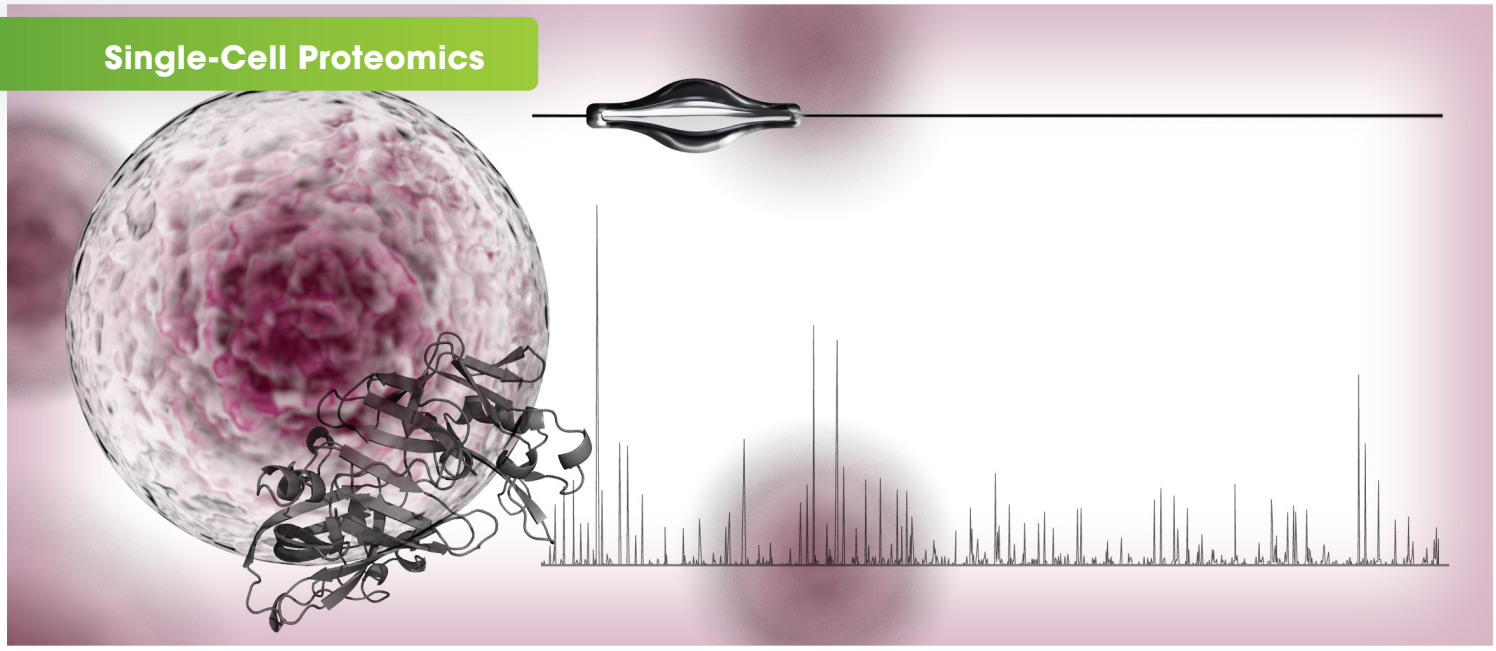

Advances in LC-MS Workflows for Single Cells

An Interview with Nikolai Slavov

Single-cell analysis has become one of the most exciting developments in biology over the last decade. The field of proteomics has seen many advances facilitated by improvements in the sensitivity, speed and robustness of mass spectrometry (MS) instrumentation, as well as an expansion of new workflows and applications. The move towards single-cell proteomics analysis has enabled a greater understanding of the complexity underlying cellular diversity.

Dr. Nikolai Slavov, Associate Professor at Northeastern University, has developed methods for high-throughput single-cell proteomics by MS and was recently named an Allen Distinguished Investor. An award that saw him receive a $1.5 million, three-year grant for his research in this field. In this exclusive interview, Nikolai describes his journey into single-cell analysis, the contributions his lab has made and the technological advances that have enhanced the role of liquid chromatography-MS (LC-MS) for single-cell proteomics.

Q: When did you first start your journey in single-cell analysis and why?

A: I first became very interested in single-cell proteomics via MS in 2012, when I was a postdoc in a lab that was using single-cell nucleic acid analysis. There were many exciting results emerging from the field, however they were confined to measuring DNA and RNA molecules.

Knowing enough about MS – and not buying into some of the orthodox opinions at the time – I believed that MS had the sensitivity to quantify thousands of proteins and single-cells, yet at the time I did not have the freedom and the resources to focus on developing these approaches.

When I started my lab at Northeastern University in 2015, I began actively developing these approaches, based on the exciting, but unproven, idea that we could enhance the sensitivity of MS to single-cells using an isobaric carrier. To give some context, MS instruments are very sensitive; they can detect relatively few copies of ions from any peptide. However, they need a much larger amount to determine the sequence of the peptide. I knew that while we would be able to quantify many thousands of peptide ions, if we didn't deliver enough material to the instruments, we would not know their identity and therefore it would be very hard to draw connections to the biology. To enhance the identification of peptide sequences and minimize losses during sample preparation, and other steps of the workflow, we employed what we call the isobaric carrier. This is a small bulk sample of 50 - 200 cells from the same population of single-cells you are analyzing. Since they are prepared as a small bulk sample, they are lysed and digested to peptides in the same way that we treat single-cells. Those peptides are then labeled with isobaric mass tags, such as the Tandem Mass Tag Reagents (TMT), and mixed together before analysis using tandem MS. This strategy mitigates losses due to surface adhesion, as many of those losses will be taken by the isobaric carrier material. However, due to the way that MS works in the presence of isobaric mass tags, many of the fragments that support confident peptide identification will also be pulled across the single-cells and the isobaric carrier, allowing us to confidently identify peptide sequences.

This was the basic idea that we started with in 2015; at the time I had no access to MS instruments, however, I was fortunate to have a good friend at Harvard University, Bogdan Budnick, who was a manager and director at the MS and Proteomics Resource facility, so I recruited him to work with me. We were both very enthusiastic. I also recruited a couple of undergraduate students who shared our enthusiasm to work in sample preparation. It was a very exciting – and risky – adventure, but the results from this experiment very early on were highly encouraging. It was clear that we were able to detect peptides from single-cells and that the signal was quantitative. Having detected some signal, we then had enough to keep improving the method until it became more quantitative. Despite not having many resources, we made excellent progress and had the time of our lives.

We received support from the National Institute of Health (NIH), when I was awarded the NIH Director's award, which gave me funding for this kind of exploratory, high risk/high reward work. I was finally able to purchase an LC-MS system for my laboratory, and I began recruiting more people to work on advancing these methods. The instrument that we purchased was very much an affordable system, however I believe that this was an advantage. While our results may not showcase the utmost capabilities of the latest MS technologies, if they work with a system that is not as powerful, they will work even better with the other systems. Most importantly this demonstrated to researchers that don't have the resources – or the funding to buy the latest instruments – that they can still do this analysis with relatively modest resources and more affordable technology.

Q: Can you describe the recent advances in LC-MS workflows for single-cell proteomics?

A: One very important aspect has been sample preparation. This comes as no surprise since traditional approaches are adapted for many millions of cells and they tend to use detergents that are not compatible with MS analysis, and when we remove those detergents, we also lose peptides and proteins from the bulk sample. If we were to apply this to single-cells, losses are disproportionately larger and, in many cases, this can prohibit analysis.

Our strategy has been to change sample preparation by avoiding chemicals that are not compatible with MS and thus reduce the requirement to clean. Colleagues are developing complementary approaches that use different detergents that are relatively compatible to minimize volumes, Ryan Kelly is making particularly good efforts in this area.

We have also worked hard to minimize volumes, and there are multiple versions of sample preparation methods that exist now to minimize volumes. The version that we first developed – and that is now relatively widely adopted – is preparation using 384 multiwell plates. All of the isolation during sample preparation happens in sub microliter volumes between 0.5 and 1.5 µL. The advantage being that it is widely accessible, as it uses equipment that most laboratories have, and is relatively inexpensive. The downside is that these volumes are still relatively large. We have developed a new droplet-based sample preparation approach which uses a system for piezo electronic dispensing of small volumes of 300 pL so that lysing, digesting, and labeling individual cells can happen in volumes below 20 nL. Both systems – the one that uses multiwell plates and the piezo electronic system – are fully automated, which is a significant advantage given that we must analyze tens of thousands of cells objectively in a way that doesn't reflect how diligent or not the particular student is. Students are better off focusing on the conceptual questions in the intellectual aspects of their research rather than preparing an infinite number of samples.

Techniques such as liquid chromatography (LC) and capillary electrophoresis (CE) are important for sensitive MS analysis, especially of single-cell samples. Different aspects [of what] have been improved by either enhancing the resolution by using high-performance separation of the peptides or through the use of lower flow rates that ionize peptides more efficiently. We haven't done a tremendous amount of original innovation ourselves; we found a commercial supplier of nanoLC columns (IonOpticks), which provides excellent reproducibility across batches and performs very well – good enough to enable single-cell sensitivity. We made a conscious choice to use a commercial solution so that it was accessible to anyone who uses the same type of chromatography. My vision from the very beginning has always been to develop a technology that can be widely adopted instead of developing something that only my lab can use.

Q: How have innovations in MS and automation aided your research?

A: A lot of the innovation with regards to instrumentation is driven by big companies as it requires millions of dollars of investment. While some of these innovations can be showcased by academic researchers, a lot of the research and the real progress is being done by instrument manufacturers such as Thermo Fisher Scientific. Our analysis has gained quite a lot from clever experimental designs and optimized parameters. Alongside the isobaric carrier, which I mentioned earlier, another example that has been very powerful is the development of automated pipelines which set various instrument parameters for the experiments in a way that maximizes the sensitivity and throughput. In the early stages, I would often get results that made no sense and would have to spend the week trying to figure out what went wrong. As I was going

through that routine it was obvious that we were doing the same type of analysis week after week, so we decided to completely automate that step with a software package so that we could quickly diagnose where the problem was. The emphasis here is on specific diagnosis; for example, we don't just want to know that we had lower peptide identification than usual, we want to know *why*, as that gets us closer to being able to take action and fix the problem. Although this is a relatively simple analysis, it has allowed us to quickly identify limitations and make improvements.

Additionally, there have been advances with regard to data interpretation. I already mentioned that you must detect enough peptide fragments to determine the peptide sequences and although the isobaric carrier helps us to do this, there may still be times where we didn't quite detect enough peptide fragments to confidently identify a sequence. There are other informative features such as retention time or ion mobility, which are widely used approaches in MS. However, they were not particularly well implemented, and I could not find anything that incorporated those features in a principled way to determine the probability of error and success. Some methods, such as match between runs typically work, but we cannot estimate the confidence of the peptides identified. We therefore needed to develop more rigorous frameworks that allowed us to compute the exact probabilities of having correct or incorrect peptide sequences from the data, and these are some of the main areas that my lab and others have contributed to.

Q: You and your lab have developed Single-cell ProtEomics by MS (SCoPE-MS). Can you talk us through how this technology works and why it was needed in the field?

A: We developed this technology in 2015/2016, as an accessible and relatively high-throughput way to analyze proteins in single-cells. We were hoping to have a method that could be used by any MS facility, or anybody who is able to perform quantitative analysis of bulk samples to be able to analyze single-cells as well. While developing SCoPE-MS, our focus has been on making this analysis accessible for as many people as possible.

Although multiplexing using tandem mass tags significantly increases the throughput, it is not as high as I would like it to be. Since the first report of SCoPE-MS, we have increased throughput by a factor of six and we are hoping to increase it further. That said, I wouldn't necessarily call this or any other current method for MS analysis of single-cells, high throughput. We currently can analyze about 200 single-cell per day, and I would like to be able to analyze thousands.

As to why it is needed in the field, at the time there were no other methods that I knew of that could analyze hundreds of proteins across single mammalian cells. MS has been used traditionally to analyze single-cells, but only in very special cases, such as for the analysis of hemoglobin in erythrocytes, as hemoglobin is significantly more abundant in erythrocytes than any protein in a mammalian cell. There were attempts to use matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry however, it is less quantitative. This was really our first foray into developing an electrospray ionization (ESI) method that could quantify hundreds and thousands of proteins across more typical mammalian cells.

Q: What are some of the key insights that you've gained when applying these platforms and methodologies to proteomic research?

A: As with any method and new technology, a lot of the initial progress has been focused on the method development itself and then, as the method becomes more robust, we begin to apply to other applications. That being said, there are certainly biological insights that we have been able to make with the current technologies. This is important, because saying that the technology is good is not nearly as convincing as demonstrating it. Even if that demonstration is in the early stages and more limited than what we think the real potential is.

Our first application for biology was to study the differentiation of pluripotent stem cells into different lineages. We started with pluripotent embryonic stem cells and triggered their differentiation before taking samples on different days and analyzing their protein composition. We found that, similar to RNA sequencing methods, SCoPE-MS could identify the lineages of different cells and find clusters of cells that had similar biological functions. We also found co-regulated proteins; these are proteins that form complexes and tend to cooperate, as one would expect, which was also part of the method validation. Using a paired joint analysis of proteins and RNAs, we found that there were some groups of structural proteins that were regulated more transcriptionally, while many of the proteins involved in the development of the regulatory functions exhibited more post-transcriptional regulation.

We then decided to investigate macrophage heterogeneity. We were interested to see whether macrophages might differ if they originated from a clonal population of cells. To give a bit of context, macrophages are innate immune cells that are present in all of our tissues, therefore they have a remarkable diversity of functions, including participation in tissue homeostasis and immune function. They can either attack and kill cancer cells, or they can protect cancer cells from the rest of the immune system, depending on their polarization. They also stimulate angiogenesis – the formation of new blood vessels and a very important process in the context of cancer. So, there is a huge variety of functions. We were interested to learn

about the molecular underpinnings, we wanted to know which protein networks regulate polarization and whether we could see polarization emerging even in the absence of external regulatory cues, such as polarizing cytokines.

We started with a monoclonal population of monocytes, which are precursors for macrophages, and stimulated them to differentiate into macrophages. We then analyzed the monocytes and the macrophages produced from this experiment using the second generation of SCoPE-MS, SCoPE2. Based on the data we were able to validate the methodology by clearly distinguishing monocytes and macrophages. Yet we also observed, quite unexpectedly, that the macrophages originating from the system were heterogeneous. Despite having originated from similar and more homogeneous cells and having not been stimulated with different cytokines, they still showed quite a bit of heterogeneity. Since we were able to quantify a significant number, about 3,000, of proteins in the cells, we were able to identify the basis of heterogeneity simply by seeing which proteins were more abundant in certain clusters from one group of cells versus another group of cells. We found that this heterogeneity lies along the previously characterized axis of pro- versus anti-inflammatory macrophages or classically activated versus alternatively activated macrophages. Interestingly, heterogeneity existed in a continuous gradient rather than discrete clusters. This is very difficult to analyze without single-cell analysis if you do not have the markers needed to sort the cells at the start. If you have the markers to sort them, you might be able to isolate distinct populations, but you cannot know that those existing populations form a discontinuous gradient, without single-cell analysis.

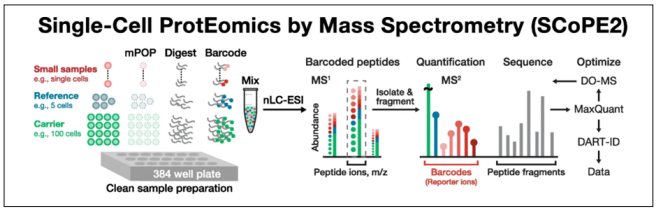

Figure 1. Conceptual diagram and workflow of SCoPE2. Cells are sorted into multiwell plates and lysed by a Minimal ProteOmic sample Preparation (mPOP). The proteins in the cell lysates are digested with trypsin; the resulting peptides labeled with TMT, combined, and analyzed by LC-MS/MS. The LC-MS/MS analysis is optimized by Data-driven Optimization of MS (DO-MS), and peptide identification enhanced by Data-driven Alignment of Retention Times for IDentification (DART-ID). The sample preparation can also be performed by automated nano-ProteOmic sample Preparation (nPOP). nPOP uses piezo acoustic dispensing to isolate individual cells in 300 picoliter volumes and performs all subsequent preparation steps in small droplets on a hydrophobic slide. This figure is adopted with permission from Specht et al., 2021.

Q: What impact do these findings have on the field?

A: Another interesting aspect of this work, in terms of the inferences being made by single-cell RNA sequencing analysis, is that variability across the single-cell proteomes is substantially lower. This is an interesting observation because transcriptional measurements are strongly influenced by counting noise, which is inherent to low copy number RNA molecules being captured with relatively low efficiency. Much of the variability that is observed in those datasets, both technical and stochastic, may not be as biologically meaningful. With regards to protein analysis, we were able to analyze about 20-fold more copies of the gene at the protein level compared to what is possible using the transcriptomic method, resulting in less technical variability while estimating protein abundances. Proteins with longer lifetimes can potentially average out some of the stochastic (transcriptional bursting) noise in RNA levels and provide a more stable indicator for the biological functions of those different cells, which was an important aspect of the analysis. Through joint protein and RNA analysis and, with better data in many more cells, – we were able to go deeper and further than what we could do with SCoPE-MS. In particular, we were able to look at the degree to which messenger RNA (mRNA) or protein abundances for transcription factors can predict their activity. We found that the transcript levels for transcription factors weren't very informative of their activities, while the corresponding protein levels, reflected to a greater degree the changes in the targets of the transcription factor. The data provided a first view into the possibility that we could use the variability between single-cells to infer regulatory interactions between both proteins and transcription factors regulating RNA production on a global scale.

Dr. Nikolai Slavov
Associate Professor,
Northeastern University