The Sequencing Continuum for Clinical Research: From Sanger to Next Gen
Webinar
12 March 2014

[0:00:00]

Slide 1

Sean Sanders: Hello everyone and a very warm welcome to the *Science*/AAAS webinar. My name is Sean Sanders, and I'm the editor for custom publishing at *Science*.

The advent of DNA sequencing technologies fundamentally changed many aspects of basic research in the life sciences. As our understanding of genomics and genetics has improved, sequencing technologies have moved into the clinical research arena.

Sanger sequencing long-known as the gold standard is now being superseded by newer Next Generation sequencing technologies that have enabled rapid advances in gathering an analysis of genetic information.

However, with these advances have come additional challenges involving validation, as these technologies become more widespread and move closer to future clinical application.

In today's webinar, our expert speakers will discuss the relative benefits of Sanger and Next Gen technologies, and their application in different fields of clinical research now and in the future.

It's my pleasure to introduce those speakers to you now. Dr. Miguel Quiñones-Mateu from Case Western Reserve University in Cleveland, Ohio, and Dr. Volker Endris who joins us all the way from the University of Heidelberg in Germany. Thank you both so much for being here with us today.

Dr. Quiñones-Mateu: Thank you Sean.

Dr. Volker Endris: Thank you Sean.

Sean Sanders: Before we get started, I have some important information for our audience. Note that you resize or hide any of the windows in your viewing console. The widgets at the bottom of the console control what you see. Click on these to see speaker bios, additional information about technologies related to today's discussion or to download a PDF of the slides.

Each of our guests will give a short presentation followed by a Q&A session during which they will address questions submitted by our live online viewers. So if you're joining us live, start thinking about some questions now and submit them at any time by typing them into the box on the bottom left of your viewing console and clicking the submit button.

If you can't see this box, click the red Q&A widget at the bottom of the screen. Please do remember to keep your question short and concise as that will give them the best chance of being put to the panel.

You can also log in to your Facebook, Twitter or LinkedIn accounts during the webinar to post updates or send tweets about the event. Just click the relevant widgets at the bottom of the screen. For tweets you can add the hash tag #sciencewebinar.

Finally, thank you to Life Technologies for sponsoring today's webinar.

Slide 2

And I'd like to introduce our first speaker, Dr. Miguel Quiñones-Mateu. Dr. Quiñones-Mateu has completed his PhD in Molecular Biology at the Autonomous University of Madrid in Spain, and his postdoctoral training in the Department of Medicine at Case Western Reserve University prior to becoming an assistant professor in the Department of Virology at the Cleveland Clinic.

Recently and following a five-year hiatus in the corporate world, Dr. Quiñones-Mateu returned to the academia as an Assistant Professor of Pathology at Case Western and Scientific Director of the University Hospital Translational Laboratory. He is also co-director of the Virology and Next Generation Sequencing Core in the Center for AIDS Research at Case Western.

His current work focuses on understanding the mechanisms and clinical consequences of drug resistant viruses, including transmission and

pathogenicity studies using next generation sequencing technology. Welcome, Dr. Quiñones-Mateu.

Dr. Quiñones-Mateu: Thank you very much Sean and thank you very much to *Science* for the opportunity to discuss our research in this forum.

Slide 3

And so today we're going to be talking about the Development of an All-Inclusive HIV Genotyping and Coreceptor Tropism Assay based on deep sequencing that we call DEEPGEN™HIV.

Slide 4

First, a disclaimer and then…

Slide 5

So today we'll be talking about the rationale why we're doing what we're doing with this particular test and this particular technology. How we decided to choose this particular deep sequencing platform and how we developed the assay per se, DEEPGEN™ and the type of study that we're developing right now using this technology in our laboratory.

Slide 6

So many of you may know at this point, patients infected with HIV can live a long period of time, many years today, and this is basically a thank you to antiretroviral treatment, and these patients are taking a combination of antiretroviral drugs.

Slide 7

So today we have 29 different drugs that can be used to treat patients infected with HIV. These particular drugs target different steps in the lifecycle of the virus. We're talking about anything from blocking the ability of the virus to enter the cell, like binding inhibitors and fusion inhibitors, or even blocking the replication of the virus, like RT or Reverse Transcriptase inhibitors, integrase inhibitors and protease inhibitors.

Slide 8

The problem that we have with all these is that the virus eventually is going to escape. The virus replicates. It's what we call an HIV

Quasispecies Population. So there's millions and zillions of virus replicating in any given individual. So the virus is basically exploring any potential opportunity to escape any pressure, from the immune system, and of course, from the antiretroviral treatment.

[0:05:06]

So in this slide, what we have is basically a representation of the amount of virus that we may have, of course it's a cartoon, in patient or in an individual infected with HIV.

So these quasispecies, all these different viruses that we have there are reduced basically undetectable in the blood by using this combination of antiretroviral drugs. But eventually, in a certain period of time, they could be monitored, it could be years, the virus is going to escape. It's going to be able to develop and select the variant that is resistant to this combination of drugs.

And then as you can see here in the red, kind of reddish viruses, and eventually it's going to repopulate the population, the quasispecies population, and become resistant to the antiretroviral drug. And at that point, the clinician will have to switch the antiretroviral therapy to another three or four drugs; a combination of three or four drugs.

So how do we do these? The first thing that we need to do is to study that particular virus and see if the virus is resistant to one or two or all three drugs. And how do we do that?

The first thing that we need to do or we could do is perform what we call an HIV Phenotyping Assay, and that is taking a sample from the individual, and basically PCR amplify and create a recombinant virus that then is going to be grown in the lab in the presence of the antiretroviral drug that you want to test.

And by doing that, you will test the ability of the virus to replicate in the presence of that drug, and you will be able to see if the virus is resistant to that particular drug, and if you will have to change it or not.

Another way to do that will be performing what we call an HIV Genotyping Assay that is basically sequencing the virus and look at the mutations that we already know that confirm resistance to those drugs after 20 something years of studying HIV. We basically know every single mutation that confirms resistance to all 29 drugs that are being used at this point.

So, we perform that and then we create a profile and we can estimate if this virus is resistant to drug A, B or C. And as you can imagine, so far we have been doing this using Sanger sequencing for the last X number of years. Even today, all the assays that are being used are based on Sanger sequencing.

Slide 9

When I was a graduate student in Spain -- and this is just to show you an example of how technology has evolved and the sequencing has evolved. Back in 1996, in order for me to sequence 145,000 nucleotides, it took me more than three years to do that, and rounding these gels that you can see here, a dozen of gels and perhaps even hundreds of gels, and to sequence 145,000 nucleotides, and that was enough for actually eight publications which was very good at that time, I guess.

Today we can do this way better. We can sequence more than 600 million nucleotides in one single run. It takes only five days and that information is enough for one publication pretty much. As you can see, I don't know why -- '89 I was a graduate student. In a different time, I should be a graduate student nowadays.

Slide 10

So HIV population, like I said before, the virus replicate as a quasispecies. So it's important to detect these minority variants within the quasispecies because those are the ones that eventually are going to be selected, like I said, either by the immune system or because of the antiretroviral drug.

So by Sanger sequencing, typically the limit of detection that we have is around 15-20%. That means that we can detect roughly 80% of the population, in that particular sample by Sanger sequencing. If you want to go even deeper than that, typically what we do in the lab is molecular cloning, which is basically clone these particular viruses into bacteria then pick individual clones using Sanger sequencing. And, of course, the limit of detection will depend on the number of clones that you pick.

So if you want to detect 1%, in theory you should pick at least 100 clones. In the practice, as you can imagine, you need to sequence way more than that. But then it comes the deep sequencing. Deep sequencing is so powerful now that you can do that really easy. In fact the limit of detection depending on the assay, in our case it's 1%, it's really good. So you can go really deep into that quasispecies and eventually you can

identify this variant; this minority variant is much easier using deep sequencing.

Slide 11

But why is it important to detect these minority variants? We know now that patients are failing these antiretroviral drugs. And these drug-resistant HIV minority variants, as low as 1% of the population, could be clinical relevant.

[0:10:04]

And it makes sense that eventually those viruses are 1%, 2%, 3% in the population that you're going to see using Sanger sequencing. Those viruses eventually are going to be selected and outgrow the wild-type viruses, and the virus is going to become resistant, so that's important to detect that. The sooner you do it, the better for the patient, of course.

So HIV minority variants can be detected by a series of different ultrasensitive assays. And a lot of people have been using Allele-specific PCR, oligonucleotide ligation assays, and of course a lot of people now in the middle of four to five years have been using deeper Next Generation Sequencing.

Related to HIV, deep sequencing has been used to detect, of course, the drug-resistant variants that we're talking about. And the CXCR4-tropic minority variants and these are the variants or viruses that are able to enter the cell using one of the coreceptor, CXCR4, and that's important and we'll talk a little bit more about that later. But this is important for one of the drugs, that is being used right now, and the patients you need to detect this particular variant.

But then go back to the deep sequencing. We can definitely do a much better job detecting these minority variants using deep sequencing. And this particular technology is expected to transform the way molecular diagnostic tests are designed and performed in the clinical research laboratories and eventually in the clinical lab to monitor patients infected with HIV.

Slide 12

So for all these reasons we decided to develop a test that could select the right antiretroviral treatment using one single assay. And something that I forgot to mention from the beginning is, that to monitor all these

patients infected with HIV, you typically have to perform, using Sanger sequencing, two or three or even four different assays to cover all the different drugs and to cover all the different genes. So we decided to use one with deep sequencing and using just one single test.

Slide 13

So the first thing that we needed to do was to select the platform that we were going to use to do this. And as you may know, there are different methodologies now, different companies that develop deep sequencing. We have the 454 from Roche, we have Illumina, we have Ion Torrent from Life Technologies, and of course we have Pacific Biosciences. I'm not going to go in detail on this table, but all of them have their pros and cons as you can imagine.

Slide 14

So like I said before, most of the study that had been done related to deep sequencing and HIV has been performed using 454, and I think that this is basically a founder effect because 454 was the first methodology that was available. But more and more people have been using it and more different labs are using Illumina, Ion Torrent and Pacific Biosciences. It's just a matter of time for these to catch up.

So the first thing that we did then was to compare these four technologies: 454, Ion Torrent, Illumina and PacBio. And to do that, what we did was selected samples that we tested with all four different methodologies or platforms to determine HIV Coreceptor Tropism.

Slide 15

The first thing that we did then was to quantify the sequencing error, the RT-PCR and sequencing errors. And as you can see in the upper right-hand of slide, basically they have a rate for all four platforms which are very similar. We're talking insertions and substitutions. It was very similar among all of them.

Illumina had a little bit less rate related to deletions as you see it here. But the bottom line was that we didn't see any difference in determining HIV tropism with any of the four platforms. So this was the first study comparing the ability of the four different current leaders in this sequencing to detect minority variants in HIV.

Slide 16

Basically, all four NGS platforms successfully detected the same genetic variants, for instance at high frequencies. Any of the current NGS platforms, like I mentioned before, will be effective in a genotypic test to determine HIV coreceptor tropism and, of course, antiretroviral treatment too. So because of that, we decided to go with a PGM from Ion Torrent for several reasons.

Slide 17

One of those is because we decide we could just put batteries in PGMs depending of the volume of the research that we will be doing. It's really easy to use, and the turnaround is really fast which is vital for us, and the ability to generate data very quick. But the most important actually was the cost, not only with the instrument but running the actual instrument, the reagents and so on. That was critical for us as well.

Slide 18

So after deciding to go with the PGM from Ion Torrent, we decided, okay, let's develop the test. And this particular assay was available in the last year or so and the paper has been published in antiretroviral, antimicrobial agents and chemotherapy, and is going to be available now in April. So if you want more details I'm just going to refer you to that particular paper.

[0:15:09]

Slide 19

But here I'm going to try to summarize this study. You don't need to read the whole slide here but I'm going to walk you through very quick to the way that the assay is performed.

So basically we take the sample from any tissue or blood sample that we need to assess the virus. We do PCR amplification, and something that is key here, is that we cover basically every single target out of the 29 drugs that are available right now. So we go all the way from the 3' end of Gag, we call the protease, RT and integrase, and of course around 500 nucleotides of the C2V3 region of gp120 in the envelope gene.

So we basically take those three amplicons or PCR products, purify those, quantified and would mix them together to shear and to basically chop all the DNA, all of these together, purify again and then barcode. And so,

every single sample has a single barcode so we can track back and identify that particular sample later on. We purify again, select the size of the fragment that we're going to be sequencing, purify again. So we purify a lot. And then we equalize the libraries before we go to the sample preparation.

Right now we are able to sequence in a single run for one sample to 96 samples in a single round, which is really amazing to me. So we go into the sequencing, do the analysis with a survey. We have proprietary software to do the actual interpretation of the data. We'll talk about that in a second. And then we just generate a report and see what kind of data we got.

Slide 20

So going to the Bioinformatic Analysis, and again, this is all well-described in the paper, so I'm going to work very quick about this. We would get the data, the actual raw data from the server, from the Ion Torrent server in a FastQ file and we put it in our DEEPGEN™ software tool that we developed particularly for this assay.

Aligning and mapping is performed with this assay. The variant calling, basically calling every single nucleotide or every single amino acid in every single position of these, you know 4,000 nucleotides that we observe first then we sequence and that information is used together with an HIV drug-resistant database from Stanford to determine if the virus has mutations or resistant to drug A, B or C.

At the same time, another output of our pipeline is that we generate a FastA file with a V3 sequences and that together with a Geno2Pheno tool we can determine if the virus is X4 or X5 or dual-tropic. And again, all that information together goes into the report that we generate.

Slide 21

So in order to verify and characterize the assay we had to do an experiment. The first one was to characterize the RT-PCR step just to be sure how sensitive the assay will be. If we cannot PCR amplify, we cannot sequence, of course. So right now if the sample has a viral load of 1,000 copies/ml, we are able to sequence and to PCR amplify into sequence.

We have a success rate of approximately 50% if the viral load is less than 1,000/ml and we're continuing to improve that. The RT-PCR step is really reproducible, it's specific and it works with any HIV-1 subtype from the

group M, and it doesn't have any cross-reactivity. It works well only with HIV. It doesn't work with any other viruses there are.

Slide 22

Now, with any deep sequencing assay, it's really important to determine the intrinsic error rate of the test, of the particular assay that you're developing, because that will tell you the limit of detection that you can go. In our case, like I mentioned before, is 1%. So in this particular experiment what we did was we picked 10 identical clones from a virus, a wild-type virus.

So in theory all these 10 clones should be identical, it shouldn't have any mutations in between them or among them. So this is basically the metrics from that particular run. It was close to 5 million in rate and then based on that we calculated the error rate, the intrinsic error rate of our test.

So in general, as you can see on the top of the table, for the pool reading, I'm talking about protease, RT and integrase, and the amount of Gag actually too. The error rate was roughly 0.4% in that case. It's mostly indels as you can there. The indels is 0.22% and the error rate based on substitution is less than 0.2% as you will see there. For the V3 it's something similar also.

But as you can see in panel D in the table, in the bottom part of the table, there were certain positions that were a little bit higher than the average and that was a concern for us. And as you can see, most of those were related to indels, meaning insertions and deletions, because the substitutions were kind of low.

[0:20:13]

So we wanted to understand a little bit more of the problem that we're having here. As you can see in this particular position in the integrase 193, the error rate was quite high. So we went in and tried to understand this more deep.

Slide 23

And as you can see, there is a decrease in the coverage in that particular position, it then of course affect the error rate as well. And this is due to an extended homopolymeric region as you can there. This is typical or has been described before not for only for 454 but for the Ion Torrent

platform too. We have been able to fix that issue bioinformatically, so we basically get rid of all the indels and we just keep the substitutions. But more importantly we're working very close with Life Technologies and Ion Torrent to fix this problem using a new chemistry, using Hi-Q™ enzyme which looks really promising at this point.

Slide 24

So the next thing that we did was to understand the assay reproducibility. Basically you need to be sure that every time that you do these tests that you get the same result. And, again, without going too much into detail, basically what we did here was we sequenced four different viruses. The first two are wild-type, the last two are drug-resistant viruses. Well, we sequenced each one of them 48 times.

And again, we PCR amplified three times and barcoded 12 times, so basically 48 times we sequenced, and we are supposed to get the same result every time that we do this. We did some genetic analysis to be sure that everything was correct. There was no cross-contamination between any of the samples as you can see there.

Slide 25

The first thing that we did was to compare the nucleotide, every single nucleotide, every single of the 4,000 nucleotides that we sequenced, every of the four viruses, and every of the 48 times. And as you can see, the correlation here with regards to the analysis is it looks really tight. This looks really good.

What we're trying to say here is every time that we call it an A, a C or a T, it was an A, a C and a T 48 times which is really good. It's amazing. We extended that to the reproducibility in actual amino acids, in the codons.

Slide 26

And as you can see in the next slide, every time that we collect an amino acid, whatever the frequencies above 1%, it was really reproducible. If you can just pay attention to the bottom of the slide there, and it's basically all 48 times that we got an average 3.8%, if I can read this here, anything between 2% and 6% that we detected. So the acid is really reproducible.

Slide 27

So next, of course, we wanted to compare it with the Sanger sequencing. And like I said before, Sanger sequencing, the limit of detection is roughly 20% and we can go even deeper and detect more mutations. So the most important part of this slide here is that we compared a lot of samples from this particular cohort in Seville, Spain that every single of the mutations were detected by Sanger sequencing, we detected with deep sequencing analysis.

So we didn't miss any of the Sanger sequencing, 20% and above, which we expected but it was important to show. And more important then, we detected a lot of mutations below 20%, between 1% and 20%. This was the same for another cohort of individuals from Madrid and from another study that we're doing with another pharmaceutical company, we detected the same thing. We are always detecting more mutations that we expected, of course.

Slide 28

You can see in this slide, and this is the paper that we just submitted for publication, that in these particular 12 viruses, population sequencing or Sanger sequencing detected a series of mutations as you can see in the second column there. With our assay we were able to detect all those mutations and more. The ones highlighted in red, they are only by deep sequencing, different frequencies, anything from 3% to 17% as you can see there.

And something that is really important at least for me is, if you see sample 08-177 with Sanger sequencing, basically what is telling you there is there is a mixture of an N, an H, and the position 155 but you don't have a way to quantify that. Here we assigned a number to that particular mixture. We have close to 80% of the H and 20% of the N. That's something that you can only do with deep sequencing. You cannot do it with Sanger sequencing.

Slide 29

The next step was to do the coreceptor usage that I mentioned before, and I'm not going to stop too much here, but the acid correlated pretty well with all the phenotyping and genotyping assay and we were able to detect those CXCR4 user, all tropic viruses in both cohort of HIV infected individuals.

[0:25:04]

Slide 30

So at the end of the test, what it does is produces this kind of report where we have all the drugs that are available right now. We detect, we see if the virus is resistant or susceptible to that particular drug, and then the last two columns covers the frequency of those mutations in that particular virus.

So the first column it goes anything between 20% and 100% and the last column to the right shows all the mutation that are present in that particular sample, in that particular virus between 20% and 1% as you can see. And this could be really important for the physicians in the future to determine if they want to continue with that particular drug or not because you have a minority percentage of that virus that could be selected in the future and then the virus is going to become resistant and the patient is going to fail the therapy.

Slide 31

So in summary, we have developed and characterize a novel all- inclusive HIV genotyping and tropism assay, that we call the DEEPGEN™ that is able to detect minority drug-resistant variants and non-R5 or CXCR4 tropic viruses at low frequencies, not more than 1%.

It has the ability to multiplex up to 96 samples, which is really good. It has a turnaround time between 5 and 15 days depending on how many samples we have. It's low cost. It's really comparable to Sanger sequencing, depending again of the number of samples that you want, and we have proprietary software to do all the analyses to do that.

More importantly, we are performing -- actually we are in collaboration with a pharmaceutical company doing a longitudinal study to better understand the role of these minority variants in the clinic and see if this is something that we have to use.

Slide 32

I don't want to spend too much time on these, but we are using this technology, not only DEEPGEN™ but deep sequencing on different studies, from studying HIV in the brain to HIV superinfection.

Slide 33

And I just want to finish thanking a lot of people actually. Our group in the University Hospitals in Cleveland, Ohio, without them, this assay wouldn't be possible. Of course our colleagues in the University of Manchester, Dr. David Robertson, John Archer and Felix Feyertag who developed the pipeline that we're using, and a series of people here that helped us in developing the assay.

Especially, thank you to Jihad Skaf and the rest of the team in Ion Torrent. They have been really helpful in the troubleshooting. So with that, thank you very much and we're open to questions after.

Slide 34

Sean Sanders: Great! Thanks so much Dr. Quiñones-Mateu. It has generated a lot of great questions so far. We're going to get to those after our second presentation but keep the questions coming. We'll be looking at all of those during our second presentation and that's going to be by Dr. Volker Endris.

Slide 35

Dr. Endris completed his PhD in the Department of Human Genetics at Heidelberg University working on the genetic causes of intellectual disabilities and continued his work in that department as a postdoctoral researcher.

Since 2012, he has worked as a research associate at the Department of Pathology at Heidelberg Medical Hospital, heading the next generation sequencing facility section. His current work focuses on the implementation of amplicons-based next generation sequencing for the routine molecular diagnostics of formalin-fixed, paraffin-imbedded tumor material. Welcome Dr. Endris.

Dr. Volker Endris: Thank you very much for the nice introductions, Sean, and also for the opportunity to talk in this webinar series here.

Slide 36

So what I want to share with you today are the experiences that we may have made in our institution when it comes to next generation sequencing, using panel sequencing, especially on the type of material we are using for formalin-fixed and paraffin-imbedded tissues from cancer samples, and how we can bring this technology towards molecular diagnostics in the future.

Slide 37

So a disclaimer…

Slide 38

So I'm working at one of the biggest academic pathology institute in Germany and we are serving many hospitals around in our area for histopathologic examinations.

Slide 39

But we also have a large molecular diagnostics department where we offer a whole variety of different molecular testing including, of course, mutation analysis of cancer, prognostic and predictive markers, and we mainly use Sanger sequencing here, especially in the field of oncology research.

Also, development on other cancer drugs is rapidly a developing field and it is expected that in the near future many novel drugs will be on the market and each of these drugs might need its own personalized signature of mutations and it can be applied to the patient.

That gives us in molecular diagnostics the problem, that we have to keep track of this development, and currently we are using Sanger sequencing, so every single assay has to be sequenced. In the future, we might need to switch to the multiplexed assays to keep track and have the diagnostics on time.

[0:30:17]

Slide 40

As mentioned before, the main materials that we have are samples that were fixed with formalin and then embedded in paraffin. Although this material is stored for a long time, the DNA suffers from this treatment; the formalin introduces DNA strand breaks leading fragments of roughly 150 to 200 bp.

We have a problem with FFPE material. Through the formalin fixation, the base exchanges introduced into the DNA, especially C to T or G to A transitions, and our problem with that is the FFPE material, or the material that we have is that sometimes we only have small amounts of

tissue especially when we have a fine-needle biopsy material we have very limited and low amount of DNA only available.

Slide 41

Now, I want to share with you today are different steps that we have done to see -- and multiplexed assay using next generation sequencing can be done on this material. We developed a workflow, we test it if it's compatible with FFPE and we test several parameters, and to see if we can bring it in the future to molecular diagnostics.

Slide 42

The workflow that we're using starts with automatic DNA extraction, out of the FFPE material, followed by a twist of DNA concentration measurement, on the one hand side a fluorometric assay using the QuBit instrumentation. And then the second step we use a quantitative PCR approach. The reason why we do is that our door to QuBit measurements is much better than spectrophotometric analysis.

Slide 43

It still overestimates the amount of usable amplifiable material in the samples due to this highly degradation of the sample in the FFPE material, that's why we in the second step apply a qPCR approach which gives us a better measurement of amplifiable material, and we use both to also measure how highly degraded the sample really is in this types of later QC steps.

Slide 44

We use a standard AmpliSeq approach and we use different panels. On one hand side, we use a commercial AmpliSeq Cancer Hotspot Panel v2. But we also have custom-made entity specific panels. Each of these panels consists of roughly 150 to 200 different amplicons and we do it in a single assay. After barcoding and sample multiplexing, we normally use the one-touch equipment multiplex PCR, and later sequencing on the Ion Torrent PGM.

Slide 45

Now, we usually multiplex 8 to 12 samples on a single run. We use the 318 chip for our sequencing and this gives us approximately 2000x

coverage for each amplicon. The throughput that we can obtain with the FFPE material on-hand is roughly 700 to 800 megabases.

And what you also can see in the lower panel is that due to low quality of DNA available, we have a little bit higher ratios of low quality reads compared to FFPE prepared DNA, which can also be seen in read length histogram distributions.

Slide 46

We then compared DNA coming with good quality control, but you also see a shift towards the left, shorter read lengths when the DNA quality is not so good and this also help us in discriminating whether a sequence was really successful or not.

Slide 47

Multiplex assay has one requirement, that you have a high uniformity across all the different amplicons in your panel as shown here in this graph. The AmpliSeq Cancer Hotspot Panel, for example, has a high uniformity across all different amplicons. It looks similar when we use the custom-made panels, although, sometimes it might be amplicons that do not perform very well, but in general the uniformity is high enough.

One problem associated with the lower read length that are only available from the FFPE DNA is that sometimes you do not get a complete coverage of a complete axon. This is something that we use for Sanger sequencing.

Slide 48

In case of oncogene, for example, in this picture you see KRAS gene and the two hotspots, G12 and G14. The position of the amplicon as shown in the box is fully sufficient.

[0:35:12]

But if you are interested in sequencing, for example tumor suppressor genes and you want to have almost a complete coverage of the axon -- sometimes the positioning of these amplicons can be suboptimal. And in these cases what you can do is that you decide to use a two primer pool panel where you can use overlapping amplicon pairs, but only with a disadvantage, that you have to run two different reactions in the very

first library preparation steps. But then you get a much better coverage throughout different amplicons.

Slide 49

So in the routine setting and having different samples on hand, we sometimes have the problem that we only have very little amounts of tumor cells. We usually try to enrich tumor cells by using macrodissection; but sometimes this is not possible. And we ask, what is the sensitivity of our approach? And also, we want to compare it then with different tumor cell contents.

Slide 50

To simulate that situation, we took one sample, roughly 80% tumor cell content and having three different mutations and we then titrated this with corresponding normal DNA and sequenced all of them to see detection limits, what we can still detect and what you can see. First of all, is that you have a high reproducibility of the assay, for example 40% tumor cell content, still all of these mutations are still recognizable.

Remarkably when you only 10% tumor cell content in your sample left, you are still able to detect a hotspot mutation in the EGFR gene with Allele frequency of roughly 3.5% which is very sensitive. How does this compare to Sanger sequencing?

Slide 51

Actually, this is another example, again with the KRAS gene, the hotspot mutation at position 12 with a frequency of 5.8%. We need to carefully look at the Sanger reads. You can still see a minor peak at that position. In that case, when we just do Sanger sequencing, we would still be able to detect this because it's a hotspot mutation and we would be confident that this is a true mutation.

But it can be more difficult if you are having such a small peak at the novel mutation, so not a hotspot then you can get, of course, problems in interpretation if this is a correct mutation or not. But the same holds also true for next generation sequencing. So when the frequency really drops down you're not really sure if this is correct or not.

Slide 52

Another problem that occurs with this low frequency mutations is, as I've mentioned before, we have in the FFPE material fixation artefacts. These fixation artefacts usually are in a range of an Allele frequency of, maybe, 1% but with very low quality DNA on hand, and in case you don't even reach 10 nanograms of DNA, that you can use for library preparations. These fixation artefacts can make up and it will come up in frequencies of 3-5% and then are recognized by the variant callers.

So how to discriminate now if this is a false positive call or not? We usually look at the reads, and as shown here in this example, sometimes you observe a G position that transition to A or a C position that transition to T. In these cases we would consider a 3-5% variant that was coded as a false positive. But there's something one has to keep in mind when it comes towards low frequency mutations with FFPE material that these might be a fixation artefact.

Slide 53

Just to summarize the sensitivity screening, we would say that the panel sequencing approach with the coverage we are using is reliable and can detect mutations from frequencies down to 5%. It's uncritical usually for hotspot mutations but can be difficult when other variants have been found. And then you also have the problem, how to actually validate these mutations, Sanger sensitivity might be too low. So we have to maybe switch to Digital-PCR.

[0:40:02]

Slide 54

What we next wanted to test is specificity. So how reliably can we really detect mutations in our samples?

Slide 55

And we took several steps to do that starting from evaluation studies where we sequenced each of the mutations with the next generation sequencing approach. We also did a cross-platform testing as a multicenter research study with our laboratories in Cologne and Ahlen where we shared in a blind fashion of samples and we sequenced them on different platforms, on the Ion Torrent PGM, on the Illumina MiSeq and Roche GS Junior.

We also participated in a Germany-wide RAS family round robin trial in colorectal carcinoma. And what's on currently is a Germany-wide NGS round robin trial also involving several multicenters. We use samples from lung cancer, from colorectal cancer, breast cancer and we sequenced them also in a blind fashion.

Slide 56

And just to make a long story short, we found very high concordance between the Sanger and the NGS data. We found a high concordance of the different NGS platforms, and also between laboratories. So we can conclude that the panel sequencing approach and the next generation sequencing is really promising and it can replace Sanger Sequencing in the future.

Slide 57

But next generation sequencing can even give you more information. For example, you can easily identify compound heterozygous mutations when the two different mutations are on different read strands, or if two mutations for example are on a single read strand you can identify them even as complex mutations.

Slide 58

Or in that case we would rather consider these two mutations as one single deletion-insertion event.

Slide 59

And an additional plus of next generation sequencing approach is that you can even detect amplifications as shown here for three samples with an EGFR gene amplification. We also found a resemblance, for example, for two gene amplications and we retested them also by either using qPCR measurements, or even normalized the chemistry, and we found a high correlation of this data.

Slide 60

So want I want to present in the last slide is what do you really need in your laboratory before you can move on from Sanger Sequencing to Next Generation Sequencing in a more clinical setting.

Slide 61

Now, compared to Sanger sequencing, the workflow for NGS panel sequencing using Ion Torrent is much more complex. Although sample management might still be the same between Sanger sampling before and NGS. You have to keep track of much more information during your workflow starting from DNA concentration measurements to library preparation, so what barcode has been used, which panel has been used and what was the concentration level, and so on.

Also the entire bioinformatics is much more complex compared with Sanger sequencing. What you also have to keep in mind is that data management and especially data storage has a much higher requirement than with Sanger sequencing.

Slide 62

What we have done actually in Heidelberg now is that we set up our own NGS database LIMS system where we can keep track of all the different information coming from sampling input to library preparation and so on.

Slide 63

And we assemble our surface where we can feed in, for example, concentrations, or which panel was used.

Slide 64

So we always have a complete overview of what happened to a Sanger sample.

Slide 65

And we even feed it into annotated variants so we can simply choose one sample and look at the variants that were identified.

Slide 66

Also, the infrastructure that you need for this kind of analysis is that here you see are just many computers and servers are connected to our network just for using the NGS. We also have, of course, a larger storage capacity now for at least storing BAM files for example. We are not actually keeping, currently, the raw data files because they are just too big. But when you run them for several years, you really have a higher amount of data that you really need to store for the future.

Slide 67

To summarize our results is you can use NGS. It's feasible on FFPE material. You can use it on resection specimens. You can use it on biopsy material. There is a high concordance of NGS and Sanger data, and there's a high reproducibility not only with experiments but also between laboratories. You can even detect amplifications and when it comes to sequence, and a lot of different targets, it's also very cost efficient.

[0:45:07]

What you have to keep in mind is that the workflow is much more complex so you have to set up it better in your normal workflows. The analysis time also is longer than with Sanger sequencing. Usually it would take two to three days for a normal Sanger sequencing and up to five days for Next Generation Sequencing, but we gather much more information. And the requirement for the infrastructure is much higher.

Slide 68

With this I want to, here in this slide, thank the people in my lab but I also want to thank our collaborators and project partners that we have and where we have a lot of sequencing done. And by this, I want to finish my presentation and thank Life Technologies.

Slide 69

Sean Sanders: Great! Thank you so much Dr. Endris, and many thanks to both of our speakers for the wonderful presentations.

Slide 70

We're going to move right into questions submitted by our online viewers. A quick reminder to those watching us live, that you can still submit your questions by typing them into the textbox and clicking the submit button. If you don't see the box in your screen, just click the red Q&A icon and it should appear.

Slide 70

We've got a number of questions in. I'm going to start off with you Dr. Endris and I'm going to ask you, do you find that in addition to the

fixation artefacts that you talked about, any variability due to PCR and if so, how do you compensate for this in samples with low tumor content?

Dr. Volker Endris: The PCR introduced artefact is something that could, of course, happen when -- but it only would really pop-up if this PCR introduced artefacts would come in. It's one of the very first cycles actually of the library preparation step.

This is something that you cannot completely control I would say, but it would be the same also for Sanger sequencing because Sanger sequencing applies on a first PCR, so you could also then amplify these mutations and see it by Sanger.

Of course, if it comes to low level variants, low Allele frequency, this is difficult to answer. But if I would only have 10% tumor cell content and I find mutation with 60-70% Allele frequency I would be a little bit curious and careful actually.

Sean Sanders: Excellent! Dr. Quiñones-Mateu, a question for you. Your DEEPGEN™ assay, are you enriching for HIV virus in your samples? And, if so, do you see any bias or drop out in the assay?

Dr. Quiñones-Mateu: We do enrich for HIV virus actually. What we do is whatever the samples we're getting. It could be plasma sample. This is typical, what we get from individuals infected with HIV. We do concentrate and basically RNA purify and PCR amplify only HIV. So it's not supposed to be contaminated with any human DNA or RNA, something like that.

The second part of the questions was? If -- what's the second part?

Sean Sanders: Oh sorry, I'll find that for you. Do you see any bias or drop out in the assay?

Dr. Quiñones-Mateu: No, not really. As I've mentioned during the presentation, if we are able to RNA purify and we are able to PCR amplify, and we don't see any drop outs we should be able to sequence that, and so we don't have any problem with it.

Sean Sanders: Excellent! I'm going to come back to you Dr. Endris. It's actually a question for both of you, but we'll start with you. Could you talk a little bit about the software that you use in your analysis and whether you have any recommendations about the best software packages? Maybe you've tried a few different ones that you could suggest.

Dr. Volker Endris: So what we actually use for variant calling is we use the Torrent Suite™ Software version 4.0 at the moment. In addition for annotation purposes we use, for example, CLC Genomics Workbench or we use the Ion Report™ Software. We're also testing currently Ion Report™ Software, compared with the Torrent Suite™ Software in variant calling and so on. But it looks like that both have very good performance.

We tested also, for example, mapping and variant calling with the CLC Bio software platform and also with our variant caller, but there were some slight differences, and we are more confident actually that Torrent Suite™ Software works best for mapping variant calling actually.

Dr. Quiñones-Mateu: So in our case what we do -- and HIV of course is slightly different, and for cancer and genetic diseases, it's easier to use a human genome as a reference and it doesn't change that much.

[0:50:00]

With HIV it's so variable that the key here was to select which reference we should use to do the mapping and alignment. And this is something that we work a lot with our colleagues from the University of Manchester, like I mentioned before, with Dr. Robertson and Archer.

What we decided to do actually was to develop our own pipeline and to use this and the mapping and alignment against the best reference that is selected among 100 of the most common HIV isolates out there. So that way we reduced a little bit of the variability, but that's embedded in the actual software.

So for HIV it's a little bit more complicated, and not only us. What people are doing is they develop their own pipeline. I am looking forward in the future to see other people doing it and see how it compares to what we have. I have to say that the one that we have right now is working just beautiful.

Sean Sanders: Great! Also a follow up to that, how important is the software in the work that you're doing?

Dr. Quiñones-Mateu: Oh, it's extremely important. As you can see the amount of data that we get, 630 million nucleotides come in one single run, even more when you do just cancer and genetics. Without bioinformatics and software and all the databases that we have to use and data bank and storage, there is no way. So it's extremely really important.

Dr. Volker Endris: And you always have to look really at the reads again. So we trust the variant caller so far especially for hotspot mutations. When we are asked, for example, each of our gene mutation analysis, we really look at the reads for all the amplicons in each of the genes to really have a visual look again and that we don't miss something that the variant caller is not really calling.

Sean Sanders: Dr. Quiñones-Mateu, I'm going to come back to you with a question. I guess a follow up to the earlier one asking, are you sequencing and genotyping from cultured virus or are you sampling from blood, and how do you enrich for virus sequence for your experiment?

Dr. Quiñones-Mateu: Well, depending on the research study that we're doing but we're sequencing from everything. We can sequence from blood, from plasma, from PBMCs, meaning cells from the patient, from cultured viruses, from tissue that is infected as well. So at the end what we try to do is purify the viral RNA from there to minimize the problems, and then do the RT-PCR with specific primers. So, it should be and it is actually various specific.

Sean Sanders: Great! And Dr. Endris, I'll start off with you for this question for both of you. Coming back to the bioinformatics side, how do you store the data generated from your Next Gen studies? And can you recommend what data to store FastQ, BAM, VCF files, et cetera?

Dr. Volker Endris: This is a good question. This depends a little bit on the legal issues that are not really solved. We do not actually store the raw data because they're really too big. What we store is just the next step that comes out, for example, from the Torrent Suite™ Software where you have at least mapping and also the process files, the basecall files. This is something that we store on our server where we actually can store all these data, which has a mirrored hard drive and we also some backup capabilities.

In addition, we store also variant caller files. They can also be stored on this server but also on a different platform because then you are safe is something happens with the software for diagnostic purposes.

Dr. Quiñones-Mateu: So in our case -- I mean the first thing that I will recommend is yes, you have to invest. Talk to your IT people and invest, and buy equipment, buy back up storage and all that.

For HIV what we're doing right now is we store the raw data, the FastQ files mainly because it's evolving and this is a moving target right now. So

I can see myself in the future going back to the raw data and analyze it again.

We're lucky enough that for HIV the files are not that big, so it's between one or two gigabytes. So we're storing for the last year, so every single room we have storage and backup, but I can see how that could become a problem in the future.

Sean Sanders: So I'm going to combine two questions that I've got in for you. The one is, is there a place for Sanger sequencing in the future? Because I'm guessing right now Sanger sequencing is the only modality you can use except for -- I think there's one, a new Illumina machine that's now being used for diagnostics, but Sanger is the only diagnostic for a lot of people in technology.

So is there a place for that in the future and did the data that you collected using your sequencing modalities contribute to any translational research outcomes? Dr. Quiñones-Mateu?

[0:55:05]

Dr. Quiñones-Mateu: I would say that in the future -- depending on how you define the near future. My guess is that in the future deep sequencing is going to replace Sanger sequencing. How long is this going to take? Well, it depends. Many of these companies are developing all this technology.

One of the things that will have to happen is to increase the read length of what we're sequencing right now. That it goes anything between 150 to 400 nucleotides. That still shows with what we get with Sanger Sequencing. That doesn't count with Pacific Biosciences. They can thousands of nucleotides but that's a different issue. So that's one of the things.

But eventually, the cost is going down so much. The amount of information that you get with deep sequencing is amazing compared with what you get with Sanger sequencing, so I can see how in the future is going to replace, totally replace it.

What we're doing right now, and I'm sure that Volker is doing the same, is many times you confirm your data using Sanger sequencing, again, because this is also new. But in the future, in the next X number of years, it is going to be replaced, even in the clinical lab, and not only in the research arena but in the future, in the clinical lab as well.

Dr. Volker Endris: Yeah, I think from the technology side it's not a problem. So Next Generation Sequencing is as specific as Sanger also, that's something we would say currently. For cancer research, I think it mainly depends on the development on the pharma side. So if there are really novel drugs coming out and we have to test even more targets, then NGS approach will be much more cost-efficient.

Currently, we could still live with Sanger sequencing because there's just a handful of targets that need to be screened. But as I've said in my talk, when there are really novel targets coming out and we have to test even more and more, then Next Generation Sequencing with panels and so, will profit I think, then it will be the future technology.

Sean Sanders: So have you any of your current results contributed to translational research outcomes as of now or is that still in the future?

Dr. Volker Endris: It's difficult to say. I mean, we have done several research projects using panel sequencing of large cohorts, of tumor cells, from different tumors actually. There are interesting results coming out and we are going to publish them, of course. Now we can see, you can maybe better separate groups of tumors and have a better outcome.

Dr. Quiñones-Mateu: In our case, for HIV, it's obvious like I discussed during the presentation, that these minority variants are really important or may be really important for the treatment of the patients and the outcome of the patients. But up till today that -- we didn't have the tools to do that before, the Allele specific reactions or oligonucleotide ligation assays were too cumbersome and too expensive or too tedious to do.

Now, with DEEPGEN™ we may be able to answer that question, and who knows? It could be done in a year or two where everything is going to be replaced by deep sequencing, and those particular mutations are going to be really important and the physicians are going to look at that particular mutation, 1% or 2% is important, just switch the drug.

Sean Sanders: So I'm going to squeeze in a couple more questions. We're coming to the top of the hour. One quick question for you Dr. Quiñones-Mateu. This viewer asks if there's any software specifically for HIV sequence analysis available free to researchers.

Dr. Quiñones-Mateu: There are a few on the web. You can go and download those for free that people have developed for research-only use. I guess ours will be one of them. But to my knowledge, there is no commercial test available right now.

Sean Sanders: Okay, great! So one final question for both of you and I will start with you Dr. Endris. What do you see the future holding for this area and how do you see the technology changing in the future and how will the weight is being applied change?

Dr. Endris, I know that's a crystal ball question but maybe you can give an idea based on the work that you're doing now, how you see things changing in the future.

Dr. Volker Endris: I think in the future what has to develop is that it can be even more sensitive with respect to the amount of DNA that you'll need and maybe that you can omit actually, some steps like the emulsion PCR. So doing real-time next generation sequencing because then you also lose this potential error of PCR introduced artefacts.

And, of course, the chemistry needs to be improved, like what he had said with Hi-Q™ chemistry, we are also looking for that to have better reads through the indels. So let's see how in the future that will be.

[1:00:05]

Dr. Quiñones-Mateu: I'll say it very quick. The future is happening already. A year or two ago we couldn't even think about doing what we're doing right now. And the fact that we have a small machine like this and it's generating an amount of information that it's getting right now.

Like I mentioned before, longer reads, less expensive, which is kind of cheap right now compared a couple of years ago but less expensive, and make the protocol so the workflow is easier.

But I can see how in the future everybody, just like now, everybody has a PCR machine in their lab; everybody in the future will have deep sequencing in their lab.

Sean Sanders: Fantastic! Great final comments. Unfortunately, we are out of time. On behalf of myself and the viewing audience, I wanted to thank our speakers very much for being in the studio with us today, Dr. Miguel Quiñones-Mateu from Case Western Reserve University and Dr. Volker Endris from University of Heidelberg.

Slide 71

Please go to the URL now at the bottom of your slide viewer to learn more about resources related to today's discussion and look out for more webinars from *Science* available at webinar.sciencemag.org.

This webinar will be made available to you again as an on-demand presentation within approximately 48 hours from now. We are interested to know what you thought of the webinar. Send us an email at the address now up in your slide viewer, webinar@aaas.org.

Again, many thanks to our panel and to Life Technologies for their kind sponsorship of today's education seminar. Goodbye.

Thank you so much.

[1:01:36] **End of Audio**