

A dream of single-cell proteomics

As single-cell proteomics emerges, perhaps labs can avoid the need to infer protein levels from mRNA abundances.

Vivien Marx

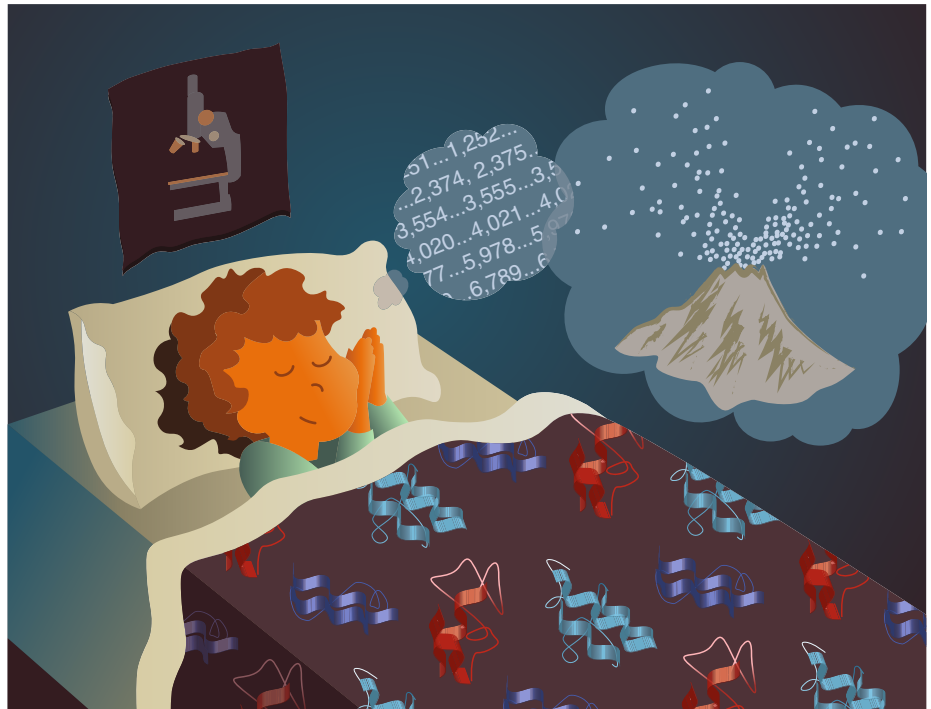
Nowadays, labs can generate massive sets of single-cell genomic and single-cell transcriptomic data. In proteomics, high-throughput single-cell methods have not yet arrived. The nascent field of single-cell proteomics (sc-proteomics) is bringing change, and perhaps helping to avoid the need to infer proteins from cellular mRNA levels. It's early days, but it's not a distant dream to be able to tally the proteins in single cells, says Ruedi Aebersold, a proteomics researcher at ETH Zurich and the University of Zurich. To make that dream an everyday reality, labs push hurdles out of the way. Proteins are tougher to work with than RNA or DNA, for example they're stickier, but eventually researchers might be able to integrate single-cell mRNA and single-cell proteomic measurements.

Over 20 years ago, says Aebersold, Richard Smith at Pacific Northwest National Laboratory and his colleagues characterized hemoglobin from a single red blood cell with a technique called capillary electrophoresis-electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry¹. The red blood cell was a special case, says Aebersold, since it mainly consists of hemoglobin. But this was single-cell analysis.

Fast forward to a recent approach that Aebersold and his colleague Ben Collins call a “marriage across the ages”². Developed in the labs of Edward Marcotte, Eric Anslyn and colleagues, at the University of Texas at Austin, the technique yields the amino acid sequence of individual proteins in a highly parallelized fashion³. A spinout company, Erisyon, has been launched to commercialize a single-molecule protein sequencer.

The approach involves Edman sequencing, with which proteins were sequenced in the days before mass spec. Proteins are cleaved and the peptides are labeled with identifying fluorescent tags; the tagged peptides are then immobilized on a glass cover-slip. Successive rounds of Edman degradation chemistry remove one amino acid at a time and the peptides are imaged at each round. The team found dyes that handled the process but some mishaps occurred: dyes fell off, didn't attach well or provided inadequate fluorescence.

In their published study, the team analyzed a zeptomolar mixture of proteins, but they



It's not a faraway dream to be able to tally proteins in single cells. Credit: S. Larochele, E. Dewalt, Springer Nature

note the approach is “inherently single molecule” and thus “there are reasonable prospects for decreasing sample volumes and protein abundance requirements.” They state that once fluorescent labeling of low-abundance proteins with fluorescent tags can be achieved, this method has application potential, such as for single-cell proteomics experiments.

It's far from single-cell analysis, says Aebersold, but “the method certainly has potential to do single cells, potentially much faster and in ways that the mass spectrometer would have a hard time doing.” As a student, he used Edman sequencing and he is intrigued to see its revival for single-molecule fluorescence. The method piggybacks on flow-cell technology used in high-throughput genome sequencers, he says. It will take work to make this a routine application, he says, and other labs have such approaches in their sights, too. This study is an important proof of principle.

Single molecules

It's hard to speculate how fast technology for sc-proteomics will develop, says Harvard Medical School researcher Peter Kharchenko, but “I am most excited about the ability to quantify phosphorylation and other modifications.” This would move the field “beyond simple abundance-based models to more accurate dynamic descriptions.” Humboldt University researcher, Rune Linding, hopes such approaches might open up new ways to analyze phosphorylation dynamics.

Proteomics and genomics labs pursue different questions “but they clearly provide different views on the same system,” says Aebersold. “In proteomics, we're always kind of limping a while behind the genomics field,” he says. Single-cell genomics and transcriptomics can capture “a kind of genealogy of the cell,” and track cells as they evolve and change through mutations, he says. Proteomics labs can now analyze



It might be a combination of different advances that lead to near single-cell proteomics, says Christine Vogel. Credit: S. Sajanlal

small numbers of cells. “It’s a beginning,” he says. This is how single-cell transcriptomics started, followed by massive multiplexing and barcoding strategies that allowed resolution of large numbers of cells. Such trends will take a while to develop in proteomics.

Labs are exploring cytometry-related ways to reap single-cell data. A team that includes researchers at the University of Ottawa and the University of Oxford, used single-cell mass cytometry (CyTOF) to capture the cell-fate decisions during hematopoiesis, and tracked how transcription factor expression changed during a cell’s lineage commitment at 13 time-points. They measured 27 proteins simultaneously in single cells. Another single-cell CyTOF effort by a team at the University of Zurich, along with colleagues at other institutions, profiled tumor and immune cells from 144 human breast tumor samples.

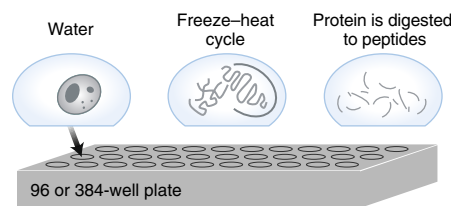
In mass cytometry/CyTOF, cells are prepped for analysis with isotope-conjugated antibodies. In sc-proteomics, it will likely be key to integrate different technologies and physical principles, as with the combination of Edman sequencing and microscopy, says New York University researcher Christine Vogel, who was a postdoctoral fellow in the Marcotte lab.

Rethinking sample prep

“My ideology is to make this as accessible as possible,” says Nikolai Slavov, at Northeastern University, about his approach to sc-proteomics methods development. In keeping with his training in genetics and biology, and interests in math, chemistry and physics, Slavov runs cross-disciplinary sc-proteomics meetings: mass spec veterans attend along with researchers lacking such expertise, as well as physicians, computational scientists and industry researchers.

Slavov is happy to see how CyTOF, single-cell Westerns and immunoassays are enabling quantification of proteins in single cells. To move the possibilities for identification and quantification beyond these techniques, his lab developed single cell proteomics by mass spectrometry (SCoPE-MS)⁴ for identifying and quantifying peptides from mammalian

cells with liquid chromatography and tandem mass spectrometry (LC-MS/MS). They rethought sample preparation: cell lysis, protein purification, digestion and clean-up. And there’s mass spec instruments’ aversion to the clean-up chemicals to consider, says Slavov. In standard mass spectroscopy, sample is always lost as it can, for example, stick to the sides of chromatography columns. Bulk sample analysis helps labs cope with that. But, when assessing single mammalian cells and their few hundred picograms of proteins each, little if any cargo should go missing.



The Slavov lab is testing their mass spec sample-prep method to lyse cells in pure water, using a freeze-heat cycle. Credit: Slavov lab, Northeastern Univ.

Slavov, his graduate student Harrison Specht, and the team, hunted a different way to extract proteins efficiently for LC-MS/MS analysis. “We kept trying different approaches, most of which didn’t work,” says Slavov. Sonication to lyse cells with focused acoustic waves led them to develop SCoPE-MS. When they validated the method, a student in the lab held the tube in the sonicator to lyse cells one at-a-time. The team mixed labeled peptides with labeled ‘carrier’ peptides to avoid the “never-ending chase” to quantify sample losses from the clean-up process. They used isobaric tandem mass tags (TMT), which bind to all peptides, including the carrier peptides. The TMT tags all have the same molecular weight so “when they enter the instrument, they’re going to correspond to a single peak in *m/z* space,” says Slavov.

Vogel likes SCoPE-MS and says it’s still necessary to test new buffers and optimize the approach, which her lab is currently doing. In SCoPE-MS, carrier cells act as a kind of internal reference, says Linding. “It works,” he says, but eventually labs will want to avoid a ‘carrier proteome’. Other methods may emerge for analyzing proteins and modifications, such as phosphorylation. Right now, sample preparation and labeling technology are limiting sc-proteomics, he says, and he believes CyTOF will play an important role in validation and imaging.

Because Slavov wanted a more affordable method, in SCoPE2 the team replaced

sonication by lysing cells with a freeze–heat cycle in pure water⁵. The team is testing the efficiency of this sample preparation technique and “it appears to be at least as good, if not better than, urea lysis,” he says. He and his team used this method to quantify 2,000 proteins in 356 cells — a sample containing both monocytes and macrophages.

Vogel says that pure water might avoid artifacts from chemicals, but she wonders how soluble proteins without ion content are in pure water. Speaking more generally about sample preparation for sc-proteomics, she says that proteins are trickier than RNA and DNA: they cannot be amplified, they’re sticky and they degrade easily. “Until someone invents something to ‘amplify proteins’ that’ll always be the problem,” she says. Many methods, including hers and others, address sample preparation by using techniques such as hydrostatic pressure or engineered surfaces for peptide enrichment.

Multiplexing

With TMT, around 20 barcodes can be used. But labeling a protein or peptide introduces complexities, says Aebersold: the reaction must be just right, excess has to be removed and there’s clean-up. Multiplexing is a scale-up that makes workflow more complex. Transcriptome analysis is readily available to biologists with commercially well-supported techniques, while proteome analysis is mainly done in expert labs and “cannot easily reach the throughput, robustness and reproducibility of transcriptome analysis,” as Aebersold and his colleague Ben Collins point out. But it doesn’t have to stay that way.

The 200–300 picograms of protein in one mammalian cell cannot yet be tallied, says Aebersold. A properly tuned and optimized mass spec instrument can detect all or most proteins expressed in a cell only when many cells are analyzed concurrently. His lab has detected 500 to 1,000 distinct proteins in a single cell from a sample equivalent to a single cell using SWATH-MS, a type of data-independent analysis. They have not yet ‘processed’ a single cell but they injected a proportional fraction from a small number of cells into the mass spec. The goal is to eliminate sample handling losses, he says, such as material sticking to the surface of a microtiter plate or Eppendorf tube. The team hunted for the least absorbent material, chose polydimethylsiloxane (PDMS) and built microfabricated devices that “work quite well,” he says.

Cells flow into the device with one cell per compartment, which is confirmed with imaging. Cells can be manipulated and lysed,

proteins can be washed, and the sample can then be worked up for mass spec. The team is still testing the approach. Aebersold likes that it only requires cell sorting — no chemicals needed.

“Eventually, in single-cell analysis, each cell is a singleton,” says Aebersold. No two cells are entirely identical. They might resemble one another closely in terms of biochemical function, but appear to be dissimilar due to differing cell-cycle phases. To address such variability, Linding says he and his team try to synchronize cell cycles in their cell lines. Even then, cells are “highly heterogeneous,” he says, which makes analysis tough. Much understanding about the cell cycle is based on population-level data but sc-proteomics-based measurements might deliver new insights about cell cycle stages.

Until his and other sc-proteomics techniques mature, and labs can analyze large numbers of single cells quickly, the techniques will not yield biologically interesting results, says Aebersold. That is why he and his team pursue an intermediate goal: analysis of small numbers of cells. They cluster cells using multi-parameter fluorescence-activated cell sorting (FACS) with 8–10 fluorescent colors, and group them according to similarity. “Rather than doing single cells and then averaging them or combining them, we combine them first and then measure the average,” he says.

A year ago, the team needed around 20,000 cells; now the method works with 100 cells or less. “Any cell population that can be sorted with a FACS sorter, even with very low numbers, is proteomically accessible,” says Aebersold. A lab might be looking at a rare type of cell or one that is only available in low numbers. For example, his team plans to work with a neuroscience lab to analyze mouse neuronal stem cells. One can obtain around 100 such cells per animal and the goal is to use as few animals as possible.

Aebersold acknowledges this intermediate approach may seem less exciting. At an annual conference on mass spec or other technology-focused event, a lab presenting protein measurement from single cells might make a big impression. At a cancer meeting, however, an sc-proteomic analysis from a tumor might reap a different reaction. “They’re not going to be impressed, because they’re going to say: ‘so what have you learned?’” In experiments with small numbers of cells, such as when exposing cells to a drug or other perturbation, some similar cell types might disappear and others appear. “This is interesting information,” says Aebersold.



“My ideology is to make this as accessible as possible,” says Nikolai Slavov about the methods he develops in sc-proteomics.

Credit: A. Glanzman, Northeastern Univ.

Computed proteins

Slavov says that many researchers in single-cell transcriptomics want to adapt their software for sc-proteomics. All tools in this space will need to be benchmarked, he says, “so that people don’t fool themselves,” he says. Labs need to diagnose data quality and identify what needs trouble-shooting. He and his team have developed data-driven optimization of MS (DO-MS)⁶ to visualize and analyze data. It’s programmed in R, built as a Shiny app and available [here](#). It can help, for example, when elution profiles are not sampled at their apex, which is needed to maximize the number and purity of ions. Slavov foresees much development work ahead for computational tools in sc-proteomics, as well as the need for standards and benchmarking to make sure quantification is well-performed.

At Harvard Medical School, Peter Kharchenko and his team have been developing a computational tool for single-cell RNA-seq data analysis to address heterogeneity issues. These challenges have grown now that RNA-seq is being applied in complex study types involving many measurements, on numerous samples, from different people. A [graphing tool](#) from his lab, written in R, is clustering on network of samples (CONOS)⁷, which tracks cell types across these heterogeneous datasets and clusters similar cell sub-groups. It can also be applied to sc-proteomics, says Kharchenko.

“We’ve designed it to be very tolerant with respect to diversity of samples,” says Kharchenko, so users can do joint analysis related to perturbations or across different tissues. A number of integration methods are emerging. At the same time, integration has ‘subproblems’ he says, such as technical variation, variability across individuals or tissues, molecular modalities or species. With sc-proteomics data analysis, the “relative nature of the signal” is challenging and these data have more complex dropouts than transcriptomic data.

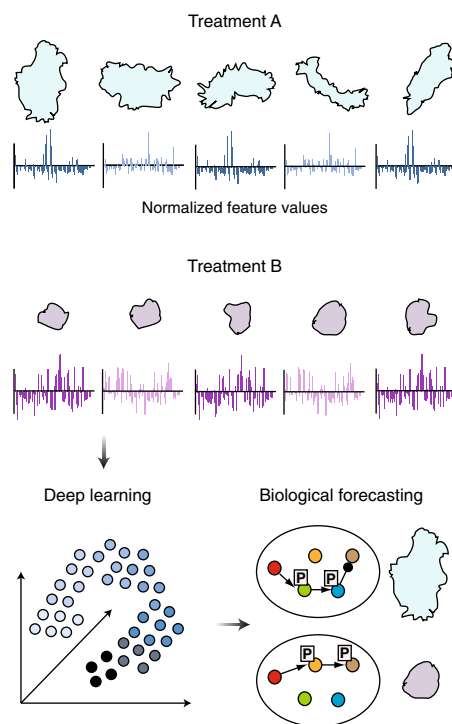
In Linding’s view, given that sc-proteomics data are ‘richer’ than single cell RNA-seq data, “what you need, is an error model.” Cellular

proteins are plentiful but in classic proteomic analysis, a cell’s single peptides are often thrown out. He’s addressing that with a new algorithm for making more accurate error assessments⁸.

In proteomics, it can seem that insufficient numbers of the cell’s proteins are detected or that only the most abundant ones are seen, says Linding, but mass spec does detect plenty. Unlike increased sensitivity detection for mRNAs, which are much less abundant than cellular proteins, even small increases in the sensitivity of protein detection make a big difference.

Given that sc-proteomics is in its infancy, many challenges also remain for computational analysis, says Jürgen Cox from the Max Planck Institute of Biochemistry. In sc-proteomics using mass spec, isobaric labelling is quite promising, as several single-cell channels can be multiplexed for a single mass-spec measurement. Including additional channels, such as multi-cell samples, enhances signal detection in the mass spectrometer and helps establish quantification standards.

It remains challenging in computational proteomics to interpret the multiplexed quantification channels, given that “isobaric labeling techniques are notoriously plagued by co-fragmentation signals,” says Cox.



In single-cell proteomics, tools are emerging to quantify and integrate data about cell behavior, signaling and regulatory networks, says Rune Linding. Credit: Linding lab, Humboldt Univ.

Fragmented peptides are “measured involuntarily” and add unwanted contributions to the signals from peptides of interest. “We are working on normalization methods and signal modeling that will improve the situation,” he says. Missing values in the data matrix are inherent to all single-cell technologies and usually need more attention than with bulk ‘omics data. Taken together with isobaric labeling, he says, special algorithms are needed for these issues.

Big physics

When physicists work with particle accelerators, probabilities are applied to the likelihood something is detected. “That, I think, we are lacking in biology,” says Linding. A mass spec instrument is little like a particle accelerator: when a signal of a certain size is detected, it might or might not be a true signal. He asks, “Can we assign uncertainties to events?” Labs juggling big data should work with probabilities, not averages, he says. Models can then help with data integration, and with explaining cellular behavior, to tease out how different quantitative data are related or to determine causality. “To do all this, we need probability-based models,” he says, also for predicting behaviors of protein networks.

More labs can now use large datasets to, for example, predict how proteins are interacting. Beyond measuring protein abundance or concentration, cancer research labs will want to track kinase activity quantitatively, learn how many phosphorylation sites there are and how many target proteins can be phosphorylated. “It is the activity of the molecule that is important, not necessarily the abundance,” says Linding. Biology is changing: more tools are emerging to quantify different aspects of signaling and of the regulatory networks in and between cells — sc-proteomics helps with that. Traveling between data captured at different scales is non-trivial, says Linding, which calls for new algorithms, and for machine learning combined with more traditional mathematical models.

Kharchenko agrees with the need for probabilistic data interpretation. In many cases, the probability of detection will be close to 1, he says, but the uncertainty in abundance will remain. It will take more experiments to figure out the structure of variation in such measurements. Models are needed to estimate this resulting uncertainty.

Counting mRNAs and proteins

At one point, labs will want to integrate mRNA and single-cell protein measurements. That will bring a longstanding discussion to the fore. The transcriptome correlates with



Proteomics and genomics labs pursue different questions but provide different views on the same system, says Ruedi Aebersold.

the proteome but, given a certain mRNA concentration, it’s hard to model how many proteins from that RNA are present. Beyond cell-to-cell variability in both transcription and translation, and varying levels of protein turnover, there are phenomena such as transcriptional bursts and many external factors that influence cells, mRNAs and proteins. Much about the mRNA–protein relationship has yet to be determined.

Aebersold, along with Jürg Bähler of University College London, and colleagues, quantified the transcriptome and proteome in fission yeast by looking at two cell-states: quiescence and rapid proliferation. “There were a lot of surprises,” says Aebersold, that shed light on the relationship between gene regulation, transcription, protein production and physiology. When fission yeast shifts to dormancy, proteins and mRNA are down-regulated, each in specific ways.

Proliferating cells had around 41,000 mRNAs per cell, with a mean of around 3 transcripts per gene. Protein-coding genes produced a mean of around 5,000 protein copies per cell, with a dynamic range covering five orders of magnitude up to a total of around 60 million protein molecules per cell.

Quiescent cells had a much-reduced transcriptome, with a little over 7,400 mRNAs, and around 31 million protein molecules per quiescent cell. When adjusted for the lower cell volume in quiescent cells, protein numbers dropped by around 10% of the levels in proliferating cells. The mRNA levels drop, and the protein levels less so: when a good food source comes, “they’re ready to run,” says Aebersold.

The shift to quiescence is accompanied by proteome remodeling—nearly half of all proteins changed their copy numbers around twofold. Protein levels drop but not indiscriminately, and those the cell will need when it emerges from quiescence stay at higher levels. Overall, in quiescent cells, mRNAs were between 10,000 to 60,000-fold less abundant than the corresponding proteins, with 1 to 10 mRNAs per gene. The results highlight an sc-proteomics challenge, says Aebersold. For example, a twofold

difference in mRNA level can lead a lab to interpret the cells as different, which they might be, or their cell physiology may have merely shifted, temporarily and stochastically.

A human cell contains hundreds of mRNA copies with proteins numbering in the tens of thousands, says Aebersold. One might find only a few copies per cell of a lowly expressed mRNA, which makes it ever more important to determine what kind of “stochastic time-space” a cell is in, he says, and the risk of drawing false conclusions again rears its ugly head. “I think there’s a lot of pitfalls in the interpretation, from a biological point of view, of single-cell data, which one would have to aware of.” Techniques matter, but when developing and using them one must heed “what they’re good for,” he says. “Then it gets interesting when they can uncover something new.”

Sc-next

As sc-proteomics emerges, labs take a diverse set of paths. As Vogel says, “it might be a combination of different advances that will lead to near single-cell proteomics.” One needs efficient protein extraction from cells, enrichment via specific surfaces, sensitive mass spec needing less and less sample, tagging tricks to enhance identification, and advances in mass spec data acquisition and computational processing. Nevertheless, Vogel is happy about some recent developments, including new mass spec data acquisition methods and more libraries of known spectra, which increase the possibilities of mined data.

“There is something to be gained by having different approaches, particularly in the early phase,” says Linding. Given that its early days in the field, he says, this is not the time to bet only on one horse. □

Vivien Marx

Technology editor for *Nature Methods*.

e-mail: v.marx@us.nature.com

Published online: 12 August 2019

<https://doi.org/10.1038/s41592-019-0540-6>

References

- Hofstadler, S. A. et al. *Anal. Chem.* **67**, 1477–1480 (1995).
- Collins, B. C. & Aebersold, A. *Nat. Biotechnol.* **36**, 1051–1053 (2018).
- Swaminathan, J. et al. *Nat. Biotechnol.* **36**, 1076–1082 (2018).
- Budnik, B., Levy, E., Harmange, G. & Slavov, N. *Genome Biol.* **19**, 161 (2018).
- Specht, H., Emmott, E., Koller, T. & Slavov, N. Preprint at *bioRxiv* <https://doi.org/10.1101/665307> (2019).
- Huffman, R. G., Chen, A., Specht, H. & Slavov, N. *J. Proteome Res.* **18**, 2493–2500 (2019).
- Barkas, N. et al. *Nat. Methods* <https://doi.org/10.1038/s41592-019-0466-z> (2019).
- Robin, X. et al. Preprint at *bioRxiv* <https://doi.org/10.1101/621961> (2019).
- Liu, Y., Beyer, A. & Aebersold, R. *Cell* **165**, 535–550 (2016).