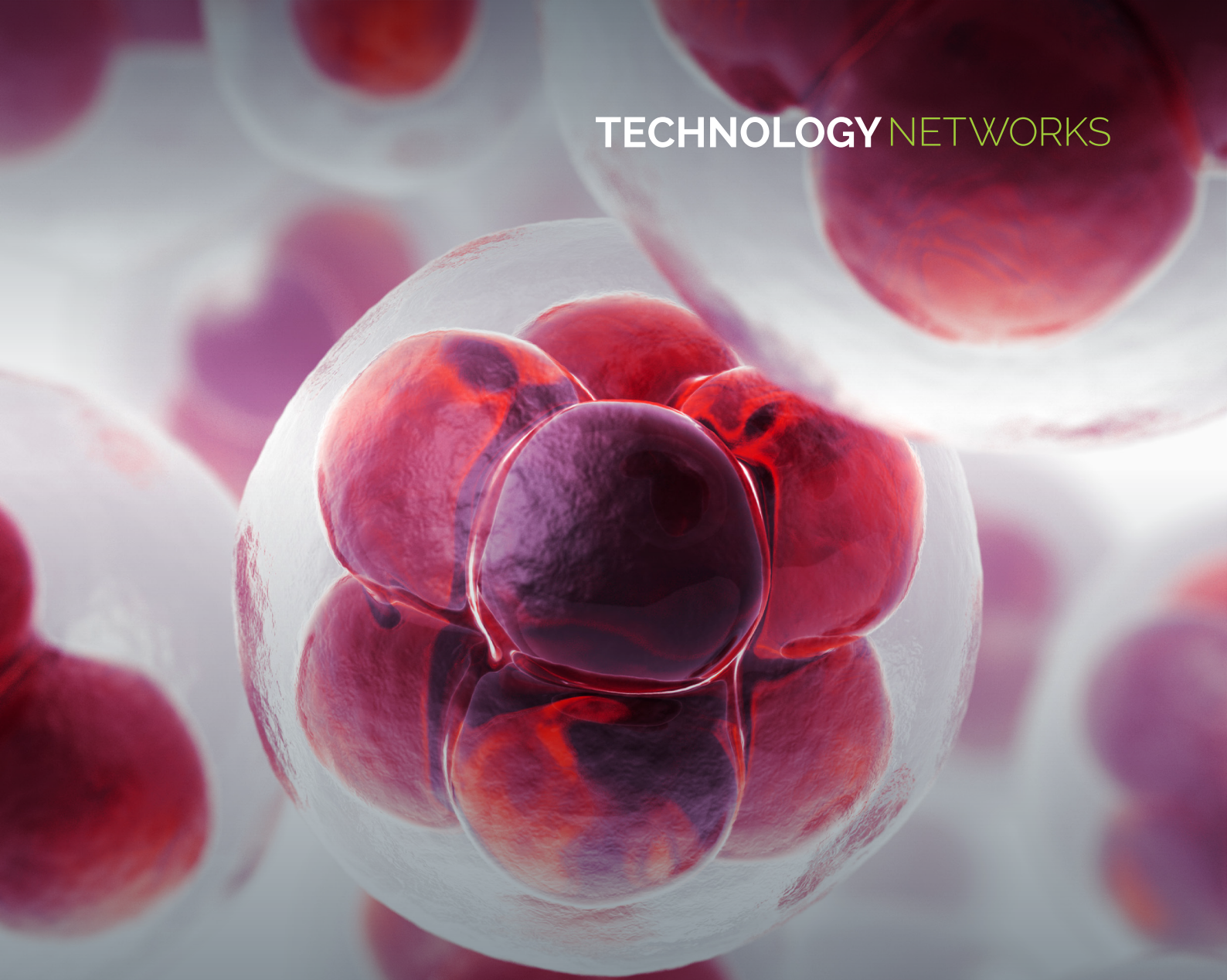


TECHNOLOGY NETWORKS

Making Single-Cell Proteomics Mainstream:

Perspectives from Key Opinion Leaders

An Introduction to Single-Cell Proteomics

Methods for Sample Processing

Striving for Sensitivity and Throughput Gains

SPONSORED BY

ThermoFisher SCIENTIFIC

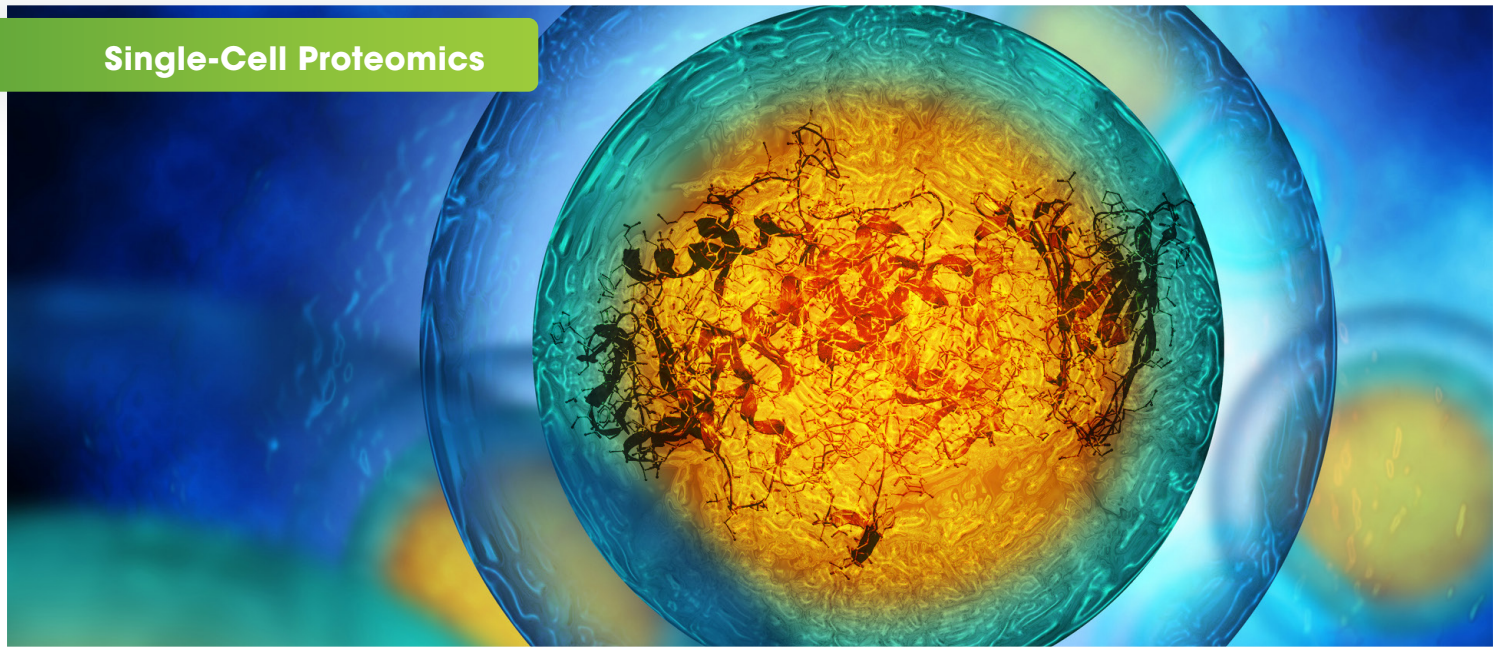
Contents

Foreword

The development of *-omics* technologies marked a significant step away from the study of a few hundred proteins, genes and metabolites into a field where scientists can examine populations of these molecules, over thousands of molecules in an unbiased manner. While the combination of different *-omics* approaches with bioinformatics and computational biology has furthered the study of networks of genes, proteins and metabolites and the mechanisms underlying their interactions, biology is a science of complex, heterogeneous systems.

Proteomics has been one of the most challenging of the *-omics* fields to apply to single-cell applications, however recent technological advances are enabling researchers to investigate biological systems at the single-cell level.

In this eBook, key scientists who are driving the field of single-cell proteomics forward, will share their perspectives on several topics and how to democratize this application, making it accessible to everyone.



Single-Cell Proteomics Becomes a Reality

with Daniel Lopez-Ferrer

As the final stage in the flow of genetic information, proteins execute a variety of functions encoded in DNA, from catalyzing the storage and release of energy, transporting molecules from one side of the cell to the other and ensuring that the genetic code is faithfully copied from cell to cell and from parent to progeny. Approximately 20,000 genes encode for different proteins in the average vertebrate genome, including humans.[1,2] However, alternative splicing and post-translation modifications such as phosphorylation increase the diversity of proteins in an organism. Therefore, it is important to think not only of the 20,000 or so protein-coding genes, but the myriad "proteoforms" in which each protein can exist.[3]

Proteomics aims to study the entire set of proteins and their proteoforms and how their relative abundance influences biological processes in health and disease. Researchers can study subcellular proteomes focusing on the proteins of a subcellular system – such as the mitochondria – whose proteome is highly studied in cardiovascular research. However, there are also researchers that only study parts of the proteome; focusing on either a set of proteins as part of a biological process – like the redoxome – or proteins that are modified post-translationally, such as the phosphoproteome. Given its broad scope of interest and interdisciplinary nature, proteomics has enabled advancements in various fields including cell signaling, gene regulatory networks and drug discovery. Moreover, proteomics has provided a platform to discover proteins that can serve as robust biomarkers to evaluate healthy and diseased cells, detect pathogenetic agents and to monitor response to therapeutic interventions.

This introduction by Daniel Lopez-Ferrer, Senior Manager of Proteomics at Thermo Fisher Scientific, will explore the impact that advances in mass spectrometry (MS) have had on single-cell proteomics, with reference to the research performed by leading scientists in this field.

The move towards single-cell analysis

The first proteomic methods developed to analyze tissue samples allowed quantification of one or several proteins.[4] Bulk methods such as one- and two-dimensional protein electrophoresis, protein sequencing by Edman degradation, MS, protein arrays and affinity columns, offer important insights into the abundance of key proteins or their modifications during biological processes of interest.[5] However, it has become clear that cells, even those with identical genotypes, exhibit significant heterogeneity in gene expression and protein levels.[6] Such heterogeneity is lost in bulk proteomics since the approach necessarily averages the signals from individual cells. Moreover, bulk proteomics is incapable of detecting features of rare cells since their signal is lost in the protein haystack.

Single-cell proteomics has emerged in order to account for cellular heterogeneity and gain access into rare cell types.[7] The key difference, compared to bulk proteomics, is that cells are first dissociated and processed separately, so that proteomes can be interrogated individually. The variety of emerging single-cell proteomic techniques include methods that detect and analyze secreted proteins (e.g., Droplet microfluidics[8,9] and Microengraving)[10,11] or cell-surface proteins (e.g., CITE-seq[12] and REAP-seq).[13] Additionally, a number of single-cell proteomics

methods have been developed to detect secreted and/or cell-surface proteins in addition to proteins within the cytoplasm (e.g., flow cytometry[14] and mass cytometry – CyTOF).[15] According to Daniel, "The granularity of information that single-cell proteomics enables is unique. While bulk analysis is something that has been already quite democratized and standardized, dealing with small samples or single-cells is technically quite difficult, as there are limited off-the-shelf tools. Therefore, current teams need significant expertise from many disciplines including protein chemistry, cell biology, flow cytometry, automation and bioinformatics."

The invention of methods to tag green fluorescent protein (GFP) to proteins of interest was a milestone for single-cell proteomics. The subsequent development of GFP variants with different stabilities or excitation spectra allowed researchers to quantify the levels of several individual proteins with single-cell resolution.[16] An outcome of these studies was a new appreciation for cellular heterogeneity, since it was normal to find a GFP-tagged protein expressed at different levels even in genetically identical bacterial clones. A limitation to fluorescent tagging is the limit on how many fluorescent proteins can co-exist in the same cell without overlapping in their spectra.[17] For this reason, single-cell proteomic methods that rely on GFP and similar tags cannot study more than a dozen different proteins at the same time. Another limitation to GFP is that it requires genetic engineering, which prevents its use with clinical samples. On the other hand, identifying proteins using antibodies can be done without transgenesis.

Antibodies have played a major role in proteomics but – generally speaking – they face two main challenges: specificity and scalability. Important advancements addressing low antibody specificity include single-cell Western blotting, which uses electrophoresis to separate the proteome of an individual cell and thereby allows for better exclusion of false positives (i.e., antibody binding to proteins of the wrong size).[17] Additionally, the development of proximity extension assays (PEAs) provides a way to reduce non-specific background since it requires the binding of two different antibodies to a target of interest in order to result in a signal.[17] As is the case with fluorescent proteins, multiplexing fluorophore-tagged antibodies is limited by overlapping spectra.[18] The development of CyTOF, which tags antibodies with rare elements instead of fluorophores, before detecting them using MS, has expanded the range of multiplexing to around 40 targets. Multiplexing antibodies was further expanded with the development of CITE-seq, which tags antibodies with oligonucleotide barcodes and detects them using next-generation sequencing, in conjunction with transcriptome sequencing. While CITE-seq is limited to membrane-bound proteins it has offered a significant improvement in throughput and a platform for seamless integration of single-cell proteomics and transcriptomics.[12]

Low input single-cell proteomics and challenges for research

Compared to single-cell genomics and transcriptomics, single-cell proteomics faces a number of unique challenges. A primary challenge is technological; the ability to replicate DNA using polymerase chain reaction has accelerated nucleic acid research and applications. However, an equivalent technique does not exist for proteomics, therefore it is not currently possible to replicate proteins in the same way. Without the ability to copy proteins, researchers cannot amplify its signal to a level that facilitates detection. A second challenge is that compared to nucleic acids, proteins are stickier and are more prone to sample loss during their processing.[19] While proteins are more abundant than their encoding mRNAs, their stickiness and our inability to amplify their signal are key challenges to single-cell proteomics. Therefore, a common tradeoff in single-cell proteomics occurs between the number of proteins that can be detected (i.e. sensitivity) and the number of cells that can be analyzed (i.e. throughput).[17] Flow cytometry can analyze thousands of cells, however, it is limited to detecting only a handful of proteins in parallel. Conversely, although MS can detect hundreds of proteins, it remains a low throughput method when applied at the single-cell level. Notably, important developments have started to allow sample multiplexing before MS in order to increase its throughput. Multiplexing via independent barcodes has been instrumental in the meteoric rise of single-cell transcriptomics and its application to proteins promises similar benefits. A different challenge faced by the single-cell proteomics community is the realization that analytical frameworks developed for bulk methods are not necessarily applicable to single-cell analyses. Therefore, while the techniques and experimental procedures that enable single-cell proteomics continue to mature, the computational methods that are necessary to interpret this new data are in need of further development as well.[7]

The impact of MS

Of all the methods used for single-cell proteomics, MS offers the most unbiased approach and has the potential to identify and quantify hundreds of different proteoforms in a single-cell.[20] However, MS methods are currently the lowest in terms of throughput (~10 cells per hour) and accurate protein quantitation remains a challenge. Nevertheless, single-cell MS represents the next frontier for single-cell proteomics and will impact various domains. For example, in personalized medicine, single-cell MS is uniquely suited to aid in adoptive cell therapies. By allowing comprehensive identification and quantification, single-cell MS can facilitate this rapidly emerging anti-cancer approach, whereby the patient's own immune cells are engineered to express T-cell receptors or chimeric antigen receptors that target the cancer in question.[21]

Daniel and his collaborators are developing technologies to better understand cell populations in the context of cancer. Their hypothesis is that "since tumors are formed from different cell populations and the response to therapies is influenced by the number and types of cells in the tumor" they can use single-cell proteomics "to see if they can better stratify patients to better assign the best available therapy for them."

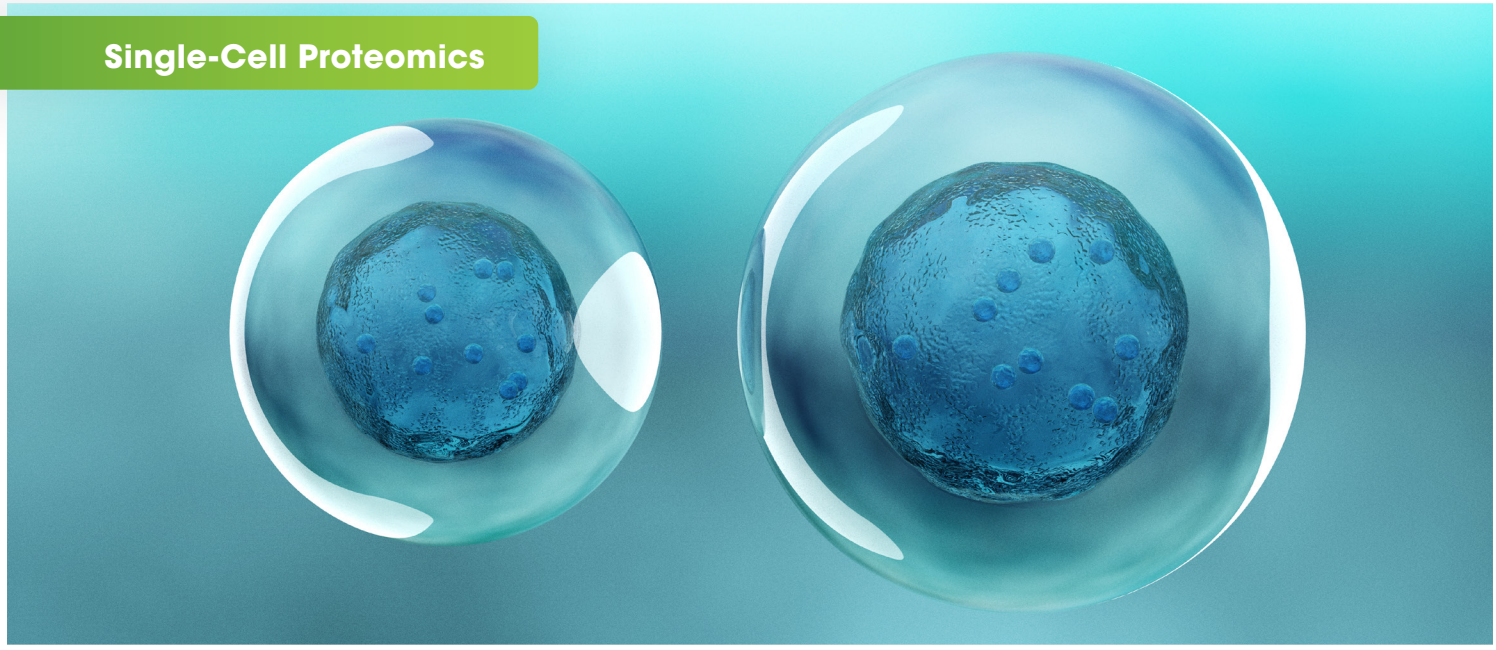
Another area of impact is spatial proteomics, where the localization of proteins and their dynamics are determined at the subcellular level.[22] By synergizing with cutting-edge microscopy and machine learning, MS is beginning to reveal protein dynamics at the single-organelle level, which makes it a powerful tool in unraveling disease mechanisms. Indeed, the fact that proteins and their modifications are major executors of biological function – compared to mRNA transcripts, which are copies of genetic information – is why single-cell MS will become a major catalyst in identifying molecular mechanisms that underlie health and disease.[23]

Daniel adds that, "Single-cell MS-based proteomics is still a very niche application, but I envision a very bright future for it. There is no other technique that can provide accurate quantification of the protein expression at the single-cell level and its application in other fields will revolutionize biology."

Daniel Lopez-Ferrer
Director of Proteomics,
Thermo Fisher Scientific

References

1. Blomme T, Vandepoele K, De Bodt S, Simillion C, Maere S, Van de Peer Y. The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol*. 2006;7(5):R43. doi:10.1186/gb-2006-7-5-r43
2. Consortium IHGS. Finishing the euchromatic sequence of the human genome. *Nature*. Oct 2004;431(7011):931-45. doi:10.1038/nature03001
3. Smith LM, Kelleher NL, Proteomics CfTD. Proteoform: a single term describing protein complexity. *Nat Methods*. Mar 2013;10(3):186-7. doi:10.1038/nmeth.2369
4. Dupree EJ, Jayathirtha M, Yorkey H, Mihasan M, Petre BA, Darie CC. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of this Field. *Proteomes*. Jul 2020;8(3)doi:10.3390/proteomes8030014
5. Graves PR, Haystead TA. Molecular biologist's guide to proteomics. *Microbiol Mol Biol Rev*. Mar 2002;66(1):39-63; table of contents. doi:10.1128/mmbr.66.1.39-63.2002
6. Altschuler SJ, Wu LF. Cellular heterogeneity: do differences make a difference? *Cell*. May 2010;141(4):559-63. doi:10.1016/j.cell.2010.04.033
7. Labib M, Kelley SO. Single-cell analysis targeting the proteome. *Nature Reviews Chemistry*. 2020;4(3):143-158. doi:10.1038/s41570-020-0162-7
8. Dhar M, Lam JN, Walser T, Dubinett SM, Rettig MB, Di Carlo D. Functional profiling of circulating tumor cells with an integrated vortex capture and single-cell protease activity assay. *Proc Natl Acad Sci U S A*. 10 2018;115(40):9986-9991. doi:10.1073/pnas.1803884115
9. Brouzes E, Medkova M, Savenelli N, et al. Droplet microfluidic technology for single-cell high-throughput screening. *Proc Natl Acad Sci U S A*. Aug 2009;106(34):14195-200. doi:10.1073/pnas.0903542106
10. Love JC, Ronan JL, Grotenbreg GM, van der Veen AG, Ploegh HL. A microengraving method for rapid selection of single-cells producing antigen-specific antibodies. *Nat Biotechnol*. Jun 2006;24(6):703-7. doi:10.1038/nbt1210
11. Ogunniyi AO, Thomas BA, Politano TJ, et al. Profiling human antibody responses by integrated single-cell analysis. *Vaccine*. May 2014;32(24):2866-73. doi:10.1016/j.vaccine.2014.02.020
12. Stoeckius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and transcriptome measurement in single-cells. *Nat Methods*. Sep 2017;14(9):865-868. doi:10.1038/nmeth.4380
13. Peterson VM, Zhang KX, Kumar N, et al. Multiplexed quantification of proteins and transcripts in single-cells. *Nat Biotechnol*. Oct 2017;35(10):936-939. doi:10.1038/nbt.3973
14. Krutzik PO, Nolan GP. Intracellular phospho-protein staining techniques for flow cytometry: monitoring single-cell signaling events. *Cytometry A*. Oct 2003;55(2):61-70. doi:10.1002/cyto.a.10072
15. Bandura DR, Baranov VI, Ornatsky OI, et al. Mass cytometry: technique for real time single-cell multitarget immunoassay based on inductively coupled plasma time-of-flight MS. *Anal Chem*. Aug 2009;81(16):6813-22. doi:10.1021/ac901049w
16. Cranfill PJ, Sell BR, Baird MA, et al. Quantitative assessment of fluorescent proteins. *Nat Methods*. 07 2016;13(7):557-62. doi:10.1038/nmeth.3891
17. Levy E, Slavov N. Single-cell protein analysis for systems biology. *Essays Biochem*. 10 2018;62(4):595-605. doi:10.1042/EBC20180014
18. Bendall SC, Nolan GP, Roederer M, Chattopadhyay PK. A deep profiler's guide to cytometry. *Trends Immunol*. Jul 2012;33(7):323-32. doi:10.1016/j.it.2012.02.010
19. Marx V. A dream of single-cell proteomics. *Nat Methods*. 09 2019;16(9):809-812. doi:10.1038/s41592-019-0540-6
20. Altelaar AF, Munoz J, Heck AJ. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet*. Jan 2013;14(1):35-48. doi:10.1038/nrg3356
21. Lombard-Banek C, Schiel JE. MS Advances and Perspectives for the Characterization of Emerging Adoptive Cell Therapies. *Molecules*. Mar 2020;25(6)doi:10.3390/molecules25061396
22. Lundberg E, Borner GHH. Spatial proteomics: a powerful discovery tool for cell biology. *Nat Rev Mol Cell Biol*. 05 2019;20(5):285-302. doi:10.1038/s41580-018-0094-y
23. Slavov N. Single-cell protein analysis by MS. *Curr Opin Chem Biol*. Feb 2021;60:1-9. doi:10.1016/j.cbpa.2020.04.018



Striving for Sensitivity and Throughput Gains

A Interview with Ying Zhu

Advances in single-cell proteomics are providing critical biological insights into cellular heterogeneity, spatial organization and gene expression regulation that are not possible with bulk-scale analysis. However, as a field that is still in its infancy, researchers are confronted with technical challenges including issues with reproducibility, proteome coverage, sensitivity and throughput. The development of more robust, sensitive, and high-throughput technologies are enabling scientists to address these challenges and enhance our understanding of disease pathogenesis.

Dr. Ying Zhu is a senior bioanalytical scientist at the Environmental Molecular Sciences Laboratory of Pacific Northwest National Laboratory (PNNL), and has over ten years experience conducting ultrasensitive bioanalysis using microfluidic techniques and mass spectrometry (MS). Working closely with the Integrative Omics group from the Biological Sciences Directorate, Dr. Zhu is making significant contributions to the advancement of single-cell proteomics. In this interview, he provides an overview of the techniques used in his research, and highlights some of the challenges that must be overcome to further advance the field.

Q: Can you tell us about your journey into single-cell proteomics with reference to your research focus?

A: I was trained as an analytical chemist and microfluidic engineer, however I have a very diverse research background, including microfluidics, robotics, microscale separation, MS, proteomics, genomics, microarrays, and optical detection. Most of my research focuses on developing highly sensitive and high-throughput analytical methods and instrumentation. In my early career as an independent investigator in Zhejiang University of China, in collaboration with Professor Qun Fang, I co-led the development of an automated nanolitre liquid manipulation system called sequential operation droplet array (SODA), which combined robotics and microfabrication to achieve robust and flexible nanolitre liquid manipulations. Because my research style is technology-driven, I always look for new applications where the SODA technology can be applied. The system has shown to be useful in many different fields such as protein crystallization, drug screening, protein-ligand screening, single-cell and single-molecule gene analysis, and sample preparation for MS. During single-cell gene analysis, I demonstrated that miniaturization of reaction volume can significantly improve the analysis sensitivity. This research inspired me to further develop the technology to single-cell proteomics.

After I joined PNNL in 2016, in collaboration with Dr. Ryan Kelly and Dr. Richard Smith, I co-led the development of an ultrasensitive proteomics platform by coupling microfabricated nanowell chips and robotics with liquid chromatography-MS (LC-MS). We called it nanoPOTS, nanodroplet processing in one-pot for trace samples. Compared with SODA, nanoPOTS completely removed the oil phase and employed environment control (humidity and temperature) to minimize droplet evaporation. After demonstrating that nanoPOTS could significantly increase the sensitivity of proteomics, I devoted myself to improving the throughput and robustness, as well as coupling it with different cell

isolation technologies. I developed the nanoPOTS autosampler to enable the direct couple of nanowell chips with automated LC-MS systems. I also integrated isobaric labeling approaches (e.g. SCOPE-MS) with nanoPOTS to multiplex single-cell proteomics and developed a new-generation nested nanoPOTS chip. To broaden the application of nanoPOTS technology, I developed many interfacing approaches to couple it to fluorescence-activated cell sorting (FACS), image-based single-cell isolation (IBSC), and laser capture microdissection (LCM).

Currently, my research focuses on three directions:

1. The development of new microfluidics-based sample preparation and sensitive LC-MS methods for single-cell proteomics. I have strived to improve the sensitivity, throughput, and robustness to achieve large-scale and in-depth single-cell proteomics.

2. The development of new cell isolation methods for spatial single-cell proteomics. Here I am trying to leverage different engineering approaches to directly dissect small tissue voxels and transfer them to nanoPOTS for spatial mapping of the proteome in three dimensions.

3. I am passionate about collaborating with biologists and clinical scientists, and helping them to address their scientific questions using single-cell and spatial proteomics technology. I always find these collaborations inspire me to develop better proteomics technologies.

Q: Which techniques do you use for sample handling and processing?

A: Most of my research is related to the analysis of small samples or single-cells with MS. For sample handling, I usually use three types of cell isolation technologies: FACS, IBSC, and LCM. FACS is a universal single-cell isolation technology. It can quickly map the possible cell populations and identify cell types using different fluorescent channels. We have demonstrated the coupling of FACS with nanoPOTS for high-throughput single-cell proteomics. The FACS has sufficient sorting precision to place single-cells on our typical 1-mm nanowells. IBSC is a relatively new cell isolation technology. Compared with FACS, IBSC takes images of every cell and then dispenses cell-containing droplets into the container for processing. Instead of using an electrical field to move the droplets, IBSC moves the capillary to the top of the container and dispenses the cell by acoustics. Thus, the sorting precision is much higher. We found that it could sort single-cells into microfabricated wells with diameters as small as 0.3 mm and with high success rates (>99%). Since both FACS and IBSC require dissociated single-cells, you can lose the spatial position of these single-cells, which will mask critical biological information about cell organization, cell-cell interactions, and the tissue microenvironment. LCM perfectly addresses this problem, because it can isolate single-cells and tissue microstructures directly from thin tissue sections. It can also be coupled with immune staining and high-resolution optical imaging to get more phenotypic information from the isolated cells.

When processing the isolated cells or tissue voxels, I always use the nanoPOTS technology. NanoPOTS uses microfabricated chips combined with a nano-pipetting system for on-chip cell lysis and protein digestion. Each sample's total processing volume is from 30-200 nL. Compared with standard centrifuge tube-based methods, the volume is reduced by a factor of 500-1,000, and the surface contact area is reduced to <0.4% of conventional methods. Due to the volume of miniaturization, protein and protease concentrations can be increased by > 500-fold, resulting in efficient tryptic digestion kinetics. So far, nanoPOTS technology has enabled us to quantify >1,500 proteins from single mammalian cells and map >2,000 proteins on thin tissue sections at 100 µm resolution.

Q: What techniques or approaches do you use to achieve higher sensitivity and how have they advanced over the years?

A: To achieve higher sensitivity during single-cell proteomics analysis, all workflows from cell isolation, sample preparation, sample injection, liquid chromatography (LC), MS, and informatics analysis need to be optimal.

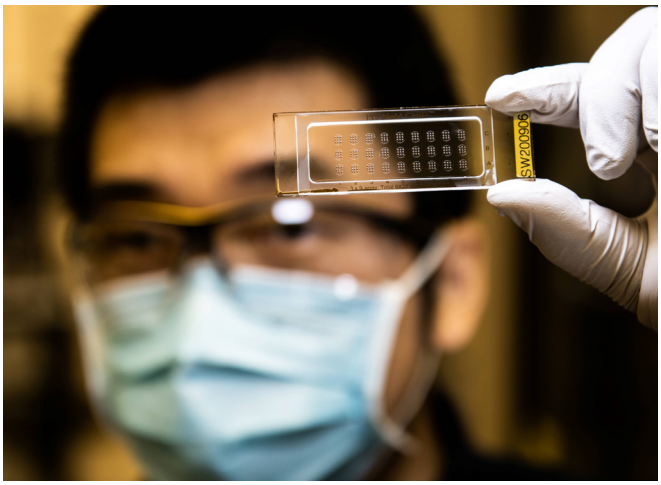


Figure 1. The nested nanoPOTS chip. Credit: Andrea Starr, PNNL photographer.

We always use nanoPOTS technology to prepare single-cells or small samples. The most recent advance of nanoPOTS technology is the nested nanoPOTS (N2) for isobaric labeling workflow, which we reported in a preprint. The N2 chip is distinct from previous

nanoPOTS chips due to the fact we cluster an array of nanowells in high density and use each cluster for one multiplexed tandem mass tag (TMT) experiment. We designed 9 (3×3) nanowells in each cluster and 27 (3×9) clusters, resulting in a total of 243 nanowells on one N2 chip. We reduced the nanowell diameters from 1.2 mm to 0.5 mm, corresponding to an 82% decrease in contact areas and an 85% decrease in total processing volumes. The nested chip design also significantly simplified the TMT-based isobaric labeling workflow by eliminating the tedious sample pooling steps. We have shown that the N2 chip significantly improved the sensitivity, throughput, and robustness of single-cell proteomics.

After the samples are ready, they are directly injected into an LC-MS system without transfer or dilution, as any additional steps will cause sample loss. The direct sample injection from the nanoPOTS chip to an LC-MS system is performed with a home-built autosampler. Such fully automated workflows not only maintain high sample recovery, but also improve system robustness and overall analysis throughput. In addition, we usually add a non-ionic surfactant (N-Dodecyl β-D-maltoside, DDM) to our sample as a carrier to avoid sample loss during sample injection.

In most single-cell proteomics studies, we employed the Orbitrap Lumos Tribid MS or the Orbitrap Eclipse Tribid MS for data acquisition. Since we published the label-free single-cell proteomics study, we have dramatically improved MS sensitivity by introducing Field Asymmetric Ion Mobility Spectrometry (FAIMS). We observed that FAIMS could remove most single charged species and fractionate ions from many populations based on the compensation voltage (CV). Because of the low background contamination, FAIMS allows elongated ion accumulation (200 to 500 ms) at the MS^1 level to enable ultrasensitive peptide detection. We developed a new MS acquisition method, TIFF (Transferring Identification based on the FAIMS Filtering). Using the TIFF method, we can quantitatively identify >1,000 proteins from single HeLa cells and apply it to study the immune activation of macrophage cells in a time-resolved fashion.

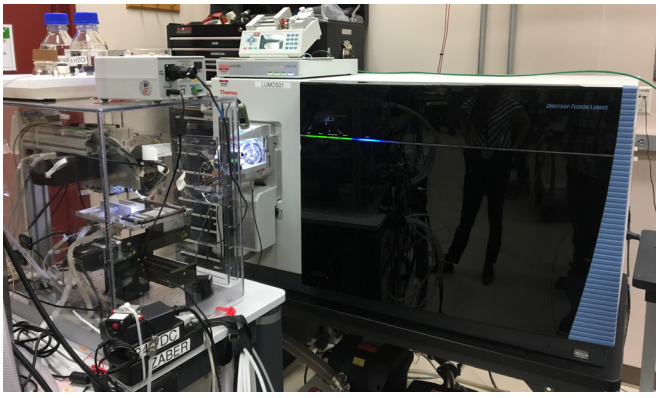


Figure 2. The NanoPOTS auosampler in front of the Orbitrap Lumos Tribid mass spectrometer. Credit: Dr. Ying Zhu.

Q: Are there any limitations to these techniques?

A: We realize there are still many limitations and challenges for these technologies. Single-cell proteomics is a very young research field, so it will take some time for it to mature. Researchers in this field are working very hard to make improvements, however some of the current limitations include:

1. With regards to single-cell preparation, nanoPOTS is highly efficient, however it requires specific microfabricated devices and a robotic system. Such requirements limit its broad dissemination to the research community. There is a strong demand to distribute the technology through a commercial company.

2. Although both LC and MS have greatly improved in the past several years, they still show significant limitations in sensitivity and throughput. Single-cell proteomics can identify and quantify an average of ~1,000 to 2,000 highly abundant proteins. It may seem amazing, but such proteome coverage is not sufficient to answer many important biological questions involving low abundant proteins/pathways, such as gene regulation and cell signaling. Moreover, the major limitation of LC-MS is low throughput. Current single-cell transcriptomics can routinely characterize >10,000 single-cells. However, single-cell proteomics is limited in its ability to characterize hundreds to thousands of single-cells. The cost is already much higher than single-cell transcriptomics, therefore, to make single-cell proteomics more powerful, the throughput should be further improved, and the cost should be reduced.

Q: What challenges remain for this area of research?

A: There are many challenges remaining. First, the proteome coverage and throughput of single-cell proteomics still require significant improvements. We expect to see new sample preparation devices, isobaric labeling reagents, automated robotics systems, LC columns, and MS systems to advance the area in the next five years.

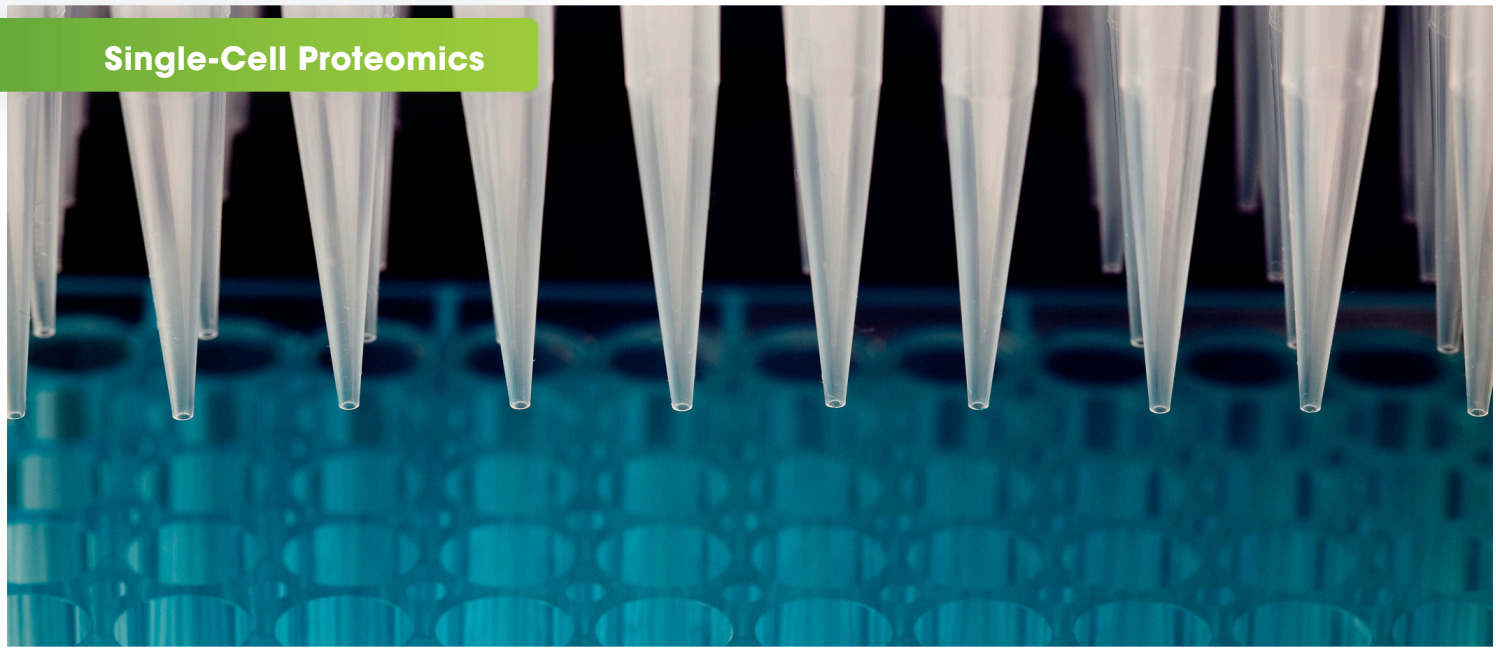
Second, spatial single-cell proteomics is still in its infancy. Although we showed the feasibility of in-depth spatial proteomics at 50 µm resolution, it is not single-cell level yet. The isolation and analysis of single-cells from tissue sections is technically challenging. We do not have a robust method to label cell membranes and segment single-cells for precise dissection, and the dissection and collection of single-cells suffers from low throughput and limited success rates. Informatics

tools are required to understand tissue organization and cell-to-cell interactions based on the spatial proteomics measurement.

Third, most single-cell proteomics studies are focused on global proteomes. However, protein functions are regulated by post-translational modifications (PTMs) such as phosphorylation, glycosylation, ubiquitination, nitrosylation, and acetylation. Such modifications are usually low in abundance and require specific enrichment, which is not feasible to study at single-cell level now. Most importantly, all the proteins are present as proteoforms in cells; single genes can produce many functionally distinct proteoforms due to mutations, polymorphisms, and PTMs. Thus, the final goal to measure proteins is to directly measure proteoforms using top-down proteomics approaches. Many areas should be advanced to enable single-cell top-down proteomics, including sample preparation, chemical separation, MS detection, and data analysis.

Finally, we know that the development of single-cell and spatial transcriptomics has transformed biological and biomedical research in many ways, however, single-cell and spatial proteomics is still in the developmental stage. Many questions remain: what is the unique value of single-cell and spatial proteomics? What are the killer applications? How can you disseminate these new technologies to most proteomics laboratories for routine analysis? Will single-cell and spatial proteomics ultimately replace transcriptomics in the future? We are working to establish answers to these research questions.

Dr. Ying Zhu
Senior Bioanalytical Scientist, Pacific Northwest National Laboratory



Novel Methods for Limited Sample Analysis

An Interview with Karl Mechtler

Single-cell proteomics seeks to complement transcriptome and genome data by generating comparative protein expression profiles of individual cells. The field has been aided by technological advances in mass spectrometry (MS), as well as the development of approaches such as Tandem Mass Tags (TMT) which have improved sensitivity and sample throughput. However, researchers still face the challenge of analyzing large numbers of ultra-low input samples.

Karl Mechtler is the Head of Protein Chemistry Facility at the Research Institute of Molecular Pathology, Institute of Molecular Biotechnology/Gregor Mendel Institute of the Austrian Academy of Sciences and Head of MS at the Vienna BioCenter Core Facilities. His primary research focus involves the use of MS to analyze proteins and peptides, specifically post-translational modifications. In this interview, Karl discusses the importance of establishing robust workflows and the impact that industry partnerships have had on his research.

Q: What is the focus of your research and what are some of the workflow challenges you face?

A: Our general focus in the Protein Chemistry Facility here at the Vienna BioCenter is technical development for our customers. We are convinced, however, that each aspect of the workflow influences the outcome of single-cell proteomics experiments. During recent years we have therefore focused on the entire workflow starting from sample preparation through to data analysis, all of which we have recently reviewed in great detail. In general, we are running one of the biggest core facilities in Europe with a focus on two techniques: protein-protein cross-linking and single-cell proteomics.

For single-cell proteomics we believe it is important to introduce new sample preparation workflows, as this is currently one of the most difficult aspects. Since there are currently no commercial solutions available, it is challenging for some research groups without experience in handling small volumes to reproduce such experiments. For this you need an industry partner. We partnered with Cellenion – a company that specializes in liquid handling robots which can pipette down to 400 pL. Together, we have developed a semi-automated, chip-based workflow, which can be directly connected to a standard Thermo Scientific autosampler. The proteoCHIP has been quite successful and will be the first commercial product for single-cell proteomics that is available for everyone as of Fall 2021.

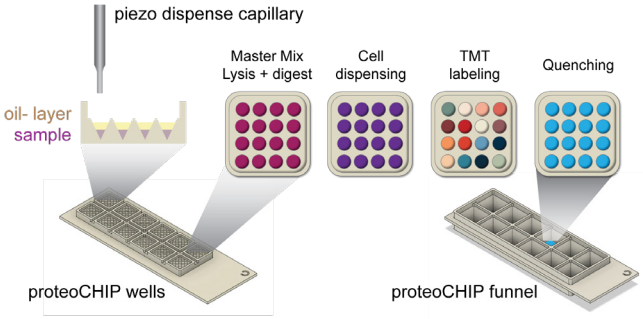


Figure 1. One proteoCHIP and its funnel-shaped lid allow to completely process up to twelve sets of sixteen single cells inside the cellenONE® without manual sample handling in a temperature and humidity controlled environment. The piezo dispensing technology enabled the miniaturization of digestion volumes down to 40 nL, which are covered with a layer of oil to overcome evaporation and remain constant enzyme:substrate concentrations (image courtesy of Karl Mechtler).

Q: What samples are you using in your research and how have you optimized your workflow to overcome any limitations that they impose?

A: I think it is most important to establish standards, so everyone can reproduce the data. We have optimized all parts of our workflow using HeLa and HEK293T cells or diluted HeLa and K562 bulk digests to ensure stability and reproducibility of the evaluations. Our aim is to benchmark the optimizations in the best way possible and provide a 'manual' for everyone to be able to reproduce the work in their own labs. This way we ensure that every lab can test our workflows on their setup and see how our approaches work in their environment. Our campus is very diverse with several interesting model systems and organoid technologies, some of which we started to profile using our established protocol. Nevertheless, we are still in the development phase and we are constantly tweaking our workflow to achieve the best possible results.

Q: What are some of the sensitivity limitations of existing proteomic workflows?

A: My team and I believe that sensitivity limitations occur due to suboptimal sample preparation, data acquisition and analysis. We started to optimize chromatographic separation very early on in close collaboration with Pharmafluidics using their µPAC columns. From previous work we had already observed unprecedented retention time stability and sample recovery, but with the second generation µPAC column we were able to push that by a factor of 10 compared to Stadlmann et al., 2019.

The combination of these state-of-the-art micro pillar columns with the FAIMS Pro Interface and Orbitrap Exploris 480 mass spectrometer, allowed us to identify ~1500 proteins from only 250 pg HeLa digest – which is close to the theoretical protein amount of a single-cell. In this study, we found that polyethylene glycol (PEG) spiked into our trace samples overcame adsorptive losses of the sample to the vial surface. This was also shown to be the case when samples are left in an autosampler for extended periods. Since then, we have used PEG not only for our dilution experiments but also during our single-cell sample preparation workflow. By minimizing the adsorption of peptide material to the sample vial surfaces we enable reliable processing of the protein digest.

Another important limitation of any field involving single-cell analysis is the throughput of the measurements. To characterize biological samples, one has to analyze at least hundreds, if not thousands of samples, which by far exceeds the number of multiplexing reagents currently available. My lab – and many others – have found that when performing multi-batch analysis of TMT-labelled or label-free samples in a data-dependent manner, the replicate overlap is dramatically reduced and requires a heavy data input. This is especially noticeable in the analysis of single-cell samples and has been most successfully addressed with data-independent acquisition (DIA) strategies, which – despite improvements in the comprehensiveness of data-independent methods – were previously limited by the analysis of label-free samples. We combined TMT-multiplexing with DIA to simultaneously increase the throughput and multi-batch overlap in the analysis of trace samples. This allowed us to generate highly reproducible and quantitative batch and cell type independent proteome signatures to classify hundreds of ultra-low input samples without any data imputation.

As mentioned before, our collaboration with Cellenion has allowed us to reduce our sample volume more than 5-fold and reproducibly prepare label-free and TMT multiplexed single-cells. As a result, we have largely overcome surface losses and avoid bias via error-prone manual sample handling. This, in conjunction with FAIMS Pro Interface and the Orbitrap Exploris 480 MS resulted in the quantification of around 900 protein groups per multiplexed single-cell experiment without any carrier at remarkable signal to noise. We are therefore confident that this optimization of sample preparation together with our highly sensitive instrument setup enables scientists to push the boundaries.

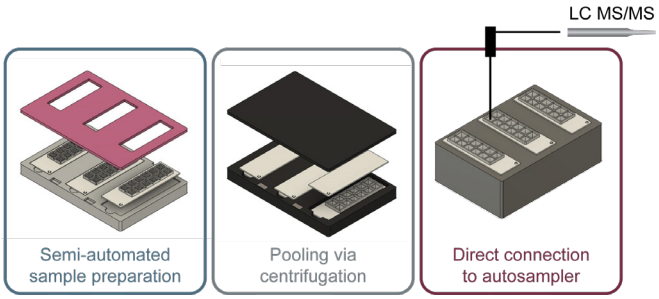


Figure 2. The semi-automated sample preparation inside the cellenONE® is complemented by the sample pooling of individual TMT sets via centrifugation into the proteoCHIP lid. Without transfer to another vessel or error-prone pipetting steps the proteoCHIP lid is directly interfaced with a standard autosampler for loss-less LC-MS/MS acquisition (image courtesy of Karl Mechtler).

Q: Liquid chromatography-MS (LC-MS) is increasingly used in single-cell proteomics. Which part of the workflow have you worked to optimize and what still needs to be addressed?

A: By now we have worked on almost every part of the workflow and at the current stage of single-cell proteomics it is hard to predict which parts of the workflow still need optimization. We are extremely glad to see that the field is growing by the minute and more scientists are interested in

this area of research. Many groups are working to increase the sensitivity of chromatography and MS approaches. This opens up quite a small community to interdisciplinary scientists, which will help us learn from diverse experiences and further overcome limitations in each step of the workflow. For example, bioinformaticians who recently got involved with single-cell proteomics have already proposed several new tools that can significantly improve our scarce data structure. In my opinion we have only seen the tip of the iceberg and I am looking forward to new developments which will address each aspect of the workflow.

Additionally, positive ions from the surrounding air or singly charged contaminants in the samples could not be removed prior to analysis, however doing so dramatically amplifies the signal by overcoming ion suppression. The improvements of the second generation of FAIMS, FAIMS Pro Interface, was really a big surprise for our lab as we did not expect a factor of 10 gain in sensitivity. With our current setup of the second generation µPAC, specialized MS methods and FAIMS Pro Interface, we have significantly enhanced sensitivity. Nevertheless, we look forward to continuing to work with industry partners to further improve our approaches for single-cell proteomics.

Karl Mechtler

Head of Protein Chemistry Facility, Research Institute of Molecular Pathology

Methods for Sample Processing in Single-Cell Proteomics

with Ryan Kelly and Erwin Schoof

When sorting cells, regardless of whether you are examining diseased or healthy tissue, it is important that scientists can isolate cells that represent the phenotype of interest. Single-cell integrity (including viability and intactness) must be maintained during cellular isolation and sample processing, so that a cell does not lose its protein content before intentional lysis. Otherwise, the technical process of cell isolation and transfer becomes a confounding factor with a negative effect on the accuracy of the analysis. This article will discuss the best practices and quality metrics used to maintain cellular integrity during sample processing, with reference to insights from both Dr. Erwin Schoof (Associate Professor, Technical University of Denmark) and Dr. Ryan Kelly (Associate Professor, Brigham Young University and Pacific Northwest National Laboratory).

Traditional ways to sort cells (FACS)

A common method for transporting individual cells from suspensions to microplates is fluorescence-activated cell sorting (FACS). FACS is a mature technology that allows fast sorting of single cells across 384-well plates. An important advantage of FACS is the ability to monitor integrity through staining with specific markers or other specific parameters (such as granularity). Granularity can be measured using the side scatter (SSC) parameter, which is used to measure how much of the incident laser is scattered by particulates inside the cell – it is an indication of internal complexity. Since different cell types have different granularities, monitoring this parameter allows researchers to evaluate cell integrity before further processing.

Erwin Schoof uses FACS in conjunction with proteomics to discover new markers for quiescent cancer stem cells. With a focus on the hemopoietic lineage, Erwin leverages known surface markers of the specific blood subtypes within the blood hierarchy (e.g., myeloid, lymphoid and erythrocyte cells) to sort each type. Then, his group record (i.e. index sort) these markers for subsequent overlaying on their single cell data to uncover further heterogeneity. He says, "We're trying to use single-cell proteomics to find new markers of these cells and find heterogeneity within what was previously deemed a homogenous population." When processing cells using FACS, Erwin uses 'Single-Cell Purity' mode to make sure each cell is intact and is placed in the middle of the droplet, "Then, because it hits lysis buffer on impact it opens up, so all the proteins get denatured and we are ready to start analyzing those proteins at that steady state."

Traditional methods for sample preparation (digestion/recovery)

Cell suspension is a preliminary requirement for FACS. However, although the preparation of cell suspensions from liquid tissues such as blood and bone marrow is relatively easy, most tissues require disruption to degrade the extracellular matrix and isolate single cells. A common approach to preparing single-cell suspensions is to dissect out the tissue of interest and then subject the tissue to mincing, enzymatic digestion and mechanical dissociation. However, this destructive process must take place without compromising cellular integrity and while minimizing cell death and aggregation. Best practices for this approach include empirically determining the best balance between

digestion temperature (the lower the better) and digestion time (the shorter the better) and single-cell yield (the more the better).[1]

The more thorough the digestion, the higher the yields obtained. However, this can come at the expense of cellular integrity, especially since surface proteins are subject to enzymatic cleavage, leading to false negatives in downstream analysis. In order to protect proteins of interest from the collateral damage of enzymatic digestion, care must be taken to ensure that the proper enzyme cocktail is used. For example, dispase is a commonly used protease isolated from bacteria and digests attachments between cells and the extracellular matrix. However, dispase can also cleave antigens critical to T cell analysis and therefore using this enzyme could lead to loss of valuable epitopes when studying T cells.[1]

Quality control of single-cell suspensions using enzymatic and mechanical dissociation includes three critical parameters: cell viability, absence of cell debris and absence of aggregates, all of which can be assessed using light microscopy or a FACS instrument. For example, SYTOX Red Dead Cell Stain (Invitrogen) binds to double-stranded DNA but does not cross the cell membranes of intact cells. Dead cells, however, are permeable to SYTOX and exhibit increased fluorescence. Viable (non-fluorescent cells) can then be selected for downstream processing. Other common viability markers include propidium iodide (PI) and trypan blue, both of which label dead cells. Additionally, using nuclear stains such as the cell-permeant dye DRAQ5 can help discriminate between intact cells and debris, since an intact cell will retain stained nuclei. Oftentimes, adding DNase I to digestion cocktails can improve cell suspension quality by degrading DNA released by lysed and dying cells. Extracellular DNA enables cell aggregation and single-cell suspensions should include minimal amounts of released genetic material. Another method used to prevent cell aggregation is to include ethylenediaminetetraacetic acid (EDTA) — a chelating agent that sequesters divalent metal ions, which in turn compromises cell adhesion, facilitates cell dissociation and prevents aggregation.

A different approach to preparing single-cells is laser capture microdissection (LCM).[2] By using ultraviolet or infrared lasers, LCM provides a method to isolate individual cells from heterogeneous tissue sections under direct microscopic visualization. The captured tissue regions are then processed as single-cells for proteomic analysis. Despite its low throughput, this approach is important as it preserves the spatial relations between adjacent cells and allows the construction of a protein expression map that is related to the original tissue. Ryan Kelly develops technological solutions for single-cell proteomics that enable the analysis of smaller samples and single cells, as well as facilitating spatially resolved measurements. According to him, "Laser microdissection allows you to address questions related to the micro-environment within certain tissues, such as solid tumors, enabling you to investigate for example, hypoxic versus non-hypoxic regions. It's a phenomenal technology."

Custom solutions for high-throughput sample processing

Commercial integrated solutions that combine the high-throughput nature of the 384-well plate format with nanoplate volumes and dimensions are currently not available. Therefore, researchers have set out to develop custom solutions for high-throughput sample processing that maintain cellular integrity and minimize sample loss.

For example, Ryan Kelly's group has developed a robotic sample processing and analysis platform that uses nanoliter pipetting and nanowell chips.[3] Dubbed "nanoPOTS" (Nanodroplet Processing in One pot for Trace Samples), this platform increases the efficiency and recovery of proteomics processing by downscaling preparation volumes to the nanoliter scale. "These are much smaller and they have less surface area than a conventional 384-well plate and we do this all in a one-pot workflow to minimize the losses associated with sample clean-up." AutoPOTS is a follow-up solution developed by Ryan Kelly that maintains the one-pot sample preparation concept but with volumes in the microliter range, thereby facilitating automation.[4]

The SCoPE2 system (Single-Cell ProtEomics by MS) developed by Nikolai Slavov and his group, has a throughput of 150 single-cells per day and a sensitivity of nearly 1000 proteins per cell (when applied to monocytes and macrophages)[5]. SCoPE2 departs significantly from bulk sample preparations by radically changing several steps including cell isolation, lysis, and sample preparation. For instance, instead of lysing cells by focused acoustic sonication, SCoPE2 uses a freeze-heat cycle that extracts proteins in pure water, thereby avoiding the need for sample cleanup before MS analysis. The innovations in sample handling allow SCoPE2 to reduce lysis volumes to the 1-10 µL range and reduce costs of consumables and equipment over 100-fold. SCoPE2 also leverages Tandem Mass Tag (TMT) labeling so that proteins from a given cell are uniquely labeled. Tagging the proteomes of individual cells allows pooling up to 14 cells in the same run and demultiplexing the data during analysis.

Commercially available solutions

A few promising commercial solutions for single-cell proteomics are on the horizon. Isoplexis offers different solutions for single-cell proteomics; the Single-Cell Secretome solution uses IsoCode Chip, which is fabricated with microchambers that capture single cells. Then,

using a proprietary antibody barcoding system, over 32 cytokines can be analyzed in a single run. Additionally, the Single-Cell Intracellular Proteome solution relies on an IsoCode Chip that lyses single-cells to detect a panel of over 15 intracellular proteins. Cellenion, a Cellink company, is adapting its controlled cell dispensing technology for integration with downstream MS analysis. Its platform, cellenONE aspirates a cell suspension into a glass capillary and generates drops on demand. The drops are monitored using automated imaging to ensure that drops contain single cells before being dispensed into selected targets. The CellenOne can now be combined with the ProteoCHIP, a tailored sample preparation platform for ultra-low volume single cell proteomics analysis with enhanced sensitivity. Cytena, also a Cellink company, offers an automated platform for plate-based single-cell analysis workflows that can be tailored to proteomics. In addition to excluding dead cells using fluorescent markers, their single-cell dispensing technology is recoded by a series of images and provides documentation of cell integrity.

Looking to the future

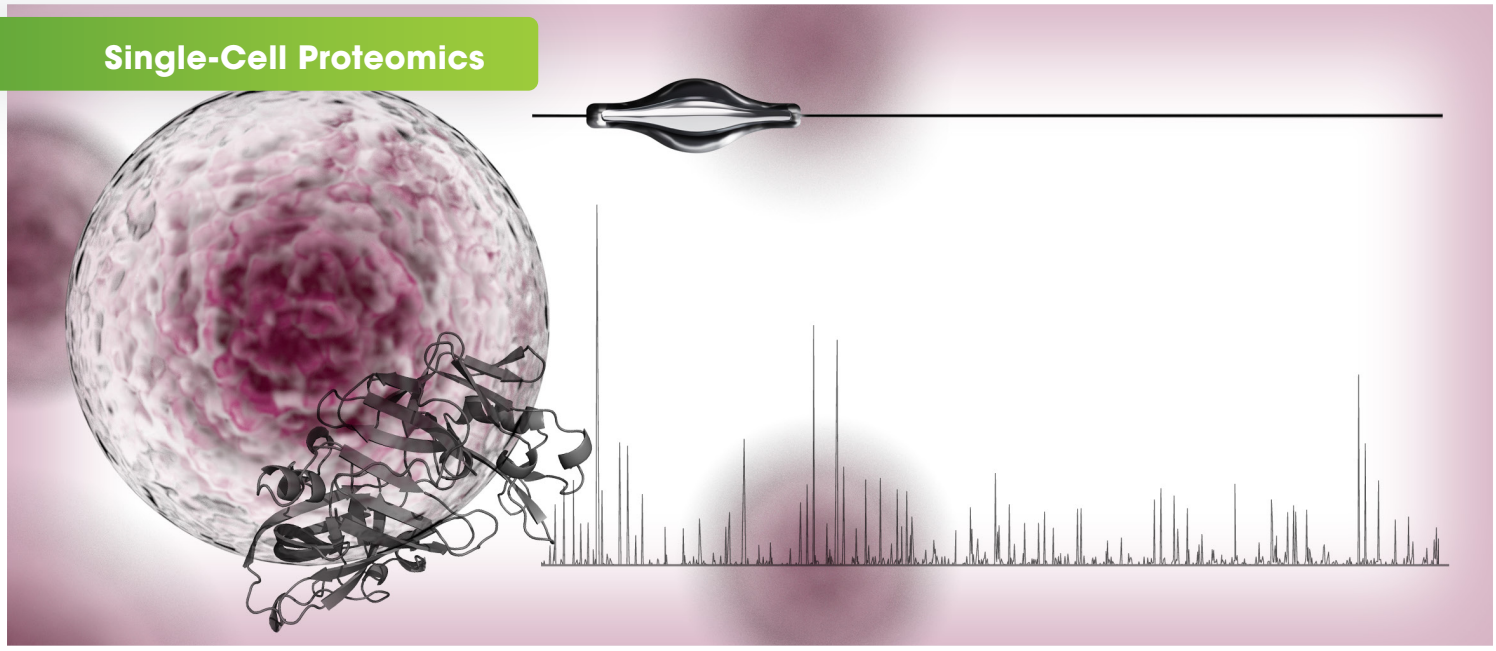
"We're still in the early days of single-cell proteomics", says Ryan Kelly, "and if we look at each step along the way from cell isolation and sample preparation, to separation, MS analysis and the data analysis following data acquisition, I think there is a lot of room to improve each one of these areas and the combined improvement across the entire workflow is going to be growing an order of magnitude over the next decade or so." Although single-cell transcriptomics may continue to have the advantage of higher throughput and popularity, single-cell proteomics will have the undisputed advantage of providing insight into post-translation modifications. Looking forward, Erwin notes that this is "probably where single-cell proteomics will be the key distinguishing factor from RNA-seq. For example, really studying phosphorylation dynamics has to be done using a protein level approach." As single-cell proteomics continues to mature, we will witness new products and systems that ensure cellular integrity during sample preparation such as optimized digestion cocktails, improved buffers to maintain cellular homeostasis, expanded barcoding systems and automated LCM systems integrated into sample processing. A near milestone for the field is reproducibly detecting over 1000 proteins per single cell. Another milestone further down the road is to apply such throughput to over 1000 cells per day. While the road ahead may be long, the combination of academic and commercial investment, in addition to the clinical promise, all demonstrate that we are on our way to reaching these milestones and learning more than ever about the proteomes of single cells.

References

1. Reichard A, Asosingh K. Best Practices for Preparing a Single Cell Suspension from Solid Tissues for Flow Cytometry. *Cytometry A*. 02 2019;95(2):219-226. doi:10.1002/cyto.a.23690
2. Zhu Y, Dou M, Piehowski PD, et al. Spatially Resolved Proteome Mapping of Laser Capture Microdissected Tissue with Automated Sample Transfer to Nanodroplets. *Mol Cell Proteomics*. 09 2018;17(9):1864-1874. doi:10.1074/mcp.TIR118.000686
3. Zhu Y, Piehowski PD, Zhao R, et al. Nanodroplet processing platform for deep and quantitative proteome profiling of 10-100 mammalian cells. *Nat Commun*. 02 2018;9(1):882. doi:10.1038/s41467-018-03367-w
4. Liang Y, Acor H, McCown M et al. Fully Automated Sample Processing and Analysis Workflow for Low-Input Proteome Profiling. Anal Chem. 2020;93(3):1658-1666. doi:10.1021/acs.analchem.0c04240
5. Specht H, Emmott E, Petelski AA, et al. Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol*. Jan 2021;22(1):50. doi:10.1186/s13059-021-02267-5
6. Schoof E, Furtwängler B, Üresin N et al. Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. Nat Commun. 2021;12(1). doi:10.1038/s41467-021-23667-y

Dr. Erwin Schoof
Associate Professor,
Technical University of Denmark

Dr. Ryan Kelly
Associate Professor, Brigham Young University and Pacific Northwest National Laboratory



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.

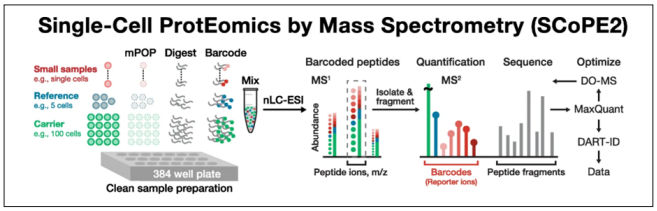


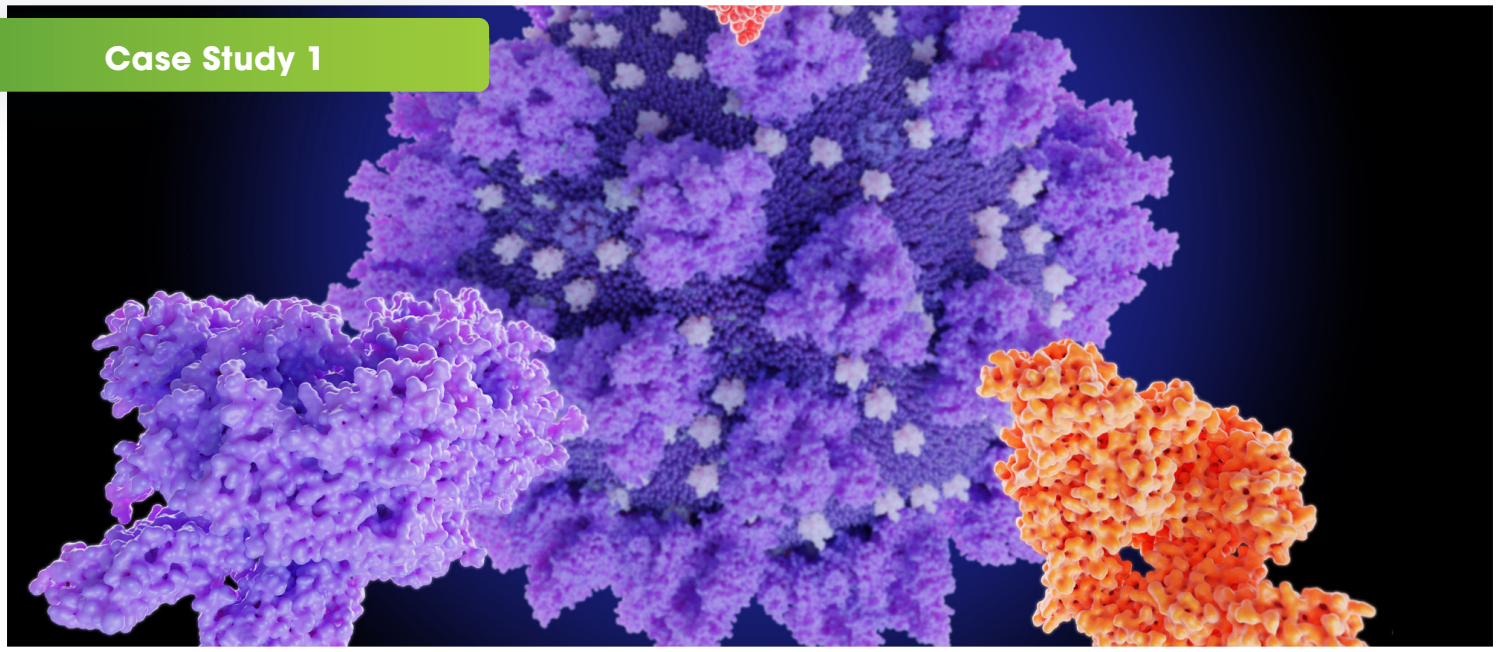
Figure 1. Conceptual diagram and workflow of SCoPE2. Cells are sorted into multiwell plates and lysed by a Minimal ProteOmic sample Preparation (mPOP). The proteins in the cell lysates are digested with trypsin; the resulting peptides labeled with TMT, combined, and analyzed by LC-MS/MS. The LC-MS/MS 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

Case Study 1



Defining the Carrier Proteome Limit

With Christopher Rose

The 'carrier proteome' is typically a mixture of cells or tissues that mimics the experimental samples and is added at a high level to enable peptide selection and identification and the analysis of low-level samples or post-translational modifications. The recent development of single-cell proteomics by mass spectrometry (MS) (SCoPE-MS) has prompted the inclusion of carrier proteome at 25× to 500× in recent single-cell experiments. However, the underlying technologies (isobaric labeling and mass spectrometry (MS)) have technical limitations that affect data quality and biological interpretation when using high levels of carrier proteome. Dr. Christopher Rose, a Principal Scientist in Discover Proteomics at Genentech, and his lab develop and implement new approaches that overcome limitations in current proteomic technologies to help researchers advance research and therapeutic discovery.

Study rationale

A carrier proteome enables more peptides to be identified, however high levels of carrier proteome may adversely affect quantitative accuracy and biological conclusions. While SCoPE-MS is an exciting step forward for researchers wanting to quantify multiplexed single-cell proteomes, early implementations were not robust enough for researchers to accurately measure a single-cell signal in the presence of a large carrier proteome signal. Chris and his team therefore set out to develop a workflow that provided guidance on experimental design, data collection and data analysis, enabling other researchers to obtain high-quality data and avoid passing on misleading results.

> "I think the biggest impact of this paper is that it shows people how to examine the data to understand the underlying data quality and then adjust the instrument parameters accordingly. It also enables them to collect the data in a way that ensures it is more high-quality."

Dr. Christopher Rose
Principal Scientist,
Discover Proteomics,
Genentech

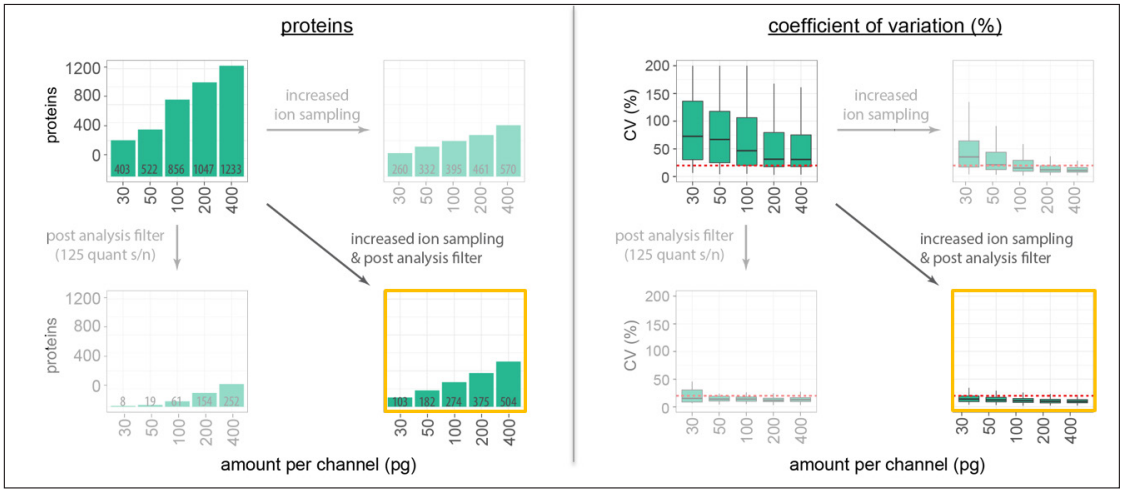


Figure 1. SCoPE-MS analysis of a sample with 100x carrier protome analyzed and filtered with different parameters. Data collected with typical instrument parameters (top left) identifies a large number of proteins, but the underlying has a median CV above 20%. This is due to an undersampling of 'single cell' ions. Data accuracy can be improved by increasing the number of ions sampled and applying a post analysis signal filter (bottom right, yellow). Image courtesy of Christopher Rose.

Methodology

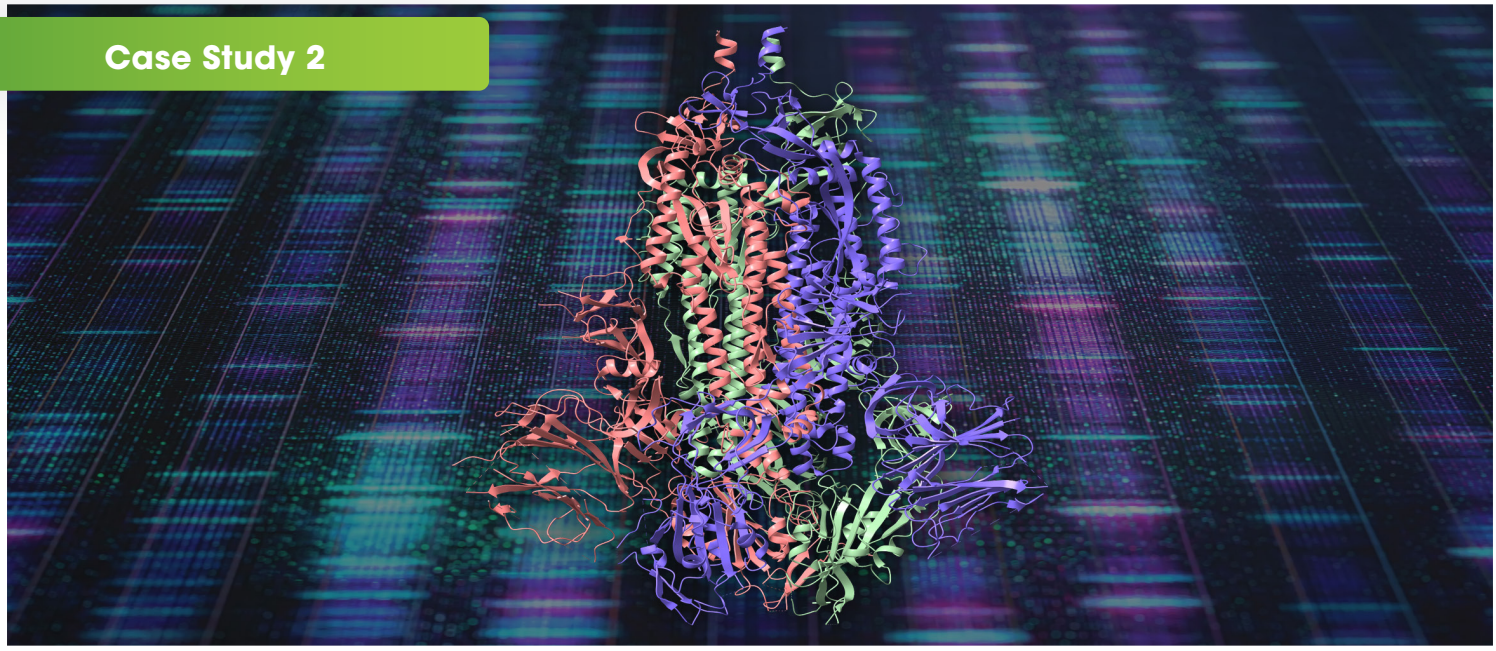
Chris and his lab performed controlled experiments with increasing carrier proteome amounts to evaluate quantitative accuracy. To investigate the variability of measurements and how quantitation deteriorates with increasing levels of carrier proteome, they performed simulations with different instrument parameters controlling sampling of ion clouds made of equal single-cell populations and increasing carrier proteome levels. The results demonstrated that higher carrier proteome levels increase measurement variability, but this effect can be overcome by sampling more ions. They went on to confirm these findings experimentally using equally mixed, non-single-cell bulk samples in the presence of increasing carrier proteome levels. By examining the signal-to-noise ratio, Chris and his team were able to examine the relationship between the quantity of ions measured and the measurement variability to determine optimal parameters that could be transferred across different instrument platforms. They created a program called SCPCompanion that enables the rapid evaluation of single-cell proteomic data and recommends instrument and data analysis parameters for improved data quality.

Results and implications

Their results demonstrated that an increase in carrier proteome level requires a concomitant increase in the number of ions sampled to maintain quantitative accuracy and that – even with sufficient ion sampling – the accurate quantitation of single-cell proteomes may still be compromised by limitations in ion coalescence and space charging. They therefore recommend limiting carrier levels to ~20x and applying the appropriate quantitative signal-to-noise ratio (SNR) filtering when using common MS instrument parameters. Overall, while this study focused on single-cell proteomics, the samples and signal ranges used are applicable to most experiments, therefore the paper itself can apply to any experiment that uses a carrier boost channel to help scientists understand the limits.

Read the full paper

Case Study 2



Quantitative Single-Cell Proteomics

With Erwin Schoof

Large-scale single-cell analyses are necessary to capture biological heterogeneity within complex cell systems, however they have largely been limited to RNA-based technologies. By exploiting a leukemia culture system, Dr. Erwin Schoof and his team used an experimental and computational pipeline to create a comprehensive mass spectrometry (MS)-based single-cell proteomics workflow. As a result, they quantified hundreds of proteins across thousands of single-cells and extracted data about cell-specific proteins and functional pathways, laying the foundation for further global studies using single-cell proteomics.

Study rationale

The study was conducted as a proof of concept to evaluate the ability of single-cell proteomics to discover heterogeneity in a well-defined system. In terms of biological interest, leukemia stem cells (LSCs) are largely responsible for a lack of treatment response in patients; they are quiescent and thus do not respond to chemotherapy – which targets cycling cells. It is of great therapeutic interest to gain a better understanding of the molecular signatures of LSC, and the possible targets of key proteins and their pathways. As the field is extremely young, researchers are still to prove beyond a shadow of a doubt that measuring the proteins in cells, at single-cell level, can provide more insights into biological processes than single-cell RNA sequencing (sc-RNAseq).

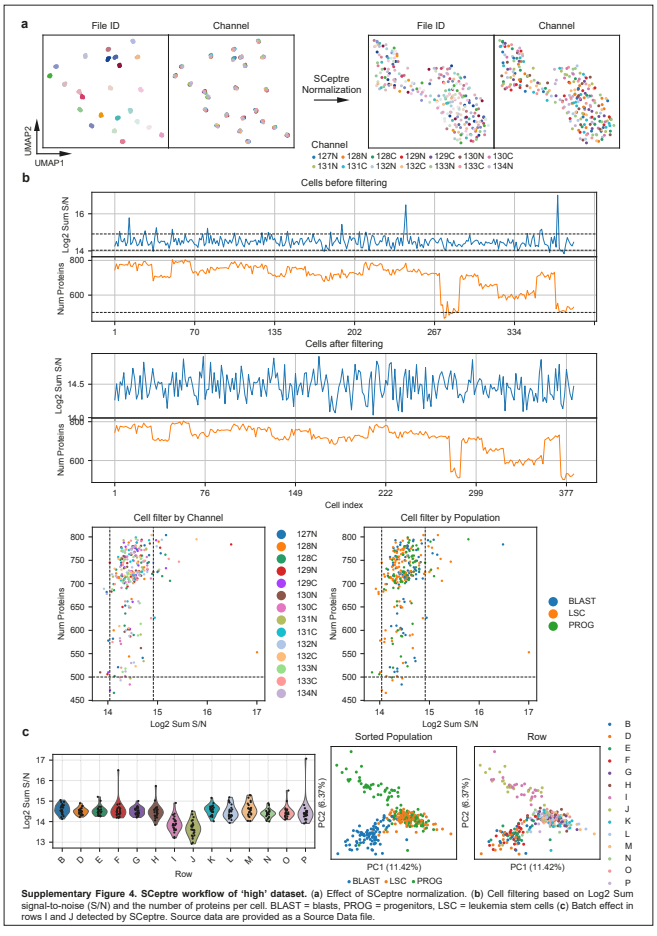
Methodology

Building upon previous work by Budnik and Slavov, Erwin

> **“As the field is extremely young, researchers are still to prove beyond a shadow of a doubt that measuring the proteins in cells, at single-cell level, can provide more insights into biological processes than single-cell RNA sequencing (sc-RNAseq).”**

Dr. Erwin Schoof
Associate Professor,
Technical University of Denmark

and his team established a semi high-throughput workflow using only standard consumables, with the inclusion of sample prep automation for greater throughput. Using a TMTpro 16plex labeling reagents multiplexed approach they simultaneously measured 14 single-cells and combined this with a booster channel of 200 cells to add more signal for the MS instrument to do peak-picking. Applying fluorescence-activated cell sorting (FACS), in a 384-well plate, while recording the immunophenotypic markers commonly deployed during flow-based sorting (termed "index-sorting"), they were able to sort thousands of cells in a matter of hours. They also built a computational workflow, termed SCeptre (single-cell proteomics readout of expression), that integrates key functionality from SCANpy (a leading sc-RNAseq software package) and tailors it for single-cell analysis data searched by Proteome Discoverer software. This allowed the team to extract cell-type-specific proteins, visually represent the highly multi-dimensional data (normalizing and correcting for batch effects) and overlaying of FACS and protein levels upon the single-cells in any embedding of choice. With a throughput of 8-10 injections per day, per instrument, MS analysis still represents the limiting factor at this stage, however, developments are being made to double the cell throughput at similar proteome coverage.



Supplementary Figure 4. SCeptre workflow of 'high' dataset. (**a**) Effect of SCeptre normalization. (**b**) Cell filtering based on Log2 Sum signal-to-noise (S/N) and the number of proteins per cell. BLAST = blasts, PROG = progenitors, LSC = leukemia stem cells (**c**) Batch effect in rows I and J detected by SCeptre. Source data are provided as a Source Data file.

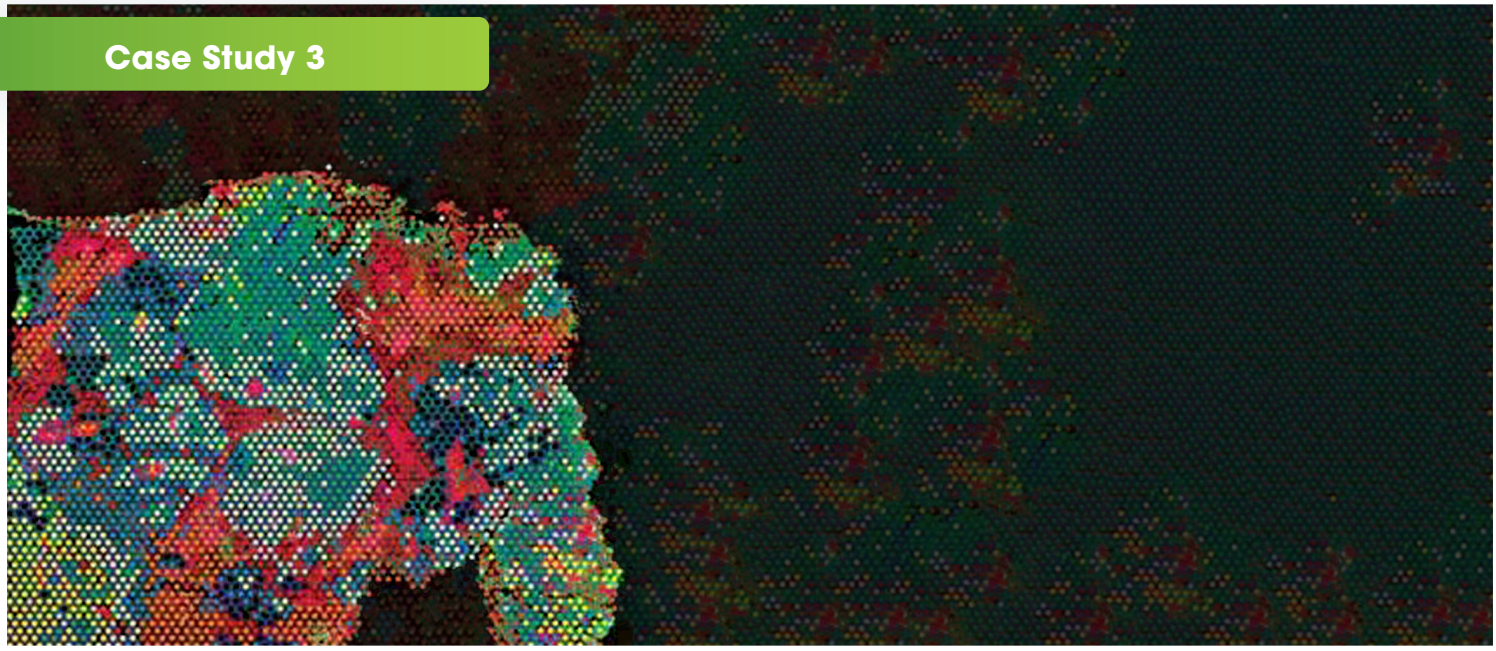
Figure 1. Effect of SCeptre normalization on signal deconvolution of individual TMT batches (image courtesy of Erwin Schoof).

Results and implications

Erwin and his team were able to detect and pinpoint rare cell types (LSCs), even amidst a bulk cell background, which holds great promise for real biological impact to be derived from single-cell analysis when used on primary samples. They also became the first research group to link surface markers from individual cells to their intracellular protein levels using MS. Their results suggest that LSCs can differentiate into progenitors and blasts, challenging previous knowledge that LSCs can only differentiate into blasts once they have become progenitors. This study also demonstrated that cell populations within a heterogeneous culture can be separated and subsequently validated using previously known FACS surface marker expression data. A bulk single-cell approach like this, using single-cell analysis, had not been attempted before.

Read the full paper

Case Study 3



Spatial Protein Mapping with Automated LC-MS/MS

With Kristin Burnum-Johnson

Imaging mass spectrometry (MS) is a powerful tool used to map the spatial distribution of biomolecules across a tissue of interest. Dr. Kristin Burnum-Johnson and her team from the Pacific Northwest National Laboratory (PNNL), used a bottom-up nanoproteomics imaging approach to analyze tissue voxels and generate quantitative cell-type-specific images for >2000 proteins with 100-µm spatial resolution across thin (10-12 µm) tissue sections. As a result, they were able to successfully map proteome heterogeneity across the pregnant mouse uterus, marking a huge step forward in the imaging MS field.

Study rationale

Biological tissues represent some of the most complex assemblies in nature. To better understand these tissues, researchers need to visualize with high spatial resolution the location of each biomolecule, including proteins, and how they combine to carry out functions. Imaging MS is emerging as a powerful way to map these molecules across tissues, however technical challenges have limited its application to proteins.

Methodology

The team extended previous nanoPOTS (nanodroplet Processing in One Pot for Trace Samples) developments by Zhu and Kelly to create proteomic images. They focused on small regions of tissue, or voxels, each measuring 100 microns long x 100 microns wide and just 10 microns thick. NanoPOTS was used for all sample processing, protein extraction, reduction, alkylation, and proteolysis,

> “Our protein images were able to characterize unique tissue microenvironments within the same cell populations by visualizing the gradient expression increase of stroma proteins along the mesometrial (top)–antimesometrial (bottom) axis of the uterus.”

Dr. Kristin Burnum-Johnson
Senior Scientist and Team Lead of the Environmental Molecular Sciences Laboratory's Metabolomics Group, Pacific Northwest National Laboratory

on each voxel in just 200 nanoliters of fluid, then liquid chromatography (LC)-MS/MS was used to measure levels of more than 2,000 proteins across each sample. With reference to molecular mapping, the peptides resulting from nanoPOTS processing of each voxel were transferred into 96-well PCR plates. For each well, a homebuilt LC system automatically performed sample injection, sample cleanup, LC separation and MS/MS data acquisition for 97 minutes on a Q Exactive Plus hybrid quadrupole-Orbitrap mass spectrometer. To help interpret the data, molecular maps were created by colleagues Bramer, Stratton, and Webb-Robertson, who are experts at aggregating and analyzing large amounts of data into forms that can be interpreted more readily and accurately. They used Trelliscope, an open-source platform that PNNL developed for data visualization and management to convert the mounds of numbers into a portrait of protein abundance. Now that these foundational methods and tools exist, the team is well-positioned to apply nanoPOTS imaging to other biological tissues.

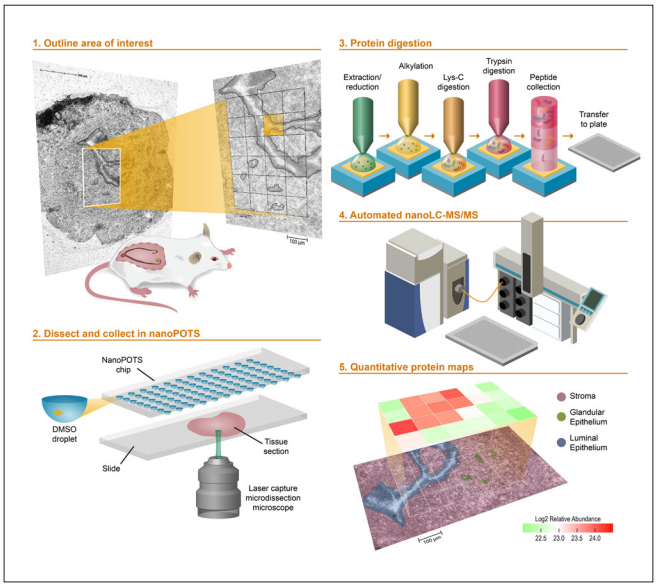
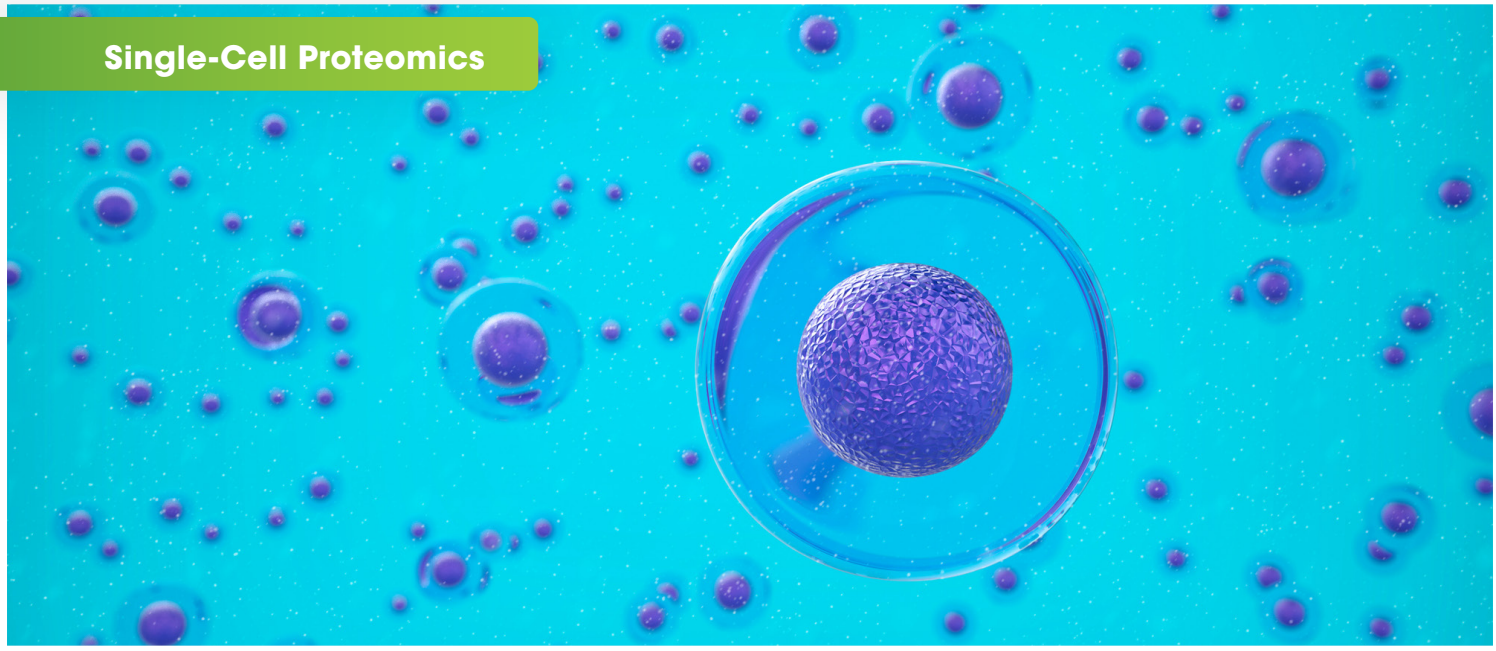


Figure 1. Schematic workflow for high-throughput, spatially-resolved proteomics using the nanoPOTS imaging platform (image courtesy of Kristin Burnum-Johnson).

Results and implications

The next steps with this research are to improve throughput and spatial resolution of the nanoPOTS proteome imaging platform by incorporating multiplexed isobaric labeling. Tandem Mass Tag (TMT) labeling is an approach developed for multiplexed identification and quantification of proteins from multiple different samples in a single LC–MS/MS analysis. We are already showing the feasibility of this approach with preliminary data generated by Zhu and Piehowski utilizing TMT10plex Isobaric Label Reagent. These initial experiments give 10 times higher throughput. Because 10 different voxels are combined in a single LC-MS analysis, the proteome identification sensitivity is greatly improved as well, which facilitates the mapping of much smaller tissue voxels without compromising proteome coverage.

Read the full paper



The Future of MS for Single-Cell Proteomics

Perspectives from Nikolai Slavov

Single-cell RNA sequencing (scRNA-seq) is a widely used technique that allows researchers to analyze RNA at the single-cell level. Thanks to the industrious efforts of researchers across the globe, new methods and techniques have been developed which enable the comprehensive and high-throughput profiling of hundreds to thousands of proteins in individual cells. As a result, researchers are gaining a more comprehensive understanding of complex and rare cell populations, the regulatory relationships between genes, the trajectories of distinct cell lineages in development, and – ultimately –of physiology in health and disease.

Dr. Nikolai Slavov, Associate Professor at Northeastern University, has developed methods for high-throughput single-cell proteomics by mass spectrometry (MS) and was recently named an Allen Distinguished Investor. An award that saw him receive a $1.5 million, three-year grant to his research in this field. In this exclusive interview, he shares his insights into the use, and impact, of high-throughput MS in the field.

Q: Could you start by talking about the future potential of high-throughput MS and its use in single-cell analysis?

A: Most of the development that has occurred in biology has happened in the space of transcriptomics with single-cell RNA sequencing. However very early on, I thought that MS had the potential to not only be much more quantitative, but also provide biologically and physiologically relevant data that cannot be inferred from transcriptomics. This includes protein abundances, but also the various modifications in proteins, how proteins interact with each other, their localization in the cell, and other important aspects that are much closer to biological function than just the abundance of molecules. At the time that thinking was very unorthodox and against the grain, because many of the prominent leaders in MS believed that this was not possible. They thought that MS simply didn't have the sensitivity to do that analysis. However, that thinking has now completely changed; there are dozens of leading laboratories trying to develop their own methods for measuring proteins in single-cells. When approaching more directly your question about the throughput, it has long been clear that many of the questions that we are asking in single-cell biology require the analysis of a large number of cells, and I say many because it is not all. There are certainly questions that one might be able to answer by analyzing relatively few cells – perhaps 100 - 200 cells – but for the most part, to be effective, single-cell analysis must be able to analyze many thousands of single-cells.

The main expense is MS time, that is instrument time. We cannot make instruments cheaper, therefore the way to lower costs and increase throughput is to analyze more cells per unit of time. And that's where multiplexing comes into view; it is a very effective way to make it possible to analyze many cells in a short period of time and therefore decrease the cost. I should also mention that throughput depends on how many proteins we analyze within a single-cell. One way to increase throughput is by decreasing the coverage, but of course that's a trade-off that is not always desirable – and in many cases it is undesirable. At the moment, we can analyze about 200 single-cells per day using multiplexed approaches. I anticipate that we can increase this significantly in the coming years, both

by coming up with higher multiplexed approaches and decreasing the mass spec analysis time per sample, e.g., per single-cell or per labeled set of multiplexed single-cell samples. The flip side of that is that our current throughput of 200 cells per day per instrument is already high enough to apply the technology to many biological questions. I am currently investigating the interaction between immune cells and cancer cells, looking at three-dimensional protein maps at single-cell resolution of both healthy tissues but also disease states – in particular how aging results in senescent cells, and how their proteomes are different compared to normal cells. So, although the throughput of multiplexed approaches is not as high as single-cell RNA sequencing, it is high enough that we can already do a lot of exciting science while simultaneously continuing to come up with even higher throughput approaches.

Q: How can we make single-cell proteomics a more accessible, mainstream technique?

A: That's probably the Achilles heel of single-cell proteomics at present; despite already being very powerful, it is not widely accessible to everybody who would like to use it. I am inundated daily by requests from colleagues who would like to collaborate with us and as much as I want to help all of them, I cannot collaborate with all of them. The solution is to develop methods that are not only high-throughput and quantitative, but also robust and accessible. We should also be able to implement these methods on commercially available equipment, however not all the methods that exist have these qualities. This has been one of the guiding principles for everything that my laboratory does, and sometimes we have had to sacrifice throughput or the depth of protein coverage so that we can keep the methods accessible to everybody and to make them as robust as possible. It is going to be a process of helping people adopt the methods once they are robust enough. We will need to have accessible resources such as detailed protocols – we recently published a very detailed protocol of our workflow for example. It is also important to explain in detail how these analyses are performed at conferences and workshops, and we host such meetings. I think adoption is going to grow from there and that is important for single-cell proteomics to have a large impact.

Q: Do you have any concerns for the field?

A: The main concern that I have is overly enthusiastic, unfounded claims of either higher accuracy of quantification, or higher throughput. Once exposed, such claims can result in negative associations and undermine the credibility of the field. This is really the only concern that I have. To some extent, this sort of thing can happen with any new field when there is a lot of excitement, however as long as this doesn't happen too often to steal the limelight, I think we are in good shape to move forward.

Q: What challenges do you face when you're developing new MS-based techniques for proteomics?

A: There are of course many technical challenges. We have been fortunate to be very well funded because there is a broad appreciation for the need of the technology that we are developing. I am very grateful and appreciative of funders – in particular the Allen Frontiers Group who has shared our vision and given us a substantial amount of funding to develop these technologies. The technical challenges in need of new conceptual solutions are the ones that we scientists like to face. There are also the more standard technical issues that include maintaining instruments with the least amount of effort, ensuring reproducibility, rigorous benchmarking, and being able to rapidly diagnose problems so that we spend less time troubleshooting and more time developing our ideas. I can speak a lot on the technological challenges; however I have described them in fairly conceptual and detailed form in my review.

The biggest challenge I have faced is communicating to a broad enough group of scientists that are interested in our technology, how it works, what its challenges are and its potential. MS is not a field that is integrated with biomedical research as well as it should be. There is a lot of jargon, and there are different languages being used, so I think communication remains an important challenge. We have made a lot of progress, in part by organizing the annual single-cell proteomics conferences, but I think we have a lot more progress to make on that front. We also need to attract more students and postdocs who are brilliant and share our passion for advancing both the technology and its biomedical applications in transparent and accessible ways.

Q: What aspects of proteomics research currently excites you the most, and what are some of your personal aspirations for this space in the future?

A: I think measuring proteins allows us to explore globally post-transcriptional regulatory mechanisms at the single-cell level for the first time. So much of the efforts in biomedical research focuses on transcriptional regulation because it's accessible to measure, and it is clearly very important. Yet post-transcriptional regulatory mechanisms are also crucial and very important, but they are almost completely unexplored at the single-cell level due to a lack of tools. More specifically, I hope that we can have the ability to infer direct causal interactions between molecules underpinning biological functions. I

gave an hour-long talk on this at the Broad Institute. A lot of the associations that we have identified in biology, such as those from genome-wide tumor analysis or genome-wide associations are indirect. Even if we assume that everything is correct, they are very difficult to interpret because they tell you how a DNA locus – a particular DNA polymorphism or a gene variant – is associated with a disease. However, that association is very indirect and there are hundreds of different molecules that interact in the path leading from the DNA mutation to the disease. In different people those interactions might happen in a different way, so even a direct causal association that is true in one population may not be true in another population (the association itself is consistent with an infinite number of models, and we cannot find the correct one). It is not a question of coming up with the right algorithm, it's a question of the information not being there to make the inference. Measuring proteins across enough single-cells may support the inference of more direct molecular interactions; the more direct the associations that we find, the more we can constrain the possible models consistent with the data and eventually find the invariant representations. This can then help us both understand basic biology and design more effective treatments.

Dr. Nikolai Slavov
Associate Professor,
Northeastern University

Compendium

Karl Mechtler	**Profile**	https://coreforlife.eu/platforms-technologies/csf-vienna-biocenter/karl-mechtler
	Paper	The Rise of Single-Cell Proteomics.
	Paper	Comparative Proteome Signatures of Trace Samples by Multiplexed Data-Independent Acquisition.
	Paper	Improved Sensitivity in Low Input Proteomics using Micropillar Array-Based Chromatography.
Ying Zhu	**Profile**	https://www.pnnl.gov/people/ying-zhu
	Paper	High Throughput and High Efficiency Sample Preparation for Single-Cell Proteomics using a Nested Nanowell Chip.
	Paper	Automated Coupling of Nanodroplet Sample Preparation with Liquid Chromatrography-Mass Spectrometry for High Throughput Single-Cell Proteomics.
	Paper	Nanodroplet Processing Platform for Deep and Quantitative Proteome Profiling of 10-100 Mammalian Cells.
	Paper	High-Throughput Single Cell Proteomics Enabled by Mutiplex Isobaric Labeling in a Nanodroplet Sample Preparation Platform.
	Paper	Proteomic Analysis of Single Mammalian Cells Enabled by Microflidic Nanodroplet Sample Preparation and Ultrasensitive NanoLC-MS.
Erwin Schoof	**Profile**	https://orbit.dtu.dk/en/persons/erwin-schoof
	Paper	Quantitative Single-Cell Proteomics as a Tool to Characterize Cellular Hierarchies.
	Article	Down to a Single Cell.
	Article	Scientists Detect Diseases Based on Proteins in Single Cells.
	Webinar	Multiplexed Single Cell Proteomics: A marriage of Sensitivity and Throughput
Ryan Kelly	**Profile**	https://www.chem.byu.edu/faculty/ryan-kelly/
	Paper	The Emerging Landscape of Single-Molecule Protein Sequencing Technologies.
	Paper	Ultrasensitive Single-Cell Proteomics Workflow Identifies >1000 Protein Groups Per Mammalian Cell.
	Paper	Single-Cell Proteomics: Progress and Prospects.
	Paper	Improved Single-Cell Proteome Coverage using Narrow-Bore Packed NanoLC Columns and Ultrasensitive Mass Spectrometry.
	Webinar	One Cell at a Time: Single-Cell Proteomics Becomes a Reality.
Kristin Burnum-Johnson	**Profile**	https://www.pnnl.gov/science/staff/staff_info.asp?staff_num=7264
	Paper	Automated Mass Spectrometry Imaging of Over 2000 Proteins from Tissue Sections at 100µm Spatial Resolution.
	Paper	High Spatial Resolution Imaging of Biological Tissues using Nanospray Desorption Electrospray Ionization Mass Spectrometry.
	Paper	High Spatial Resolution Imaging of Mouse Pancreatic Islets using Nanospray Desorption Electrospray Ionization Mass Spectrometry.
	Webinar	Breakthrough in Spatial Proteomics: Automated High Resolution LC-MS/MS Tissue Imaging Workflow.
Nikolai Slavov	**Profile**	https://slavovlab.net/index.html
	Paper	Multiplexed Single-Cell Proteomics using SCoPE2.
	Paper	Single-Cell Proteomic and Transcriptomic Analysis of Macrophage Heterogeneity using SCoPE2.
	Paper	Unpicking the Proteome in Single Cells.
	Paper	SCoPE-MS: Mass Spectrometry of Single Mammalian Cells Quantifies Proteome Heterogeneity During Cell Differentiation.
	Webinar	Using Single-Cell Proteomics and Transcriptomics to Understand Gene Regulation.
Christopher Rose	**Paper**	Defining the Proteome Carrier Limit for Single-Cell Proteomics
Thermo Fisher Scientific	**Website**	https://www.thermofisher.com/us/en/home/industrial/mass-spectrometry/proteomics-mass-spectrometry/single-cell-proteomics.html
	White Paper	Challenges and Emerging Directions in Single-Cell Proteomics
	Application Note	High-Throughput Single-Cell Proteomics Analysis with Nanodroplet Sample Processing, Multiplex TMT Labeling, and Ultra-Sensitive LC-MS.
	Technical Note	Label-Free Proteomics Performance with the Oribtrap Exploris 480 Mass Spectrometer with Single-Cell Sensitivity.
	Technical Note	Ultra-Sensitive LC-MS Workflow for In-Depth Label-Free Analysis of Single Mammalian Cells with Nanodroplet Sample Processing.