

Winning the Translational Race: Making Good Choices in Biomarker Assay Development for the Clinic

Webinar

9 Oct 2013

Slide 1

[0:00:00]

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

Slide 2

In today's webinar we will focus on the process of discovery and validation of biomarkers for possible clinical use.

Despite the recent surge in biomarker research and optimism around their application, relatively few are in routine clinical use today. The major reasons for clinical failure of biomarkers are poor experimental design and inappropriate choice of assay. It is therefore critical to carefully construct biomarker discovery and validation studies using the most appropriate assays and to develop these into validated and analytical methods suitable for clinical practice.

Over the next hour, we will be reviewing the range of technology choices available for biomarker assays and providing insight into the challenges confronting researchers when selecting the right biomarkers to pursue and the best assay to develop for the clinic.

I'm very pleased to welcome our two speakers for today, both from the United Kingdom, Dr. Richard Kennedy from Almac Diagnostics in Craigavon, Northern Ireland; and Dr. James Timmons from Loughborough University in Leicestershire. Thank you both so much for being with us today.

Slide 3

Before we get started with today's webinars, as always, some information that our audience might find helpful. You can change the size

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 the speaker bios or additional information about technology related to today's discussion or to download the PDF for the slides.

Each of our speakers will give a short presentation about their work, after which we will have a Q&A session during which our guests 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, just click the red Q&A widget at the bottom of the screen.

Please remember to keep your questions short and concise as possible as this will give the best chance of being put to our panel.

You can also log in into 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 Affymetrix for their sponsorship of today's webinar.

Now, I'd like to introduce our first speaker today and that's Dr. Richard Kennedy.

Slide 3

Dr. Kennedy is the McClay Professor in Medical Oncology at the Centre for Cancer Research and Cell Biology at Queen's University of Belfast and the Medical Director at Almac Diagnostics where he is currently responsible for the internal research programme. He also heads a CLIA compliant diagnostics laboratory that supports patient stratification for clinical trials on behalf of several large pharmaceutical companies. At Queen's University, Dr. Kennedy leads a laboratory and hospital-based research group focused on various aspects of stratified medicine. He is also a medical oncologist at Belfast City Hospital and continues to manage cancer patients.

A very warm welcome to you, Dr. Kennedy. Thank you for being with us.

Dr. Richard Kennedy: Thank you for your introduction, Sean, and the invitation to speak today. So today I'm going to talk about some of the lessons we have learned over the last eight years at Almac and hopefully provide some

information for people who are hoping to develop biomarkers in this competitive area.

Slide 4

What I've done here is just outline some of the areas where biomarkers are being used in cancer medicine at the moment.

Slide 5

But today, I'm going to focus particularly on predictive biomarkers.

Slide 6

So predictive biomarkers are used to predict the benefit from specific therapies in the cancer clinic. The last time I looked, over 15,000 manuscripts have been published reporting predictive biomarkers for cancer management. However, very few have ever made an impact on clinical practice, and we'll talk today about why that may be.

Slide 7

So when we think about biomarker research, we should really consider the different treatment phases. We have discovery and development phase and the validation phase. I'm going to talk about some of the issues around each of these phases.

Slide 8

So first of all, looking at discovery and development, we need to consider what are we actually measuring.

Slide 9

So most biomarker discovery really follows this pattern where we gather material, and it can be from various samples from a patient. We apply some kind of technology, usually high-throughput; and from this, we identify a biomarker to distinguish between two groups.

[0:04:55]

Slide 10

The biomarker itself is either qualitative. So this is something which is a yes or no result. So the difference could be a mutation such as the KRAS mutation or BRAF. Or it could be an expression or no expression of a particular gene such as c-KIT.

The other types of biomarkers are quantitative. So these are score-based where a positive or negative result depends on the score, usually with a threshold defining positive and negative.

Examples of this would be immunohistochemistry that is used for the estrogen receptor in breast cancer, or the more novel qPCR or DNA microarray-based signatures such as Oncotype DX or MammaPrint that are being used to predict the prognosis of breast cancer.

It's important to appreciate the differences between qualitative and quantitative biomarkers as these may influence the validation approach which we'll discuss later.

Slide 11

So the next thing to consider in the discovery and development is the actual discovery step.

Slide 12

So what I've described here are the three main approaches that we see at Almac. That is the preclinical model systems, the retrospective archived tissue approach, or prospective discovery.

Slide 13

So first of all, looking at preclinical model systems.

Slide 14

These are typically either human cell lines or animal models, particularly murine-based systems. The advantages of these systems for biomarker discovery is that they can be performed early in the drug trial process so that we can discover biomarkers before giving a drug to patients and also allowing the behavior of the drug to modeled in a specific molecular complex where we can vary expression of genes and look at the influence on the response.

The disadvantages of these systems are that kits or animals are maybe of different physiology to human. Particularly this could be an issue in drugs targeting elements of the stroma such as antiangiogenic agents.

There can be little genetic variation between animals, so we can see the same level of genetic heterogeneity that we see in humans. There is no immune system. Particularly in the xenograft systems this can be an issue, because we know that the immune system can have an influence on response.

In the case of cell culture, there is no tumor stroma, which, again, can influence response.

And also, any cutoffs that we use for quantitative assays may not translate from these preclinical model systems into actual material.

Slide 15

However, we have had success in this approach at Almac where we've used isogenic cell line modeling and xenograft data to identify the biology which is underlying the SRC activation, and this is now being used -- being validated in the clinical trial where patients have received SRC inhibitors for ovarian cancer.

Slide 16

The second major approach is using retrospective archived tissue.

Slide 17

In this case, we are typically using tissue from tumor banks. The advantage here is that we're using relevant human material. We are using full clinical annotation including outcome, which can be helpful in designing the biomarker. There may be large numbers of samples available, and this is particularly where we are able to analyze formalin-fixed, paraffin-embedded tissue. And from this then we are able to perform clustering analysis to identify biologies that underlie response to drugs.

Also, these datasets are used to set population-based cutoffs for quantitative assays. So for example, when we've developed a score for an assay, then we can define what the top 25% of the population would be for that cutoff.

The disadvantages, however, is that it's unlikely to have material for novel therapies available. The tissue may not have been collected appropriately, and this can be an issue. So we've seen cases where a tissue has been stored very poorly and it has been degraded or there may have been a long time to fixation from collection, which in turn affects the biomarker discovery efforts.

Also, archived tissue can degrade over time, so sometimes this could be misleading.

Slide 18

However, we have used this approach, and this is data we've presented at ASCO two years ago where we did a large clustering effort using DNA microarray data from 300 high-grade serous ovarian samples, and we were able to identify molecular subgroup that represented angiogenesis. From that we were able to develop a predictive biomarker which is now being validated in the medical research ICON7 study on antiangiogenic agents of ovarian cancer.

[0:09:53]

Slide 19

The third approach is prospective discovery.

Slide 20

So this is where we're going to analyze tissue from patients who are responding or not responding on a clinical study.

The advantage of this is that the material obviously is relevant to the drug in question. The problem is that all preclinical trials give new drugs in combination with other therapies, so it can be difficult to develop the biomarkers that are specific to the therapy.

Also, this approach can require large numbers of patients to be adequately powered.

Slide 21

So this is an example of this kind of study where patients have received the new treatment. We have responders and non-responders, and from that we generate the biomarker.

The issue, however, is in our experience, to develop robust biomarkers, we typically need 50 patients in each arm. And if we have a new agent, with a predicted response rate of maybe 10% in a non-selective population, then we would at least 500 patients in this kind of study in order to develop a worthwhile biomarker; and obviously, that would be very difficult.

However, novel adaptive trial designs which are now coming out and being published may be the answer to this.

Slide 22

The next things to think about are technology and reagents.

Slide 23

So when you're in the biomarker discovery phase and you're thinking about going forward into the clinic, it's important to think what kind of material will be required for the biomarker.

If you mandate fresh material, this can sometimes be challenging in the clinic where it may be difficult to collect fresh tissue, and this may also be difficult to ship in a stable manner.

But we also have to think about patient safety and comfort. So if the patient has, for instance, lung or liver metastasis, these can sometimes be difficult to biopsy and actually may be hazardous to the patient. We know from clinical trials, that if we mandate a biopsy prior to novel treatment, we can expect the patient enrolments to fall by 30%. So, obviously, that's going to slow down the ability of the trial to complete. This is why there's a big push now to move towards biomarkers from blood, which is obviously a much easier material to get.

It's also important think about the technology. You want to be using high-quality technology, preferably CE marked or meet good manufacturing practice. The equipment you use to discover biomarkers must be properly maintained and should be calibrated and scheduled regularly.

And also, it should be a practical technology. So you do not want to use technologies can be extremely expensive and have a very long turnaround time, and that is unlikely to make it into the clinic. It should be something that's relatively easy to use.

The other thing to think about is is the technology actually going to be around in a few years or is it going to be obsolete? Which means that the even if you have developed a good biomarker that maybe no equipment could deliver it.

Slide 24

Reagents, I think, are certainly something that we should some time thinking about because many of the laboratory reagents we use are research use only. There can be considerable variation in performance batch to batch with these reagents. The problem is if you developed a biomarker using a single batch reagent, the biomarker may become batch dependent and will only ever work with that batch of reagents.

So ideally, you should use GMP reagents which have been batch tested and quality controlled. But if that's not possible, the other possibility is to pool batches of RUO reagents to try to smooth batch to batch effects.

Slide 25

Just to show how this sort of problem can arise, here we have material which has been analyzed at two different time points. So this is what's called a principal component analysis that's basically a summation of the genetic information from a tumor. We have samples that were run in one year, which are marked red in this diagram. And the same samples run a year later, marked in blue. And you can see there has been a large shift in the data.

The reason for this in this case was that the supplier of the amplification kit had changed their supplier of DNA polymerase and the efficiency had changed slightly and that was enough to change the data. So you could certainly see from this data, had you discovered a biomarker in the red data set, that it would be unlikely to work a year later in a blue dataset.

Slide 26

The other thing to think about is the lab operator effects. So we know that biomarkers are discovered by a single lab operator often when we work with that individual. That may be because this individual is particularly good in the lab and may have modified lab protocols which they use themselves, or they may just be very experienced.

It's very important to adhere to strict standard operating procedures from the discovery phase right through to the validation.

And also, ideally, you should randomize samples between several operators to take out lab operator effects.

Slide 27

So here we have another example. This again is a principal component analysis, and we have a set of samples from two different operators, red and blue. And you can see there's been a shift in the data between the two. This is a fairly extreme example.

[0:15:00]

What has actually happened here, the red operator here was actually a trainee, and they were missing one of the steps. It was sort of extreme, but you can see more subtle effects between experienced operators.

Slide 28

The third thing to think about are centre effects.

Slide 29

So, unfortunately, there are numerous examples in literature of biomarkers that have been discovered from a single centre, but they haven't been applicable anywhere else and there could be various reasons for this. Sometimes we never really get to the bottom of it.

But some more obvious examples, in some centres, surgeons may have specific surgical approaches which are different to other centres. So, for example, a surgeon in one center may choose to tie off the blood vessels in the tumor early which may lead to hypoxia of the tumor, where in other centres it may be done later and, therefore, the tumors are slightly different when it comes to analyzing transcription or indeed other elements of hypoxic response.

The other thing to think about between centres is that there may be variations in fixation protocols when we're looking at formalin-fixed tissue.

So ideally, you need to use material and clinical data representing response or non-response from multiple centres during the discovery phase.

Slide 30

And here we have an example. This was from we did a few years ago when we were developing an assay for prognosis in colon cancer. Again, this is a principal component analysis and you can see in this diagram the samples from the centre marked in blue are outlied compared to the other samples.

The reason for this, the centre, the surgeon did surgery on a Friday, the samples were sent to the pathology lab and then left in formalin over the weekend and fixed on a Monday morning. However, in the other samples, the entire process was done in one day, and this is enough to change the data. And this is not just unique to transcription. They also saw differences on the histochemistry level as well.

Slide 31

Another thing to consider is population effects. So it's very important to ensure that the population used for biomarker discovery is relevant to the population to which the biomarker will be applied.

Slide 32

So, for example, we know that Afro-Caribbean populations have variations in the genetics of prostate or breast cancer compared to Caucasians. Similarly, the Asian populations can have variations in lung cancer biology. So it's very important to make sure that there's adequate representation that is going to be applied in these populations.

Slide 33

The other thing to be very careful about are convoluting factors.

Slide 34

So what I mean by that is that if we take a population of responding patients and non-responding, but it's just a pure chance, for instance, maybe there have been more males in the responding; more females in the non-responding, and we develop the biomarker in these populations, there's a danger that we may develop a biomarker for gender rather than response. That was just an extreme example, but there's a number of these things that can convolute a biomarker. And if you're not aware of these factors or aware of the design, there's a danger that you may develop the biomarker for the wrong thing.

Slide 35

So the next part of the biomarker research process is the validation.

Slide 36

So within validation, I think it's worth discussing the regulatory landscape. But this, on its own, could be a talk for an hour, so I'll not spend much time on this.

Slide 37

However, it's important to be aware that there are different levels of regulatory approval, and the choice level of regulation is influenced by the type of the biomarker and the intended use and risk to patients.

Companion diagnostics. These are biomarkers that are used for the licensing of drugs for clinical use, are the most stringent, and typically have to go down the premarket approval route of the FDA.

The relevant bodies to be aware of are CLIA, FDA in the US, and EMA in Europe.

I think it's important for any researcher to have some awareness of the regulatory landscape because they need to start thinking about this early in the development process to make sure that they fit in the criteria for the regulatory bodies.

Slide 38

Most validation can be broken down into analytical and clinical.

Slide 39

So first of all, thinking about analytical validation, the two major areas in this are precision and accuracy. There are other things we can talk about as well, but these two merit discussion.

Slide 40

For precision.

Slide 41

This is a measure of biomarker repeatability. Loss of precision can occur because it's inherent to the technology. So for example, if we use immunohistochemistry for phosphoproteins, we can quite often see quite a large amount of variability just because it's inherent to the approach.

[0:19:56]

Also, it can be due to problems with reagents, equipment or technique as we discussed earlier.

Something though that we should think about and I'll discuss a little bit more is the influence of the ratio between normal, stromal, and malignant tissue in the sample.

Slide 42

So here we have prostate cancer. We analyzed three prostate cancers representing high result, median result, and low result. And we could see here on the left one we used full biopsies, we're getting quite a lot of variations so there is a loss of precision in this assay, and this is a PCR-based assay.

However, if we macrodissect the tumor, so this is where when we cut away the normal tissue and try to enrich for the tumor, we see that we can considerably improve the precision of the assay. So this will be much more acceptable for a clinical test.

Slide 43

The other important factor to look at is accuracy.

Slide 44

So this is a measure of how close the result is to the known truth. So truth can either be the result from a reference lab or it may be the gold standard technology such as Sanger sequencing permutations.

Accuracy can be affected by the site or type of biopsy and one thing you should be aware of is the possibility of heterogeneity for your biomarkers. So there can be variations for biomarkers throughout the tumor, so it's important to take biopsies from different parts of the tumor to make sure you get the same result.

Also, there can be incorrect sample fixation or lab techniques that can result in loss of accuracy.

Another thing to think about though is the possibility that the patient could have had further treatment since the original diagnostic biopsy resections taken.

Slide 45

So here we have an example which we did at Almac where we looked at ovarian cancer. We took a biopsy of the recent treatment and then a further biopsy after recurrence, and these patients receive six cycles of platinum-based chemotherapy.

And probably unsurprisingly, when we looked at the second biopsy, there was quite a variation in genetic expression. So we saw almost 500 genes that were greater than twofold differentially expressed between the original tumor and the tumor following recurrence. And, obviously, if we went to give a treatment not based on the original biopsy material, it could be very misleading.

The genes pointed out in this have currently been published as potential biomarkers in ovarian cancer.

Slide 46

The last thing to think about is clinical validation.

Slide 47

So it is an absolute requirement for companion diagnostic assays, that they are clinically validated if they're going to be used in clinical practice.

The strategy needs to be agreed with the regulatory authority up front. There are two main types of study. There is a simple biomarker validation or more complex.

The purpose of this is to show that the biomarker can adequately stratify patients in terms of sensitivity and specificity for binary or hazard ratio for patient survival related to progression pre-survival.

Slide 48

Here we have an example of a simple biomarker validation study. The patients enroll, the biomarker is performed. Typically, for this kind of study, this would be a mutation or something which is quantitative.

And then the biomarker is applied to the positive population and then we look to see: are able to predict response by comparing the outcome between the experimental drug and standard therapy.

Slide 49

The other type of biomarker study is the treatment interaction design. Now, this is a much more complex kind of design. I've given a reference here to Mandrekar and Sargent who published this in the *Journal of Clinical Oncology*. The idea here is the biomarker is applied and the predicted responders and predicted non-responders are then randomized to receive the experimental drug or standard treatment. And from this it's possible to calculate the sensitivity and specificity and also test the cutoff for qualitative updates.

Slide 50

So the important thing to consider for predictive biomarkers used in the clinic, I think, is first of all, are you using the right approach for discovery and development? Do you have the correct discovery dataset? Are you using the correct technology and reagents? Are you sure that your biomarker is not convoluted by other known factors?

And for validation, are the regulatory requirements being met by your process? Do you have good enough analytical performance for your assay, particularly precision and accuracy? And do you have a validation strategy in order to demonstrate the clinical utility of the assay?

So thank you very much for your time.

Sean Sanders:

Thank you so much, Dr. Kennedy. Fantastic talk and a really great start to this webinar.

We're going to move right along to our second speaker today and that's Dr. James Timmons.

Slide 51

[0:25:00]

Slide 52

Dr. Timmons is chair of Systems Biology at Loughborough University where his work focuses on translational medicine projects to cover a variety of age-related diseases including type II diabetes, cancer cachexia, osteoarthritis, and muscle aging, specifically the evaluation and development of informatics strategies to integrate genomic and clinical data to yield biomarkers and predictors for personalized medicine. Dr. Timmons has been extensively published and his work has been featured on the renowned BBC documentary program, Horizon.

A warm welcome to you, Dr. Timmons. Thanks for being with us.

Dr. James Timmons: Good afternoon and thank you very much for the invitation to speak today. My talk today will build very much on what Dr. Kennedy has just told you about the entire process of biomarker development.

Slide 53

The examples I will use, however, are not in the field oncology where most of the biomarker activity has taken place. But instead, I will give you some examples from biomarker development within lifestyle and age-related medical areas.

The way I want to start first today is to just raise the point that the concept of needing a biomarker to understand individual patient responses is not a new idea. It goes back at least several hundred years where in the quote that I show you in the moment from Caleb Parry, a GP in Bath in England, he was very clear that he had to know much more about the particular patient who had a disease than to know what kind of disease the patient had. So in his mind already, it was how the patient responded to the disease and the patient characteristics that were most important.

That's what we were really trying to do when we develop a biomarker. We're trying to peel out some unique information about the patient that will tell us how they're likely to respond to the disease or how they're likely to respond to the drug or health recommendations.

Slide 54

Just to give you a brief background, I've spent time both in industry and within academia, and so my biases are broad, and I also have spanned physiology through to genomics and bioinformatics. The point I would

make here that I think is very important that you either have multidisciplinary people or assembled multidisciplinary teams when you're thinking about how to develop a biomarker.

Slide 55

So what stimulated my interest in biomarkers was really my time in industry, and here is a relatively old slide now that it gives you a relationship between the amount of money industry was spending during the two decades, and the rate through which they were producing new small molecule therapies. This is obviously not a very successful model, and we've seen the consequences of that.

Slide 56

One of the reasons for this was simply speaking the preclinical models of disease and biomarker activities were not predictive of what was happening in the human, and there have been various academic publications in recent years detailing some of the reasons why this is the case.

So if you were going to focus on personalized medicine, then I would say that it should be based on translational research done as much as possible in humans in order to solve some and avoid some of these problems.

So why do we need the biomarkers for lifestyle-related advice? Well, it's becoming extremely clear that lifestyle-related advice in a one-size-fits-all type approach will not yield the benefits that we expect. So this recent publication in the *New England Journal of Medicine*, the Look AHEAD trial, the intervention worked in the sense that key biomarkers were changed. There was less diabetes. There was greater fitness and less fat on average in the subjects after 10 years follow-up. However, the real aim of the study was to demonstrate that this would impact on major cardiovascular events and also mortality. And, unfortunately, that is not noted and in fact the trial was stopped early because there was exactly the same event rate in the control versus the intervention group.

And this is really important because what this tells you is that the biomarkers of diabetes, for example, or fatness in this case are not translating to predict the lower rates of stroke, for example, and that stroke had been making really big cost burden on the health service and also a huge impact in the quality of the life of the patient. This is an

example where public health strategies perhaps haven't quite got the right biomarkers.

[0:30:00]

Slide 57

But there's another interpretation, and that is maybe the problem of the Look AHEAD trial was the fact that the prescription of the lifestyle advice—and you can replace that idea with “drug”—was not appropriate. There is some data to suggest that this might be the case. So we look at the benefits in biomarker response that we see in humans in response to supervised lifestyle intervention, in this case, exercise. What we notice is that there's a huge range in response. Some people benefit dramatically in terms of fitness gains. Some people, a minority, will actually have a lower fitness. Some people will gain muscle tissue and strength, and others won't. And some will gain insulin sensitivity and again, others will show an adverse response.

So this leaves us with one of the major challenges, which is how much exercise should an individual be told to do to get the best benefit or any benefit from that activity, and that's something we're addressing in the Metapredict trial and study over the next couple of years.

Slide 58

So the first example I'm going to give you is the use of a biomarker strategy to predict the fitness gains. It's a proof of principle that you can use this type of approach and predict rather complicated physiological changes in humans.

Slide 59

So everyone will be familiar today listening to this webinar of the central dogma of molecular biology, which is the idea of information flowing from DNA through RNA to protein/metabolite and ultimately physiology of the individual.

And I highlight this because there are clearly many points in the system that you could choose to measure to generate a clinical biomarker. So for example, you could do DNA sequencing and try and try link that to clinical variation. Technically, what you need there is very large sample sizes in order to discover a DNA marker that links to the clinical phenotype.

You could also take an approach where you're studying other variations and metabolite concentration of protein. Now, metabolites are looking particularly promising. However, the technology used to measure them is expensive, low throughput, and very specialized.

As far as protein biomarkers are concerned, I think there's a real problem there. By measuring the abundance of a protein, they're probably gaining very little insight into the function of that protein. There are so many ways proteins can be modified, changing its location within the cell and, therefore, changing what abundance means. That's a problem.

Instead, what we focus on is variations in RNA. RNA has a number of different advantages. One is we have technology measure RNA very specifically and also in a very quantitative manner.

The other aspect of RNA is that typically the RNA is in a relatively few places in the cell. Increases in RNA and decreases in RNA are more tightly linked to biological variation.

So let's look at a few examples.

Slide 60

So this is a scheme that we have developed for the validation of biomarkers for lifestyle-related research. So we have the discovery phase on the left, we have a very critical validation step, and then we have an optional step where we can use some of the candidates from the RNA work to move forward into DNA biomarkers. I won't touch on that today. What I will do is focus on the first two steps.

So, much like Dr. Kennedy mentioned, there are typical desirable sample sizes that were around 50 for starting off. We then build our response predictor. But then really the key is to take the gene list you discovered there and move it forward into at least two independent samples, clinical studies of a similar nature to validate your data.

And what I'm going to show you now is an example of one research study.

Slide 61

So, in this case, we're using the Affymetrix gene chip. We're using the U133 Plus 2 chip, in fact, and just to point out that for many studies, we're using blood and micro tissue samples, very small samples,

relatively non-invasive with low gene scores to generate your global transcription signature.

[0:35:15]

So when it comes to depth in this case, the fitness outcomes in a set of humans.

Slide 62

We have a multitude of baseline determinations of their fitness. We recently established the baseline in fine detail. We then supervise the intervention. We make sure to take the drugs if you like and the fact that they actually do the intervention required. And then we gather a number of what could be considered confirming factors biochemical and physiological.

And really the key aim here is to predict who is a non-responder from the baseline RNA profile.

And so what we did was we built a quantitative model. We weighed the baseline RNA to the outcome.

Slide 63

This gave us the linear quantitative model you can see in the center. Now, if you compare the performance of that initial prototype with, for example, baseline fitness in relationship to outcome, you can see there is very little indicator of baseline fitness predicting the aerobic fitness. However, you will note that the RNA classifier had a very high ability to relate to the clinical outcome.

Now, most importantly, what we've then done is taken this classifier to a number of independent dataset on the right-hand side there, which is Group 2. You'll also notice that these studies are relatively small. So the original classifiers are relatively powerful. It's capturing a lot of variance and it does replicate an independent studies used in different labs with different populations.

So that's just one example of developing prototypes and we're now applying those methodologies to much larger clinical studies with several hundred patients in each arm.

Slide 64

The second example I'll give you today is an example of a diagnostic for healthy aging. In this case, we've relied on something like 700 clinical samples.

Slide 65

The schematic for building and testing this particular biomarker, on the left-hand side you'll see that we start with two sets of human tissue. A very, very critically and probably powerful success here is that we're using tissue that has been snap-frozen, full control over its storage, and then detailed profiling on Affymetrix chips.

We haven't relied on so-called biobank junk where you have very little idea of the phenotype of the subject or the QC over the tissue handling.

We then build in this case a KNN-based classification model, and I'll go through some of the validation steps of that. But then having developed a qualitative classifier, determining whether the tissue sample is from a healthy old or a healthy young sample, we then create a quantitative classifier, which is that we create a gene ranking metric from the classifier genes in the discovery phase and apply it to longitudinal study. In this case, a 20-year follow-up study where we have samples from the baseline and taking a 1993 stored *[indiscernible]*.

And then we ask if the signature's prognostic.

Slide 66

So just to run through, probably one of the most critical things for biomarker development and I think probably one of the main reasons why predictive or claims of predictive diagnostics fail to replicate, and that is really the way the mathematics and the computer science has done in the first place.

So I've got a schema up there which basically is a looped system where when you're building your classifier, you're holding out one sample within your classification dataset, your test dataset for identifying good RNA predictive biomarkers and you're holding at the second in order to ensure that you are not overfitting or biasing your correlation or association.

What we do is we go through each sample within our test dataset, removing it one by one and regenerating a rank of RNA, in this case,

probe sets, which are predictive during each loop and take the best performing probe sets forward to a final classifier list.

[0:40:15]

So at that stage we've created a prototype. On average, it's somewhere between 50 and 150 group sets.

At that stage, what we do is we throw away that first dataset. We've build the classifier on that dataset. There is no rationale for using it during the validation step, particularly as you will simply bias the validation step. So, at that point, what you absolutely need is two further independent clinical datasets to validate the classifier performance. I think this is a very critical step, and most studies don't do this.

Slide 67

So how is each classifier working? Well, the green bar on the left is the performance in the training dataset. So an average, each selective probe set was about 90% successful at identifying a young from an old sample.

We then take that probe set list through another five independent datasets. These interestingly have been developed and obtained partly in our lab and partly in other labs that we saw. This is RNA which has been collected in a similar manner: snap-frozen tissue. But the laboratory regions, the technicians, and even the Affymetrix process, equipment that has been used to profile each of the independent datasets has been different. And there you can see a range of performance with an average of about 93% accuracy in a total of 300 subjects.

In terms of confounding factors which again Dr. Kennedy mentioned, what we were interested in doing was ensuring that none of the markers were related to common factors that might change with the age or indeed differences in gender.

So the question that we next is: is this biomarker going to operate in other organs? So, again, we take the same 146 RNA molecules and indeed we can tell the accuracy of the age of human brain samples up to 91% accuracy. So there really does appear to be a common biomarker in this case.

What's interesting here is that the handling of the clinical tissue for these brain samples will have been less ideal than for the muscle samples. And I think one of the key things there is that if you've build your classifier

using very good, well-handled material, then those are good potential it will work in noisier data, and that seems to be the case.

Slide 69

So we have a new diagnostic. The ROC plot here of sensitivity versus one over specificity. In the case for most clinical cohorts demonstrates very high specificity and selectivity.

Interestingly, just as a parallel, most of the 146 genes have not been previously linked to ageing, and that's probably a reflection of most ageing molecular biologies generated in model organisms.

Slide 70

So the question is can we now apply this diagnostic in a quantitative manner? So we have selected 146 genes as described and now what we are going to do is we're going to take a group of 108 subjects, all who are 69 years of age chronologically in terms of their passport. And what we're going to do is we're going to score them for the expression of each of the 146 genes and create a rank order. And then we're going to compare that rank order with their medical outcome over a 20-year period.

Slide 71

So muscle tissue was obtained in 1992. It was stored at -80, in the freezer. Subjects ranked first for low median gene score had failed to activate their gene expression program. Subject ranked 108th of 108 had a high median score and had strongly activated the age biomarker.

And we noted at least a fourfold range in gene score despite all of the men being 69 years of age and of comparable lifestyle and fitness at that time.

[0:45:00]

Slide 72

So what we're able to demonstrate in this first clinical longitudinal study is that this biomarker for healthy aging did relate to decline in renal function during 12 years. However, it's most strikingly also relating to risk of death for the total of 20-year period, where if you fail to switch on this particular healthy ageing biomarker, you were far more likely to have

died during the 20-year follow-up period than had you switched on strongly.

So here's some example of where you can take both a qualitative and a quantitative approach to biomarker discovery and apply it to a longitudinal study. You can integrate data from different laboratories. And perhaps the most important reason for that is the computational methods and selection processes in the beginning for picking your biomarkers and ensuring they are independent.

Slide 73

So what should the future look like? Well, in terms of our interest in lifestyle advice, what we hope to be able to do with these types of biomarkers is essentially give each person a health and fitness risk profile where we'll know how they're likely to respond in terms of fitness, drug response, for example, in terms of a blood pressure lowering drug or a statin.

Muscle strength and tuning to combat frailty. Blood pressure to combat stroke, et cetera, et cetera.

Slide 74

And just to wrap up in terms of general conclusions. In my experience, RNA provides a better promise than DNA in general as variance in expression integrates both genomic, epigenomics, and environmental influences.

Proteomics is far behind in cost and reproducibility. And I think there's a key misunderstanding here. People believe that proteins are closer to the action in terms of functional biology, and that may be true; but protein abundance may never be a good functional surrogate for genuine protein activity because there are just so many postranslational modifications that dictate where the protein is.

Technical and cost barriers currently prevent widespread RNA diagnostics in much more routine areas in clinical medicine, perhaps not oncology, but for sure in other routine areas of medicine, but I believe this will change.

And a good biomarker is characterized by its technical performance, not by how much we think we understand its link to the disease. And in many, many conversations, people have been very focused on trying to

understand why the molecular marker is a good biomarker instead of looking at how it was developed, how it was independently validated, and for those then shown to be of general use in a wide range of clinical samples.

And that's me finished. Thank you for your attention today.

Sean Sanders: Thank you very much, Dr. Timmons. Fantastic presentation. Very interesting results there that I hope will generate a lot of discussion amongst our viewers.

Slide 75

Many thanks to both for our speakers for their presentations. We're going to move on right on to questions submitted by our online viewers. Just to remind you that you can still submit questions. Just type them into the text box at the bottom left of your viewing console and click the Submit button. And if you don't see the box, just click the red Q&A icon and it should appear.

So the first question I'm going to put out to both of you and we'll start with you, Dr. Kennedy, is the terms "translational science" and "translational research" have become quite popular phrases in the last few years. What is your opinion on how much the academic research community really understands what this mean and how that might impact the way they do their science?

Dr. Richard Kennedy: Yes. Well, I think the majority of people who are performing basic science are doing it with the hope that it will impact patient care. One of the issues, I think, is the understanding of the development pipeline. So from the basic scientific findings through to the clinic, specifically the regulatory requirements, which mimic the opportunities on technology and reagents that can actually be used in the clinic.

So I think it's probably important for the basic scientists to at least have some idea of what the restrictions are, all those things that can actually be used in the clinic, and factor this into their experiments early in the process.

Sean Sanders: And Dr. Timmons, something to add?

Dr. James Timmons: I would certainly agree with those points. I think one of the main problems with the area of translational medicine is much like systems biology and it means many things to many people, and it can be

convenient to mean many things to many people because it unfortunately drives the fashion within public funding. And so in some cases, I believe people perhaps stretch what real translation means. To me it very much means that the basic research that you're doing should within one or two steps be potentially applicable in the clinic. You should be able to test the hypothesis of it, and it could be useful for treating human disease or improving healthcare provision.

[0:50:25]

And I think what we really need to do is improve the general understanding of what barriers people face both in terms of therapy development but also in terms of health provision right at the coal face so to speak in the clinical wards or other things that make the biggest difference. And then I think we'll have a better and more common understanding of what translational medicine means.

Sean Sanders: I'm going to stay with you, Dr. Timmons, for the next question. Considering, as you mentioned I believe at the start of your talk that successful translational science requires quite a range of skills and experience, what are your thoughts about the importance of collaborations and partnerships in the field and what advice might you provide to researchers out there on how they might develop collaborations?

Dr. James Timmons: That's a good question. I've been involved in coordinating several newer applications and currently coordinate a multidisciplinary research FP7. I think it's very challenging—certainly my time in the industry where I had to work with chemists, drug metabolism people, clinical development, marketing—it really does take a great deal of effort to try and understand what everyone's agenda is and their perspective.

I think that we have a barrier to developing more individuals that have multidisciplinary skills partly because research and career tends to reward people with focus in a particular topic, become well known for that topic and develop a system sort of grand funding stream for that topic. And that puts them in a very narrow silo, I believe. It's much more difficult to jump around in terms of technology but also in terms of research topic.

For example, if you spend a few years learning about oncology and cancer research and take that back to, for example, diabetes research and vice versa, we may actually get a lot of cross-fertilization of ideas.

So when you put them together, in the absence of a system that encourages multidisciplinary scientists, I think what you have to do is be very broad-minded from putting together a consortium that's going to tackle a medical problem that requires multidisciplinary skills. You really have to think actually and bring in people that perhaps may not necessarily be interested in your medical problem.

Sean Sanders: So let me ask the next question of you, Dr. Kennedy. Could you talk a little bit about some of the research that's been done on cell lines that have been maintained for a number of years in culture in which this has likely changed their biology?

Dr. Richard Kennedy: Yes. I think that's a very important point. I think most of us in the field would agree that it's highly unlikely that we could just pick cell lines representing a specific disease, treat with the drug, and look at the genetic differences between them and then apply it to a patient population. For the reasons which you pointed out, there's drift of time, there's selection for growth in a plastic environment.

A number of the cell lines that were established years ago are not actually the correct pathology. For example, we have looked at ovarian cell lines and we found that a number of them are actually probably gastrointestinal rather than ovarian. So, obviously, any hypothesis generated from them may not be correct in ovarian cancer.

The way we tend to look at the cell lines is that we use them to modulate gene expression so we target them to overexpress or underexpress, and we look at the genes that are regulated, then we use that pattern to go into archives materials to identify the same molecular groups and archive education material and then we develop the biomarkers from that material rather than the cell lines.

Sean Sanders: Perfect! Dr. Timmons, anything to add?

Dr. James Timmons: I have some experience of colleagues who have been rather successful in developing pharmacological drug response signatures from standard cell lines and then taken those into prognostic performance within clinical trials...

[0:55:11]

So they can make a medical prognosis, for example. And I think, again, one of the key challenges is to try and understand why things have been failing in the past, and I think that again I come back to the computational

approaches that have been taken for picking out candidates. In some cases, there is a strong focus on differential expression and with differential expression we have two artificial cutoffs. One is the degree of change; the second is typically our P value cutoff, which tries to control for false positive and false negative rates. But, of course, it's an approximation.

I give you a very clear example now: that for some of the ageing biomarkers which are extremely precise in their particular studies, they may only change about 10% to 15% over five decades. So the change in signal is rather small, but it's so consistent in terms of that, that can be a very useful biomarker gene. So I think there are certainly developments on the computational side that improves some of the historical experiences.

Sean Sanders: Another question for you, Dr. Timmons. I believe this is referring to something you've showed toward the end of your talk. And this viewer asked whether you're saying that as long as a biomarker is validated extensively, it's not required to understand its actual biological meaning or function. Is this correct?

Dr. James Timmons: Yeah, that's correct. I think I've become rather frustrated with the discussions that I've had particularly in my time in the industry, the pharmaceutical industry, where people were very tuned in to trying to understand the biology of a disease, and I think that we perhaps overestimate our abilities to understand in a classic sense how disease works.

If you just take the example of a single gene or a single protein, the function of that piece of protein will be very dependent on the context, will be dependent on where it is in the cell, which cell type it is in. And in many cases, you can describe examples where a protein has an opposite function whether you compare it in, for example, a cell line versus a primary cell; certainly in adipocyte biology we've noticed that with a number of transcription factors?

So I think it's a very tricky issue to focus in on the described biology of a gene nest. I think it's perhaps much more important to focus on the math and statistics and computer science, unfortunately.

Sean Sanders: All right. So we're at the top of the hour and we need to wrap up this webinar. So I'm just going to put out a quick question to both of you, looking towards the future of biomarker technology and particularly the use of next generation sequencing. Dr. Kennedy, I will ask you first.

With next-gen sequencing being used more and more in basic research as a tool for biomarker discovery, when do you think it might become a routine clinical application?

Dr. Richard Kennedy: Yes. I think certainly the way we're pushing now is the use of next generation sequencing in the clinic, that's what we're aiming to get to. In fact, there is a large CRUK-sponsored study at the moment to test the feasibility of that. It's in the UK.

The issue we've got is the more detailed the sequencing we do, the more information we get. But we still have to validate our findings. We still have to show that each of these particular mutations or gene expression profiles predict the patient with sensitivity and specificity with the result. So it doesn't get around the issue that we need proper validation.

So we have to be careful that we don't get ahead of ourselves with the data and not spend time on showing the clinical utility of the data for specific drugs.

Sean Sanders: Dr. Timmons?

Dr. James Timmons: Well, if you're referring to high-throughput sequencing technologies for DNA variance, then I think obviously in particular niche diseases, that will have its place. If you're referring more to the technology in terms of RNASeq, then I think that there are so many fundamental challenges that it is very far away from being able to be used as a diagnostic tool. First and foremost, reproducibility of the signature that you might obtain from one sample to the next is nothing like its closest preexisting gene chip technology.

[1:00:04]

I think that the other problem which was just highlighted by Dr. Kennedy and I is that the volume of data that you get is enormous. So if you consider one thing and that is the first step of building a computational model of which transcripts are most predictive in your test dataset.

If you work with a U133 chip, an older chip from Affy, you can run your analysis on a Macintosh, a desktop Macintosh.

If you move to the latest generation Exon chip, you typically have to use a high performance computing cluster to do the same analysis.

The scale up would be the amount of data you start with with next-generation sequencing. And so actually, in reality today, you have to filter that data dramatically before you can run some of the computational models that we rely on, in which case you're already introducing a very sort of arbitrary filtering process that we need to throw away the most valuable data.

So I think there are a number of laboratory technical costs and real practical issues, and I think that actually the technology should be seen as a technology for something quite different which is just exploratory biology. And we should rely on proven technologies for biomarker development.

Sean Sanders: Fantastic! Well, I think that's a great place to end. So many thanks to both of our speakers for providing such interesting talks and discussion, Dr. Richard Kennedy from Almac Diagnostics and Dr. James Timmons from Loughborough University.

Many thanks to our online audience for the questions you submitted to the panel. My apologies that we didn't have time to get to all of them.

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 view again as an on-demand presentation within approximately 48 hours from now.

We'd love to hear what you thought of the webinar. Send us an email at the address now up in your slide viewer, webinar@aaas.org.

Again, thank you so much to our panel and to Affymetrix for their kind sponsorship of today's educational seminar.

Goodbye.

[1:02:31]

End of Audio