IPQ INTERNATIONAL PHARMACEUTICAL QUALITY
Inside the Global Regulatory Dialogue

NGS SESSION PANEL DISCUSSION AT CASSS CMC STRATEGY FORUM EUROPE

The following is most of the panel discussion that took place after the presentations at the session on NGS use in adventitious virus control of vaccines and biotherapeutics at the CASSS CMC Strategy Forum Europe. Participating in the discussion were presenters Sebastiaan Theuns (University of Ghent), Carine Logvinoff (Sanofi Pasteur) and Domenico Genovese (Italy National Center for the Control and Evaluation of Medicines), and moderators Robin Levis (FDA) and Emmanuelle Charton (EDQM). The NGS issues that were addressed related to: ● capability vs. cost ● long reads vs. short reads ● method accuracy ● continuity of data quality when new technology is introduced ● quantifying bioburden as well as adventitious viruses ● ensuring complete DNA/RNA extraction ● ease of use of different platforms ● emphasis on virus standards ● discussions with regulatory bodies besides FDA ● product-specific validation during method transfer ● accelerating vaccine development ● confirmation of positive results ● use of nuclease treatments, and ● pharmacopeial monographs. The comments from the panel discussion have been edited for clarity and brevity.

Capability vs. Cost

Levis: I have a couple of questions for Sebastiaan which are technical. You gave a slide of all of the technologies that are available and the cost analysis going from iSeq being the cheapest all the way to NovaSeq. So my simple question is: Do you get what you paid for? I mean, is there really a defined quality differences or can we do good analyses with iSeq if that is where your resources lay?

Theuns: I think that is a good question. Basically, in the quality there is not a lot of difference. The only thing that these devices are different in is the amounts of data that are being generated. So if you want to sequence the human genome, then of course, you will not do this on an iSeq, because the data that it can generate is only one or two gigabases. If you do it on a NovaSeq, you can generate up to six terabases. So then you can analyze at the same time, a lot of human genomes.

Long Reads vs. Short Reads

Levis: Just one follow up question, because we were talking about the technologies: We were able to do long reads versus short reads. If all we are looking at is a virus database, which relative to all of the other genomes you described is relatively short, is it okay to use a short read system or are there advantages to also use long reads for looking at a viral database?

Theuns: Of course a viral sequence is sometimes much shorter. If you have a circovirus, we have a genome of 1.5 kb [kilobase]. So that is quite short. Then, it should be possible to have it in one read, if you have good protocols available.

The thing is, if you have some very short fragments that are present in your data sets, and you end up in your results with this, it might be difficult to place these results. So, if you have a short amplicon of 150 base pairs, and you see it in your data set, you do not really know the origin of it. So then you need to really have good bioinformatics tools to place it.

If you have more context – you can have a full read of thousands of base pairs – then it might give you more context to look into the bioinformatics data sets to see, is this really a true viral read or has it got contaminants? So it can give you advantages. Also for the assembly, there are a lot of advantages to use longer reads, because it will just make the puzzle much easier to solve compared to short reads.

Method Accuracy

Levis: The first one is for Sebastiaan. Can you further explain the accuracy of the method? Do you use defined reference standard DNA and RNA sequences? And, what are the main and potential biases in accuracy between the varying technologies you described?

Theuns: Yes, there is actually a general quality system for this and how to compare quality between different technologies. It is the 'Q-score system.'

So when you are analysing this data, you will get a certain Q score, which is the quality. A Q score of 10 means that you have 90% accuracy on a single read. If you have 90% accuracy on a single read, for instance, then it means that on-e in ten nucleotides can be wrong. So, if you are making your assessment on a single read, then you have a lot of chance that you will also have some wrong annotations.

The thing is, if you have a lot of data, you will compile a genome that is much more accurate, and these mistakes will be resolved. So that was what I was showing with slides of the Oxford Nanopore technology. The R10 flow cell had Q50 scores for a bacterial genome. This makes sense because then, if you have a lot of data, even if the single read has some mistakes, you will in the end make a genome that is very accurate. So that is something to take into account.

I think also for viral sequences, we should be not too critical. The thing is: What is your goal? If you are looking for the tiniest mutation, then you need a lot of very accurate data. But if you want to identify an adventitious pathogen in your sample, then you are probably not looking for one single nucleotide mutation in the genome. You just want to see if it is there or not. And if then there are ten mutations in the 30 kB genome wrong, you will not be affected by this in your bioinformatics analysis.

In the end, yes, the things that are on the market will do this now, so that will not be a big issue.

Continuity of Data Quality when New Technology Is Introduced

Levis: The next question is for everybody. It was actually one of the questions that Emmanuelle and I had for general discussion, too. But I will point it to Carine first, because I have a specific question about it. It has to do with a very quick evolution of all of the technologies and the databases and how they are used

- How do you really ensure continuity of the quality of your data and the validation package if all of a sudden new technology comes out?
- And Carine for you at Sanofi, what are the pressures to keep up with the newest technology versus utilizing a well validated platform that you have?

Logvinoff: This is actually the true challenge within the GMP environment to keep on continuing the evolution of the technology. The advantage of the validation approach that we choose with working with a segmented validation will allow us to follow the evolution of the technology. So we will not have to revalidate everything at the same time.

Fortunately, Illumina told us ahead of time when they will stop the support of some of the equipment. So that is providing us some time to prepare the next generation of equipment that we want to work with. But this is a challenge within the GMP environment, that is for sure.

Quantifying Bioburden as well as Adventitious Viruses

Levis: It seems that the technologies, especially Nanopore technology, could revolutionize adventitious virus and also bioburden testing. I would just point out to you that that is what we are hoping for.

The question is: Could bioburden in adventitious viruses be detected or even quantified in one run? Sebastiaan, maybe that is for you because you have had field experience for getting that. Then we can also hear from others about how they would think about the use of this across the testing platform.

Theuns: I think with every platform, a kind of quantification is possible [and] the pathogens that are present at the higher loads will likely induce the most severe disease.

Levis: Carine, have you at Sanofi thought about using next generation sequencing in the realm of microbiology and for bioburden analysis?

Logvinoff: Not up to now. We are wondering…if the reagent could be contaminated with some phages or things that may interfere with the detection at the end. So, we have not approached it this way. For the moment, we are limited to virus. We have not developed NGS as a quantitative assay right now. We have developed it as a limit test. We don't see linear progression. So for the bioburden aspect, if we want to quantify, we are not there yet to make sure that we will be able to quantify it.

Theuns: I think that is a good point that Carine made about the contamination of reagents – especially when you are going to look for bacterial assessments. For a lot of these, like extraction kits and resources like water, there are nice papers about this. They can harbour a lot of bacteria or bacteriophage contaminants. This is why it is important to also include negative controls in your assays.

Ensuring Complete DNA/RNA Extraction

Levis: Carine: How do you ensure a complete extraction of all of the DNA/RNA species present in the sample?

Logvinoff: For the extraction, we have worked on different kits to make sure we will recover all types of nucleic acid. I think we have evaluated 11 or 12 different kits. We decided that the best approach we choose is using two kits so that we will have both total nucleic acid but also double strand DNA that were quite difficult to have in certain cases within the sample. We need these two extractions to get the sequencing from that. So, this has been described within a *npjVaccine* paper. I think it was two or three years ago, maybe.

Levis: You described a manuscript that is going to be coming out, and when will that be published?

Logvinoff: It has been submitted. So hopefully this year, this will be available. Within this publication, we are presenting the HTS against the data that were generated within the *in vivo* study from NIH. We use both viral seed metrics and cell-bank metrics, because these are the two main steps that we want to apply HTS to. I have shown you within the slide only the viral seed, but we did the same approach on the cell bank.

Ease of Use of Different Platforms

Levis: That will be nice to see. Sebastiaan, how easy it is to use or how much expert knowledge is required to run the [Oxford] Nanopore or the PromethION. Can anybody do it, or do you have to get some training?

Theuns: We provide some basic training on it. You need a lot of preparations in well-equipped labs at this moment. Also for the data handling, the bioinformatics, you need people that are really aware of what they are doing and really know what you are analyzing.

I think these are issues that a lot of companies will be struggling with. Generating the data is not that difficult, but analyzing that is especially difficult – making something out of the data that really does make sense. We see this often in diagnostics as well.

Emphasis on Virus Standards

Levis: Domenico, there is a question from Martijn [van der Plas, Netherlands CBG-MEB], about one of your comments. The question is, while the NGS will require validation, why so much concentration on the virus standards, when we have accepted the results of *in vitro* and *in vivo* virus assays for many years actually, where we don't have international reference standards? So, why is the difference applied to NGS?

Genovese: I think that for the moment, it is a necessity to have a wide range of control. This is why we need a lot of different viruses with different characteristics. There are too many variables. I think we need more experience on NGS.

Levis: I would also add that as we have these new great technologies that lend themselves to really being able to validate and to look at specificity and look at the sensitivity, we should take advantage of that because then it broadens the scope and use of the product.

Genovese: Yes, absolutely.

Discussions with Regulatory Bodies Besides FDA

Levis: So here is a question that can be split between Domenico and Carine: It is mentioned specifically from Mats [Welin, Sweden MPA] to Carine, but I think Domenico might have a point, too. You had talked about your discussions with us at CBER on the introduction of NGS. Have you had any discussions with EU regulatory bodies?

And then I would extend the question to Domenico in general, which is also another one: Do you know of any discussions through the EMA or EU on this? But we can start with Carine in terms of other outreach Sanofi has done.

Logvinoff: We did not have an in-depth discussion as we had with CBER. But we did present a strategy to substitute the *in vivo* by HTS on seed lots with PEI [Paul Ehrlich Institute] – the argument being the technology and how we handle it, and also those 3R. Based on that, PEI, the German authorities, approved the approach.

It is mainly because as a specification test, we only apply it to our new vaccine for the moment, and most of our clinical studies are run in the US. So that is why it is mainly CBER that we met with.

Genovese: To my knowledge, in Italy at the moment we have no presentation of an application with the use of NGS. At the moment, we have no experience about this. We are preparing, but we have no presentation.

Product-Specific Validation During Method Transfer

Levis: Carine, can you comment on the type of product-specific validation that needs to be performed to use the method, to transfer the method to new products, or to transfer to a new lab?

Logvinoff: Transfer to a new lab – that may not be easy. Transfer to a new product – the approach that we are following currently is to run generic validation on the matrix type, I would say – on viral harvest on cell bank. We do in depth validation with that using the 16 model viruses that are presented coming from the NIH study, the one that we link with the *in vivo* test.

From that, we have a solid validation based on the 16 model viruses. We show that we were able to detect them at 10^4 genome copy per ml – this is for viral seed. Then if we need to apply to the same family of products, then we will do what we call a 'verification.' So we will select the model viruses that were the most challenging as far as validation to be detected. We want to make sure that within this new matrix, but of the same family, we are still going to detect them. So, we will do verification runs to make sure that we are detecting these model viruses at the level where we are supposed to detect them as per the initial validation.

Transferring the technology to another lab is something complex that we are not sure yet we are be able to do in an easy way, because that requires a lot of expertise. This is not a simple PCR that we are dealing with, so it is far more complex. To set it as a routine test, it might be really difficult. And maybe that is why for the moment, we are only considering this test as a substitution of the *in vivo* and at the seed stage, so cell bank or seed lot. But we are not considering it as a routine test, because it is not an easy test to run for the moment.

Accelerating Vaccine Development

Levis: Here is a question for Carine: How much do you think the utilization of the high-throughput sequencing will speed up vaccine development and manufacture?

For Sebastiaan, I would add: In your field use of HTS, is there a difference in terms of breadth of detection and speed using the NGS versus traditional techniques used in the veterinary fields?

Logvinoff: How it can accelerate – I think you gave an example with the coronavirus vaccine that you are assessing these days. If we were to run the traditional testing package, it is a minimum of six months that you have to consider, to conduct the *in vivo* [tests].

You have a lot of investigation also associated with this test, because a dead animal does not mean you have contaminants. It means that you have to investigate to understand why this animal died initially. So that is why these tests were really long.

For sure, if you run the HTS, maybe we have to run an investigation, but at least you know the contaminant you are looking for. Because if there is a signal, you know [what] that belongs to, and you can further investigate. Also you gain confidence within your material because you now have a broad agnostic detection of whatever may be there. So in that sense, I think you are saving some time in terms of investigation and the knowledge that you acquire within your testing package.

Theuns: First of all, if you looked at the speeds, I think we are not there yet – that we are at the same speed as PCR or real time PCR. There you have results in three hours. For instance, if you are now being tested for coronavirus in hospital, you will get some results in three or four hours. But we are not there yet. So currently, we can give results to veterinarians in a couple of days, which is already extremely fast, but still not that competitive compared to PCR or something else.

But regarding the breadth of the analysis – yes, basically if you have good methods, bioinformatics, right lab procedures, and you have validated everything, you should be able to detect all viruses just within the analysis. If it is there, you will pick it up.

Also novel viruses, if your algorithms allow you to really look for distinct viruses, you will be able to detect them. So that is the unique thing – the vets or even human doctors if they start implementing it, they will not have to make a prior selection. You will just sequence at random. They will not add any targeted primers to really amplify your DNA

compared to a PCR. You will just put everything in the sample in the system, and then you will look with the computer into the data and see which viruses are in it. So you will have no bias on that. And that is the power of it. It will replace all the PCRs that are being done nowadays for diagnostics. In the end, it will be replaced by sequencing, I am convinced. This is the way to go.

Confirmation of Positive Results

Levis: Then you are going into another one of our questions that is related to the confirmation of positive samples. So you will be able to utilize the NGS to identify, but then you will subsequently have to show that a positive is a real positive, right?

Theuns: Yes that is true. In the end, that is the discussion also with PCR. We have a positive result, but is it genetic material? Is it a virus? You don't know. There are ways to see this.

Also in the veterinary field, if we detect a new virus, there has to be a follow up like we have to do them in the field of vaccine development. What does this signal mean? If it is a certain virus, is it because of the bioinformatics database giving a wrong hit? Is it a real virus? Can we detect it in the animal? Then we have to go back to the postulates of Koch and really show if the virus can replicate. Do some animal experiments, maybe, to demonstrate that the virus is really of importance.

So I think the technology will be there and will be helping us to give new insights and identify new pathogens. If we would not have NGS, yes, we would not have detected the coronavirus as well, not that fast. It allows us to really build a specific real-time PCR, to really scale up the diagnostics in a very fast way. So this is really, really helping science. But you have to be critical. I think that is an important conclusion.

Levis: Right. I agree. So Carine, do the regulators – probably meaning CBER at this point – request a demonstration of the equivalence between the *in vivo* assay and high throughput sequencing? You kind of looked at that study, but can you comment on that – whether you think it's necessary? That is an important thing for everyone to think about.

Logvinoff: This is also linked to the reference. The link to the *in vivo* helped us to show that the method was showing the performance that we were hoping to reach. So it was validating. To use this approach was a way to validate the future – the assay but also the substitution of the *in vivo*.

I think we provided evidence that, yes, HTS is consistent with the *in vivo*. And, on the second part, we showed that the way we are running the HTS is correct as we are able to detect this viral contaminant – at least at 10^4 genome copy per ml. So I think it is both.

But for the dossier what was really important, was the level of detection that we were able to reach for most of the model viruses that we use. Maybe that is where there is a link between why we will need some reference standards for this technique versus we did not have that for the previous compendial method.

Maybe since we do not have a common approach – and I do not think this is the way to go, to have only one way to deal with NGS – it is good to all have an ability to assess the performance of the assay that everyone may run. So that is the goal of the reference standard, in my point of view. While in the past for the compendial test also, there was no organization among the different regulatory bodies, the overall tests were pretty similar.

Levis: Well, I would just say, here is a real opportunity for us to introduce this new technology and really try to get harmonization across the regulators with respect to this, and across industry, in terms of how it is done. It is a good test package for us to show that we can actually do that and to bring the 'one-health' view, and have it done across the veterinary and animal vaccine platforms as well.

Use of Nuclease Treatments

Levis: So we have one more quick question – whether or not you are using nuclease treatments to enrich the selection of encapsidated nucleic acids, to decrease risk for false positives, or any other technologies used in that regard.

Logvinoff: No, we do not right now at Sanofi Pasteur. We are not using any enrichment.

Levis: Okay, and Sebastiaan for your assays with your purifications, are you?

Theuns: We do a nuclease treatment, yes. Because you have a lot of free-floating nucleic acids that can come from your host, and it is better to discard them.

Levis: And I think that the source material for your samples are very different than in manufacturing settings.

Theuns: Yes.

Pharmacopeial Monographs

Levis: Okay, so in the last few minutes, I would like to turn it back to Emmanuelle, if she has any comments or questions or to any of our speakers – if you would like to have a minute or so to conclude or any comments you would like to make.

Charton: No, for myself, I was very pleased by all the questions asked by participants and also by the three complementary presentations. I think that a lot of things are being done. And as you said, Robin, this is a nice package and opportunity for future harmonization on how to approach these techniques in the future.

Levis: I have a question for you, Emmanuelle. From an EDQM point of view, would this be a technology that would warrant a monograph, or would it just be referred to, just in terms of working towards the harmonization of how it is used?

Charton: Yes, I mean, surely not as product-specific tests. I don't envisage this in the future – but maybe more guidance on how to validate the techniques. This is currently being discussed in the vaccine group – how to further encourage the use of these methods in the future.

But since, as it was noticed also by participants, the technologies themselves evolve so quickly, I do not envisage the reference to any specific equipment – more guidance on how to validate and how to demonstrate equivalence. And there is lots to work on. Lots of things have been published, and these things deserve some consideration.

So I will say we have opened the doors. I was pleased to see from the presentations from Carine, and also Domenico, the chapters of the pharmacopoeia, which opened the door to these techniques. So it is the first step. We have heard from the meeting at Ghent last year that mentioning them in the European Pharmacopoeia is very much appreciated. You heard from the stakeholders that this was a very positive sign from the European Pharmacopoeia.

Of course, it is always a question of what to write and what not to write. Because if we write too much, then you could be seen as preventing innovation. So, that is surely not what we want to do, but to continue to encourage the use of these techniques for the characterization or the identification of adventitious agents in biological medicines.