RO Podcast - Speaking of Mol Bio - Season 1 Episode 3

Transcript

Steve Lewis 00:09
Welcome to Speaking of Mol Bio, a new podcast series about molecular biology and its trending applications in life sciences. I'm Steve Lewis.

Dr. Gabriel Alves 00:18
And I am Dr. Gabriel Alves.

Steve Lewis 00:20
In our first season of Speaking of Mol Bio, we're focusing our conversations on four exciting application areas: CRISPR cell engineering, multiomics, exosomes, and single cell analysis. And today, we're returning to multiomics with Dr. Steven Williams.

Dr. Gabriel Alves 00:37
Steve boasts over 30 years of experience in molecular biology, including time at Pfizer, and the National Advisory Council on Biomedical Imaging and Bioengineering. He's currently the Chief Medical Officer at SomaLogic, where he oversees the SomaScan® platform and contributes to commercial business, assay development, and bioinformatics. We hope you enjoy our conversation.

Dr. Steven Williams 01:06
What I'm trying to do, or what we're trying to do, is to change the way medicine is practiced. So by measuring your individual biology, understanding what's going on in your body right now, effectively treating the proteome as if it was the body's internet. Can we interrogate that to say, what's you know, you might be healthy now, what's coming up, what might need to be changed, what drugs somebody might need to be taking that they are not taking today, the idea of having a being a liquid health check, is part of the original vision of our founder, Larry Gold. So, that's what we're trying to do. And you can imagine, you know, thinking of the multidimensionality of that, there are environmental causes. There are dietary causes, there are genetic causes. No disease that kills most people is a single, unidimensional cause. So, the multidimensionality there is pretty established. The challenge is, where's the biology? Because measuring more of it across more platforms, of course, costs more money. And, of course, if you, if you're an academic, and you can afford it, and you've got a big grant, then you say, well, I don't need to guess where the biology is, I'm going to measure all of it. I'm going to do sequencing, I'm going to do proteomics, I'm going to do metabolomics, I'm going to do the microbiome. And that's expensive. And I kind of think of it as being a little bit like, I've sometimes been mean and called it, lazy multiomics. It's lazy if you've someone else is paying, and you've measured all those things. And it's lazy if you then still treat them one at a time. Because if the biology is really multidimensional, what we should be trying to do is to say what combination, what's the best combination of measurement techniques that's going to actually do what we want. So, what I'm trying to show is that if you can't afford to measure everything, you're going to have to take some

bets on where you think the richest biology is. And at the moment, today, I think that's genetics and proteins.

Steve Lewis 03:22
Based on your description, it sounds very much like systems thinking is a strong part of your approach to biology. What are some of the gaps right now in your mind for that kind of integrated thinking about biology?

Dr. Steven Williams 03:39
I think as far as systems thinking goes, in some ways, I've realized that with the I, and we are systems ignorant. We humans cannot possibly grasp systems, to the stories, we make up little stories that give us a warm feeling about what a system is doing. And what we do today is rely on high-powered computing and machine learning to tell us about the systems. Because if you're measuring things at scale that no one's ever measured at scale before, then the diagram that came from the literature, or people's favorite tissues in the literature, there are there are so many biases out there, that I'd rather start from scratch and use the computer and machine learning techniques to tell us what the system is. And even though we end up often with a mathematical equation that might relate genes and proteins to make a prediction, I actually don't know why the machine, if you like, chose the components of that equation. So, I have to kind of let go of the wish to be able to tell a physiology story. Letting go of a physiology story requires you to get more data and do more validation. But I think that in the end, that's where we're going; that we won't necessarily be able to understand the multilayer interactions in multidimensional space that are actually going on inside us all the time anyway.

Dr. Gabriel Alves 05:17
Pulling back your final goal, which is to change healthcare, how healthcare is practiced; how would you translate that, for example, if someone walks into a doctor's office right now with some suspicion of hypertension? How is that? How will proteomics would change medicine?

Dr. Steven Williams 05:41
The analogy I would like to make is that protein network patterns in our bodies or rather like the internet, that they've evolved to transmit information from one biological system to another in multidimensional space, as we talked about. And so, when I describe what proteomics is, it's really like interrogating your body's internet; that the signals around what your body's doing and what your current health is, and what paths you're on for future changes in that health, they're in there. You've just got to interrogate them. And if you're interrogating the internet, you need two things, you need bandwidth and a search engine. And that's what proteomics is. The bandwidth is, you have to measure thousands of things at once. And then the search engine is, well, you have to actually use machine learning to say what are the keywords, what are the protein patterns, that link a particular change in the protein patterns to a clinical truth, which brings me on to how you would use it in healthcare and actually just had my own SomaScan done. And here it is. Each of these little boxes is a result. And each of these results came from measuring thousands of proteins in thousands of people, in people where the truth was

already known. And then you use machine learning to say what protein pattern relates to the clinical truth. And that's what each of these models was developed on. So, this is how we think that, when you're thinking about how medical care could be changed, depending on these results, a physician might be able to prescribe me a drug that I really needed. Some of the studies we've got going on right now are people with diabetes; there are some great new drugs to protect the heart. But not many people are taking them. The problem with them is they don't change blood pressure or cholesterol; they don't change the things you can measure. They work on new mechanisms. So, how do you know who really needs them? So, part of the reason the uptake of these drugs is poor is because they're expensive and they have some side effects, albeit modest ones. But you have to kind of guess who might need them. And there are there are ways of trying to enrich the population, maybe people with bad kidneys or people who've had an event in the heart in the past, but they're not very good. And so, we're saying if you look at the proteome, you can actually overcome all of that noise and get a much more accurate prediction of who really needs these new therapies and who doesn't.

Steve Lewis 08:22
One of the things that stuck out to me was the idea that biomarkers, identification, and even understanding of biomarkers might be one limiting factor. Where does that play into some of the protein studies that you all run? Is it all pretty much biomarker-based that you know, or are you discovering new biomarkers in order to make certain approaches to these clinical tests?

Dr. Steven Williams 08:52
My bias is, as I've explained, is to make clinical tests. But around the world, researchers are using the SomaScan technology for a whole variety of different purposes. I like to focus on readily available matrices like plasma and serum, but you can apply this to cancer tissue. People have looked at brain, liver, pretty much any organ, they've even looked at ground-up teeth. And so, we're looking for biomarkers. What we provide is—we always measure, at the moment, seven thousand proteins. And so, we know what we're measuring. The question is for any given researcher, how does that relate to what else they're measuring or the scientific question that they're trying to ask and answer. There's been some wonderful breakthroughs in terms of cancer tissue profiling, in terms of genetic variants. But they don't answer everything. They don't, for example, they don't necessarily tell you about the host response. What they're great at is telling you what's different in the tumor genetics. But as we're realizing in immunotherapy, a lot of that depends on the host. And so, that's where proteomics comes in, I think it can complement some of the genetic changes that are unique to the tumor with things that the host is or is not doing to resist or not resist the tumor effects.

Steve Lewis 10:25
Do you have any thoughts around the recent gene therapies that have been developed? I think, with what you just kind of described around, specifically immuno-oncology, do you have a lot of hope around some of these gene therapies that are being developed and are significantly in the pipeline right now for the FDA.

Dr. Steven Williams 10:52
So, I think that what we've seen people use proteomics for there is: can you predict or detect early toxicity? So, some of these therapies have quite nasty adverse effects. And if you knew, if you could predict upfront, who was going to get one of those, and if that pattern is encoded in the proteins, either beforehand or an early response, then it might help you head off some of those adverse effects. The other way that you mentioned, the FDA, that they're using the SomaScan technology is to look at the downstream effects of different therapies. They're using it for biosimilars. Whereby you might, there's a biosimilar product which purports to be equivalent to the original product. How do you know it's completely equivalent? And the FDA is looking at using the pharmacodynamics, using seven thousand protein measurements. There's a pharmacodynamic fingerprint of any one of these therapies, and they're looking to say, you know, does the biosimilar have the same pharmacodynamic fingerprint as the original? But as for gene therapies, I think, yeah, will be, if they work, they're going to change the downstream proteome, not just in the target protein, but do they change any other proteins that might represent an off-target effect?

Dr. Gabriel Alves 12:31
How would that change drug development in the future?

Dr. Steven Williams 12:35
So, I think that the way we think that proteomics helps the productivity equation is right from the beginning, in healthy volunteers or in early patient studies, you can look mechanistically, again, casting a very wide net, seven thousand proteins, you can say, is the mechanism I expected changed and are there any off-target mechanisms that we don't like, or that we do like, that might be unexpected benefits of a product? So, measuring seven thousand biomarkers at once helps you do that. And then there's the tests. Because the SomaScan tests that I showed you, they ride on top of the seven thousand measurement platform. And the reason that drugs costing a billion or two to develop isn't because any one drug costs that much, it's the 90 out of 100 failures that cost that much. So, if you can reduce the cost of failure by taking it earlier, finding you're not going to get the effects you wanted, or you're going to get the effects that you didn't want, if you can find that out early, or if you can find out good things early. Or if you can choose the dose more precisely because you could look at the mechanistic effects on the proteins. So, I think all of those things are ways that proteomics is already helping drug development.

Steve Lewis 13:58
We hope you're enjoying this episode of Speaking of Mol Bio. We wanted to take a quick moment to tell you about the Invitrogen School of Molecular Biology. It's a great educational hub for molecular biology with rich and reliable technical content designed for new and experienced molecular biologists alike. Check it out today at thermofisher.com/ismb. And now back to our conversation.

Dr. Gabriel Alves 14:30

I would like to switch gears a little bit and ask you some questions in regards [to] validation. First of all, the molecular aspects of validation, what kinds of molecular biology products do you use to validating and in terms of applying this to healthcare, applying this to drug development. How is the validation or paperwork going on?

Dr. Steven Williams 14:55
The interesting thing about our technology is it's really underpinned by a new kind of reagent, the aptamer. So, they're sequences of DNA that have been evolved if you like to bind to a particular epitope on a protein. And the interesting thing is, we measure seven thousand proteins at once. But people are naturally much more suspicious of a new measurement technique like aptamer-based proteomics, than they are of antibodies or antibody pairs. So, there's a demand for orthogonal validation that are you actually measuring the protein that you said you were measuring? So, there's the faith that people have over antibodies [that] doesn't translate to a new technique, and so people tend to say, well prove it. And, of course, proving it is quite hard because of the seven thousand proteins we measure, probably, at least five thousand of them aren't measurable by other common techniques. So, it's difficult. What we've had to do then is to take the reagents one at a time, and to work our way through the menu, trying to get orthogonal validation; that we are measuring the intended protein. So far, we've got, I think it's four thousand nine hundred fifty-four of them, have got at least one alternative method showing that the reagent binds to the right protein. So, what are those methods? Well, we do pull downs, we do gels, on all of them. So, all seven thousand have got gels, well, we've pulled out the protein, the source protein, and put it on the gel, and shown that the molecular weight of the protein is exactly what you would expect. But that's not really counted as orthogonal validation. But we do that for every reagent. So, we've got that for seven thousand. We've got gels for all of those. Orthogonally, though, it's some other technique that says that the reagent you're using is measuring is binding to the protein that you want. So, one of the ways of doing that is through genetics. So I think two thousand two hundred of the reagents have a cis-PQTL (protein quantitative trait loci), where in a large genetic study where the genetic variations are known, you can show that where a particular genetic variation is present, and in or around the gene, our protein measurement is different. So, the genetic variant leads to a change in the protein measurement. That's a cis-PQTL, and if the name of the protein lines up with the name of the gene, that's a way of orthogonally validating that, yeah, you're measuring that protein. Another set of methods is your mass spectrometry. I think we have over two thousand mass spectrometry–confirmed protein measurements that we are measuring the right protein. And then there are immunoassays. So, there's there much of the seven thousand in the menu, there aren't very many immunoassays, but there are some, a couple of hundred immunoassays. We're taking it seriously that people have doubts about new measurement techniques. And therefore, we're being quite rigorous. In a way, we're being more rigorous than people with antibody-based techniques are being because they get in, they get grandfathered in on faith. Whereas because we don't, we have to use these alternative methods for validation.

Steve Lewis 18:40

Commented [RA1]: You may want to provide the full term for the abbreviation (Protein Quantitative Trait Loci)

One area of interest for us, and I'm interested to hear your perspective, is we do offer gene optimization where you can modulate or up-regulate the expression of a protein based on optimizing a specific sequence related to that protein. When you're doing clinical development, do you have any of those kind of processing challenges that you might take into perspective, in terms of how a gene is expressed and how much of a gene is expressed?

Dr. Steven Williams 19:22
Yeah, there are challenges, I think, in that what we're measuring with the aptamers is a three-dimensional shape charge epitope on a protein, and that can change if the amount of the protein is different, but it can also change if the proteoform changes. So, if there's a change in the shape or charge that might be mediated by a genetic variant, then our measurement will change. And if that measurement, if that genetic change is physiologically interesting, then it's actually useful. And what we observed there in the clinic is we often see a tri-modal distribution of the measurement of the protein. So, the people who may be homozygous wild-type, maybe have, usually have, the highest level because we've designed the reagent to bind to the unmodified protein. And then at the other extreme, there'll be people who are homozygous with the variant, which will have the lowest level, it may be absent, or it may be the aptamer may not bind to that protein at all. And then in a population level, there are people in between there, they've got some of each kind of the protein. So, you see, you often see a tri-modal distribution in a population of a protein due to those genetic variants. If they were all physiologically relevant, and that the measurement change reflected a change in physiology, then you'd be happy, wouldn't you? You'd say, these changes are useful. The issue I think that we're facing is that not all of those changes are physiologically interesting, and that there are some changes in the genetic variation that do change the shape of the protein that will change the aptamer binding, and they will create a mode, if you like, in the population, but it's noise. What they're doing, if the irrelevant variations are actually causing us a bigger spread in the population measurement of proteins that you wouldn't otherwise have. And I think this is partly where proteogenomics comes in, in that if you knew the genetics and the proteome, then you'd actually have three normal ranges in a population. And then you'd be able to actually measure whether any, any physiology changed when one of those modes changed or when someone in one of those modes changed. Whereas today, I think that that looks like noise. Because you've got this genetic variant which is picked up by the protein measurement, and because without the combination of the two, and without resetting the normal range to be into three modes, you're not seeing, you can't see, the physiology. You can't see whether it's really happening.

Dr. Gabriel Alves 22:23
Yeah, the technology that is showing to be very valuable for the present and upcoming future and talking about future, Steve. I would like to hear from you, where do you see proteomics, genomics going in the next five years?

Dr. Steven Williams 22:39

Well, I can certainly go on to the where the cutting-edge of proteomics is going. So, more measurements. How many measurements is enough? People often say, well, we've heard people say, well, three thousand was enough. But it turns out, we've been able to prove that three thousand isn't enough, because we've had different generations of the SomaScan assay its start, when I joined SomaLogic, was eight fifty, then eleven hundred and thirteen hundred and three thousand and four thousand, you know, now it's seven thousand. I don't know where we will stop it. At some point, it will be diminishing returns. We know there are roughly twenty thousand genes. So, when we get to, maybe when we go from nineteen thousand to twenty thousand, maybe the amount of biology won't keep on going up. You only know that when you look back in time. You only know that, oh, when we added the last two thousand, three thousand new measurements; actually, we didn't we didn't enable any interesting new biology to be found. We'll know that one day. But actually, today, we know that we have not yet reached that plateau. So, twenty thousand native unmodified proteins, but then you think, well, how many modified proteins would you like to measure? When we develop an aptamer to a protein, we pick the best one. But we sequence hundreds of them. There are families of aptamers that also bind to that protein, but they bind to different parts of the proteoform. So, would it be useful to have more than one reagent to probe different parts of the proteoform at the same time? About six hundred of them are present in the current assay, and we know that sometimes machine learning chooses two of them. Because they're not completely correlated. Machine learning only chooses reagents if there's new signal, not noise. So, if they're perfectly correlated, it doesn't choose them. So, we know that there's biology and probing more of the surface of the protein. So, that would be another angle that you would go for, I guess is, as going on beyond one reagent per protein and probing multiple parts of the surface that might tell you about different biological aspects of the proteoform. So, I think that's where proteomics is going in terms of content. And then the other thing that needs to happen or is happening is, you know, that you need to engineer down the cost. And of course, Illumina is helping us with that in terms of the sequencing, we're also working on automation of the assay and cost reduction there. But yeah, we'd like for it to be scalable and used on all of us. We'd like it to be more scaled and cheaper and more globally distributed. So, I think you'll see all of those things happen as well. And we are working on the global distribution, we're installing the SomaScan setup kit sites, we call them, but basically centers that want to run the assay themselves. That is kicking off and scaling up in the next—from now onwards. So, I think you'll see that too, that people want, at the moment, we run most of the assays centrally, and you'll see that global distribution enabled over the next few years, as well.

Steve Lewis 26:24
One of the areas related to aptamers that that I'm somewhat fascinated by, and one of the areas that Thermo Fisher Scientific has been trying to emerge a little bit is in the cryo-EM space. You mentioned the charge or even, I believe it's the topology of the protein. How can a technology like cryo-EM change the landscape of drug discovery, especially when you're talking about aptamers that I believe you said it was it was a proteoform. Yes. I'm interested to hear about how new imaging technologies like that will maybe even open up the field even more.

Commented [BJ2]: There is an assay mentioned here that I am not familiar with/could not quite hear.

Commented [OS3R2]: Edit was made to address this comment.

Dr. Steven Williams 27:08
Yes, I don't know for sure, so I'd have to speculate. We develop our reagents to recognize the proteins in the context of our assay. So, in human plasma at the temperatures that we operate at, but we, it's certainly possible to develop aptamers that bind to proteoforms in different circumstances, you just would select them differently. We select them under the same environmental conditions as the assay will run, and by select what I mean is, we don't design the aptamers. We take a protein, and we—a highly purified human protein—and we incubate it with a library of 10^{15} random variations of aptamers. And we keep the ones that bind, get rid of the ones that don't, we amplify them back up, we do it again, and that process is called the select process. That's how we choose the binders from the massive library; by sequential amplification. Even though the aptamers themselves are modified, we've chosen enzymes that are tolerant to the modifications, the modified side chains in the aptamers. So, the selection process does use PCR amplification. Where I'm going with this is to think that if you, if you wanted to choose aptamers that bound to proteoforms or shape charge epitopes, and it doesn't have to be just proteins because they made aptamers to bind into caffeine and theophylline, for example, what you're looking for is something that has a consistent shape charge epitope. And if you can select the aptamers under those conditions, then you could find reagents that bound under those conditions, but I'm just being rather speculative. We know that, for example, with tissue microscopy, formalin-fixed tissues, the aptamers generally don't bind very well because the proteins have become crosslinked. But people have published on undoing, reversing the crosslinking, and then you can get the aptamers to bind, or you can get them to work in that kind of process on fresh frozen tissue because the epitopes, the proteoforms, are preserved. But I do think here I've speculated again, that if we did selection of proteins in the formalin-fixed state, then you might find an aptamer that actually recognized the formalin-fixed version of the protein. But because that's not how we did it, then the reagents generally don't bind. So, I think that that's a select process. And again, people have published cell selects, where you're not selecting to a known protein, you're incubating a cell, with a library of aptamers, and you're seeing which ones bind, and then you find out later what their proteins were. So, I think this this selection process of finding aptamers that bind to epitopes that you're interested in, that could have a whole set of new applications, but those are outside the scope of what we do today.

Steve Lewis 30:34
Got it. And you are alluding to membrane proteins there, right?

Dr. Steven Williams 30:41
Yeah, the cell selects is basically bound by finding the surface epitopes of membrane bound proteins.

Dr. Gabriel Alves 30:52
Steve, what would you say is the most important ingredient to your success.

Dr. Steven Williams 30:59

The company's been going for twenty years. And I think that for the first ten or twelve years of it, it was Larry Gold, our founder, trying to make the technology work. And so, I would say that his persistence in the face of sequential failures was the most important ingredient to actually—to be scientifically persistent, and to get people to keep on funding the technology for that period of time, enabled the assay that we call today, SomaScan, to actually be developed. All the iterations since then, after the first ten or twelve years, have been improving the basic formula, but it took a whole set of blind alleys and failures, but overcome by creativity and persistence to actually make it work. But it took a long time. And it took a founder who had that vision and was persistent enough and could keep getting funding for long enough. So, I think that it's a human quality that he had that meant that this technique was actually developed.

Dr. Gabriel Alves 32:22
That was Steve Williams, Chief Medical Officer at SomaLogic. If you're interested in hearing even more of today's conversation, you can view the extended video version of this interview by visiting the URL in the episode notes and consider sharing this episode with a friend or colleague. We love sharing our conversations with more great people. This episode was produced by Matt Ferris, Sarah Briganti, and Matthew Stock.