key analytical studies, really are accuracy, limit of detection and precision in reproducibility.

[00:08:00] As I mentioned in the case with the Cobas test, contrived samples may be used to supplement clinical specimens, but provided that there is an appropriate functional characterization study. So this study is important in order to demonstrate that the performance of the contrived samples is similar or representative of the performance that you would see with clinical specimens. And so one potential study design for such a functional characterization study is shown here, where you can take clinical samples and contrived samples which have the same mutations and then evaluate them at four to five dilution levels between the limit of detection and the limit of blank or LOB of the assay.

[00:08:30] So what about follow on claims to tissue tests? This question is actually very tricky. In general, follow on companion diagnostics are expected to demonstrate

comparable level of performance to that observed for the original companion diagnostic test. Based on the information that we know for liquid biopsy tests,
[00:09:00] there are performance differences as opposed to tissue tests. For example, with respect to the sensitivity of the assays. It is ideal that there is clinical outcome data based the liquid biopsy test. In the example that I gave with the Cobas test, patients can be enrolled into the trial based on tissue testing, but then it's important to also collect the blood samples and then additional considerations
[00:09:30] are needed with respect to the study sample, the study population, as well as the timing the samples are obtained, relative to when the tumor samples are obtained for testing.

Finally, are there additional validation considerations for liquid biopsy and GS panels? At a very high level, I think that we recommend that all variants with clinical claims should be included in the validation studies. Based on the
[00:10:00] performance that is known and has been demonstrated for MGS tests. A representative variant approach for validation for single nucleotide variants as well as insertions and deletions is acceptable. However, this may not be acceptable for fusions and copy number variants. I think that if a sponsor is interested to use a representative approach for those variant types, then it's important that they provide data to show that that approach would actually be
[00:10:30] acceptable. So really, to show data that performance of detecting a particular fusion could be extrapolated to all of the fusions that are detected by the panel. So that kind of support would be needed.

In addition, the range of variant types that are detected by the test, should be
[00:10:51] included in the validation studies. And consideration should be given to the context of as well as the context of the genomic regions in which the variants arise.

[00:11:00] With that, I will just leave you here is a couple of resources to that companion diagnostic FDA guidance document as well as some FDA websites and I'll end there.

Thank you very much.

Reena Philip: He's going to talk about examination of analytical factors impacting concordance of plasma versus tumor testing by next generation sequencing.

Carl Barrett: Thank you. I'm going to address two topics; the one was the topic that was given to me. That is; when can liquid biopsy replace the tumor biopsy. The answer to that, you've already heard actually. The EGFR test has been used in this context
[00:00:30] with certain provisos where we have to identify the plasma positive patients and demonstrate their response.

I'm not going to dwell on that very much. I will use this as an example to point out one principle of how we validated next generation sequencing panel test. And I'll explain how we use this approach to sort of understand concordance
[00:01:00]

between plasma test and tissue testing. And that's where the majority of my talk will focus.

I've listed here my take home messages, just as a starting point. Plasma testing with a cobas assay you've already heard from Eunice, you've heard from [Posse 00:01:11], is already approved by the FDA for patients. I want to emphasize today is that we used ... We were very concerned about tumor heterogeneity, and that's obviously a big thing when you're talking about concordance between
[00:01:30] tissues and plasma, and we use, and I show you the data we use to illustrate that this test could pick up these EGFR T-79 mutations with a great deal of specificity, and we could confirm that when there was a discordance between the tumor and the plasma tests, this was not a technical artifact of the assay but rather, you know, the biological heterogeneity which I think everyone now accepts.

There are a lot of data that we have around next-gen sequencing tests for EFGR
[00:02:00] as well as other digital PCR assays. They're highly concordant with the FDA-approved cobas test, but they're not currently approved by the FDA but they are commonly used in clinical practice and perhaps we can discuss that.

But what I really want to focus on is in contrast to this high degree of concordance which you heard for AKT from Nick, you heard from Posse about
[00:02:30] EGFR, you heard from Scott about KRAS. We did a study where we compared four different next-generation sequencing tests using the same samples for each of the tests, and we looked at the concordance between the tumor and the plasma, but also the concordance between one plasma test and another. And what I'll show you is a very high degree of non-concordance in this test, in this evaluation, and then I'll go into some details about some of the analytical factors
[00:03:00] that underlie this and [inaudible 00:02:58] with some suggestions for how to go forward to deal with what we think are some potential issues and pitfalls of NGS testing.

Now again, I don't need to describe this, this has been well-documented in the course of the day. We all know that most drugs are approved on tumor testing, and we consider tumors the gold standard. I might raise some issues about whether that's the right approach or not. We know that there are two NGS
[00:03:30] panels approved by the FDA for tumor-based testing. And we also know that tumors are heterogeneous, so we expect discordance between a tumor and a plasma tissue. Plasma's unit set can be readily obtained. It can be obtained at the same time as the patient is undergoing treatment. It's less invasive, as Scott and others have talked about. Eunice mentioned the cobas test for EGFR testing has been approved, and next-gen sequencing panels, as I have said, are commonly
[00:04:00] being used, but the question is how valid are they.

So first I'd just like to really briefly discuss how, the process by which we use to evaluate plasma C-DNA assays for EGFR mutations and how we deal with this issue of tumor heterogeneity and how we try to eliminate the possibility of technical artifacts. These were mostly PCR-based assays which we know have a

[00:04:30] certain degree of false-positive rates and how we are able to convince ourselves that these assays were picking up real positives.

This simply is a cartoon that illustrates the problem that I think everybody now recognizes, that a lung-cancer diagnosed with a driver mutation, an EGFR synthesizing mutation, can be readily detected and we can analyze those by a variety of techniques, but after first-generation EGFR inhibitors there is selection
[00:05:00] for resistant mutations, T-79M being about 60% of those mutations as Posse mentioned earlier. Obviously tumors will contain both mutant as well as non-mutant regions, and you can very easily biopsy a region that is positive or negative in this case. And therefore you have a hard time really knowing whether the test is the variable or the biology is the variable.

So very early on in this process where we were trying to decide upon whose
[00:05:30] assay we were going to use, we undertook an evaluation of four different platforms, all PCR-based platforms: the Roche cobas assay, the Inostics BEAMing assay, a Qiagen therascreen assay which has been used previously, and a digital PCR-based assay. We actually had, in the case of this lung cancer trial we, at time of treatment, we had biopsies of every patient that was treated with
[00:06:00] Osimertinib. So we had plasma and very nearly concurrent biopsies from the same tissues, so we avoid the issues of temporal differences between a biopsy taken years ahead. We also had the matched tissues and we were able to do this analysis. We eliminated, we picked the top two in an early study that we did, a
[00:06:30] preliminary study, and we did more detailed analysis of both the cobas test and the beaming test from Agnostics.

And what you can see here is that, again, we're looking at three different mutations, there are the two most common sensitizing mutations of L858R and the exon 19 deletion, and also the acquired mutation of T-79M. And you can see that both assays gave us a very high degree of concordance. When we looked at, then, the sensitivity and the specificity, as was mentioned before, sensitivity was
[00:07:00] quite high as it is for KRS at AKT, the specificity is quite high. The sensitivity is less than 100% because not all the patients shed DNA so therefore you're going to have a lower pickup rate of these mutations, but they are very comparable.

We got down to the T-79M mutation, which is heterogeneous, then obviously the sensitivity drops, but also the specificity drops because you will have
[00:07:30] heterogeneity of the tumors reducing the sensitivity, but also you will have tumor negatives but plasma positives, which decrease your specificity.

What we did to try to rule out that these differences were not technical artifacts is that we compared the patients that had the T-79M negative tumors with the T-79M positive plasma using these two technologies. And what you can see here
[00:08:00] is when we did this for 18 patient samples, 14 out of the 18 were picked up by both assays, convincing us that using this orthogonal assay approach could convince us that it really wasn't an artifact of one assay but rather the biology of the two. And we furthermore looked at a large number of non-small cell lung cancer patients that have KRAS mutations which are generally mutually exclusive

with EGFR and we didn't pick up any positives with either assay in that case. So
[00:08:30] based upon this, we picked the cobas test, because they're experienced in developing diagnostics and Eunice told you the rest of the story in terms of getting this approved by the FDA.

And Posse actually showed this graph as well which shows that the response rate on patients selected on plasma-based testing as well as tumor testing are essentially identical. And the assay was approved about this time last year.

[00:09:00] So now I'd like to segue from these digital PCR assays which perform quite well, to NGS panel testing, and what about for EGFR testing, and what about for other mutations at other genes as well. So as I said, we have a very high degree of concordance between NGS and the cobas assay, and other digital PCR assays like Biodesix, this data will actually be shown next week at the World Lung Cancer conference so I'm not going to be able to show it today. We did find, I think it
[00:09:30] was Posse mentioned, or somebody mentioned, that these tests are more sensitive. The cobas test is actually not that sensitive, so we pick up a lot of cobas negatives that are indeed positive. We think these are true positive because again we can confirm them by orthogonal assays. But now what about NGS panel testing, and what I'm going to describe to you is what we fondly call the AZ AstraZeneca Plasma Seek bake off assay.

And so what I'd like to talk about is the issues with panel testing, that not all
[00:10:00] assays, not all mutations are the same, and what is the gold standard? Is it really tumor testing?

In part this was stimulated by our own observations, but also by a publication last year from [Kudero 00:10:13] et al. in JAMA Oncology, and they presented a rather disturbing observation, where they did actually, they looked at a number of patients, I think it was 14, it was a small number. But they had data from
[00:10:30] Guardant Health plasma tests and they had Foundation Medicine tests at the same time. And what they reported that of the mutations picked up in either test, only 22% of them were concordant, but more disturbingly, when they tried to make clinical decisions based upon the actual mutations that were reported, they only came to the same conclusion one out of four times. So this obviously creates a major issue for clinicians trying to use these tests to make clinical decisions for patients, late stage patients.

[00:11:00] And there are other studies that have reported similar discordant levels as well. So the question, when I talk to the companies doing this, they say "oh, it's just tumor heterogeneity." Maybe. But is there also some technical reasons for that as well, and this is what we attempted to sort out in our study.

So our study consisted of looking at a comparison of four commercial vendors,
[00:11:30] we know them as A, B, C, and D, so we're not naming the names. For two reasons: one is that one of them had a contractual obligation we not do so, but also the purpose of the study was not to validate any one company, but rather to

validate the robustness of this test across companies, and see what we can learn from that process. So I think the names can remain anonymous in this case.

[00:12:00] So we have 24 samples we purchased commercially from different vendors, they're along the variant of breast primarily, a few prostates. They were selected, and the criteria is that we had both tumor tissue and normal adjacent tissue for each of these, and we had enough plasma to send each of these companies an aliquot of two mils of plasma which contained enough DNA for
[00:12:30] them to run their assay according to their criteria. Each vendor extracted the DNA themselves, and they got varying amounts based upon different results, but pretty much similar amounts. They returned all the data to us, including their calls, which I'm going to show you about the mutation calls, but also the QC data, the raw data, the band files. This was a major accomplishment in my mind because all the companies will tell you they cannot give you the band files for
[00:13:00] proprietary reasons, so we said "okay, fine, either you provide it or you don't play." So they all agreed to that in their final reports.

Now I'm going to show you the data. T's a little bit complicated when you're looking at lots of mutations across four different assays. So let me describe to you what our definitions are, because you're going to have to understand this to understand the data. So the things that we consider true positives, that is our
[00:13:30] concordance, are anything that's positive in a tumor and positive in a plasma. That makes sense. But realizing that tumor heterogeneity is an issue, if it was positive in two plasma tests out of four, we called that a true positive. You can argue about that, but that was our predefined definition. If you accept that, then the false positive is a single variant reported in only one assay, that was not in the tumor, and not in any other vendor's assays. And a false negative was not
[00:14:00] reported by one vendor, but was detected in the tissue in at least one other vendor's results, so this has to be a shedder by definition if that's the case.

So here are the data. It's a kind of funny looking graph but let me walk you through it, it's not that hard to understand. So, first let me point out this group in yellow here, which are all germ line mutations. We know that because we have
[00:14:30] the normal tissue, so we know exactly which are germ line or not. They were reported as somatic by the vendors, so in fact it was misreported. One vendor didn't report it, so that was fine. But given the fact that these are easy to call, quite easily recognizable, we didn't knock anybody for that. So we just took that out of further consideration.

Then you can look in ... the graph represents the alele fractions, and this line
[00:15:00] across the bottom is 1% alele fraction, which will become obvious why we chose that shortly. If you look then at the true positives, which are the greens in all of these four columns, you see the greens range from very high alele fractions down to well below 1%, around 0.2% in many of these assays. If you look at the
[00:15:30] false positives, however, the majority of them are below 1%, and the same can be true for the false negatives as well. Now note that vendor B has fewer dots because it has fewer genes, so we didn't knock anyone that didn't have the gene

on their panel. And vendor C had only one false negative, and I'll come back to this in a moment.

So here are the summery, and then I'll go through some real data with you so
[00:16:00] you can understand the conclusions. Germ line mutations were reported differently, so we excluded them as I said. Most of the discordant calls were observed at low alele fraction in the plasma, less than 1%. However, in this study, 60% of all the true positives were also less than 1%. So you can't simply say, "we're going to make that cut off and ignore everything else." Interestingly, many of the low-alele false positives in these studies were what we called novel,
[00:16:30] because you look at the databases and you don't find any other reports of these markers. They also have strange sequence changes that I'll show you in a moment. But this, again, supports the fact that these are indeed, perhaps, false positives.

There are two major categories of discordant calls. One is biologic, that is, tumor heterogeneity, in case the tumor result is a false negative and the plasma is a true positive, and I'll show you some examples of that, or the tumor is a false
[00:17:00] positive, and I'll show you one example of that as well. But the real interesting thing which hasn't been reported before is some of the technical, analytical changes in the calls.

So here is the summary of the sensitivity and positive predictive value. We don't know what truth is so we can't really do sensitivity and specificity, but the positive predictive value which is the true positive so the true positives plus false
[00:17:30] negatives, it gives you a sense of that. So you can see the sensitivity for the four vendors, overall, varies a lot, from 40% to 95%. The positive predictive value also varies. But then when you look at it by alele fraction you can see that the sensitivity drops for all but one vendor, and the positive predictive value drops as well, suggesting that herein lies the rub.

[00:18:00] Alright, so I'll show you some real data. This is the interesting part. So here is a false negative, PIK3CA mutation, an A to G, not called by vendor B but called by vendor C. And I hope you can read this, that if you look then at vendor C, there were 5 reads out of 2000 reads, or a 0.2% allele fraction. Vendor C had five,
[00:18:30] vendor B had four. Vendor B did not call this, but vendor C did. So basically the data are the same, but the bioinformatics cutoffs are obviously set differently by these two companies.

Now here is a false negative in TP-53 an A to G missed by vendor D, and here the interesting thing is if you look here at vendor C, you can see that there are 6
[00:19:00] reads out of 1400 or 0.4% alele fraction used to make the call. Vendor D sequenced at a depth of 42,000. So deep sequencing is good, but it also creates problems. You can see here the noise. So in this case, the number of gene calls is 43, but you also get 12 Cs, 18As, and 445 "we don't know what they are" base calls. So, a lot of noise created by this deep sequencing. So bioinformatics is key.

[00:19:30] Another case, a false negative in P-53, missed by vendor A as opposed to vendor D, and you can see here the same thing, 43,000 reads, and a lot of noise, with 28 versus 9 versus 24 reads, where as vendor C has a very clean 2000 overall reads and 19 calls of that base. So sequencing quality varies a lot.

[00:20:00] Here is another case of a false negative, due to poor sequencing reads. So here, vendor A did not call it, there are four Gs, but two Cs and two Ts, where as there are just 6 Gs, the right call from vendor C.

[00:20:30] Now the other interesting thing is that if you look then at the novel mutation calls of these false positives as said before that in many cases are not previously reported, and the fraction of T to A and A to T mutations is disproportionally high which is a red flag for sequencing.

[00:21:00] This is another example. This is a triplicate measure of the same vendor, where we sent them seven samples, and looked at how reproducible the samples were. And you see only two out of the seven samples gave exactly the same assay, and the other ones showed a lot of variability in at least some of the reads with each of these samples. So the reproducibility of these low allele fractions can be quite low as well.

[00:21:30] Now this is the vendor C false negative, which turned out to be a BRCA2 mutation, it's somatic, not germ line, it's at a prevalence of 17% allele fraction, so presumably they thought this might be a germ line sequence even through it's not 50%, and they decided therefore not to call it, but it's clearly identified in their data.

[00:22:00] Just a couple of quick examples of tumor heterogeneity here is a case of a discordant EGFR variant, which is not seen in the tumor, but seen in two different vendors, suggesting that the tumor is wrong but the plasma is right. Clean data. And in this case, there's a [inaudible 00:21:57] where you do see it in the tumor but it's at less than 0.5% allele fraction which we would never call, based upon the tumor. But clearly the plasma actually confirms this low call within the tumor, interestingly.

And in this case, is a total sequencing artifact of an insertion at a repeat sequence in mTOR that was called by the tumor but also by the plasma as well. So again, you can get misdirected by some of the calls as well.

[00:22:30] So just to conclude then, we believe that the plasma DNA Seek enables patient selection for well-qualified markers like EGFR, AKT, KRAS. The panel tests are variable in their performance. Most are reliable at allele fractions greater than 1%. Our best practices recommend that you should be cautious at allele fractions less than 1%. You should look at the data. Data is important. When suspicious of

[00:23:00] low allele frequency variance, consider mutation bias, whether or not you've got A to Ts, and whether it's a novel variant. Seems like there should be some policy on reporting germ line mutations. Each of these companies seem to have their own view of that, and I would think that, the BRCA2 mutation would be eligible

for a PARP inhibitor, but it wasn't called for this, so it was actually a missed opportunity.

[00:23:30] Sequencing a lot of normal DNA is important and you will find artifacts, and most companies have their blacklist of genes they take off or they just don't report, and I think as someone said in the last session, you should be careful to base your clinical decisions on the clonal events, and not these rare, potentially artifactual mutations. You can enhance the confidence by doing reproducibility, looking at the data, looking at the sequencing quality, checking the publicly

[00:24:00] available databases, and looking at mutational bias. Of course we need more data in each of these areas. I'm out of time so I'll skip this. But I do want to acknowledge Dan and Brian and their team at AstraZeneca who did these studies. Thank you.

Reena Philip: Going to talk about pre-competitive deferred to deliver reference materials for ctDNA assays.

Mickey Williams: Well I'd like to begin with thanking the organizers for the opportunity to speak to you today. I'm also going to thank Carl a couple of times because I think he set us

[00:00:30] up very well for this project that we've got going. So, I'm here speaking on behalf of the foundation, NIH, which is a non-profit organization that actually has really made some major and impressive advances, I think, in the area of biology, and in my interest oncology. We're working with them and have been for about a year now and a lot of it was done underground. But as of about a month ago, we got

[00:01:00] the official approval from the executive committee to make this an actual project, so we're moving forward.

I think everybody is here because we see the potential of ctDNA and liquid biopsies. As we've already heard, it's being used in clinical trials and also being used to actually manage patient treatment. We all know that everybody has their own favorite assay. There's commercial companies. There's academic

[00:01:30] institutions and their clinical labs. There's people doing research. Pharma companies, obviously have their assays and free agent manufacturers are beginning to give us panels now that we can use in laboratories for testing.

There's a big problem in this phase where we're looking for the needle in the haystack. Good clinical specimens are hard to get a handle on, or get a hand on.

[00:02:00] And therefore to be able to do reproducibility studies, to do platform comparisons and the like ... It's very difficult unless you're fortunate enough to have well-banked material to do this. And as Carl was showing you, it takes a lot of effort to do these cross platform studies and therefore, it's obvious that I think we need some sort of reference material to move into this phase. The project that I'm going to talk about ... It's endgame is to deliver reference materials.

[00:02:30] We're going to call them reference samples until we actually demonstrate the performance and can make the claims that we need to. And the one thing that I think I'm concerned about, and probably others too, is if we don't these reference materials into people's hands soon, there's a potential that we could

do as much to the field as we do good when studies start to conflict with one another because we don't what the accuracy of the assays were.

[00:03:00] So, again, thank you Carl. You set us up wonderfully. I think this is a really good take home message. Right now the field is growing. We've got really interesting data coming out and it's exciting, but we need to have faith and trust in the assay's performance. So we've got a working intended use for these reference materials right now and I just wanted to go over that because we aren't going to claim to be everything. But we do want to have material that is demonstrated to
[00:03:30] be comparable in its performance to real, circulating tumor DNA specimens derived from patients as well as normal donors. We believe that if we can demonstrate the commutability that this will be an aid in the establishment of performance characteristics of ctDNA assays. In other words, they could help with testing sensitivity specificity and reproducibility.

I would say that it would never take the place of a real clinical specimen, but it
[00:04:00] could certainly aid in these activities. And we hope, also, that they could be used for testing an assay's performance overtime. Just to be certain the assays is performance reagent lot to reagent lot. Operator to operator. And if we're really lucky, they could potentially be used for proficiency testing at some point in time. Again, based on the fact that we would need to be able to demonstrate
[00:04:30] that they behave similarly as does a clinical sample. This is a working intended use and obviously we're going to be having discussions with the FDA and anybody else that will talk to us to challenge this and be certain that we deliver data that will suggest that these materials do behave the way that we think they should.

When we got together, we decided that couldn't handle the entire process so as
[00:05:00] far as the upfront collection, and pre-analytic processes, there's data coming out now and also ASKO and CAP and AMP are working together to deliver a white paper on this topic. We decided to take it from reference materials ... Actually controlling from the level of extraction all the way through the assay and reporting of results. That's really where we're intending to focus on. The
[00:05:30] materials that should come out of this will all be, either in synthetic plasma or in human plasma, such that they can be extracted prior to actually going in to the assay. The other thing that we had a lot of discussion about is reference materials. What do we need to demonstrate that these are actually fit for purpose. So we're focusing on what I call analytical validity, and I don't know if
[00:06:00] that's the appropriate term, but it's the one that we're using. And really important, we need to demonstrate that these are stable reagents. That they're accurate and they do reflect a clinical specimen and that they're commutable. And that's what we're intending to do when we go through the process of testing for these materials. At the end I'll talk very briefly about a clinical pilot that would be the final stages of what we intend to deliver.

So here's a working list of what we've got for the different variants that would be included in this panel. There's a total of twenty variant listed here. The ones in
[00:06:30] yellow are the ones that we're considering critical. Right now we're very

fortunate. This is a pre-competitive effort and in the public domain we've got three commercial companies that are actually active in reference material space that are at the table with us. They are Horizon, Sara Care and Thermo Fisher, and we've asked that the three of them include the thirteen minimal number of
[00:07:00] variants. Their panels will all probably be larger than this, but at least they'll have these thirteen variants in common.

What we set out to do with this is we went through several rounds of discussions and we tried to pick variants that would be clinically relevant, that would represent different types of variant, meaning SNVs, insertion [inaudible 00:07:18]. We've got some complex variants that we believe could be difficult for
[00:07:30] some assays to test and we also wanted to include copy number variants, as well as gene fusions. This is a small panel. We don't think it's going to inclusive of everybody's favorite variant, but it's a starting point and we hope that as we go through this, we'll build a roadmap that other variant can be added as needed for particular types of application. But this would be a good, broad testing panel that could, again, go across different tissue types in oncology. If you don't see
[00:08:00] your favorite variant, please come and see us. We're all ears. It might not go into the first version of this, but obviously we're interested in, you know, what variants that would be critical that might be missing.

We will have three reference laboratories. The molecular characterization laboratory at Frederick National Labs for Cancer research will be one. [inaudible 00:08:20] ct at the Harvard will be another one of the reference labs, and Ken
[00:08:30] Cole from NIST will be providing reference assays.

The assays that we'll take in to doing the initial testing will all be validated assays. We'll have at least two next generation sequencing panels. Hopefully we'll have one that is an amplicon-based, another that would be probe capture-based, and then obviously we'll be using droplet digital PCRs, another assay platform. So again, I want to make the point that the reference assays will be analytically validated and characterized.

[00:09:00] The commutability is actually, we realize, going to be the toughest aspect of this, and it's probably the most critical aspect, to be able to demonstrate that these materials would behave as a clinical specimen would. So we're going to initially start off by using a serial dilution approach described in a CDC manuscript, where take clinical specimens ... Take serial dilutions of the clinical specimens, and then
[00:09:30] mimic those serial dilutions with reference materials to show that we get linear regression that is comparable between the two materials. There's one area that I don't want to get into, but it's one that's obviously on our radar screen, which is the coefficient of conversion. How well does this material mimic fragmented DNA from plasma? Something that we're aware of and that we'll be paying attention to as we go through these commutability studies.

[00:10:00] So, the plant experiment that we have right now, and again, we will be discussing with the FDA. We thought we were going to have a pre-sub, but we're talking with them. We do have folks that attend our working group meeting from the

FDA and we want to be certain we're taking input from them. We want to be certain that we're delivering a product at the end of the day that has taken in
[00:10:30] their concerns and input, so that it is a product that would be used by laboratories. So we'll have the three manufactures providing their own reagent sets. Again we'll have the two NGS assays and droplet digital PCR, run in three different reference labs. And we'll be testing the actual materials for the sensitivity of detection on each of the platforms. The specificity, a very critical
[00:11:00] thing that they'll be a wild-type diluent that serial dilutions will be made into. We want to be certain that, that wild-type diluent actually serves as a good specificity control, too.

We'll be looking at, again, the reproducibility and linearity of these materials. The next part of this testing will be in the commutability. And again I've described how we're intending to do that. We'll be running lots of replicate, we
[00:11:30] intend to work in a real clinically relevant ... Concentration somewhere twenty-five to fifteen anagrams totaled, nucleic acid DNA input and variant allele frequencies, starting around 5 percent and going down to at least .1 percent. Some of the folks that we've talked to, that run service ctDNA assays for clinical use, have told us to go below this so that we can really break the assays. So we'll certainly do what we can to provide a good allele frequency working range to test.

[00:12:00] This is just another way of looking at the actual experiments that we intend to do, and again, this is a drawing board snapshot right now. This will be modified as we get deeper into the project. But certainly in the phase one, we intend to show that these materials are valid by running the replicate and serial dilutions
[00:12:30] with at least three assays. Then we'll move into commutability. Then the phase two will be the clinical testing. And I think I have another slide, here, in a second that will get into that ... Yes. So again in the clinical testing, we intend to try to get at least ten different labs in a mix of commercial labs that are offering ctDNA test, as well as academic clinical centers that are developing ctDNA test. And give
[00:13:00] them blinded specimens that they would test and then very much like what Carl and AstraZeneca did with real clinical specimens. To pull that data in and look to see how well every body is doing. So that would be the endgame of this, but along the way we hope to demonstrate that these really do have the properties that they needed to be used for reference materials.

This is just for anybody that's interested, kind of the timeline as we mapped it out and presented to the executive committee. The first phase, is obviously
[00:13:30] getting the assays up to snuf and demonstrating that they're analytically ready to go. The next phase will be testing the materials and doing the commutability. And then the last phase, which comes into the second year shown here in yellow, is the actual clinical testing that would be done. Now I can tell you the one thing that the executive committee at the FNIH told us when we were presenting this, that they would fire us. That we weren't going fast enough. And we hear.
[00:14:00] Obviously we're going to do everything that we can to accelerate this. We've got a bunch of motivated laboratories that have already stepped up to the plate that said they're interested in working on this. So really, we're planning on jump

starting this very soon. Probably in the beginning of next year with testing of materials.

And then I'd just like to finish up by thanking the people that are part of the working group here. Bob McCormick, who's not here today, is really I would
[00:14:30] consider the major driver. He keeps us all marching and moving forward. There's Ken Cole from NIST, Chris Karlovich from the Frederick Labs, Cloud ... We've got Carl ... Actually he's very motivational when he brings data to the meetings like he just showed you. We've also got Jeff Oxnard. And then we've got Stacey Adam
[00:15:00] from the FNIH, Dana Connor, who I'd like to acknowledge. Then Susan Keating, Caroline Seygman and Linda Dudee from CCA Associates. They've beem very, very instrumental in getting this moving. It's an official project now. We'd be happy to talk to anybody that might be interested. Anybody that has comments because again, we want to deliver something that will be useful for people that would be in this phase in the clinical laboratory. With that, I think we have one more.

Reena Philip: Last talk for this session is from Dr. Meijuan Li about statistical studied science for an early detection claim.

Meijuan Li: So first I want to thank the organizer for the opportunity. I'm glad I'm not the
[00:00:30] only statistician in the audience. But I promise I will not talk about the p-value or statistical formula. Just the general concept for study design consideration for early detection claim call validation and there are some unique challenge for liquid biopsy. For if the one varies a diagnostic device, we know that the same
[00:01:00] technology could have different indication for users or intended users statement. That's we vary the assay. We actually look at the combination of the IUs, IFUs such as, for example, you could have diagnosing or risk prediction monitor and treatment, prediction, stuff like that. Because even if you have same exact test that because the IU is different, the study design may be different, too.

[00:01:30] We look at the combination of the IU and the IFU and also physical product system and reagent or salt, stuff like that.

I'm from, so we talk a lot of validation. I think said, they did not specifically say validation, but we talk a lot about validation. In my talk I use validation a lot.

A validation study for diagnostic device is not just one study. It's many, many
[00:02:00] different studies. So for example, you have pre-clinical study, you have a non-clinical study, you have a clinical study. And I'm going to talk about clinical study. There are two different types of clinical study, depends on your intended users statement again. Called clinical outcome study. Which typically be used for company diagnostic devices. For this type of study, we actually look at the device
[00:02:30] output, whether they impact patient outcome. That's called a clinical outcome study. Another one is called a clinical performance study. This is more common.

For my talk today about early detection. Actually we typically use a clinical performance study. Some people said this is observation study. Actually it's not

exact a so-called observation study. Because actually investigated... do get some involvement... it's not merely observe the real life setting. We called... we have
[00:03:00] unique name for this type of study, so-called clinical performance study. For example, we want to look whether this device, if this device is used for lung cancer diagnosis. The objective of the study, actually to evaluate the device's sensitivity specificity. This is related to the clinical performance of this study. Of the device, I'm sorry.

Before you conduct a clinical study, we will look at the clinical study protocol.
[00:03:30] This study protocol should include a lot of different components. For example, you have to define clinical objective. Based on your RU and IFU. Align our statistical goals with your objective.

And there are lot of different components, just to point out.

In my talk today I will just focus on the study design. This is just one component
[00:04:00] of the study protocol we typically look at. OF course, for a given device, there are several different types of study designs you can use to validate a particular assay or device. The appropriate clinical study design depends on, again, its indication for IU or IFU or proposal objective and also depends on the device output.

[00:04:30] Even though you have same exact indication for users or the technology. For example, if your device output is quantitative versus qualitative. The study designs are actually different. It's not just your device the output. Actually, the comparative device output. For example your device output is qualitative, but
[00:05:00] the comparative actually has a quantitative device output. Then the study design is different from both have qualitative output.

In choice of study design is driven by a combination of scientific clinical studies also sometimes ethical considerations. For example, if you have a lot of negative patients. It's sometimes, it's harder to get, for example biopsy results for
[00:05:30] negative test results patient may not be ethical. That's why I said it's a combination of all kinds of information together.

Now I just want to say to basic users of non-statistical language. What do we consider in the study design? Two components. One is collect the data. What does that mean? You're supposed to have no bias in your sample. What is a bias?
[00:06:00] It's a deviation from the truth. So-called a systematic error from the truth. And the bias can be introduced from study's subject correction, selection, study design, study conduct, or data analysis procedure. In my talk assuming data analysis position we have no bias. We have to make sure there is no bias in the design stage, or the study conduct stage.

[00:06:30] There are many, many different types of bias. Even for a subject selection bias, you could have a spectrum bias, you could have verification bias, there are all kinds of biases. Due to limited, constraints on time, I will not talk about this today. Just so you know that there are many different kinds.

[00:07:00] I want to give an example because we often see in the study, in our view, and sponsor correct the data, just two very extreme. Control is a health donor patient for example, and the other patient is already diagnosed with cancer patient. That's not a correct study design because for example for early detection, do you really need to correct a patient that does not have a symptom?

[00:07:30] A normal health patient is just part of a patient of who do not have diseases, but it's not entirely non-diseased study population. Also, with the cancer already diagnosed patient is this not a non-symptomatic patient, they already diagnosed the patient. Make sure you do not in the study design... avoid those kind of biases.

[00:08:00] And the second major point in the study consideration is the correction to sufficient data. Sufficient data basically we talk about in statistical terms is called a sample-sized determination. A larger sample could provide more data so that your estimates has less calibrated. We said less calibrated meaning that your estimates are, the fitting is more tighter, is tighter, is more precise. Sometimes, for example, if you do not define your clinical significance values correctly, you could increase your sample size, for example, infinitely. You could lead a significant finding, but it's not clinically meaningful at all.

[00:08:30] You have to balance about those. In a whole, larger is large enough. You need a sufficient sample size for both non-disease and the non n-disease patient. So it's basically two sample size calculations. This is different from conventional like, a particular type of device.

[00:09:00] Now I just want to give an example. Here's an archived blood sample since the talk today is about the liquid biopsy. Just to use an example, because we haven't seen in the summation that sponsor asked whether that can use an archived sample, because it's harder to do, like a perspective where you collect a larger data set for the blood at the base of the assay. In order to do so, from a statistical point of view, basically you have to do, to avoid a so-called bias. Again, it's about the bias issue.

[00:09:30] And, of course, has covered a lot of good points from a scientific perspective, but here I just point out some additional points you have to consider in terms of bias. For example, sample has two list to represent intended user population. How to define a population based on disease spectrum, aging, gender, or other factors. It's factors defining your population. Then you hopefully your study population will cover those factors. Consistent to the device intended to be used by the population.

[00:10:00] Sample collector, you have to follow pre-specified protocols. Although you cannot perspectively collect a sample, but you could perspectively design the study. So we call the perspective design the perspectively conducted study. No selection bias again. Inclusion/exclusion criteria have to be pre-defined and fully described.

[00:10:30] I just want to say, last point, you have to mask the results. For example, your clinical truce is already a variable, is that right? When you do the test results by the new assay you have to mask results, information to the operator. For example, if the patient is a cancer patient. Hopefully not biased if you know the patient is cancer or not cancer.

[00:11:00] Now the next point I want to talk about. Collect sufficient data. This is basically a sample size calculation. What is a performance major? There are different types of performance major. Typically for diagnostic assay, it's come as a pair, not a one. So pair for example, sensitivity specifies if it's pair. I put a definition here, used non-statistical terms. You could have a possible predictive value pair and a negative predictive value pair. Of course, the other type of performance major

[00:11:30] depends on sometimes we use ppv and pv as a primary performance major. Other time we use the sensitivity, specificity. Depends indication for users.

When you calculate sample size, you have to think about desired confidence interval. How narrow you want the confidence interval you want to be. Actually it's joint hypothesis testing. It's not just one sensitivity, specificity, you have to do both together.

[00:12:00] Again, just emphasize the sample size calculation. Because there's two uncertainty associated with your performance major. One is the number of subject in the study. That's called sample calibrating. In order to have a size estimate, you need to minimize sample calibrating. That is related to how many samples you include and load into the study.

[00:12:30] Second again, is a bias. That's related to what type of subject you loaded into the study.

That's the two uncertainties.

And sometimes when you in the conduct of the study, you have a bias, it's very hard to quantify or to collect it during the analysis stage. So you really have to think about this during your design phase and conduct stage.

[00:13:00] Again sample size, I just emphasize what I just said. A sample size calculation is dreamed by your design, statistical uncertainty bias, and also, I just said sample calibrating. And you have to think about not a point estimate, but the lower-bound of your performance major. Lower-bound is 95% confidence interval lower-bound. Think about lower-bound, not a point estimator. And this coverage

[00:13:30] in high spectrum, in the entire population? Not like I said, not like a house donor versus disease population, but the entire spectrum. And coverage sufficient number of patient allowing the cutoff, we pay a lot of attention allowing the cutoff because that's called a challenger condition.

It's always easier to detect a high negative or high positive, but it's always harder to distinct, allowing the cutoff from positive to negative.

[00:14:00] Coverage of the range of different test conditions. Because diagnostic assay, not just accuracy, but the positioning also matters. By including different locations, you're basically encoding different test conditions. Hopefully this assay produces a variety of position will be incorporated into your study design.

[00:14:30] Now, what I already said, it's probably not unique to the liquid biopsy. So here are some unique question I have tossed here. I put them up, we do not have answer, we have to look at them case-by-case. Those issue are actually unique to the liquid biopsy. First off, for example, I think many speakers already gave good discussion about tissue-based results versus the liquid-based results. And there's a difference not that in the test results, but could it be due to biological

[00:15:00] differences? How we define the clinical truce for liquid biopsy assay. This is an analog to how do we define the clinical acceptance criteria?

[00:15:30] Then if that's discordant in how we address. What have we decided? Really this patient is truly positive or negative? Truly is a cancer or not a cancer? Should all the project subject to be receive the biopsy? I already said that, for example, it's special for early detection. This is a permanent, maybe law, then you end and a lot of tests with negative results. Is it ethical to have those patients all have a biopsy? And that's probably not ethical to have all the negative patients, but how much? Or if any? What should we do?

[00:16:00] And, now next question is difference between the somatic mutation and the germ-line mutation? Can you distinguish. Of course, can you distinguish the patient, the mutated DNA from normal patient from the patient from the tumor. That's another kind of question to be answered.

[00:16:30] How can we treat a patient based on liquid biopsy results? Can we truly trust the liquid biopsy test results? And, of course, from a logistical perspective, we could think about the test results, three different potential creams you could do. For example, loaning in, loading out, or reflexing assay. For example, if you really want to consider loaning in. What that means, is that you want to make sure that if a patient who is liquid biopsy tested positive is truly tissue positive. So you

[00:17:00] want to have very high ppv. For loaning out is the opposite. You exclude the lower patient. For example, the patient has a test negative liquid biopsy. You truly believe they are negative. So you do not miss any positive one. For reflex essay, for example, you could combine liquid biopsy assay with a tissue-based

[00:17:30] assay. One example, I'm just tossing, for example, I don't want any patient who I called test in the negative by liquid biopsy assay, I don't want this patient to, just by chance, is a false negative. So I want this patient to make sure it's truly negative. So then we have to test it by a tissue based assay. So that's another sort of type of a cream you could consider.

[00:18:00] I just have some left, two references, for you in case you are interested to read it. Thanks, again.

Reena Philip: Can I have all the speakers back

Reena Philip: Now that they are all coming here I'll ask a question to Mei-Juan because now you are all in the statistical mind right now. Mei-Juan, can you give maybe quick advice to this audience in terms of ... if someone wants to bring in a monitoring assay where there's early detection because that is something which came up throughout the sessions today ... that these assays could be used for monitoring. Maybe your quick thoughts on it?

[00:00:30]

MeiJuan Li: Sure. I think ... Can you hear me? Hello? Oh, It's kind of [inaudible 00:00:49]. Monitoring and early detection is different, actually. For monitoring, I can think of five different aspects that are different from early detection. First and fundamental is that for monitoring you have longitudinal data. You are not only measuring for each patient, not just one time point. But multiple points. That is fundamental difference from the early detection. Secondly is that [inaudible 00:01:21] you have to think about what is right follow-up time that have to be clinically relevant. Is six months okay, or one year okay, or three months is okay? Depends on device.

[00:01:00]

[00:01:30]

And time interval, how often? You have total of one month, three months follow up time better ... how often ... what's the time interval? That has to be clinically meaningful too. And fourth point is that you have test results. You also have clinical truths, alright. They are supposed to make, for example, you have five different points you measure for a given patient. Ideally, for each of those five time points, the clinical truths and tests results should be measured in the very tight time window. That way you can match the clinical truths to the test results.

[00:02:00]

The last thing from a statistical point of view, of course, because you have multiple data points per patient, you have to include all those data because your from seeing patient and then it's possible those data points are expected to be correlated. That's different from early detection. Thank you.

[00:02:30]

Reena Philip: Micky, it was really great to hear that you got the funding, perfect timing for this talk. I think I had two questions from the slides. One, have you fin ... it looks like you haven't finalized the list yet?

[00:03:00]

Mickey Williams: More or less, yes. But obviously we want to hear if there are things that are missing that are critical ...

Reena Philip: When do you plan to finalize?

Mickey Williams: I would say we are more or less finalized as far as what is going to be manufactured, but obviously this is not going to be a static scenario. If there things that are clearly there, I mean, we know we don't have any prostate relevant genes on this particular ...

Reena Philip: Do you have a timeline on the first phase that ...

[00:03:30]

Mickey Williams: Well, yes. Risking being fired, we were going to try to get this done within a two year period, but we're going to try to accelerate that and have materials that have been through all of the testing in the clinical pilot ready sometime within 18 months or less.

Reena Philip: The first quarter is anytime starting from ... when will you start the project?

Mickey Williams: [00:04:00] The project officially started a month ago. The FNIH is doing their fundraising right now, but we're obviously working on the assays that will serve as a reference material, or reference assays for the materials. We're hoping to have the materials finalized and in our hands by late winter, early spring, sometime early next year so we can begin the testing, yes.

Reena Philip: [00:04:30] Thank you. Eunice, I was thinking maybe you could actually talk about what contrived samples ... right now FDA uses in ... as a substitute for the clinical samples for some studies.

Eunice Lee: [00:05:00] I think ultimately the types of contrived samples that are used are really up to the sponsor provided that there is data to show that the performance of the contrived samples are really representative for clinical specimens. Some sample types that have been used are sonicated cell line DNA spiked into plasma from either healthy patients or healthy donors, or cancer patients, to try and mimic the background a little more closely. I think that when contrived samples are used, we do request that there is a justification or scientific rationale that is provided for how the samples are chosen; if sonicated DNA is used then there is [00:05:30] a description of how that is similar to what is found in clinical specimens. And another point is when cell lines are used, it's important that if the cell lines are used to develop the assay and used as a training set, then really that should not be used as a validation set, and that validation should be conducted with an independent sample set from the samples that are used to train.

[00:06:00]

Reena Philip: [00:06:30] Thank you. Carl, in Dr. Hayman's ... one of the slides, actually he said in this molecular testing alterations from CFDNA now it's like a reality and that's being done in the ... based on the data that you presented, what are your thoughts on how that's actually ... whether it's actually good or is it real time?

Carl Barrett: [00:07:00] It's really the $64,000 question and I suppose $64 million question ... We just want to call out some discrepancies and we've met with each of these companies, we've shared our data with them. They all kind of went "Oh, we can do better. Let's try again," so they are re-analyzing. I hope that will mean that they will make improvements in their pipelines and improvements in their processes. That's about all we can do. I think ... I was going to actually ask David what does he do if he gets a discordant call between a plasma and a tumor based test, how does he make a decision from a clinical perspective. I think that's really the question that is probably facing a lot of physicians right now in terms of how [00:07:30] they interpret these tests. My recommendation was you should look for the clonal events, those are the ... so if you start getting too sophisticated looking for

the rare mutations, which actually may be not real, then I think you're probably going to make the wrong decisions. But if you have a very clear clonal event, it's a known driver within that disease, then I think you're safe.

[00:08:00]

Reena Philip: I think I'll open it up for the audience now for the Q & A.

Speaker 1: My question is for Dr. Barrett. The whole NGS assay that you use ... could you subtly explain a little bit more about the differences that you see and then ... sort of which assay actually you think actually we need ... we actually ...would be better to use and the reason why.

[00:08:30]

Carl Barrett: I can't go into the details about the QC because this is all published by the vendors. They are all CLIA approved assays, in some cases New York state approved assays. I think they've gone through their standard validation, but this is non-standard, right. That is the question we have to think as a community about how is the best way to validate these tests. For us, the orthogonal

[00:09:00] confirmation of the data was very important, the reproducibility of the data. Obviously if you do a multi ... very expensive test of several thousand dollars and you get a low or little fraction, you think that's really important, you're probably not likely to reproduce that assay as we did, but I would if it was my family member.

Thomas Khan: Thomas Khan, IRB. We are running an IMI trial, Innovative Medicine Initiative

[00:09:30] called Cancer ID and this is a five-year program, which is now running for two and a half years and we spent about a year to establish SOPs for [inaudible 00:09:36]sample handling and we have three main work packages on ctDNA, RNA, and CTCs. This is a program based in Europe with more than 37 partners and there are also a few partners from North America. My question goes to you

[00:10:00] Dr. Williams, will it be possible to share samples so we establish a European bio bank for that ... especially for this purpose? And we are running ring trials with different samples on that ...

Mickey Williams: I think that would be very exciting. I'd love to follow up with you. We are going to be looking for clinical pilot sites. We haven't specifically discussed Europe, but I think that harmonisation is a good thing in this space and having materials that we're confident in and begin to address that would be quite nice.

Thomas Khan: Thanks.

[00:10:30]

Speaker 2: Question for the FDA members on the panels. We had a discussion today about quality control and performance issues with LDTs, the FDA spent a lot of time looking into this over the last three or four years, it's kind of gone silent. I wonder if anybody could kind of summarize where you are today and perhaps are there any reports coming out, guidelines? What's going on at the FDA about

[00:11:00]

regulations of FDA that we could use today to help with quality control performance?

Reena Philip: Very good question. Actually, I do not know the answer ... (laughs) ... other than that there was a paper that was published last November I think, that is the latest, and you guys probably know more than us because things come out in the press and so ... that's where it stands right now in terms of LDT.

[00:11:30]
Speaker 3: Could I ask Mickey a question? Mickey, main indication that we're using liquid biopsy today clinically is for quality not sufficient. And we're not going after because of pan cancer assay is just not proven clinical utility, we're trying to rule in or rule out only x number of driver genes, or truncal mutations, not these ... whatever the variants [inaudible 00:11:55]was picking up. If a lab validated their
[00:12:00] assay using classical tissue validation methodology and didn't try and make these artificial specimens, but had tissue time concordance with just addressing these clonal mutations, and that specificity was up in the Cobas level with a sensitivity 70 or 80%, would you think they've done the best validation there is? Because you won't have discordance on drivers.

[00:12:30]
Mickey Williams: It's a very fascinating point that you bring up and it's my hope that clonal drivers will be much easier to reproducibly detect, I just haven't seen a great data set to convince me that we're there yet.

Speaker 3: I mean that's what you basically did with the GF assays ... That's a clonal driver,
[00:13:00] now you almost had with a [inaudible 00:12:59]screen 100% specificity with about 70% sensitivity, which means you're not selecting because you're ... 70% of the patients I doubt you're selecting.

Mickey Williams: I've seen enough difference between assay analytical performance that I'm still a little bit worried, especially when we're in this needle in a haystack space where
[00:13:30] the variants are so low in the liquid biopsy space. I hope that's where we end up [crosstalk 00:13:31] Yes, I'm hoping it's less of a problem, but I need to see more data, personally.

Speaker 4: Two questions for Meijuan, would you consider a two-step early detection marker where the first step would be the liquid biopsy and the second step versus a tissue biopsy could be a more invasive, but earlier detection imaging or
[00:14:00] colonoscopy-type thing. Is that capable of putting into an FDA pipeline?

Meijuan Li: I think Reena is a better person to answer this question. It's [inaudible 00:14:14]device indication for yours, so it depends how you define your R U statement.

Speaker 4: Gotcha, okay ...

Meijuan Li: If you say this assay actually is a two-stage process ...

[00:14:30]
Speaker 4: It would fit the similar criteria, the end point would be the second step versus the biopsy [crosstalk 00:14:32] ...

Meijuan Li: I'm probably late, Reena answer this question about clinical performance acceptance criteria whether you should sort it based on second one or based on combination both first stage versus second stage because[inaudible 00:14:49] have both stage pass acceptance criteria.

[00:15:00]
Reena Philip: And it also depends on the pre-specified criteria, right, if you have already pre-specified it versus if you're not.

Meijuan Li: I just want to add imaging is not a final clinical truth although

Speaker 4: Right, correct. But imaging is more invasive as in radiation, so to have a diagnostic mark to say, "You need imaging," pancreatic cancer comes to mind,
[00:15:30] we can't image everyone that has a stomachache. But if you had a liquid biopsy test, it would say "you need some more imaging," that would be very helpful.

Meijuan Li: I do see that one of the potential use of liquid biopsy.

Speaker 4: And the second question for Eunice, does Mickey need FDA approval for his standards and what does that need to take?

[00:16:00]
Eunice Lee: That's a great question. I know you're planning to come in with a pre-submission to discuss the regulatory pathways well as validation. Recently, for controls and calibrators, they were classified as Class 2 devices, which required 510K
[00:16:30] submission and recently they have become exempt, so they are Class 2 exempt and subject to registration and listing.

Speaker 4: Thank you.

(laughs)

Speaker 5: I have a couple questions and the first is for Dr. Eunice Lee. You mentioned that
[00:17:00] often times in ctDNA [inaudible 00:16:55]diagnostics and referenced methods not available, in fact your example is Cobas for T790M. I believe you mentioned that they actually use partially validated NGS as a comparison or just to show the data, right, and show that's what it is to concordance. My question is we noticed
[00:17:30] NGS, this [inaudible 00:17:29] with NGS probably is moreso than what you can imagine for PCR. You use that as sort of a reference and comparison with how you interpret this result knowing that that method was only partially validated and not cleared or proofed by FDA?

[00:18:00]

Eunice Lee: That's a great point and I think just given the circumstances that there are no reference methods, there is no gold standard for these types of molecular assays, we're really doing the best we can and I think that right now when a sponsor chooses a comparator method we ask for some level of validation, which will include input study, some LOD study, some precision. And then, in
[00:18:30] order to have some level of confidence in the performance of that test that they are using as a comparator. And of course the performance for that comparator has to be for the mutations that they are using it for. You're right in that is not a true accuracy study because you don't know the truth, but it is a method comparison so it is a orthogonally validated method that they're using to show you you're getting answer A with test 1 and you're getting answer A with test 2
[00:19:00] as well. It just provides that additional level of assurance.

Speaker 5: Okay, thanks. My second question is for Dr. Barrett, and Carl. You summarized that discordant and [inaudible 00:19:20] methods are caused by two major costs
[00:19:30] and first is cancer heterogenic and the second is technical issue with each platform. I'm not sure you quantified that factor, for example did you know that the hemo-heterogeneity cost like 80% of discordant or only 50%?

Carl Barrett: That's a good question. I haven't actually calculated that, but the biggest
[00:20:00] problem was obviously the false positives and those ar, we believe, mostly due to technical artifacts. These are one-off examples for enrichment for T to C mutations that are novel mutations that you've never seen, so to me they all smell like they are sequencing artifacts and those are the biggest challenges.
[00:20:30] There are cases of heterogeneity, I showed you a few cases there, there are not that many in this case study. But this again ... I mean there has been obviously lots of reports in the literature.

Speaker 6: This is kind of directed at Carl and Mickey. Quick question, on the lack of
[00:21:00] concordance between ... you have the 10mLs of plasma and then you divided it into four tubes. If it was a rare variant, is it possible that only one tube got the rare variant? That would be one thing where technically if you had a 25% chance of detecting it, the three would be false negatives, but in fact they're actually truth, right because it just never got the blue ball in the mixture. That's one
[00:21:30] question around the false negatives, and then on the false positive side, I think some nucleotides are just a bit noisier than others. T790M is more noisy in general than L858R, so to Mickey I'd be curious in the variants that you've selected, and there's some kit mutations ... I mean there are just certain mutations whether it's the sequencing context, but they may have a 1% level of background, so if you're calling a .1 it's probably a false positive.

[00:22:00] It would be interesting to think about how the control strategy will look at those and then the final piece is ... I don't recall which mutation it was, but one of them had like 42,000 X coverage, and then the others had like 2,000 X coverage so I didn't understand if they each got an equal, or roughly equal amount of input, how was one so much higher than the others or did it just not know how many
[00:22:30] unique mole ... was not that unique coverage. And I assume that it wasn't and

that's why it probably had so much PCR and sequencing derived error associated with it.

Carl Barrett: All good questions. First let me come to the stochastic issue, certainly very valid. You saw in some cases calls were made on four reads, two in the one direction, two at the opposite direction. That seems to be common in this field that people make it with that level of call, which is not well understood by many people.

[00:23:00] Certainly that's subject to stochastic changes. That's why we did the triplicate study with the one vendor, perhaps it wasn't ideally[inaudible 00:23:10]because we focused on some of the false positives and I didn't have time to really go through that data. But, if you looked at it, you'd see that there are some of them that are very high ileal fractions and yet beyond what you would expect to be

[00:23:30] just sort of stochastic variations, and it ... there is a variability between vendors as well as even within one vendor with different replicas of the same samples in terms of the yields of DNA that is used for the input ... didn't seem to correlate surprisingly to me, so that was another thing.

The other question, obviously we took the data that we got from the companies,

[00:24:00] again we just simply took their calls and then we looked at the data they provided us from the band files. It's ... I think you're right, a lot of these are not de duplicative reads. We don't know what the re ... probably one or two vendors are and one or two were not. I think that is a big part of this, but the noise is there regardless of how you do it. That is still an issue, but certainly when people

[00:24:30] say they sequence 50,000 X, I don't know if that really means they are doing 50,000 unique reads, which is a lot more than 50,000 total reads. That's something again is not standardized or less reported in my mind.

Speaker 6: Yes, it's hard to call a negative if you're ... you don't know if you have one molecule 50,000 times ...

Reena Philip: Can you go up to the microphone please?

[00:25:00]

Carl Barrett: Yeah ... I got your point.

Mickey Williams: But again, they were using their pipelines as they had validated.

Carl Barrett: Exactly, that is the point. These were there calls, right. If they have 40,000 calls and they have all this noise, in some cases they made the right call and some cases they didn't make the right call even though you look at the sequences

[00:25:30] you'll see there is more than ample reads of the right call that was just noised out with the wrong call. This again is a lot of [inaudible 00:25:37]bioinformatics, which is no surprise to anybody does NGS, right?

Reena Philip: But these all four had New York state approval, right?

Carl Barrett: Sorry?

Reena Philip: All these four had New York state approval? Is that what you said? All the four platforms ...

Carl Barrett: Have?

Reena Philip: New York State Approval?

background : New York State approval ...

Carl Barrett: No, I don't know which ones. I know some of them are, but I don't know if all four of them.

Reena Philip: Okay.

[00:26:00]
Carl Barrett: But they are all CLIA approved.

Reena Philip: Thank you, so ...

Carl Barrett: Don't laugh Carrie ...

Reena Philip: I think since there are no other questions, we will wrap up a little early and I think I would like to take this opportunity to thank AACR for co-hosting a workshop with the FDA, thank you. And I also want to thank all the session
[00:26:30] chairs, we have been in many calls with AACR to plan this workshop to deliver a great workshop like the last one. I think we did. Thank you all. I just want to ... Julia Beaver, Gideon M. Blumenthal, Carlos, and Pasi, all the session chairs, and also thank the audience for staying til the end, almost. I guess ... I believe we also have a lot of people on the WebEx. Thank you all, we'll see you next time. Maybe
[00:27:00] another conference on this liquid biopsy part three.