

An interview with Assoc. Prof. Nikolai Slavov

Hosted by
Dr Jarrod Sandow

Topic:
Single-cell proteomics and
the SCoPE2 method.



Full Transcript from March 22, 2021

Jarrod Sandow:

Welcome Nikolai.

Thank you so much for making time to sit down with us today - it's a real honour to be talking to you. To start us off, can you tell us a little bit about you, your lab and research, and in particular, what attracted you to focus on proteomics, and in particular, the field of single-cell proteomics?

Nikolai Slavov:

Thank you very much for this invitation. It's a pleasure to discuss those topics with you. One of the things that is very important for me in terms of the research topics that we choose is to make non-redundant contributions. Which means that I like to do work that perhaps will not happen now, or it will happen later in the future, unless we help give it a push in that direction. About five, six years ago, when I started my lab, I felt that the field of single-cell proteomics by mass spectrometry was in that position. That the technology created the capability to do this kind of analysis, while at the time that wasn't as actively developed and recognized. And I liked the challenge. I felt that this kind of analysis will be both feasible and also very fruitful in terms of the biological questions that we can answer with it, and the insights that we can obtain.

So I assembled a small group of colleagues and undergraduate students at the beginning to help me with the initial experiments in quantifying proteins in single mammalian cells, and for the most part these experiments worked better than expected. We had early, promising results, and we kept going in that direction. But going back to your bigger question, in terms of the research in my group at the moment, we obviously focus quite a bit on developing methods for single-cell proteomics by mass spectrometry. As these methods mature and become more powerful, we are increasingly applying them to various biological problems.

In particular, we have projects that use the technology to map protein abundances in 3D at single-cell resolution in healthy and diseased human tissues, which is more on the descriptive side of things. We also have projects that are leveraging the power of single-cell proteomics to identify molecular mechanisms that regulate biological functions. We are also broadly interested in post-transcriptional regulatory mechanisms, not necessarily only studied at single-cell level. We are also interested in ribosome modifications that can regulate protein synthesis.

Jarrod Sandow:

Yeah, well, it was really exciting to see, I guess, you've talked about being a pioneer in some of these method developments. It was really exciting to see the recent publication on your SCoPE2 paper titled Single-Cell Proteomics and Transcriptomic Analysis of Macrophage Heterogeneity Using SCoPE2 published in Genome Biology. I'd also like to note that we noticed that the first author, Harrison Specht, was also listed in the Journal of Proteome Research as a rising star in proteomics. I guess speaking to the quality of the group that you've managed to assemble. So, congratulations to you both. Can you give us a little bit more information and run through this paper, these techniques, and what are some of the important insights that you discovered during this process?

Nikolai Slavov:

Thank you. So the paper has two components. One is the development and improvement of the original methodology that we came up for quantifying protein single cells, and the other is the application to an important biological problem. I see the two as being intricately connected, because the utility of a method is best demonstrated in the field when applied to a real problem. So on the side of method improvement, the overall increase in accuracy in throughput is very significant, on the order of some steps as much as a hundred-fold, such as the decrease in cost for a sample preparation and increase in throughput. And some parts are ten-fold improved, as accuracy and so on.

These improvements came from a lot of components that interacted synergistically. Some of the significant improvements include sample preparation, as I already mentioned. We replaced the original cell lysis by a focused acoustic sonication with a freeze-heat step, which we spent a lot of time validating and making sure that we can apply in a parallel manner to efficiently extract proteins for a mass spectrometry analysis. We've completely automated the sample preparation in the sense that protein digestion, and peptide labeling and all of these steps can be performed by liquid handler rather than manually, as we used to do. This both increases the throughput, but also can add a layer of reproducibility and standardization because the quality of samples is less dependent on the person who is preparing them.

Equally important is the improvement in the liquid chromatography, the front end of the instrument. There we've used high performing analytical columns, which allowed us to obtain a very high performance separation of peptides and efficient delivery, efficient ionization into the instrument. That part, for us, actually took a while because initially we did not start using IonOpticks columns. We started using a variety of different columns, and the results weren't so great. With some columns, interestingly, some commercial columns, we were able to obtain very, very good results when using highly purified standards and peptides. But those same columns resulted in very poor results when used with real samples that had not been cleaned up. We also packed in-house columns, but I strongly preferred to use a commercial option because that would make the method much more easy for other groups to adopt, to be standardized. One of the advantages also for our own internal use has been the extremely high degree of reproducibility of retention times and performance between different batches of IonOpticks columns. Another good commercial option that we encountered in that process was PharmaFluidics. We used their columns as well, and they're a reasonable option. But we obtained better results with IonOpticks, and also considering the lower price, that was a decisive consideration to still keep using them.

Of course, we made some changes in data acquisition methods, and the analysis and interpretation of the mass spectrometry data. Again, coming back to retention times, we were able to leverage the very high accuracy of retention times as an additional feature to increase confidence in correct peptide identifications and to reduce confidence in incorrect ones. So, that's on the side of the method. Probably more details than you anticipated.

Jarrod Sandow:

No, I'm interested, in fact, in delving a little bit more into the method. As an emerging field with single proteomics, there's a number of methods that have been proposed by a variety of research groups at the moment. What do you think makes the SCoPE2 method stand out from the crowd when compared to the other alternatives that are being proposed at the moment?

Nikolai Slavov:

So another category of methods is performing label-free analysis, analyzing one cell at a time. When I see the promising results in that field, my first reaction is to think that this is a great technical feat, to achieve that level of sensitivity. But the flip side of those technical marvels is the lower throughput of analysis. Because only a single cell is analyzed at a time, the total number of cells that can be analyzed per unit of time is fewer. And the cost per cell is substantially higher, because most of the expense associated with single-cell protein analysis, with either our methods or other methods, is instrument time. Mass spec time spent in the analysis.

One aspect of the SCoPE2 method and the methods we have developed of SCoPE-MS is the higher throughput afforded by multiplexing as compared to the label-free approaches. Another aspect is the affordability of the methods in terms of using only commercial equipment that is relatively inexpensive when compared to some of the other methods that use instrumentation that is not commercially available, and therefore more difficult to reproduce.

So I think the methods that we develop are relatively easy to adopt and be widely used, and that's not a coincidence. That's, in fact, a very important feature that has always motivated our efforts at various branch points. When we had to make a decision, do we go for higher performance or do we continue to emphasize methods that are likely to be robust and widely reproducible by other groups, we've always chosen the option of the robustness and broad accessibility. Because I think that for single cell proteomics to have a broad impact, it has to be applied broadly by many groups, not the few groups that are interested in method development.

Jarrod Sandow:

I'm actually interested in a little bit more of the biology as well. That was contained within the SCoPE2 paper. I was really interested to see that you demonstrated that the macrophage populations that were actually previously thought to be quite distinct, actually exist across the spectrum when you use your single cell techniques. How prevalent do you think this is? This sort of mis-characterization of cell populations is in the literature? Just because the tools previously weren't available, people have relied on flow cytometry, or even single cell RNA techniques to define these populations, instead of something like single cell proteomics.

Nikolai Slavov:

I think there are many examples of people having assumed that there are discrete populations, then having isolated cells based on a few markers into discrete populations and having found that the data indeed confirms their assumption that there are discrete populations. Unfortunately, when one performs this kind of analysis, there is no way to reject the incorrect assumption of discrete populations because by design, one measures discrete sub groups of cells.

Now, in the case of macrophages, it wasn't entirely a surprise that there is a spectrum of polarization. While traditionally macrophages have been studied in two extreme classes of either being classically activated or M1-type macrophages, which are inflammatory or alternatively activated or M2 macrophages, which tend to be anti-inflammatory, in reality, it was known before our work that macrophage polarization is more complex and that complexity of the polarization is indeed very important in a variety of clinical settings and relates to the diverse biological functions that these cells have.

What we found that wasn't known and we did not expect is that this kind of polarization into a continuum spectrum occurred even in the absence of polarizing cytokines. It occurred when we started from a clonal cell population of cells growing in more or less homogeneous conditions, giving rise to this continuous spectrum of polarized states, which actually happened to correlate quite well to the previously defined M1 -- M2 axis of polarization. They did form a continuum in a way that we could have never identified without performing single cell protein analysis.



Jarrod Sandow:

Well, actually, I'm really interested to hear that, describing something like flow cytometry to define cell populations where you go in with the preconceived idea and sort of unsurprisingly, you find that the two distinct populations or three or however many. I mean, now that we have a harnessing of the power of single cell proteomics, where do you see some of these established techniques, flow cytometry, CyTOF - technologies like that - that are limited by the number of event things they can sample per cell? Where's the future for those, given that the techniques that we're currently developing?

Nikolai Slavov:

I think they can be applied more to clinical settings, not so much to the discovery phase, where having a method with a much deeper coverage, less biased coverage can enable new discoveries for which we did not have prior hypotheses. But when it comes to analyzing predefined markers in a higher number of cells, I think one can still use the antibody-based approaches, especially when highly specific antibodies are available, which unfortunately is not commonly the case. It tends to be more frequently the case with surface proteins because they have received more attention in terms of antibody development, and they're less affected by the molecular crowding problem associated with introducing antibodies for intracellular proteins.

So these methods have traditionally played an important role, and I think they will continue to play a role in various contexts, but certainly, the power of unbiased quantification of many thousands of molecules has been well illustrated by single cell RNA sequencing methods recently. And there is a lot of interest, certainly in industry and in big pharma, in a number of companies with which I interact, including Merck, Sanofi, and others. There is a very significant interest in being able to perform large-scale, unbiased protein analysis in single cells.

Jarrold Sandow:

One of the advantages of things like flow cytometry and I guess increasingly, single cell RNA sequencing, is the ability to analyze thousands of cells, tens of thousands of cells that in quite a short space of time. When do you think we will be in that sort of level in the proteomics field where we can characterize these enormous populations of cells and pull out very discreet and small populations out of those tens of thousands, hundreds of thousands of cells?

Nikolai Slavov:

Well, if we think of single cell RNA sequencing methods, there is a spectrum in terms their capabilities. Some methods are very good at analyzing large number of cells, but they do that by sacrificing capture efficiency and the number of copies for messenger RNA that are being detected per gene and the number of genes detected. In contrast, the multi-well plate based methods have higher capture efficiency and lower throughput.

And to some degree with mass spectrometry, I can see a similar trade-off with methods that use shorter acquisition times that might be able to analyze more cells, quantify fewer proteins and conversely, methods that are more time consuming that have higher depth of coverage.

At the moment, we are able to analyze about 200 single cells per instrument per day. So in the course of a couple of weeks, even with a single instrument, we can analyze a few thousand single cells. And I think at this point we can already use the data for a lot of interesting biological experiments. I have no doubt that throughput will continue to increase, and I think one prominent avenue for that to happen is by increasing multiplexing capabilities. To some extent, single cell protein analysis might be a primary driving force for that because it provides compelling motivation for increasing the multiplexing. It's not trivial to do it, but it's possible, and I believe that will happen.

And it's very good to think of ways in which we can do that, perhaps shortening also the duration of separation so that we can analyze a larger numbers of sets per unit time. But what I think what we have to realize -- and I want to emphasize -- is that even though the current throughput is more limited than that of single cell RNA sequencing, we can still analyze in reasonable time and at reasonable cost, many thousands of single cells. And we are ready to apply these methods to biology and biomedical questions, not just for method development and technology-oriented purposes.

Jarrold Sandow:

When developing these new technologies, these new methods, you can really only dream of the impact you might help out on biology on the community, on human health. Where do you think the end game is for this technology? What is the sky's the limit type of approach here?

Nikolai Slavov:

One very challenging, long-term goal is to use these data, to identify molecular interactions and mechanisms with fewer assumptions. So let me step back to first describe what the problem is and then how single cell protein analysis might help with that. We have a variety of methods in biomedical research that allow us to establish associations between diseases and molecules. And in some cases, even if we establish very confident causal associations, for example, between a DNA polymorphism and a disease such as diabetes, we have a very difficult time understanding what this means because the association is very indirect. It might be causal, but there are a hundred different molecular interactions between the DNA polymorphism and the phenotype of interest. And therefore, that causal association is consistent with close to infinite number of possible models. And that's a major hurdle to being able to develop therapies based on that observation.

The advantage of being able to measure protein abundances across many, many single cells is we can measure the molecules that are directly interacting with each other. And then we can condition these measurements that we have from possible confounders. For example, if we want to understand whether a kinase I phosphorylates kinase J, we can quantify the activities of these kinases across lots of lots of single cells, thinking of each single cell as representing a perturbation of a kind, as well as all of the other possible kinases that might be phosphorylating kinase I and J.

And then we can condition so that we pick a subset of cells in which the other kinases do not change their activity, and we just look at the section of the data at the activities of kinase I and J and ask whether the joint distribution of their activities is just a product of the marginal distributions or not. And the advantage of being able to do this kind of analysis, is that we can determine whether there is a direct interaction between these kinases' direct regulatory link, without needing to make any assumptions. Of course, to be able to do this type of analysis, we need to have highly quantitative data across many single cells, and I think at this point, the data that we have will be highly challenged to make those inferences with high confidence, but nonetheless, I think the potential is there in the longer term, that we'll be in the position of being able to make that kind of analysis.

Then of course, at the moment, we have focused on analyzing protein abundances in single cells because that's the entry level thing that is accessible. But much of the technology that is currently being developed is also going to generalize to many other types of analysis that you can do with mass spectrometry, such as direct measurements of protein-protein interactions, direct measurements of protein confirmations, protein localization and so on.

Jarrod Sandow:

It's interesting you mentioned analysis as one of the great challenges of looking at these large datasets, that we're likely to generate. You're really well-known and active in the global community with your online lecture series, where you actually walk through a lot of these datasets and challenges that we face. Can you tell us a little bit more about the motivations behind these initiatives and where people may be able to find out more?

Nikolai Slavov:

I am tremendously excited about the potential of mass spectrometry data to answer key pressing biological questions. When I take a large-scale 10,000 feet view of what are the challenges to realizing that potential, I see a disconnect between the questions that we ask, the data that are present and have the potential to answer them and the ability of researchers to effectively analyze the data, to answer their questions. I think that there is need to explain what the strengths and the limitations of the existing data are and how to properly analyze them with both simple and more sophisticated approaches, so that we can make reliable inferences about the underlying biology.

I have always felt that the biomedical research community can benefit substantially more from mass spectrometry based proteomics, metabolomics, if they understood better those technologies and therefore, I try to provide an introduction to those technologies in a manner that I aim to make accessible and conceptual, so that I can maximize the benefit from these really wonderful technologies for biomedical research.

I try to do my best to reduce the jargon that is an inevitable part of all fields, actually, but nonetheless that jargon can be a significant impediment to taking full advantage of the available data, the available methods. I know it was for me when I first transitioned from genomic systems biology to mass spectrometry. I had a hard time reading a paper and understanding it because every sentence I had to look up a couple of terms that were confusing for me. When somebody knows that terminology, we don't even notice it. We quickly understand it and we move on, but I still remember the challenge that I had at the time. Now I record short videos on focus topics. I try to make them as self-contained as possible and as accessible as possible, so that I can quickly introduce these topics to interested students, interested researchers, who may not have a lot of background in mass spectrometry, so that they can quickly get up to speed and know how to use the technology and the available data for their research.

All of the videos that I record are on YouTube. You can find them from our YouTube channel or just Google for them. [*Find all of the Slavov Lab videos on YouTube [here](#)*].

Jarrod Sandow:

It's interesting actually, you talk about jargon and the approachability of some of these types of analyses. When I was a younger researcher and in the same position, trying to approach these, the proteomics research was actually sort of a one-stop shop. You did the sample prep, you ran the mass spec, you did the analysis of the sample and then you handed it over to your collaborator. With these larger experimental constructs and large data sizes and complex questions, how long is that one-stop shop going to last? Are we going to just increasingly have to involve multiple people in a mass spectrometry experiment, a proteomics experiment, where bioinformatician is always involved, just to get the most out of some of these data sets that are being generated?

Nikolai Slavov:

Well, I think the people or the group of people, it can be a person, it can be a group of people, but the people in charge of designing the experiment have to understand all of the technical limitations and capabilities of mass spectrometry. I frequently find that even very well informed and brilliant colleagues, leaders in biomedical research, have misconceptions about mass spectrometry and they still think of mass spectrometry as perhaps being a semi-quantitative methods at best, something that perhaps was once upon a time the case, but they do not realize all of the strengths of current methods.

On the other hand, current methods do have weaknesses. They're not perfect, and one has to understand what these weaknesses are and how to circumvent them, to design the experiment in an optimal way. So somehow we have to find a way to put together that knowledge for all of the different aspects of the analysis. How the data will be analyzed, what will be the power of the study, how many samples we need, what kind of methods to use, what are the trade-offs between having a lower depth of analysis versus more samples, what methods we use, what is the difference between the accuracy of relative quantification versus absolute, how being able to use a stable isotope labeling approaches allows to cancel out nuisances due to variability in chromatography or ionization.

All of these are very, very important things to know for designing an experiment in an optimal way and while it is true that having all of that knowledge at once is not commonplace, I think if we are able to communicate it in a conceptual enough way, we can get closer to a situation where more and more people are able to design better and better experiments. The more we can do that, the more we can maximize what we learn from those experiments. I think in the best case scenario we have somebody who knows everything and designs perfect experiments and that is not likely to happen. But the more we can approach it, the more progress we are going to make. I think part of that is finding a way to explain in conceptual ways certain aspects of the work. Sometimes some details are not essential and it's knowing when to focus on those details and when those details can be skipped, so that one can provide the essential information for designing the experiment. I think that kind of balance is very important to strike and can be quite helpful.

Jarrod Sandow:

Additionally, from a different educational front, you're also a founder and an organizer of the annual single cell proteomics conference that you host at Northeastern University. How over the years has attendance and participation changed as the interest in the field grows from scientists, from vendors, from industry and pharma?

Nikolai Slavov:

That's a meeting that gives me a lot of delight. I enjoy organizing it. It's also time-consuming, so it's not only pleasant, but it is mostly pleasant for me. The meeting started in 2018, and we had a group of about 60 brave, interested colleagues who wanted to explore what this new field is going to be all about. And then since then, the number of attendees has been doubling and tripling every year.

I very much enjoyed the in-person meetings that we had before COVID. And with COVID, the meetings have become virtual. And this year the meeting is going to be hybrid. It's going to be both in-person and it's to have a virtual component. And I actually do like the virtual component, and perhaps we will keep that even after COVID, because it allows people who are not able to travel for various reasons to attend the conference and makes the presentations more accessible, which of course is one of my primary motivations for organizing the meeting, to make technology and progress accessible to as many people as possible.

And this meeting also emphasizes my conviction that we need to have a better communication between different disciplines so that we can enable interdisciplinary research. So, the meeting combines presentations from computational biologists, from experimental biologists, from mass spectrometrists, from systems biologists. And it's a wonderful interdisciplinary meeting that I certainly enjoy very much. And it's also a meeting that provides ample time for interactions. It is not intended to be a go to listen to meeting. After each presentation, we have 15 minutes for discussions. And usually that time is not even enough because there are plenty of questions. We always use up the time. So, there are multiple slots throughout the day where we specifically engage the community in discussions and interactions.

We also organize workshops as part of this meeting so that we can share details of experimental analysis or data analysis and provide a forum for exchanging what we have learned the hard way, what other colleagues could learn the hard way so as many people as possible who can benefit from that experience.

Jarrod Sandow:

And as a means of promoting it to the community, can you tell us when it's being held this year, the dates, and is there a specific theme for this year's conference?

Nikolai Slavov:

That actually will be the first time when I announce the change of dates that will be announced this week more formally and more broadly. But the meeting was planned for the beginning of June, but because of the delays in COVID vaccination and because of my desire to maximize the in-person attendance, I will shift the dates to August 17th and 18th. And the meeting will be held in person in Boston. It's always held in person in Boston at Northeastern University. And it will have a virtual component that all registered attendees can join.

And we usually limit the virtual attendance in the interest of allowing for more discussions between the participants. But if you cannot join the meeting either in person or virtually, you should be able to watch the recorded presentations of all speakers who agreed to record their presentations and post them on YouTube. That's a tradition of the conference, that we always ask presenters whether they're willing to share their presentations on YouTube, and if they give us permission, we upload the presentations to [YouTube](#).



Jarrod Sandow:

So thinking more broadly about mass spectrometry, LCMS, what sort of technical technological advances would you hope to see in the next five years? What are your frustrations right now? Where do you think the field is going over the next five years?

Nikolai Slavov:

Well, there are lots of things that can improve, but I think increased multiplexing is going to help tremendously to make the technology more accessible and make it higher throughput. It is common that the lower throughput of mass spectrometry limits the statistical power that we have for various biological analysis. And I don't imagine and I don't see a possible path to instruments becoming cheaper. They become more and more powerful, but not cheaper. So the only way to increase a throughput and make it more affordable is to decrease analysis time, and multiplexing is an obvious way to do that. Of course, we can also shorten the active gradients to a degree, but multiplexing appears to provide a path to increase throughput with fewer trade-offs. And that's certainly something that can be tremendously helpful for single cell analysis as well. There are certainly many advances on the side of instrumentation, increasing the resolution, which is especially important for top-down proteomics, and increasing speed of analysis. But I think what is probably even more limiting and more important to improve is the front end of the analysis, improving the robustness of separation and ionization.

"Life is a dance of protein interactions, and if we cannot measure those interactions, if we cannot follow that dance, we cannot understand it."

At the moment, one of the major limits of our sensitivity is the efficiency of ionization and the fraction of ions that we sample from those that are available that even go into the instrument, because if a particular peptide elute over a period of five, ten seconds, we sample it only for hundreds of milliseconds or maybe for a second. And the sharper those solution peaks are, the more efficiently we can sample a larger fraction of the peptides.

So we think it's both on the side of robustness of the front end and an increased ionization efficiency, increased delivery of peptides to the instrument in the form of ions that many of the significant gains can be made in terms of sensitivity and usability.

Jarrod Sandow:

And probably finally, a bit more of a question about the field in general. When you hear of a statement, something like the power of proteomics, what does this statement mean to you? What is the power of proteomics?

Nikolai Slavov:

Well, it's a very broad statement. It's to understand life. Life is a dance of protein interactions, and if we cannot measure those interactions, if we cannot follow that dance, we cannot understand it.

Jarrod Sandow:

Excellent. Excellent.

Look, I'd really like to thank you for your time today. It's been really insightful and it's been great to be able to understand your specific field of research and your thoughts on the field more broadly and some of the ideas on the technological and analytical advances that come. Yeah, really thankful for joining us and it's been amazing.

Nikolai Slavov:

Thank you, Jarrod.

For more information:

Associate Professor Nikolai Slavov [@slavov_n](#) | [LinkedIn](#)

Slavov Lab [website](#) | [@slavovLab](#)

Register for SCP2021 [website](#) | [@SCP_meeting](#)

Watch more videos from the Slavov Lab on [YouTube](#)

