Speaker 18: Couple of minutes?

Moderator: I think.

Reena: It's 1:00, right? It's 1:00. Welcome back after lunch. I see a lot of empty spaces and I'm hoping they will all be back soon. We are going to start the case studies, so as [Gidan Amensh 01:15:57] mentioned this morning, we FDA-approved the first liquid biopsy test as a companion diagnostic. It was approved right before [askel 01:16:07]. This is a reflex test for treatment with erlotinib for treatment for lung cancer patients. We are very excited to have all four partners, I mean FDA, [CD 01:16:20] and CDRH, and Genentech, and Rosch.

The order will be Dr. David [Sheams 01:16:28] from Genentech will go first. He is a principal scientist, department of oncology, biomarket development at Genentech, and that will be followed by Karen Byward. She is a master scientific reviewer in our division, CDRH, and that will be followed by Dr. Walter Coach from Rosch Molecular Systems. He's the vice president, head of global research at Rosch Molecular Systems, and followed by Dr. Erin Larkins. She is a clinical team leader who is responsible for the colAbrahamling approving this assay reflex test for erlotinibs. She's from Cedar. With that, I will get the mic to David Sheams.

David: Thanks again for the opportunity to talk about the work that we've done with our colleagues in Pleasanton on the development of the Cobos EGFR plasma test. My job is to provide some context for why we sought approval of this particular assay. To begin with that, I think many of you know there's been significant advances in lung cancer in terms of identifying patient subsets that respond to highly active targeted therapies, the first of which was Tarsiva, which was approved in 2013 for use in first-line metastatic lung cancer for patients whose tumors have EGFR mutations. The study that was used for that approval was called [Eurtech 01:18:21], showing the hazard ratio for PFS on the right side of the screen.

I think Geoff Oxnard explained more eloquently than I ever could the unmet need for why these tests are actually really important for patients, particularly in lung cancer. I think there are a couple of other elements to that unmet need, which I'll highlight briefly here. One is when we began thinking about ways to improve patient outcomes in lung cancer, we did some research around the world and found out that unlike breast cancer and colorectal cancer, where treatment rates are above 90-95%, in lung cancer, it's actually fairly surprising to note that patients diagnosed with metastatic lung cancer, only about approximately half of them ever receive any therapy whatsoever. In the EU, depending on the country, it can be higher or lower, and in developing nations or somewhat mostly developed nations, like China, that treatment rate is actually very low.

One reason for that happens to be that there isn't appropriate molecular testing and potentially a plasma test, which is much more accessible, or the substrate's much more accessible, perhaps that those numbers would improve.

Another element I think which is fairly unique to lung cancer as well is that tissue is very much an issue. I think many of you heard Fred Hirsch talk about that. It really is, in a

practical sense, where if you look at the right side of the slide, you see the NCCN guidelines for tissue testing in the US, and it starts with the diagnosis of histology, which isn't always very easy. Sometimes requires a number of IHC tests, which uses up slides, and then EGFR testing, which is recommended for patients with non-small cell lung cancer, and increasingly ALK testing as well. By that time, you've gone through ten slides, which may be all you get from a core needle biopsy. What do you do for the next molecular entities that are coming through, for example, PDL-1 testing, which may require additional tissue?

That particularly impacts the rarer mutations where you may well have active therapy, but it only makes sense to really test for the things that are most prevalent. Having a test, either singleplex or multiplex test, approved for detecting those different mutations would have significant impact potentially for patients.

With respect to the EGFR test, I remember when Tony [Mach 01:21:15] presented the initial data from the Fast Act 2 study at Asco a couple of years ago. I was sitting next to who was then one of the senior VPs in our company, and we looked at the outcome data for the plasma tests, and it really just struck both of us how incredible the outcomes were. What you're looking at here on the right side of the graph is the outcomes for PFS, for patients that received the combination of erlotinib and chemotherapy versus chemotherapy alone.

Then, as Lecia mentioned earlier in the meeting today, tissue was not mandatory for enrollment, so there was about half the patients on the study that didn't actually have any tissue available. That could have been for various reasons, but most likely is that there just was no block available. What you're looking at on the right side there are actually the outcomes in patients where there was no tissue. This is plasma test only, and you can see comparing the hazard rates that they're essentially identical. The point estimates for PFS are slightly different, perhaps indicating that patients without tissue perhaps weren't ... The tumors weren't accessible to biopsy. Maybe they had a lower performance status. To us, I think this really exemplified that we need a different way to test, otherwise these patients wouldn't be getting these active therapies at all.

There was a little bit of an unfortunate part of that. We had a very large, randomized study demonstrating the utility of this test, prospectively. The testing was actually done retrospectively, but was prospectively defined in the study. We had plasma from essentially all patients. There were two problems from a regulatory standpoint. The fast-act regimen was not approved in the US, and the company did not want to pursue registration of that for various reasons. The other issue that we talked about and commiserated about many times was that the test that was actually used for testing was a research prototype. Unfortunately, we had a beautiful data set but couldn't actually use it for regulatory filing.

We went to go talk to FDA about this, and fortunately, we did have a different test, a different trial, but it was already enrolled and already had outcomes in it, which was the Ensure study, which was essentially the equivalent of Eurtech, but it was run for registration in China. We went to go talk to them about the possibility of trying to register a companion diagnostic test based on a non-prespecified analysis of a

retrospective study. Fortunately, they agreed, and I think Walter can tell you more about how that actually happened. I think Karen may have an intervening discussion. Thank you.

Karen: Is it the green button or the red button? Okay.

As Rina said, this was the first CT DNA or liquid biopsy test we actually had for non-small cell lung cancer, and we did approve it just before the ASCO meeting started. One of the things that we'd had when we entered into the discussions with Rosch is based on some preliminary data that they indicated and they showed us that the agreements between plasma and tissue were not very significantly high. That's how we developed the reflex intent.

One of the fundamental differences we have is we did have a modification to the intended use, in this case, what's highlighted in blue are the key changes we had. These, we are looking at the defined mutations that's detected by the test and circulating free tumor DNA or FFPE. Then, the reflex is that patients who are negative by plasma should be tested and have their EGFR status determined by the tumor biopsy. We also expanded out and have analytical claims for the detection of other EGFR mutations and plasma. That's also with FFPE.

Regulatory history is Rosch actually came and contacted us in May of 2014, and it was essentially ... It was about five weeks after the expanded access program draft guidance came out. They were citing that they believed that liquid biopsy technology had met the unmet medical need for patients who may not be able to provide a tumor biopsy. It's a breakthrough technology that actually provides clinically meaningful advantages and also the premarket and postmarket allowances that were in the draft guidance. Actually, the EAP is actually really similar to what our standard operating procedures are in our division.

We ended up having multiple interactions to try to work out the analytical and clinical protocols. This is a completely different sample type than what we were typically used to. During the course of the discussions, we agreed to have a modular PMA submitted which allowed us to work on different aspects of the PMA over the course of time.

I think Walter will get into some of the key discussion points and considerations that we had, where the use of contrived specimens for some of the analytical studies, as Geoff pointed out quite early, that there's not a lot of plasma you can get out of these patients, enough to do all the analytical studies. We actually requested a commutability study to look at intact versus shared cell line DNA that's been diluted and either healthy or non-small cell EGFR wild-type plasma to function as a surrogate specimen. Also establishing efficacy where the patients were originally enrolled on tissue specimens and also the preanalytic effects and the need for well-controlled protocols when collecting and processing the specimens.

What was actually submitted is we had multiple studies, and these are actually very typical from what we normally have. We did allow some stability data to be presented post-market, although we did actually have preliminary data that was submitted pre-

market to use as a basis.

Walter: Thank you, Karen. I look forward to walking through some of the studies that we did. You saw there were over 20 there. I'm going to be merciful and not talk about each and every one of them, and focus on those that are really more unique to the sample type and the novel test that we were addressing.

Just to reiterate, in non-small cell lung cancer, although it's been the standard of care for several years now with NCCN, ASCO, ESMO guidelines, we're still not testing all of the patients. There are a number of reasons that have already been highlighted here. Some of these patients are ineligible because of health status. There are concerns about some of the morbidity associated with it. You can have pneumothorax, something which will require a chest tube. Often the biopsy sample is simply not adequate to do the testing. We know about heterogeneity of the tissue and so on. At best, at this point in time, about four out of five patients in developed countries are getting tested.

If you're an in vitro diagnostic manufacturer like we are, we work under design control. The first thing you do before you start developing a test is get customer and product requirements. You go talk to your stakeholders, the laboratorians, the oncologists, and find out what do they need. What they said is they wanted a fast turnaround time, they needed really high accuracy. They wanted the analytical and clinical sensitivity to be clinically useful, that seems obvious, and they wanted a robust and consistent test that had good cost-effectiveness. Cost will become an issue for some of these tests in the future, I'm sure.

Regulatory requirements. We are, of course, as was mentioned earlier, a global company. We're talking about FDA requirements here, but we also take these products into the rest of the world, and so we try to develop a test that we can apply generally.

Sample acquisition will be something we'll spend a good deal of time on. There simply are not enough samples, real world samples, available to do your development and technical performance evaluation. We looked at alternatives and worked with the FDA to come up with some approaches for that. Then, the question that's been brought up before, too, reference methods. How do you define truth with regard to either tissue or plasma, and go forward?

I seem to be going in the wrong direction. Green means forward. Okay.

All of this starts with a blood draw. Very simple, typical sample for many diagnostic tests. We modified the tissue-based tests that we had to have a sample prep methodology that now worked with the plasma for cell-free DNA extraction versus formal and fixed paraffin tissue. Then, a PCR set-up with three different reactions that can detect up to 42 different mutations and a reporting tool, all automated, on a 4800 thermal cycler. This uses a proprietary enzyme for allele-specific PCR to get enhanced discrimination between three prime mismatches, and that's how you detect the various variants.

Importantly, from a turnaround perspective, once the plasma sample is in the lab, to set

up and run takes about four hours. Certainly can be done within the course of a normal working day, and as Geoff was alluding to between shipping a sample, running it, and reporting back within a two- to three-day turn-around time. We felt that was adequate for what the clinicians wanted.

The sub-bullet here says it all. It's impractical to use clinical specimens for all of this development work. Here we have a situation where EGFR mutations in Caucasians are about 10% to 15% of all non-small cell lung cancers, whereas in Asians, 30% to 50%. There's a reasonably good number there, but think about those mutations that are only 1% or 2% and what the challenges will be there going forward.

In order to do the studies that I'm going to talk about, we spent, I think, in excess of $2 million for both mutant cell lines that we could use for creating contrived samples, as well as other normal and non-small cell lung cancer specimens. We had to buy some pretty large volumes, which is not easy to get from cancer patients. Where we did find those samples that had the mutations of interest, they were often at low prevalence and low concentrations, so we screened about a thousand samples to find 8%, or 80 specimens, that had the technical EGFR, that's lower than the prevalence I mentioned. Most of those only had about 200 copies per mL or less. If you're using that do to spikes, you're not starting from a good point.

There's clearly also ethical considerations, HIPAA and otherwise. IRBs don't generally let you draw more than 30 to 40 mLs of plasma, so this drove the need for us to find a way to make contrived samples to do some of the analytical testing.

The question of defining truth. We don't have validated reference methods, so we had to validate one ourself. There are no international standards as yet for EGFR, and the answer of truth is also partly dependent on when the samples were collected. Ideally contemporaneously, but often, you have a diagnostic sample and then at some later point where there could have been changes in the tumor, the blood samples.

Heterogeneity we've already talked about. We've also heard, and this is certainly recapitulated in work that we've done, that metastatic progression to later stages leads to much higher levels in blood and makes it easier to detect. The ultimate indicator really ought to be the clinical response. In this case, we're trying to predict response to TKIs or erlotinib.

The quantitative reference method needed to have a sensitivity and specificity defined over the reportable range, it needed to be able to detect these rare mutations in the background of abundant cell-free normal DNA, most of that coming from lymphocyte lysis and why it's important to treat the plasma sample carefully. We have instructions for use about that. We initially thought about digital PCR, but the FDA felt pretty strongly that a non-PCR method should be used as a comparator, and as we had used next-gen sequencing for all of our FPET work in the past, we simply, simply, I say, developed an assay for that. I'll talk more about that in a moment. Depending on how low you need to go and how much wild-type is there, the read depths can be quite significant to get to where you need to be.

For every mutation that we detect, we developed receiver/operator characteristic curve analyses to optimize sensitivity and specificity. In this case, we want to have few false positives and the best sensitivity we can under those conditions. We have a wild-type control, internal control, that tells us the test is valid. It's in another part of the EGFR gene. We evaluated hundreds of samples from healthy donors as well as patients over several thousand data points to generate this area where the cutoffs would be established for this particular mutation. The wild-type CT values that are low represent breakthrough of the allele-specific PCR and use those to help set those cutoffs, and then the EGFR mutation CT values represent the true positives.

The next question was then how to develop or determine the limit of detection with contrived samples. The FDA felt that shearing cell-line DNA was going to best mimic the short DNA that we typically find after apoptosis. We did do some comparisons with unsheared, and I have to say they look pretty similar. We did all the studies that you'll see here were sheared cell-line DNA. Those were spiked into healthy donor, initially. They wanted also to seek commutability, so if you put that same DNA now into non-small cell lung cancer plasma, did you get equivalent results at or near the limit of detection, and then finally, confirming that limit of detection in plasma panels from non-small cell lung cancer patients.

We applied the SOP of CLSI-EP17A2. You may or may not like that one, but most people in diagnostics use this for determining limit of detection. In this case, about 72 replicates across three different lots for every level being tested. This was in healthy donors. A limit of a blank determined for normals. Importantly, in this particular set, and something funny happened to the circle, but let me just point out that the background here was approximately 100,000 copies. This is not typical. We don't normally see this much unless someone's done a bad job of preparing the plasma and gotten a lot of the lysis.

As you can see, what should have been circled, about a four-fold range of anywhere from 25 to 100 copies per mL was the limit of detection that we established for the various mutations you see here in this background, which means an abundance level of about 0.025 to 0.1%. Fairly sensitive.

We demonstrated this commutability, then, by doing the same spikes into non-small cell lung cancer plasma. We focused on the most important deletion T798, T858R, and T790M mutations and the levels, at eight different levels, from 0.03 up to 10X. Importantly, the hit rates that we saw doing this with either healthy donor or cancer plasma were very, very similar, and all the statistical analyses for bias or with regression intercept tests showed that there was no significant difference. The FDA accepted these approaches as a substitute for having the requisite number of cancer patient specimens.

We then went on to confirm that in a clinical setting, and so we developed a panel with 11 members. There were five different mutations tested at 1X and 2X LOD and then one non-mutant samples. This is pretty typical testing plan: three sites, two operators, three reagent lots, different days, and then replicates for everything.

To summarize, for ten of the 11 members, there were 100% hit rates for both levels,

and the one case where it was not 100% was T790M. You've seen some others also describe this one as a little bit more difficult to pick up sometimes, and that was at 96% here.

Again, in this case, there was 24,000 copies of wild-type or normal DNA in the background, so the mutations were detected at approximately 0.3% in this study. Then, clinical reproducibility had to be established. Again, you see multiple lots of reagents, multiple instruments, multiple sites, operators, consecutive days. About 650 replicates with nine panel members. These were made at 100 or 300 copies of mutant per mL, and we combined two mutations into each of these cell line combinations. In this case, the sheared DNA was spiked into a background of about 12,000 copies per mL for the cancer plasma that we had.

It was a very low invalid rates. Only four out of over 1200 tests failed. Then, you can see the hit rates here for the various mutations at 100 copies per mL and 300 copies per mL, and no false positives at all with the wild-type samples that were tested.

Then, importantly, and I think I'm going to buy us some time, analytical accuracy correlated to next-gen sequencing was done with 128 samples. These came from some phase three clinical trials, aspiration with some supplemented med-lung and med-lab. To say a few words about the next-gen sequencing reference method. At times, it feels like it's almost as much work to validate the reference method as the test itself. I won't have time to go into all the details, but this particular test was a two-tube amplicon test with aluminum [iseek 01:41:32]. We used the same alluot that the Cobos test would see. That would be 25 microliters per reaction. Then, we used about 300 individual samples from both healthy donors as well as cancer patients without mutations to set the cutoff for the negatives, and then made pools of wild-type DNA with cell-line blends ranging from 0.03% up to 1% mutation and inputs of 2 nanograms up to 500 nanograms of DNA.

We verified and validated this both internally as well as at a GLP reference laboratory. Normal donor samples, 24 replicates of each, the mutant spikes also 24 replicates of each at four different levels. Then, once we'd done that initial validation, every subsequent use of the test included either wild-type and/or known mutants at different levels as run controls. The system, the software, also has to be validated and verified similar to a class three IBD and device system.

All right, what did those data look like? We did this. I'm showing here for [ex on 19 01:42:47] deletion and L858R with 128 samples that were paired. Thirty-two had mutation detected, and 95 had no mutation detected. The positive percent agreement was about 88%, and the negative percent agreement about 97%. We had to use this trial, because as often is the case, you simply, we've heard earlier, don't have enough sample left from all of these studies to do the various tests that you would want to do. That looked pretty good.

Clinical outcomes. Is this meaningful? We had 180 samples. As was mentioned earlier by Dave, the Ensure trial was this first-line registrational trial for China looking at patients who were randomized after testing for positivity or negativity of EGFR to either the

chemo or [lotinivarm 01:43:42], and that's what the Kaplan-Meyer curve looked like for progression-free survival. That was investigator-assessed. It recapitulated what had been seen, as you saw earlier, with the Eurtech trial that was done previously. There's four spots where tissue and plasma mutation are detected, tissue-only plasma-negative, and the intent to treat population here at the bottom, all showing very similar hazard ratios.

Last thing I think I'll share, then, is the correlation between plasma and tissue. Always the question, is tissue truth? In this case, from Ensure, we started with 650 patients; 600 of them had a valid tissue test, and then of those, about 86% or 517 had matched plasma. We ended up with 430 patients that had matched valid tumor and plasma results with 2 mLs of plasma being available for this test to be run. To cut a long story short, in aggregate, the sensitivity was 77%, meaning concordance with what was seen in tissue, and specificity at 98% with overall agreement of 88%. This has been pretty similar from other studies, also the fast-act study that we did previously, as well as what we've seen with some other tests that are available.

This is why you can't depend on a negative to say it's not there and why you do have to reflex back to tissue again. All oncologists told us they would take such a result and move forward with a plasma-positive and treat that patient.

One important thing to point out. In the case of EGFR, we have, I mentioned earlier, a big difference in the prevalence of these mutations between Caucasians and Asians. In Caucasians, in this, say 10-15% range prevalence, you can see you have a high negative predicted value and a somewhat lower positive predicted value. In contrast, over here where we would expect the Asian population to be, it flips the other way around, so your positive predicted value goes up and your negative predicted value goes down.

Just food for thought. As we start to look at mutations like BRAF in lung cancer and others that are in the 1% or 2% or 3% range, you can see the challenge that we're going to have with detecting those reliably.

Then, I'll turn this over to Erin to discuss the clinical results.

Erin: I'm just going to focus on about three things that we looked at form the clinical perspective. Most of this approval is obviously based on the validation done through CDRH. From the clinical perspective, obviously, any clinician wants a test that's going to find patients who may benefit from the treatment and rule out patients who will not benefit from the treatment. Given the specificity of the test, it's not a problem. Adding the reflex language into the lAbrahaml covers the other side of it.

What we felt we needed to see based on this was that using the plasma test could be just as productive for outcome. The problem was this was a convenience sample from the Ensure study because only percentage had enough plasma available to do the plasma testing. However, it was about 83%. We need information to make sure that that convenience sample is similar demographically and other prognostic factors to the entire Ensure population. We need enough patients to address any potential unintended bias that there may be for prognostic factors that aren't measured between

the convenience sample group and the real group, and 83% is a fairly good number.

Advice for anyone developing these. The plasma collection was prespecified up front in this study. You're supposed to collect up to 6 mLs of plasma. Obviously, it doesn't happen for every patient, but if there's any interest in using a clinical study for potential diagnostic markers later, you really want to optimize the collection of the tissue and the plasma samples so that you have as close to 100% as you can. It just makes the case a lot stronger from the clinical perspective if we know we're looking at the same population when we're comparing the PFS rates.

Speaker 18: Thanks to all the presenters. Now the floor is open for questions from the audience, if you have any questions for these presenters.

Geoff: Great to hear the story come together. I think the most astounding part of this is the development of a new, never before studied reference standard to serve as the plasma reference for the new plasma assay. Who thought of that?

Walter: Well, as I said, we had suggested considering digital droplet PCR, and the FDA felt that it was circular, it was a PCR assay for another PCR assay, although technically they're quite different. Nevertheless, since we'd already had experience with next-gen sequencing for our tissue-based testing, we said, "Okay, let's do NGS. We'll just have to sequence a lot deeper to find the low abundance that's in plasma relative to what we had in tissue."

Geoff: Did you apply your final assay to your clinical validation cohort? For example, you just nicely described how your positive [Richter 01:49:34] value dips because of apparent false positives. Are they real false positives? Did those false positives in using your Cobos assay validate with your orthogonal assay?

Walter: We didn't have enough material, unfortunately, to run both assays. We have done other studies where we've used orthogonal methods, including collaborations with some folks in Massachusetts for comparing CTCs and cell-free DNA by two different methods. Frequently, you find that the tissue was wrong and the two orthogonal methods are positive for plasma. It depends.

Geoff: The question everyone in the audience is going to be thinking is, "As I develop my assay," they say, "Do I have to invent my own reference standard, or does this co-bas assay need to be my reference standard, or is there flexibility?" Help the people out here think about what makes a pseudoreference standard, and has this science changed now that there already is one, of sorts? Does that make sense? Does it mean this new co-bas one needs to be the reference standard, or is there flexibility in that decision?

Speaker 18: Yeah, there is flexibility, and there is a mechanism where you could actually talk to us. That's a presubmission process, so we will give you feedback on that.

Geoff: Okay. But there's flexibility. Okay!

Karen: I'd actually like to also point out, we actually ... We do this in a lot of cases. We actually have the sponsor, when they come in with the reference method, that basically they

have to create for us, we actually do review that, the validation data that was performed and provided for us.

Speaker 22: Come on, Geoff, you knew the answer to that question before you asked it. [Garish Prouchem 01:51:22], MoldX. Just a really small question, I guess, which is just, help me understand. Given your experience across the board with the commutability studies, how do you think about the extendability of contrived samples for other alteration types, or is it too early to tell? I knew you'd say that. Okay, fine.

Speaker 23: Question for Walter, and that is, in terms of the ... For lower limit of detection, 25 to 100 copies of mutant per mL, you're using two mLs, and then how does it get split up across your 42 different mutations? How many copies are in an individual assay, just the way the whole format works?

Walter: Great. Great question. You could derive that, but the 2 mLs extracted end up in three reaction tubes. Actually, three-quarters of it do. There's 100 microliters of alluot, 25 microliter per each of three reactions, and the 42 mutations are spread across the three reactions. You're using one fourth of the 2 mL, or half a mL of plasma per reaction that you're testing. If you do the math, you can see you're getting into the low numbers when you're down at 25 per mL, depending on what your recovery is, 50%, 60%, 70%.

You push into [poissan 01:53:00] limits quickly.

Speaker 23: Of course. You're not using a half a mL. The mL and a half gets split across three assays, so the equivalent of half a mL per each one of those reactions, each one of those run a little specifically.

Walter: Correct.

Speaker 23: Okay, thanks.

Speaker 24: Do you think the limitations in sensitivity are due to limit of detection, or what do you think the cause of that is?

Walter: There's both. There are obviously analytical limits, which we defined here, and every assay has its own analytical limits. Beyond that, there are biological effects as well, and a lot of good reasons why you might not detect something in plasma even though it's been shown to have been there in tissue at some point.

Speaker 25: Just a couple of technical questions. When you did your NGS, did you use the same amount of plasma sample as you did it for your Cobos?

Walter: Yes.

Speaker 25: If you did not know Cobos was positive, were you able to detect [inaudible 01:54:16] NGS without knowing what the result [inaudible 01:54:19] in Cobos?

Walter: Well, they are blinded and separate.

Speaker 25: Blinded.

Walter: As you saw, the NGS actually picked up one or two more than the Cobos test did.

Speaker 25: Your analytical sensitivity for NGS was how much? What did you say? 0.1 at 1%?

Walter: I'd have to go back and look at it, what the actual numbers were.

Speaker 25: Sorry.

Walter: We actually train the algorithm for every unique site, because all mutations are not arising at the same level, and the background noise is different. That's what we did to validate it. We trained on every unique site.

Speaker 25: Okay, thank you.

Speaker 26: First of all, excellent presentations. Could you provide some insights on how you handle statistics so we get an estimate of what kind of sample sizes ... How do you decide how many reproducibility runs you're going to do? Six mLs is also a lot of plasma. You're obviously fortunate to have that. When we approach a clinical trial, we do have some statistical plan. Does that happen in the end, the beginning, the middle? I'm just curious.

Speaker 18: Sample size for the clinical validity? Is that what's your question?

Speaker 26: Yes.

Speaker 18: Yeah, so that's decided during this presubmission interactions, what is actually required for ... Depending on the intended use. Depending on the intended use will be a lot of discussions on what is required for each specific intended use.

Peter: In this case, you had a very high inflection rate, and [inaudible 01:55:57] in most accounts would be not necessarily [inaudible 01:56:00]?

Speaker 18: Yes.

Peter: I forgot to introduce myself. I'm Peter Caine with Counsel. What level of background noise did you observe in the NGS reference assay?

Walter: I didn't bring that information along. As I mentioned, every site is different, so T790M would be different than L858R. The important thing is that you establish what that background noise is by running large numbers of normal, healthy patients through. Then, based on that, you can determine what your cutoff above which you can detect.

Speaker 18: Time for one quick question because we need to move on to [crosstalk 01:56:42].

Speaker 1: Can I just ask a general question? What are going to be your guidelines at the lowest limit of specificity you're going to accept if you haven't done orthogonal testing?

Because then you're giving a drug to the wrong patient.

Speaker 18: The specificity.

Speaker 1: What's going to be ... There must be guidelines-

Speaker 18: You saw this one had a very high specificity, so that's why we improved it.

Speaker 1: Give us a lower threshold of what you think is clinically important and clinically [crosstalk 01:57:12].

Speaker 18: I think that's what we'll discuss in the next section, what is the minimum required for analytical and clinical value.

Speaker 1: Clinical specificity.

Speaker 18: Yes. Clinical and analytical we'll be discussing in the next session. Moving on. Good segue to the next session. Thanks to all the [inaudible 01:57:27].

Session three, we're going to talk about analytical and clinical validation requirements for liquid biopsy tests. You heard in the morning there are a lot of tests out there, but what is needed for a good analytical and clinical validation is what we will discuss. We did have a lot of conference calls with all of the members in the panel. There's some volunteers who actually volunteered to present their data, so you'll see that data. The order will be Abraham [Tzou01:58:08] from our division, who's a medical officer, will provide the FDA considerations, and then followed by Dr. Phil Stevens from Foundation Medicine, will talk about the analytical and clinical data study requirements they have met for their device. Then, the panelists will come up on the stage.

We have actually thought about three different areas that we think are critical. I know there are a lot of other things that are critical, but depending on the time, we will get into those. For the three different areas, we do have some slides from, again, some volunteers who are going to present, and then the panel will discuss among themselves their input on each topic. At the end of all three topics, it'll be open for the audience.

With that, Dr. Tzou from FDA.

Abraham: All right. I'm just going to give an overview talk. It's really going to be big picture concepts, no real actual data, just trying to orient people as far as possible perspectives, what sort of priority areas to think about, recurring things that come up. Some of these themes have really been alluded to directly or indirectly already during this workshop so far.

First part, I'm just going to give a very standard baseline scenario of the prototypical liquid biopsy scenario that we'll be discussing as a platform to discuss the general principles or issues that arise in liquid biopsy diagnostic test development, and then I'll try to focus and touch on three areas of common concern regarding orthogonal methods-

Abraham: … regarding use of contrived specimens and interpretation of Plasma to tissue concordant studies. As we've heard this morning much of the data and with the case study you just heard applies in patients who have already been diagnosed with cancer. It's a more tracktable setting because for patients who've not yet been diagnosed figuring out what results mean and how to follow those up. Obviously there are more complexities with that.

We're primarily interested in this discussion in solid cancers, solid tumors and the idea of course is that liquid biopsy may offer many potential advantages. However, in this case of solid tumors much of the clinical therapeutics and developmenting outcomes have been established based on tissue. If one does not have direct liquid biopsy based clinical outcomes. To what extent can one link to existing tissue based results and decide whether or not liquid biopsy results are adequate or sufficient, as a basis of selecting therapy for these patients?

The key areas of consideration, as we've heard there are some concerns or doubts some uncertainties about whether different ways of evaluating liquid biopsy results agree. Are they telling you the same thing, are they telling you different things? What is the actual liquid biopsy status for that patient? In the case of discussion of contrived specimens there is really only so much plasma you get from patients.

If you're going to resort to contrived specimens, how does one make a determination when those are appropriate for analytical validation? The question that was just posed at the end, when you have some level of discordances what does one do with that and what does that really mean?

What are the general principles? I just hopefully I'm going to just provide some perspectives of orientations of what the priorities are. Obviously there are limited and pragmatic and logistic considerations of what are the specimens you can get. The goal is presumably once trying to understand how the test would perform in the actual patients of interest. Just because specimens have to be available that's really good and that's important. If those specimens do not represent the patients of interest obviously that is a potential concern.

Obviously if you can't have enough clinical specimens to do an exhaustive reproducibility study you may need to supplement with contrived specimens. If one does not understand the extent to which performance in contrived specimens behaves like clinical specimens. That is also a potential area of concern. Interpreting these tissue concordance, discordance, agreement, none agreement. What is the difference between heterogeneity versus the relevance for clinical outcomes?

You'll see this basic schema repeatedly. It's pretty simple right? Whatever patient population you're interested in you're trying to get your device results, you're trying to understand what happens with those patients and you're trying to establish a meaningful statistically and clinically relevant association between your liquid biopsy results and the clinical outcome.

People often hear that you need to have your validation study independent from training and development. I am not sure if this is the case or not, sometimes people primarily think of that in the context of clinical study. If you have a clinical cohort and you picked your clinical cut off to optimize your performance for say perhaps clinical sensitivity and specificity in a clinical cohort. Probably there's a decent understanding that you need an independent clinical cohort to establish clinical performance. The other thing I would just want to raise is that also applies for analytical performance.

In the context of liquid biopsy where you may be on the leading edge of technology or the leading edge of performance and you're trying to go down to very low levels or detect novel variance that other technologies may not be able to perform as well in, and you are setting your cut-off in analytical performance based on known characterized contrived specimens. That using those same specimens to establish your analytical performance also is problematic.

I hope that's clear. You have your cell lines that's establishing whatever, you're showing the right cutoff, you're performing well, you detect that well and then you use almost the same cell lines again and you say, "I detected it consistently." Well, that's because you're cutoff that way. The same thing that applies to clinical outcomes does also apply to analytic device results and then not just for analytical testing that could apply for sophisticated bio-informatics, packages or so forth depending on what in silico data or

modification one makes.

I just want to put that as a general caution. The analytical bio-informatics side people sometimes don't necessarily consider what the provenance of the developments and how that relates to the validations sets. Ideally for liquid biopsy test then you would have your intended use liquid biopsy population. For discussion today it will be primarily more later staged diagnosed cancer patients by depending on what setting you're interested in obviously that would differ.

These terms are not officially defined, there's no consensus. I'm just using different terms here as placeholders of properties or characteristics that could differ or there could be discordances across different methods, but that exhibited consistent results, compare results, uniform results across liquid biopsy tests. Those are all trying to get at some concept of accurate analytical performance and that one also had clinical outcomes related to the liquid biopsy tested population.

Obviously that's ideal, the reason I'm speaking here is because that's not pragmatically actually going on. Instead of having the target intended use population of interest what one typically has instead is available samples. Are the available samples representative of the target population of interest? That's one question. When one runs out of clinical samples one often is discussing what sort of contrived specimens one is going to use instead.

Then of course are those contrived samples operating and behaving similar to the target patient population samples of interest? This question as far as if you're having orthogonal method, different methods do different things, they have different performance characteristics and there are different sources or features or properties that are not always in congruence. Which test results, who do you believe all of those issues? When you're comparing to tissue you have a clinical trial where tissue was the base of the trial. You have outcomes related to tissue positive status and only those results and those patients you have clinical outcomes on. What do you do with that?

Based on that I will just go over some of the considerations. The first one you have your orthogonal method, you're not sure which orthogonal method to pick and you don't get the same answer. The first thing I just want to point out, there are different categories

of not the same answer. Just because you don't agree you cannot agree on multiple interesting or not ways. These are just words I picked, they're not official definitions. I just picked different words because I'm trying to show that there are multiple ways of not agreeing.

You can agree more profusely or more politely or whatever. The first thing is what I personally call consistency. Consistency in the example here is this morning people always have this needle and the haystack thing. If I have haystacks and I have needles maybe when I identify a needle there really is a needle there. That's true, that maybe when it's positive, it's positive. That is not the same as saying anytime there is a needle I will pick it up.

When we say limit of detection that means when there is a certain amount of needles in that haystack I will detect it consistently. That is not the same as saying if I pick up a needle there was a needle there. The consistency is that there is a level that whenever it reaches that level I will consistently detect it. That is one aspect of performance or agreement. What I call none comparability here that could be different ranges, different reportable ranges.

For example, CTCs if you use different CTCs there's different capture methods that capture different types of cells. You can end up with different numbers of cells. Some methods will give you higher numbers, some methods will give you lower numbers. The number are not directly comparable. They may have some relative trend but the numbers in and of themselves may not have a one to one correlation. Another issue is that the range where it operates well may not be comparable.

Some ranges may work well in lower levels, another range may not work as well at lower level. It may work well from arbitrarily 10 to 100, another one goes down to 0.1, to 10. There's a gap between what part you can cover from one method compared to another. Again another term I'm just using just for convenience of uniformity. That could be that different variant types, some variant types are more challenging and you will not detect them as well because they're just technically analytically more difficult to pick up.

If I go to needles, my made up analogy would be toothpicks in a haystack. If I have wooden toothpicks in a haystack and my hay looks yellow and my wooden toothpicks look yellow. It may be more difficult to pick up toothpicks in a haystack than compared to my needles in a haystack because that's just more difficult to identify. Those are just some characteristics, obviously those are all rationales of why things could be different. There's frankly not a great solution or easy answer for that. You just have to understand what are the nature of these differences and what are the indirect methods including possible use of contrived samples to try to address why you think certain areas are more robust or more applicable for your given method or your orthogonal method.

The second area contrived specimens. There is not enough clinical specimens in order to have robust characterization of analytical performance. You need additional supplementary specimen so you go to contrived specimens. The first concept is clearly the expectation is that there is some performance in clinical specimens to resort in purely to contrived specimens definitely leaves a lot to be desired. Then the question of course is how does one contrive specimens that actually to clinical specimens?

Part of the panel discussion for example if we're talking about plasma tumor DNA. If you're looking at the target analyte of interest, how would get that in there in the way that you want ensuring does that do what you need, does that depend on how your detection method is or is not sensitive to length or where the breakage is occurring, what sort of background DNA do you want, how much should there be? Just as a prototypical example if one came in purely and said I have oligos and buffer and I detect that consistently.

People probably would have some consideration or concern of whether that is a basis to be confident in clinical performance. Where the DNA could be more fragmented, there could be a mixture of background. How does that actually translate? This is again just a conceptual diagram of this concept of commutability or how contrived samples would perform or not compared to clinical. For diagrammatic purposes this curve on the right in solid would be just an example. Detection rate is how often is it positive within increasing levels of analytes.

We'll just say the clinical falls off and then if we say limit detection to something like 95% there is this LOD for this clinical. The general concern is if one comes up with contrived specimens that are cleaner so to speak or easier. That one the behavior could

be quite different than clinical specimens and that one would lead to think that, “My test performs really well,” when that’s just because you have easier samples that it performs well. On which is not how well it performs on clinical specimens.

The concept of a comparison study is try to demonstrate how the behavior of the test performance compares in clinical specimens, compared to different options of contrived specimens. One thing I would just comment here is even for contrived A, if one only let’s say at one point on the curve where there’s high detection. One might lead to the impression that, “Just because I get high detection here that means it’s all the same.” That may not necessarily be the case.

Comparison, we have multiple discussions, we’ll also have more discussion in the panel as far as what do you do when you compare your liquid biopsy results to tissue, you do or do not have outcome? Obviously the best would be to just have clinical outcome based on liquid biopsy that’s just I think most people would agree that is the ultimate answer and we’ve heard about different trials or proposals that are in progress where that is primarily the design. The question of course is in the absence or in the interim where people do not have that. What can people do instead?

If one has tissue specimens with outcome, then comparing liquid biopsy to tissue may be a sort of stepping stone, if you don’t have liquid biopsy clinical outcome directly. The caveats are of course their differences in trials could lead to different specimen cohorts, differences in patient populations. As we’ve heard you have trunk mutations or evolution or treatment resistance overtime and then that could complicate or provide nuances of how to interpret these concordant studies.

Obviously the ideally to minimize discordance would be to have match sampling to avoid evolution or heterogeneity over time that would magnify the rate of discordance. It should be straight forward that obviously the conditions where you have less discordance it makes it easier. When you have more discordance whether it’s due to heterogeneity or something else that just leads to more question as to what the impact is if we were to use it in a clinical setting?

As you’ve heard with the case study. In the case where the liquid biopsy is negative and the tissue is positive, what the understanding that there is still clinical meaningful

benefit into that sub-group. The idea of reflexing to tissue if that were understood by users and providers would be one way to address that sub-group. The somewhat more problematic group is those that may be liquid biopsy positive and tissue negative.

One reason that's problematic is because if you have a development program where tissue positive patients were the ones that were treated and you have outcomes on, you just don't have outcomes on tissue negative patients. You really don't know directly what would happen if those patients were treated. The other thing that is potentially problematic is in the context of therapeutic trials which are enrolled based on tissue positivity.

It could be that patients who are initially evaluated in their local or community settings as being tissue positive were more likely to be refereed to those trials. Those trials may be enriched and have a higher prevalence of mutation positive patients. If you use those trials as a basis to estimate what the likely expected group size of this liquid biopsy positive tissue negative group is. You may be underestimating the size of that population and it may be more important in the actual use than you might understand from a clinical trial.

The general question of course is if you were to switch from tissue to liquid biopsy and this liquid biopsy tissue negative population you don't have direct clinical outcome is. What would be the overall effect? To the extent that there are different proposals of how to address that, different end points obviously to the extent that is part of one's liquid biopsy development program. How to factor that in is an important consideration.

For those who are interested sometimes they do not have access to the tissue results that were associated in trials so they want to go to outside cohorts. Then you don't have clinical outcomes related to the liquid biopsy and you don't have clinical outcomes related to the tissue directly either. You're two steps removed so to speak and clearly as you've heard there are many factors that could influence how much DNA or circulating CTCs may be shed. All of those may impact how you interpret the concordance of those liquid biopsy to tissue studies.

The general idea is if someone came in with later stage patients that were more earlier in treatment prior to evolution in treatment resistance. That if you have higher shed

DNA primarily looking at trunk founder mutations, driving mutation that you might see concordance. That may not translate to how you use it in the setting of actual interests. Again my basic scheme here there are these issues as far as what this liquid biopsy you're actually getting as far what is available, how you use contrived.

At least these are not all great silver bullet solutions but at least trying to get at some of the issues. If you're looking at available clinical specimens to what extend they're representative of the population of interest that's being intended for use. If you're looking at contrived to what extent you can support the commutability and similar behavior as clinical specimens.

If you're comparing your device results with an orthogonal method to what extent you can address the underlying nature or the disturbance, the disagreements or different operating characteristics. If you're using tissue outcome associated to try to support the clinical relevance of liquid biopsy tissue that the liquid biopsy negative tissue positive could be addressed by reflex and have a discussion regarding how you would support the liquid biopsy positive tissue negative population as part of the development program. All right, thanks.

Reena: Any quick question for Abraham before we move to Phil?

Male: One pretty quick question [inaudible 00:19:54] Simmons Healthcare. I was wondering how you deal with the collection method that's mostly used for this not being FDA approved, not for diagnostic use, research only.

Abraham: In the context of when we evaluate products we evaluate in the whole context of the whole system. Whatever liquid biopsy product we would look at the entire collection step would go through as part of the evaluating, authorizes part of that system. For what is being used outside of FDA setting or for research or investigational use different considerations would apply.

Male: I don't think I had my question answered from previous. We're dealing with a founder mutation where we should not have the problem of heterogeneity as distinct from resistance mutation which are currently evolved and there could be heterogeneity.

Clearly there when people have done it and the Boston group have shown there's high concordance and a high specificity when you're dealing with a founder mutation. What still is the lower threshold of specificity you would accept? There has to be some guidelines on how worried you are down to mutation and resistance mutation is a different biology.

Abraham: Founder mutations, there is a certain level performance we've seen and we've pronounced on. If someone has performance that's substantially different from that, that would identify more patients. They do not have clinical data to address how that would impact it clearly that would be a concern. Now I can't tell you exactly when it would cause some trigger, but if you're going to expand and pickup patients for which treatment would otherwise be selected and you do not have data on that population obviously that questions would be raised.

I can't necessarily off hand that there will be nobody able to provide an answer to that because if people want to do that study of course they could do that study and so like, well we identified these additional patients and treat them and see what happens. If people want to do that study of course we would entertain that.

Male: Don't you think it has to be a hard level in view of Geoff Oxnard and Passy's data. Which is 100% specific with tissue on the two founder mutations?

Abraham: Clearly that provides the context of at least the starting off expectations would be. If people would do something different then that of course they would need to explain why it is different.

Reena: Okay thank you. We'll move on to Dr. Stephens Stephens he is the Chief Scientific Officer Foundation Medicine.

Dr. Stephens: Thank you Reena and I'd love to thank the organizers for the opportunity to present today. Some of the challenges as we see of tissue based and a liquid based diagnostic and at least how we at Foundation Medicine have come up with analytic validation approach. It would hopefully set up a really lively panel discussion later. It's become increasingly clear that therapies that target the genomic alterations within an individual

patient's disease can be both more effective and less toxic than none targeted therapies.

What I think are remarkable is the rate at which this field is evolving and a nice example is illustrated here. In 2012 we and others described the identification as RET fusions as a novel oncogenic driver in non–small cell lung cancer. A mere 13 months later, I apologize there is a typo on this slide this should be March 2013, we are now colleague at Memorial Sloan Kettering published the first manuscript in which patients with RET fusions were responding to a RET inhibitor.

If one goes back and looks at the time-frame. It took us six months to write the manuscript and get it through the peer review process and some of the patients were on drugs for nearly six months. What actually occurred was within a few weeks of this novel oncogenic gene fusion being described in the literature. Memorial Sloan Kettering had written a prospective clinical trial and were sending their negative patients to us for diagnostic testing.

Now again I find that incredible but as the number of clinically relevant or targetable cancer genes increases, matching the correct patient to the correct targeted therapy presents some tremendous diagnostic challenges. It's also becoming clear that it's relatively straight forward to apply next generation sequencing and other technologies to identify base peer substitutions but finding these other classes of genomic alterations is orders of magnitude more difficult to do.

This is incredibly important because if one just looks at non-small lung cancer alone there are over a dozen genes thought to be clinically relevant with the short [inaudible 00:25:29] and the copy number alterations. There are now at least five clinically relevant gene fusions. To underscore this in a tissue based study we did with Memorial Sloan Kettering in which 50 plus patients who were negative for the markers on the NCCN guidelines were sent to us. A quarter of these patients had an alteration that was missed in the NCCN guidelines.

We have recently presented data and published data showing from looking at around 10,000 patients with non-small cell lung cancer where we've been able to get a look at the prior diagnostic testing. Around a quarter of patients with EGFR mutations and

about a third of patients without fusions and ROS1 fusions are being missed by standard diagnostic testing. This is FDA approved on lAbrahaml standard of care testing and alterations.

Why I'm I bringing up tissue based testing in a liquid biopsy session? I think many of the speakers this morning has underscored doing broad based genomic from the blood using circulating free, cell free or circulating tumoral DNA is orders of magnitude more difficult than tissue based testing. Physicians and the care team and academics and we an industry really do have to think about what is appropriate and how to use these tests.

One consideration is which patients do actually apply the liquid biopsy tested? If we look at the literature and a recent paper from [inaudible 00:27:23]. Somewhere around 40% of patients do not shed sufficient DNA into the circulation to do a broad-based diagnostic test. This study did have a reasonable number of glioblastomas in it and we know that only about one in five GBMs shed sufficient DNA.

Around a third of patients with advanced and solid disease will always be negative from a blood based test and we don't fully understand yet which patients these will be. The take home message is that even if these patients had an on lAbrahaml alteration such as EGFR and lung cancer, HER2 amplification in breast cancer. If there's insufficient DNA to profile these patients will always be negative.

The fact that at around the 60% to 70% of patients who do shed DNA from the tumor into the circulation. In many if not the majority of these patients the amount of DNA that's derived from the tumor in the circulation is infinitesimal and it's usually sub-1% and this does present some tremendous diagnostic challenges. If you look at data presented at ASCO this year and last year both from commercial labs and academically. From the blood the alteration frequencies of key drugable cancer genes such as ALK, EDRF and HER2 and in this example c-Met were wildly different than what we're finding, what the rest of the world is finding in the tissue based testing.

I've just used a single example that was recently published in Oncotarget again from the [inaudible 00:29:05] group showing c-Met amplification frequency in breast cancer at 15 fold higher than the rest of the world is finding and statistically significantly different.

When one tries to understand what is happening what this assay is almost certainly doing is discriminating polysomic C-Met, 3, 4 or 5 copies of chromosome 7 from focally amplified c-Met 6, 7, 8 perhaps 9.

We know from the many failed clinical trials in the literature that these drugs do not work against polysomic c-Met at least the current drugs. They will only work against the focally amplified C-Met and more recently against the exon 14 splice site mutations. Just to push this point if the results from this assay and this is a big if. If they were broadly recapitulated for every 15 patients in which the ctDNA assay recommended targeted C-Met therapy for breast cancer. 14 of those 15 patients would not have a targetable alteration and clearly that is of a huge concern.

In my final note of caution is a case study from a patient. She's 36 years old with advanced lung cancer. We identified a RET fusion, a P53 mutation and lots of P16. This patient was incredibly fortunate because not only was her physician able to get her onto a Ret inhibitor, she went on to have a blockbuster response with near complete resolution of the disease. We know that most patients with RET fusions do not show this level of clinical benefit.

At the same time as a tissue based test she had a blood based test from another entity, failed to find the RET fusion, failed to find the P53 mutation and loss of P16 and came up instead with a false positive EGFI mutation. How do we know that this is a false positive and not what this company say it was? Well how, this is obviously tumoral heterogeneity and this is the advantage of a blood based test.

We know because a physician phoned me up and was incredibly concerned about taking his patient off a drug that was working so effectively and switching it against an alteration that wasn't in her tumor. She said, "So what do I do?" We would develop our own blood based test I said, "Look in R&D, send us a couple of hours of plasma from this patient." We knew he had frozen in the freezer and both pieces were identified the RET fusion, the P53 mutation and there was not a single mutated EGFRE despite over 10,000x coverage. We now have dozens of examples of a similar vein.

Again this is really underscoring how challenging this is to do and the importance of a very thorough and comprehensive analytic validation. Over the last two years we've

developed and we recently have launched our own CTD in an assay. It's focused like many others on the most drugable genes. This gene content looks very similar to many other tests and we've optimized it for the extreme challenge but importantly for all classes of genomic alteration.

We've done an extensive analytic validation and presented it actually prior to the commercial launch of our test and this is what all companies should be doing. They should be able to demonstrate what the accuracy of there test is to themselves before they launch it and unleash some of these tests on patients. Because how is a physician supposed to know whether a diagnostic test meets the diagnostic and accuracy requirements for his or her patients if there is no analytic validation?

Again the first thing you want to do is look at accuracy, the sensitivity and specificity for all classes of alteration on the test. Do not just take base [inaudible 00:33:17] substitutions the easiest class of alteration to identify and extrapolate across the other alteration classes, it does not work. As many of the speakers this morning has demonstrate the workflow has to be compatible to routine clinical care. There's no point developing a test where you take blood immediately do your test and then find out when you launch it clinically that the DNA is being seen in the post or 48 hours and it doesn't work.

Precision, we've demonstrated high intra at inter-run concordance using clinical samples and importantly from a reasonable number of samples we've done orthogonal validation of genomic alterations detected by our ctDNA. What I'm showing you here is, what are the ramifications of improper sample storage. Here we have the patient Kress positive with pancreatic cancer. Looking at different storage mechanism for the blood in a similar to Geoff presented this morning.

Here we have using a pretty decent assay what we believe validated zero Kress mutations. Yet if you use the struck tubes that Geoff and others have alluded to Lo and behold you go from a completely Kress wall type patient to 26% Kress mutant a little frequency within the plasma. People keep talking about it's important to get depth of coverage. Yes, but more important is the uniformity of the coverage. There's no point having 50% of your test at 5000x if the other half of the test is present at 100x.

Really optimizing the hybrid capture or the other sample prep approaches such that you have uniform accuracy across your tests. Actually look at what the sensitivity is to detect the different classes of alteration at a wide mutant a little frequencies and importantly at the mutant a little frequencies that one would expect to find in the patient's plasma and perhaps not 100 times higher bar and then presenting those results.

What I'm showing you here, we've done this for base subs and other classes of alterations. I'm just showing you the data on short insertions and deletions. From 168 true positives we have managed to identify every single one of them down to a low minor a little frequency. We had a couple of false positives for a sensitivity of approaching 100%. You can never say you're at 100% and all the time maintaining a very high specificity required for patient care.

If one can extrapolate this curve here, once you get about 4500x coverage using our assay at least, I know you can get pretty accurate short insertion and deletion and detection going on down to 1% and below. I talked about and we've heard all morning about the importance of concordance with orthogonal platforms. We know that tumor heterogeneity exists, we know that tumors evolve after essentially targeted therapy but to what extent is this happening. If you do not have an accurate, an analytically validated assay you're going to see wild discordance all over the place.

Tissue concordance is really an important thing to do but if you find alterations in the blood that aren't there in the tissue again it's equally important that you take a meaningful proportion of those and actually see whether these are real. We took 87 of these. Over half of them at minor a little frequency less than 5% and we were able to show that everything that we found at least in this study in the plasma we could actually confirm with orthogonal testing.

Here is the result, and we've presented this publicly many times, of the analytic validation of our study. We think we do incredibly well for base pair substitutions. The short insertions and deletions and the gene fusions. We really suck at the focal amplification and why do we? The question that was asked earlier is, do you think that you'll be able to get focal amplifications from most patients with a blood based assay in ctDNA?" The answer is absolutely not.

If you think about the easiest clinical scenario perhaps HER2 amplification. Which is amplified at 16 to 20 copies and is typically deployed two copies. You can mathematically model what level of a tumor or DNA you can find these down to. I think you can get down to around 5%, maybe you can push it to 3% tumoral DNA. The vast majority of patients have much less than 3% DNA. Then when you're thinking about C-Met where it's typically amplified at seven copies or eight copies and you're discriminating from a background of three, four or five copies.

Then once you drop significantly below 20% tumor content you really are just guessing and what we have decided to do is when we statistically have enough data to make a call, we will give a call. When there's statistically not enough data to make a call, we will make a no call and call out that no call. We really have to remind ourselves that this is sick patients with cancer. Patients should be getting the standard of care diagnostics and therapeutics and there should be no room for error or emission in the care of patients with cancer.

We know that many patients shed insufficient tumoral DNA into the circulation and these patients will always be negative. Understating which these patients are and then pushing them away from the liquid biopsy which again is tremendously exciting but it is not appropriate for all patients. We absolutely can't say this enough times. This is patient care. If the company offering a clear based test hasn't done an analytic validation it does not know what the accuracy of its test is, neither did the treating physicians and the care team and neither do the patients.

Finally, in the tissue based testing we know somewhere between a quarter to a third of non-small lung cancer patients are being missed for on lAbrahaml standard of care therapies and this is likely to become much worse when we start looking at the liquid biopsy unless companies start developing their tests optimizing them and validating quite a lot. With that I will thank you for your attention and can we have a quick couple of question before and then go to the …

Male: Very nice talk I wonder if you could just very quickly comment on if in your sequencing approach you're using any molecular barcoding and if so what approach? Also along with that are you using any supplemental bio informatics that are focused at the barcoding type methodology you may be using.

Dr. Stephens: We've publicly presented this at AACR. We believe it's absolutely essential that you do have error correction and it has to be double stranded error correction as Geoff Oxnard and others may have presented this morning. When you get 5,000x coverage, this 500,000x coverage, unique coverage of double stranded DNA. It's very tempting to use half of the PCR products which may be derived from single stranded DNA.

If you do that you use your ability to error correct. I think that you cannot underscore the importance of advanced computation algorithms. Again we've publicly presented a lot of these data and we only need to submit a manuscript for peer review at the end of Q3, early Q4 this year.

Male: Thank you for your presentation. We've talked a lot this session about bridging the gap between contrived specimens and clinical specimens. It wasn't clear from your slides how exactly your company went about that. Can you comment on that?

Dr. Stephens: Yes, sure. Just correct me if I'm wrong, I'm looking around a quarter of and Geoffrey actually did the work and Travis. Rule of thumb about a quarter of them were cell line based, about a quarter contrived in half patients. I'm getting lots of nods so yeah. We did bias it towards patients but again using these contrived samples you can really do a more thorough evaluation.

Male: I'm not a molecular biologist but explain to me. You say you got close to 100% concordance on sensitivity and yet you cut off is it 1% a little frequency? It seems incredible.

Dr. Stephens: The first thing you do you have to measure a meaningful number of genomic alteration. For INDELs we measured 168 of them over a wide range of minor a little frequencies with a bias towards the lower mutation frequencies and we were able to find a 168 out of 168 with two false positives. If you get 5,000x coverage and you have a 1% mutation. You should be getting 50 reads covering that. Okay thank you [inaudible 00:43:58] is always hard for me.

It's not unreasonable that you should be able to push the sensitivity down to below 0.5% and ultimately it could get as low, as about one in thousand. The real challenge is and we may get to think about it this afternoon. When one thinks about monitoring and BCR-ABL translocations. One in a thousandths accuracy may be okay for some monitoring applications but for many of them you need to go down orders of magnitude more sensitive.

Again I think he liquid biopsy the same as next generation sequence is doing an incredibly fantastic things in the research but it's when and how you take this promising research grade technology and apply it to diagnostic assays to help patients. Because again it's been said multiple times a bad diagnostic assay is bad as a bad drug.

Reena: Perfect timing, we are moving on to session three. I invite all the panelists. As they're moving up I'll introduce them. In addition to Abraham and Phil, we have Dr. Eerkes Eerkes. She is the Vice President Strategy and Clinical Operations for Resolutions Biosciences. We have Dr. Talasaz Ali Talasaz Co-founder, President and CEO of Guardant Health. We have Dr. Koch Koch Vice President and Head of Global research Roche Molecular systems he was there for the case study. Dr. Motley Head of Clinical Development Medical Affair [inaudible 00:45:44] and Dr. Mark Sausen Vice President of R&D PGDX.

You heard the three different topics that Abraham covered in his talk and Phil also covered in his talk. What are the requirements for the plasma to tissue concordance? Then what are the analytical performance will be the second topic and what are the important considerations for establishing analytical performance. The third topic is the use of contrived specimens. When our contrived specimens suitable for analytical validation.

What we will do, we do have as I said earlier some of them volunteered to present their ideas on how they do this. They will present that and then the panel will discuss, the rest of the panelist will provide their input on that respective topic and they'll move on to the next topic. Once we finish all the three topics then we will open up for questions from the audience.

For the first topic as I said it's going to be about the plasma to tissue concordance we already had a lot of questions. What is the, why would some plasma to tissue concordant studies be more or less informative. The panelist has to comment on the relevance of specimens with or without clinical outcomes. What are the caveats of conducting and interpreting retrospective studies and what is an appropriate clinical use? We did the reflex testing, is there anything else that panelists think that will be good for this. With that I think we do have Amir from Guardant Health.

Dr. Talasaz: Great thank you very much for the opportunity to present our experience with liquid biopsy at Guardant Health. I thought maybe it would be good to first start with data that we have from a cohort of 15,000 patients that we have processed in our lab. Let me first go over the details of the cohort. These cohorts are representing 50 plus different cancer types, all solid tumors with stage of three or four. Generally, they were second line or higher in terms of the stage of the patient in the journey.

In 83.4% of cases we found some alteration in the circulation and when we look at the landscape of somatic mutation that we found. We found pretty high similarity with the landscape that has been previously published in TCGA. Except some notable differences in resistance mutations. Then we actually look at the ctDNA PBB versus tissue biopsy for the cases that we got access to the molecular part reports for the match tissue. If you look at the PPT of our findings.

First we're going to look at the [trancol 00:48:58] driver mutations. What we found was the PPT ranged between 94% to 100% and one thing to note is what we experienced was the PPT of the lower MAF variant. By lower MAF I mean less than 0.5% still we found PPT was 94.5%. This is very important since the median of a little frequency that we've seen in the cohort was like 0.4%.

One other thing to note there was a single ctDNA negative, ctDNA positive for tissue negative out fusion case. That we found a patient got treated with Crizotinib and responded to the treatment. When you look at the [subconal 00:49:45] mutation actually the story changes and the PPT drops 13% to 33%. Which is likely due to the resistance and emerging discordance mutation. When you look at another mutation like ERBB2 amplification since it could be [trancol 00:50:01] driver or it could be emerging resistance. We found 74% as the PPT.

Back in 2014 we launched our prospective clinical utility study in advanced cancer patients. Where tumor tissue was not available the patients got genotype by liquid biopsy and got matched based on what we found in ctDNA. The data that I'm showing is inter analysis for gastric cancer and none small cell lung cancer. Out of 78 gastric cancer patients 10 of them got ctDNA match therapy mainly drive by ERBB2 amplification and for none small cell lung cancer out of 72 cases. 17 patients got matched to ctDNA match TKI they were easier for out fusions Cases. We found very favorable and strong response rate for both cohorts so far.

Now to share some of our perspective in terms of plasma to tissue concordance base and the experience that we have and some of the viewpoints that we have in terms of relevance of specimens with or without clinical outcomes. We believe that clinical outcomes are necessary in establishing normal clinical application. Having said that extending existing clinical application should not require repeating the same study.

ctDNA and tissue drug genomic DNA are basically two manifestation of the same biological entity and may be expected to have similar clinical implications. A proposal here could be met with comparison to original FDA approved CDX methods, but using clinical samples that are representative of patients in the pivotal drug trial. Clinical outcome we believe may be required to explore new variants or new applications.

In terms of caveats with retrospective studies we've run many bio-bank samples and there are some experiences that we learned that I would like to share. Although they're representing the same biological entity but due to biological clinical and technological reasons. In retrospective samples we observed some discordances which could be due to treatment intervening time. Also when you collect sample plasma and tumor tissue at the same time due to the disease heterogeneity you should expect some discordances.

Then orthogonal reference method could basically adjudicate this tissue to plasma discordances. In terms of appropriate claims actually we also in agreement with many talks that we had today that there are many different reason that ctDNA levels are actually varying across different patients. Due to these variations and their effect on PBA ctDNA testing is inappropriate for definite rule out at the time being based on the current way that we're analyzing ctDNA.

A variant not detected ctDNA results should reflect to tissue based testing when biopsy or biopsy is possible. We've shown some preliminary data and there are some emerging literature so we believe variants detected in ctDNA but no tissue should be considered equally actionable as concordant calls.

Dr. Eerkes: I'd like to say thank you very much for the opportunity to participate this has been a great workshop. I won't belabor the point on our opinion on the tissue to plasma concordance. I think there are some concordance in the audience today about the fact that it's unlikely to be perfect and for the obvious reason that I think have been discussed very well so far today. That being said there really can't be an appropriate method validation without having some access to clinical specimens and some point of comparison.

There will probably always be some tissue to plasma concordance that's necessary for any approved assay. The opportunity though presents itself to think about some alternatives and we've covered most of them today or at least some ways to augment the analysis that you might do. One that I'd like to add to the pile is plasma to plasma concordance and you can imagine that this could take place in a few different varieties.

You can look at plasma that's blinded, taken from the same time point and assessed with the same technology in the same lap and two different laps or look at longitudinal analysis across different time points in the same patients. Of course that should probably be augmented with some information about clinical data such as what treatment the patient is on when they may have progressed, what specific outcomes you might be looking at.

The other augmentations I think we've gone into that everyone agrees, it's incredibly valuable to be able to look at well clinically matched cohorts, retrospectively or prospectively and compare plasma to the current standard of care which is primarily tissue in most cases and look at the real outcomes that we are all trying to measure here for patients. Which is do they respond to drugs, does it improve their overall survivability and outcomes.

Then lastly of course I think we know that once we've reached the limits which you should define well ahead of time but when you've reached the limits of your cfDNA detection technology reflexing into tissue becomes the obviously option in order to ensure that you haven't caused a false negative to affect the overall standard of care.

Reena: With that I'll open up for discussion with the panel members. Maybe I'll start from Dr. Lee.

Dr. Lee: This discussion around reflex to tissue is very important because as we talked about earlier today. There's a very heavy educational component that needs to be addressed that most practicing physicians I think in the absence of that education may inappropriately interpret a negative result and not reflex the tissue appropriately. Thereby deny patients access to the appropriate care. Hopefully that will be addressed by the other panelist is how we can as test developers in commercializing these tests actually address that educational need.

Dr. Stephens: Maybe if I can just make a quick comment. I think the standard care in most clinical scenario should be tissue testing first until proven otherwise. If you have an FFPC biopsy available that should be interrogated first. If you have an assay that has been analytally validate and you know the performance is good. There is strong rationale to use it in the resistance and recurrence of the targeted therapy and all other scenarios.

We should actually really understand what the concordance is. In thousands of patients with any given tumor type and understand how to evaluate those differences. Moving away from the tissue into the blood is going to happen but it needs to happen at the right pace and at the appropriate with appropriate regulation and done responsibly. Because if it's done too soon it's the patients who are going to suffer.

Dr. Talasaz: Actually we agree with this educational gap. What we are promoting actually the liquid biopsy now commercially is really at the time of progression and especially for the cases that you don't have access to tumor tissue and you can consider biopsy or basic blood first kind of a method that we're proposing. One thing I would to share with you is we've done it still it's not published but when you look at the community practices and it should not be surprising.

You just look at ALK and EGFR testing in non-small cell lung cancer you can figure out that just 60% to 65% of the patients are getting tested just with these two. You look at the other guideline recommended bio-markers the number sharply drops to about 10% to 15%. Just the amount of testing that we do for ALK and EGFR shows that there are some inherent barriers in the system that currently the patients are under genotyped.

One advantage of liquid biopsy could be that we can increase the standard of care for these patients but I agree that if we can find anything in the liquid we need to really communicate it very clearly with the doctors that negative test is a tumor not detected not that the variant is wild type.

Male: Just to build on that a little bit we would all agree that there are certain indications where the plasma based liquid biopsy has certain advantages over tumor based biopsy. Whether it's at the time of progression or it's too many clinical risks associated with getting that biopsy. The plasma to tissue concordance question is really a secondary question to what Abraham had previously discussed in his presentation about in the ideal scenario you would have outcome data associated with your liquid biopsy results.

Going forward we as diagnostic developers need to partners with those on the therapeutic side to be able to develop these datasets in the appropriate indications to really demonstrate the utility of this type of testing.

Dr. Koch: I Suppose in the real world, although I'd like to have it I don't know if I can, I'd like both tissue and the blood result. The reason why being particularly when there's disseminated metastatic disease we don't sample all of the lesions. There's an intrinsic assumption that whatever we have biopsied is truth and it represents some kind of truth but it doesn't necessarily capture all the biology of the cancer in that patient. It would be potentially nice to be able to do both. I just don't know if it's affordable.

Male: This argument on heterogeneity in the liquid biopsy is valuable for that it's very dangerous because we have to know what volume, what level having a clone of cells is clinically relevant. Just because you pick it up and it may not be in the biopsy you did it may not be relevant. You acting on that as one of the presenters said could be

dangerous for the patients. As I mentioned and I will swear on the bible I saw that data of Geoff's and that plasma positive, tissue negative, tissue negative, plasma negative no difference. A patient who is at least in the resistant situation where you're looking at a none founder mutation. There is no data that if you don't find it in the tissue that that patient … Geoff I'm going … Why don't we project [crosstalk 01:01:50]?

Geoff: Finish your question.

Male: This is an important point that this plasma positive tissue negative doesn't just because it's heterogeneous and may be real doesn't mean it's clinical relevant.

Abraham: Walter I'm going to echo your point. I'd love to get plasma and tissue on everyone. You know what, I'd love to get multiple biopsies on everyone, I'd love every patient to have three biopsies then I'd know about their heterogeneity right? Except the patient comes to me and I tell them, "On my trial you get a blood test and two biopsies," and they say, "I can go order a clinical blood test over here and skip your trial. Why would I do your trial?"

If I say, "To go through my science you get to have a blood test and a biopsy," they say, "But I can just get the blood test from my doc." The fact that this is commercially available from everyone means we're unable to do that definitive science that does blood test and biopsy for everyone. This is the conflict we're in. I would love to write that trial that allows me to have concordance of tissue and plasma for everyone but patients have moved on, they just want the plasma.

We have to struggle with that problem that if we require everyone to have both all the time we're just going to make trials that won't be completed and we have to come up with some kind of flexibility that addresses both. I don't know the answer there but the fact that this is such clinical reality is making it harder to meet the rigorous science that the FDA and everyone wants and we're kind of stuck. I don't know if anyone hears me on that or anyone else clinically has experience with that as an investigator but I'm having a hard time with designing those trials.

Male: Don't you think any trial should be done where we collect both and we have the information because today we haven't gotten the [inaudible 01:03:23] information to allow us to make a decision?

Abraham: I'd love to do it I just need it to be recognized. We get to heterogenotypes it will just get harder and harder if the plasma test is clear and clinical and produced for the shiny report that says 'good to go,' to tell the patient, "No, no, no hold off let me get the biopsy first." Sometimes it will work out but sometimes it won't. I'm struggling with how to make that into a prospective trial where I can predict how often patients will comply versus not.

Dr. Koch: Maybe I just make quick about both things. Our assay has been out since mid-2014 and so far just on the [inaudible 01:04:05] the up to date data is maybe we've processed over 10,000 lung cancer patients. That that's true that in the beginning we did not know if you have a driver mutation at 0.1% the patient would respond or not and still we haven't done a systematic study to show that what's the real clinical cut off in terms of response. There are a lot of anecdotal patient stories that we've heard from our partners that even at very low NAFs for driver mutations they're getting good response rates for patients.

Regarding the clinical studies as the right side of the prospective there are some challenges that we have in terms of doing the right studies. When you basically pitch these kind of prospective clinical studies that based on plasma just go and treat the patients. This is not a hot pharmaceutical study for many of the centers both on the academia and community. For this kind of studies typically we go to international sites with very small communities in order to be able to do this kind of trials. There are some kind of limitations in running these clinical utility studies even on the device side.

Dr. Stephens: Maybe to more directly address your question of patients wanting the easiest option which is blood. You're going to have the data to persuade them the biopsy is better than the blood. In a lot of clinical scenarios, we already have that data and it is the minority of clinical scenarios where it is compelling. As we start thinking about new discoveries mutation burden in non-small lung cancer, melanoma, bladder cancer, MSI, high choleric to an endometrial patients. A significance proportion of patients are actually responding and this looks to be the clearest biomarker response for immune therapeutics.

Currently you need to do about a mega base which probably in the near future is only going to be suitable for solid biopsies and not yet doable in the blood side. Over the time as we get the right data there will be some clinical scenarios what is an ambiguous yeah absolutely with a good test a liquid biopsy is great and there's going to be a lot of scenarios on the other side and it's going to be something in the middle.

Reena: Can you hold on to the questions, we have two more topics and then we can come back to that. Okay thank you. The next topic is about the analytical performance characteristics. Again Abraham discuss what are the important considerations for establishing but I would like to get the panelist's opinions of these three. What are the appropriate orthogonal method validation, how do you do the limited detection and what do you think the best way to do it? Then what are the potential for level calls and how do you select the cut off? How do you evaluate the three different topics Abraham discussed consistency, comparability and uniformity? With that I believe Resolution Biosciences have a slide.

Dr. Eerkes: Obviously the analytical performance measurements, it's a fairly broad topic. There was one thing that we thought worth observing before we talked about some of the specifics such as LOD or sensitivity specificity. That is there is a characteristic that we measure in all of our clinical specimens that we refer to as the actual complexity of library and we measure it with a unit called genomic equivalence. It really does affect the other measurements that you make and it was actually re-assuring to see the foundation data and your specificity around how many copies of the genome you were actually looking at because the GEs or the genomic equivalence is essentially that.

On average how many fragments of cfDNA actually overlie the region of the genome that you're trying to assess. The reason it matters, take for example a patient that has … this is pre-amplified input into the assay and they have simply 500 genomic equivalence. What that means is that if the circulating tumor fraction of that entire amount is only say 0.1%. Then on average you'll see one or more ctDNA fragments at any location only 60% of the time.

Which means that your lower limited detection, your theoretical lower limited detection is simply not reproducible at a clinically meaningful level. If you have a fraction of ctDNA that's 1% in that mixture, then your odds of seeing at least one molecule go up

significantly into a range that has some clinical meaningfulness. If you compare that to a patient with 5000 GEs now your ability to actually have the change of detecting a circulating tumor fragment out of your fraction of cfDNA can get much lower with much greater reproducibility.

The end result is that when you publish numbers on your LOD and the reproducibility or precision of that number it's not a one dimensional number but it's actually a two dimensional metrics that's going to vary depending on what the complexity of that specimen is going into the assay prior to any amplification. That's an important feature of the way to determine some of these other metrics because clinical specimens vary, patients vary and the amount. Not just in the mass but in the complexity of the library that you put into your assay.

Next slide please. One of the other topics for this was orthogonal comparators and we've hit that pretty well in the sense that we've talked a lot about tissue to plasma comparison. It's well worth adding plasma to plasma comparison as a ddPCR and in our paper published on our platform with Dr. Yani and Dr. Oxnard. We did look at all three methodologies and of course what we learned is what we've already said. They're all different methodologies, they all have different caveats and requirements for their performance and validation and therefore trying to use them as perfect comparators is going to be a challenging thing.

None the less these clinical comparators are very, very important and as I've already stated in addition to that you can augment some comparison work that you do with the idea of the plasma to plasma comparison and this does become especially important when you have indications where you can't collect a tissue specimen at the same time you can collect the plasma specimen or you can't collect a tissue specimen at all. This can be one useful augmentation for your overall verification of your assay.

Dr. Talasaz: Regarding orthogonal comparators again some lessons that we learned I would like to share. We believe orthogonal comparators are essential for studies in terms of comparability and accuracy but one thing to consider is orthogonal doesn't mean truth at least in the field that we're in. The traditional NGS methods, digital PCR different kind of off the shelf technologies all have some kind of distinct qualitative or quantitative error modes that we should acknowledge and consider when we're looking at those methods as a comparator.

Quantitative accuracy and reproducibility is very important for LOD studies and LOD verification. Quantitative accuracy and sensitivity particularly is very important for accuracy studies especially when we are just looking at the performance of the test at low MAFs. In terms of LODs and cutoffs probability of detection is independent of accuracy of the detection. For instance, LOD 95 what's the level that you can detect at mutation with 95% probability.

Below this MAF however the probability is not zero and this calls may be both accurate and really clinically useful for the patients. In fact, as I mentioned in over 15,000 patients the median MAF that we see is 0.4%. Some LOD calls however must be subject to accuracy studies independent to LOD or near LOD studies when you want to report sub-LOD variance. Accuracy studies that belong together all MAFs can mass poor performance at low MAF.

Also accuracy studies should be supported by independent reference methods. Always in our validations we use normal set and clinical samples both. Internally we are making standard reference materials but we believe it's going to help the field if we have standardized reference methods and reference materials. For different type of variants and context in terms of genomic context we go through a rigorous validation and studies to look at the precision and reproducibility of our calls.

Reena: Again if other panelists your thoughts on these topics.

Male: I was very excited to see the presentation from Dr. Stephens and the slides earlier here from resolution. Just indicating that many of them same challenges that all of us are facing have a number of different pass forward. Maybe one theme we should emphasize is the need for transparency. That since there isn't a right answer and a standardized approach at this moment that being very clear as to how one is establishing the metric for success on the analytical front.

The methods used contrived specimen, if it's clinical specimens, orthogonal methods. That those are clearly spelled out and that everyone has a chance to evaluate these for what they actually are. As opposed to using terminology which actually has different

meanings for different people. That would be an important thing going forward as this transparency of reporting.

Male: We spent a lot of time talking about the analytical performance and you hear a lot of this test is more sensitive than another one higher sensitivity or specificity. I still fundamentally think that the clinical correlates is much more important. As far as I know no one has ever done the following analysis for melanoma that has 50% [inaudible 01:14:49] mutation versus 5% versus half a percent. Is there a difference in the response rates across those within the tumor itself?

If that's not known it's even more difficult to associate concentrations, we find in the blood with the likelihood of response. To date for the most part if we find these driver mutations and we treat we generally see the same sort of response rates overall that have been found in tissue. We should do orthogonal comparisons to try to understand the limits of the tests that are being used but for me the clinical correlates are the most important.

Male: I agree with both of you and maybe just a quick point on both. With the clinical correlates I think a lot of the targeted therapeutic trials are now using deep next generation sequence. Which is semi-quantitative to quantitative. I think over the next coming years we will understand sub-clonality and the level of sub-clonality. That data is being generated now. What the pharma companies need to do is correct for the normal and stromal contamination and things like that.

Transparency is everything and we've got to start using the same metrics. It's appropriate to measure specificity per patient, it's appropriate to measure specificity per base but if you're measuring 100,000 base pairs it's easier to have exactly the same result, one looks 100,000 times more specifics. When we start thinking about concordance, traditionally if you're comparing ALK, let's say EGFR Exon 19 deletion to Exon 19 deletion and you find 10 out of 100 you'd say 10% concordance.

If you start looking at the negative genes on your test to the negative basis in EGFR there are ways to contrive and confuse the data. Complete transparency and using the same measurements and putting your data into supplementary tables on your

manuscript so people can actually go in and see what you've done.

Male: Just as liquid biopsy test developers run in interesting situation where we've developed these amplicon based or capture based approaches that often have sensitivities at which we wouldn't even get enough DNA to be able to evaluate mutations that occur at levels that are below one and 10,000. It's important to think about that in the context of as we do these studies where we'll looking at outcome for patients at different frequencies.

How we design those studies on the analytical side to be able to provide claims around certain DNA inputs and frequency cutoffs. That allows you to continually improve upon how much DNA you get from an individual patient because you're always going to get a certain amount of DNA from patients you screen but for not insignificant amount of them. You may get 10 times more where you'd have improved analytical performance.

Male: From Gordon we heard about sub-LOD performance that it may be potentially clinical important and that those studies would need to be characterized separately. I'd just like to get broader feedback from the panelists, their perspectives as far as pushing cutoffs lower to below LOD and how one would characterize that. If you're getting to below LOD it is a level where you cannot detect it consistently. It could be random, it could plus on characteristics.

Addressing with that technically if you do it indirectly if you have contrived you need to do dilution then of course how well do you do dilution? What are the panel's thoughts as far as just this concept and how one operationally tries to get at this? What is an expected Sub LOD detection that is not contamination, that's not noise, that's not artifact? That is potentially real but not consistently detected if it were there?

Male: As far as I know the FDA requires you to determine a limit detection that is the level that you get to report. Sorry I'm not trying to be flippant. Clearly you can do hit rates below and I showed down to 0.3x LOD and you'll detect at some frequency but not reliably enough that you want to make a call on it. That's why we do limit of detection or limit of quantification so that we know where those limits are.

Dr. Talasaz: Maybe just I make a follow-up comment here. What we are dealing with here in the field of rare genomics is something that we need to think a little bit broader. How we should define performance matrices. For an assay that meets the clinical need that you can reproducibly detect. Actually you can detect the driver mutation at the frequencies that I just mentioned. You're dealing with less than a handful of mutated molecules just for putting a copy number variation aside.

That basically a lot of performance is going to be limited by sampling errors. Just LOD 95 that 95% probability of detecting something at a specific level and you don't report below that. We are going to limit the clinical benefit of such a test with an inherent physical limitation of sampling error. We believe actually we should figure out a way to have a cutoff which is lower than LOD, but a cut off that you have very specific. What goes through rigorous validation that for the sub-LOD how specific your test is.

Male: I'm [inaudible 01:21:11] touching on some things a little close to my heart. Look the LOD is the LOD Walter you said it right? To Amir Ali's point if you have an LOD and this is what everybody in the FTA probably needs to close their ears because this is LDT world, this is the world we do because the reality is these things get reported out all the time.

I think the best that you can hope for is, and this is why you should close your ears, is that you need to provide transparency. If you're going to do that then again Amir Ali the point is when you've done the experiment to characterize what your LOD is that is your limit of your detection. If you want to report something below that and a pathologist or a physician wants to act on that that's their prerogative. I think it is at least the lab's responsibility to make it abundantly clear that's exactly what you are doing is reporting something below your validated LOD.

The problem with the transparency and maybe I'll talk about this later but it ought to be third party verified right? Not just a claim by an interested party about what the performance is. You can provide all the transparency in the world that you want but if it's your data and your claim with respect I'm a little skeptical.

Male: It's a tough act to follow. We've seen this with virtually every test that we deal with. Ten years ago PSA level 4 was considered normal. Then it goes down to one then it's 0.5 and the same thing happens with the depth of sequencing. A few years you were happy I

watched foundation report. First they have it at 500 then up to 1000 and now we're talking about 100,000.

It seems to me that with all the technology if you have a test that performs to a certain level and it's consistent why can't you just do an independent test and show that circling DNA performs at that particular level. Don't kill yourself with the lower, you'll just say what it is, the test is going to work in every patient. Here you have an opportunity to essentially get the multiple samples relative to the cost of a biopsy. If you can spend of course it's $5000 for one biopsy and you have a variable yield grade. How many blood samples can we get potentially for circling DNA?

Let's say you're in a place where there's absolutely no possibility of doing a biopsy. It seems to me listening to all these presentations we could be learning a tone if we stop beating each other up as to which one is giving different information and just see what you get with that single test. If you think about the trials that you do. In my own experience if you collaborate with pharma in a big phase 2 trial that takes forever to get answers because the correlator is always done after the trial is complete.

What people in the real world want to know is how do you actually use the drugs that you have and that's going to require standard of care approaches for which there's going to be no support. That's where we can have a discussion about the role of third party players in evidence generation.

Why not create that scenario? You've got very good technology and you keep trying to push it rather than just saying, "Look at this level we can do X and then when it gets down to a lower level we'll do X plus Y," but give us the X consistently. Not to speak for you if I can I think we'd be pretty happy because we can make real decisions based on what you do almost in your sleep.

Reena: Can I move on to the third topic and then we can come back? Sorry Prakash we can talk. Moving on to the third topic. This is about the use of contrived specimens which we struggle a lot actually. When we can allow using contrived specimens, for what analytical validations studies and again we do have some slides and then we'll have the discussion on this. This is Resolution okay.

Dr. Eerkes: We've definitely heard it said to day that everyone would like a clinical patient that is infinitely renewable, that never changes and that has ever positive analyte that we'd all like to test. That we can use as an infinite resource. That unfortunately isn't true and so we do rely on contrived specimens and it's very worth pointing out that most contrived specimens if not all of them do depend on in some way shearing of sisonicating, some form of genomic or artificial DNA.

There's a few very important differences that come out of this that you need to take into account as you work on your assay development or your validation. One is that the size ranges are very different between genomic and the punctate ladder that you see in cfDNA from patients or wild type healthy volunteers.

Another important difference that you see after you sequence is that the start site of the fragments that you make in gDNA are very different than the start sites are endogenously produced in cfDNA. There does seem to be a preference for CC dinucleotides which may relate to a biological function nucleuses and how they clive cfDNA. That is very clear when you look at the start sites of cfDNA.

What that equates to is in your assay you should expect to see differences in performance that may be region specific depending on the actual motifs of the DNA that you're analysis. Between the contrived sample based analysis that you do and the cfDNA based, the clinical based analysis that you do. That's one important point to make next slide please. That being said we don't have that infinitely renewable cancer patient that meets all of our requirements.

We would depend on contrived specimens and there are some characteristics of these that would be useful to agree on, to just point to every one's point that we would love to be able to do cross platform equivalency studies, commutability studies and proficiency studies. Having proper contrived specimens would help all of us in doing that and properly having the transparency that we require for this industry for the physicians and for the regulatory environment.

We all want them to have [snoves 01:27:28], INDELs, fusions, copy number variation

and in that I mean not just copy number amplification. Which is a primary focus but a very important lesion that I think most cfDNA assays in a robust form need to begin testing and we need contrived specimens for. In our labs certainly is copy number loss both single and biallelic loss. These are very important lesions in cancer. It's a required feature of the robust cfDNA assay to begin to explore being able to make distinctions in that.

Obviously you need to be able to distinguish your contrived specimens from your actual clinical specimens if you're using them as positive controls. Most contrived specimens are going to target some of the more common somatic lesions you find in patients. You don't want to induce any false positives or false negatives because you've made a mistake in what you call. Then of course there will need to be some harmonization so that different platforms can use these as different types of controls or proficiency testing.

We've been lucky enough to collaborate with a number of different vendors that have contrived specimen material available now. I'm showing some data from the work we did with the Mocha lab at NCI. We think that wow not all of these features are available today, most of them are and the additional ones are very close. We were very happy with the way that many of these contrives specimens performed. We're advocating that for now there's a need for contrived specimen for the reproducibility that you need to have in your analytical validation.

Dr. Talasaz: Need for clinical specimen I think it's obviously that contrived material is really necessary especially for this field. When you're trying to do a high degree of characterization, really have a high degree of manipulability. You want to really create and study rare mutations and normal mutations. Also have basically a material which is highly reproducible in terms of manufacturing and material generation. However, this contrived samples needs to be carefully controlled, especially to ensure commutability and I'm going to have some of our experience in the next slide.

At the end it needs to be studied in the context of clinical specimen. There are many types contrived materials available but based on our experience they differ substantially with the way cell free DNA really works. The commutability of cell line derived materials must be rigorously demonstrated especially at low MAF and if you don't that the performance of the test could get over quantified or under quantified. To our

experience a good model is a blend of two more cell line derived cfDNA containing none mutation spiking to the cell free DNA that's getting collected from young healthy donors.

Based on our experience shear genomic DNA simulates cell free DNA very poorly. Synthetic material and oligos actually simulate cell free DNA properties very poorly. Our limitations even on the cell line world like you want to look at your INDELs performance there's a limited number of cell lines that have different kind of INDELs available. We are doing some [encyclical 01:30:50] studies in order to supplement the data that we generate from cancer cell lines.

Reena: I would like go get some feedback from the panelists on what are the minimum studies that you think actually clinical specimens have to be used? At least in FDA we have been saying that a low de-confirmation has to be done using clinical specimens, precision repro maybe in at least some subset of variants should be actually towered by clinical specimens and accuracy study. I would like to get some feedback on that and also we have not heard. What is a commutability study? I'd like to hear feedback from each of the panelist on these. Do you want to start, maybe Walter?

Dr. Koch: Well I'll just point out again as I did earlier the practical side of things. If you can find patients of course you want to have a mix of both contrived and actually clinical specimens. We're starting to push into a realm where it will be difficult if not impossible to identify enough if even one or two patients with the sort of monies that most of us have to spend on such endeavors. Therefore, there may be situations coming up where contrived specimens alone would be necessary to find the performance characteristics.

Dr. Stephens: The way we've approached this type of question has typically been in the context of establishing the LOD with these contrived samples. Then as Reena had said validating that with clinical samples and a lot of these different assays are covering multiple different targets and multiple different alteration types but a lot of those can be categories into different ranges whether that's the genomic context, GC content.

You can establish a sort of representative set of variants that could be used across these contrived samples and clinical samples to cover some of these more rare alterations

types that are potentially detectable.

Dr. Talasaz: You also look at the LODs mainly with contrived specimen first and then we verify LOD in clinical samples but I tell the way we are doing in LOD tends definitely not perfect. Just based on our own data in clinical samples then we titrate down those clinical samples and we verify LOD based on our data because the orthogonal method is not perfect.

For higher MAF commutability studies there are some available methods like digital PCR but when you're going to 0.1% or even lower than that and you want to make sure at low MAF your assay and your contrived specimens have commutability to clinical samples. There is no gold standard, there is no good orthogonal method that you can really verify the mutations in the clinical samples and that's one of the bottlenecks.

Reena: Now we have a few minutes for questions from the audience. Thank you for your patience holding on to your questions.

Lyle: Sort of a question and a comment. The question is what LOD? I should introduce myself Lyle Arnold at Biocept San Diego. We work on circling tumor cells and circling tumor DNA. We have extremely sensitive assays that use a blocking strategy and validate orthogonally with sequencing when we see genetic events. One thing that we very much grappled with was how do we report out prevalence of mutations?

We've heard two different things here, copies that Walter puts into is copies per mil in terms of sensitivity and everybody is talking about percent. What we've actually resorted to is reported out both ways. Copies of mutant per mil, the copies of mutant per wild type. We have reported out both percent because for some people that's what really matters and that's what you should use as some sort of LOD. Probably what matters clinically is what happens particularly if you're monitoring therapies over time to the number of mutant copies per mil.

If you go and you walk in to the edge of a table and you lies a bunch of normal cells you may greatly increase with some sample draw the wild type background. All of a sudden it looks like your percent mutant is going down if you're representing as percent. We've

gone reporting both ways. Total number of wild type per mil, total mutant per mil in absolute copies and then percent mutant. Just a question for the panel. What LOD do you use?

Dr. Talasaz: I'm not sure if I'm answering your question but we think the first intended use statement for this kind of test is actually a qualitative test not a quantitative. A lot of actually clinical use cases that we are seeing is when the patients are getting treated based on what we find in liquid biopsies especially for the cases that tumor was there. Quantitative aspect of liquid biopsy still the research needs to be done. Even if we report our variance quantitatively we are not promoting our test as a monitoring tool because we don't even feel comfortable.

Male: That's probably in part if you're using quantitative real time PCR based assays or next gen sequencing, you're going to quantitative real-time PCR or digital methods quantitation becomes more feasible. In order to inform patient decisions presumably longitudinally. If you want to see if somebody's T790m is going up and I don't know exactly what the low cutoffs are that you should change a patient decision but one thing that a lot of docs are interested in is monitoring patients with time. In that case you need more quantitative assay in how you represent it.

Reena: Next question.

Male: This quantitative aspect is extremely interesting because there's a lot of opportunities there. Someone put up that there is a missing element of time, there is a lot of information in an individual's baseline level and how it varies in the setting of intervention. That sets a whole new bar for a performance of the assay. There's not that much known about variation of cfDNA within healthy individuals where in cancer patients in the absence of therapy.

What's the point to point reproducibility within and individual and to what extent can we have precise measurements of these changes that would then be clinically relevant? This all ties back ultimately to having clinical outcome data in the setting of all these results so that we can draw these connections because today we can argue all we want here, but we don't know what a 0.1% or a 0.5% of a 5% value really means in terms of benefit to patients. That data has to get collected otherwise it's very hard to compare

the performance in a clinically meaningful way.

Male: Can I echo that question before it gets missed okay. We're not talking about quantification. You're asking when you set your level of detection what metric should you use to set it? Should it be in copies or should it be in AE. I think you know the answer to this or do you have a preference tell us as to how it should be measured because in fact the assays report out number of reads, but then they get mathematically rejiggered to AFs and which are the ones that we all should be using as we try to set limits? Do you have a feeling on that?

Male: What I shifted to is terminology we've used in other submissions is performance at low levels. Low levels is not necessarily one characteristic. It could be low DNA input level, it could be low total DNA whether it's total DNA, whether it's tumor DNA those mean different just as Dr. Park was saying there's cell free DNA, there's tumor DNA, there's mutant or variant DNA of interest. Those are all different ways now for this transparency issue. We know how one defines it, what is the nomenclature when terminology one used but presumably one could address this data.

First is how much DNA you put in, how much you bucket that DNA into tumor DNA versus none tumor DNA. Whether you can do that or not that's one issue, how much you have of the target of interest? This is sometimes what I call the [inaudible 01:40:17] the numerator denominator problems. If you're doing copies that could just be called the numerator LOD that could be one metric. One is the fraction LOD. The fraction LOD depends on what you define as the denominator.

I'm just saying I'm acknowledging that there's definitely multiple ways to slice this pie to the extent that one is able to address all those and make it all out there if they were in more consensus as far as what to call these things that would be great. Well right now I just call it performance at low levels, how you define each of those metrics I'm not sure there's uniformity …

Male: We're all making it up.

Male: … but to extent people are interested in all of those things, if it's all their people can analyze it as they're interested.

Reena: Just one more question we're already running into the break time just one more question.

Male: I'll just keep this short then. This is for the gentleman from Guardant Health. What are you referring to by a cell free cell line DNA or cell line derived cell free DNA?

Dr. Talasaz: Actually based on our experience the best model material is when you extract cell free DNA from your cell lines. There are different ways you can get cell free DNA out to cell lines but make sure just to use cell free DNA instead of shearing genomic DNA and stuff because it's totally different biology.

Reena: Before I conclude the session I just want to let you know there is an FTA led [inaudible 01:41:43] consortium workshop on C2 that's the second version. There will be discussions on liquid biopsy and deep sequencing that will be held on September 13th and 14th. The flier is out there and if you would like to get more information contact me I'll get you the right person. With that I thank all the audience on the panelist there and for two presenters it was a great session.