SYSTEMS BIOLOGY

Unpicking the proteome in single cells

Single-cell mass spectrometry will help reveal mechanisms that underpin health and disease

*By **Nikolai Slavov***

Recently, the throughput of single-cell RNA-sequencing (transcriptomics) and genomics technologies has increased more than a 1000-fold. This increase has powered new analyses: Whereas traditional analysis of bulk tissue averages all differences between the diverse cells comprising such samples, single-cell analysis characterizes each individual cell and thus has enabled the discovery and classification of previously unknown cell states. Yet, the nucleic-acid–based technologies are effectively blind to an important group of biological regulators: proteins. Fortunately, emerging mass-spectrometry (MS) technologies that identify and quantify proteins promise to deliver similar gains to single-cell protein analysis. These proteomic technologies will enable high-throughput investigation of key biological questions, such as signaling mechanisms based on protein binding, modifications, and degradation, that have long remained elusive.

The abundance and activity of many proteins are regulated by degradation and posttranslational modifications (PTMs) that cannot be inferred from genomic and transcriptomic measurements. Moreover, genomic and transcriptomic sequencing cannot report directly on protein-protein interactions and protein localization, which are critical for numerous signaling pathways (*1–3*). The extracellular matrix surrounding each cell is composed of proteins whose chemical and physical properties, such as stiffness, can also play vital roles in regulating cellular behavior, including proliferation, migration, metastasis, and aging (*4*). Yet, current single-cell sequencing tools provide little information about the protein composition and biological roles of the extracellular matrix (*3–5*). Thus, methodologies are needed that can directly analyze a broad repertoire of intracellular, membrane-bound, and extracellular proteins at the single-cell level.

Single-cell protein analysis has a long history, but the conventional technologies have relatively limited capabilities (*6, 7*).

Most proteomics methods, such as mass cytometry, cellular indexing of transcriptomes and epitopes by sequencing, RNA expression and protein sequencing, and CO-Detection by indEXing, rely on antibodies to detect select protein epitopes and can analyze only a few dozen proteins per cell (*6*) (see the figure). However, many antibodies have low specificity for their targets, which results in nonspecific protein detection. Indeed, fewer than a third of more than a thousand antibodies tested in multiple laboratories bind specifically to their cognate targets (*6*). As a result, ~$800 million is wasted worldwide annually on purchasing nonspecific antibodies and even more on experiments following up flawed hypotheses based on these nonspecific antibodies (*8*). Although some highly specific and well-validated antibodies can be useful to analyze a few proteins across many cells, the low specificity and limited throughput of the current generation of single-cell protein analytical methods pose challenges for understanding the interactions and functions of proteins at single-cell resolution.

These challenges are being addressed by emerging technologies for analyzing single cells by MS without the use of antibodies, such as Single Cell ProtEomics by MS (SCoPE-MS) and its second generation, SCoPE2. These methods allow the quantification of thousands of proteins across hundreds of single-cell samples (*9, 10*) (see the figure). A key driver of this progress was the development of multiplexed experimental designs in which proteins from single cells and from the total cell lysate of a small group of cells (called carrier proteins) are barcoded and then combined (*9, 10*). With this design, the carrier proteins reduce the loss of proteins from single cells adhering to equipment surfaces while simultaneously enhancing peptide identification.

Other key drivers of progress include methods for clean and automated sample preparation, for which there is preliminary evidence (*11*), as well as rigorous computational approaches that incorporate additional peptide features, such as retention time, to determine peptide sequences from limited sample quantities (*12*). Further technological developments can increase the accuracy of quantification and numbers of analyzed cells by 100- to 1000-fold while affording quantification of protein modi-

fications at single-cell resolution (*7*). For example, the carrier protein approach (*9*) can be extended to quantify PTMs by using a carrier composed of peptides with PTMs while avoiding the need to enrich modified proteins from single cells and, thus, enrichment-associated protein losses.

Although current methods can quantify proteins present at ~50,000 copies per cell (which is the median protein abundance in a typical human cell), increased efficiency of peptide delivery to MS analyzers, e.g., by increasing the time over which peptide ions (proteins are fragmented into peptides and ionized in MS analysis) are sampled (*7, 13*), will increase sensitivity to proteins present at only 1000 copies per cell. In general, the emerging technologies offer a trade-off between quantifying low-abundance proteins with increased accuracy or quantifying more proteins. This trade-off can be mitigated by simultaneously sampling multiple peptides (*7*). Over the next few years, improvements in sample preparation, peptide separation and ionization, and instrumentation are likely to afford quantification of more than 5000 proteins across thousands of single cells, while targeted approaches are poised to enable analysis of even low-abundance proteins of interest (*7*).

MS methods have the potential to measure not merely the abundance and PTMs of proteins in single cells, but also their complexes and subcellular localization. When proteins form a complex, polypeptide chains from different proteins can get close enough to be cross-linked by small molecules. Because only proteins in the complex are likely to be cross-linked, the abundance of such peptides can report directly on complex formation and composition. Some cross-linked peptide pairs are observed only with specific complex conformations, and thus these pairs can be useful in distinguishing active and inactive complexes. Furthermore, if a protein complex is close to organelles, targeted MS analysis of cross-linked peptides between the complex and organelle-specific proteins may report on the subcellular localization. Such analysis has not yet been applied to single-cell MS, but is likely to be feasible.

Realizing these exciting prospects requires concerted effort and community standards devoted to ensuring that hype does not overshadow scientific rigor. For

Department of Bioengineering and Barnett Institute, Northeastern University, Boston, MA, USA.
Email: nslavov@neu.edu, nslavov@alum.mit.edu

example, systematic artifacts, such as contaminant proteins introduced to single-cell samples during their preparation or chromatographic separation, may result in reproducible measurements. Despite their reproducibility, such measurements do not reflect protein abundances in single cells. If reproducibility is misinterpreted as accuracy, the resulting errors may erode the credibility of this emerging field.

Single-cell proteomics will find many applications in biomedical research. Some applications, such as classifying cell states and cell types, overlap with those of single-cell RNA sequencing. Other applications can only be achieved by measuring proteins. For example, the development of cells for regenerative therapies through the rational engineering of directed differentiation

processes that lead to the desired cell types and then simulating them by using agonists and/or antagonists. Whereas single-cell RNA sequencing can identify the cells of interest, the amounts of messenger RNA are poor surrogates for the signaling activities mediated by protein modifications, such as phosphorylation or protein cleavage (2, 15). Single-cell proteomics could provide a robust means to characterize such signaling dynamics.

Another potential application is the identification of the sets of molecular interactions leading from a genotype or a stimulus to a phenotype of interest. This goal presents a substantial challenge in part because interacting molecules within a pathway are rarely measured across a large range of phenotypic states to constrain cellular network models. This limitation is particularly evi-
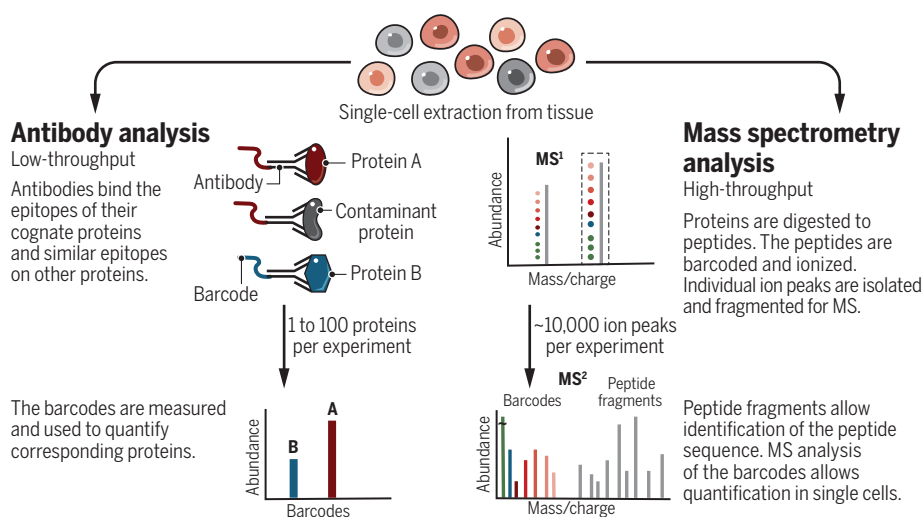
Next-generation single-cell protein analytical technologies will reduce these assumptions and thus increase the validity of inferred mechanisms. If proteins, RNAs, DNA, and metabolites are measured across tens of thousands of individual cells, it may be possible to identify direct molecular interactions without the need to make assumptions about basic aspects of the pathway. Next-generation single-cell analysis is poised to generate just this type of data, which should underpin systems-level understanding of intracellular and extracellular regulatory mechanisms.

Single-cell proteomics may also have clinical applications. Protein measurements from limited clinical samples are attractive because they afford direct measurements of deregulated signaling pathways that drive disease. Furthermore, measuring protein concentrations allows the development of assays to test therapies that induce protein degradation, which are among the most rapidly growing therapeutic modalities (15). Additionally, protein assays may be more robust than RNA-sequencing assays because protein concentrations are less noisy and proteins degrade more slowly than RNAs. Moreover, the cost of protein analysis will decrease proportionately with increased multiplexing (7, 11).

The latest generation of nucleic acid–based single-cell analytical methods has opened the door to describing the varied and complex constellation of cell states that exist within tissue. The next generation of proteomics-based methods will complement current methods while shifting the emphasis from description toward functional characterization of these cell states. ∎

Single-cell protein analysis

Traditional methods identify and quantify a limited number of proteins based on antibodies barcoded with DNA sequences, fluorophores, or transition metals. Emerging single-cell mass-spectrometry (MS) methods will allow high-throughput analysis of proteins and their posttranslational modifications, interactions, and degradation.

Single-cell extraction from tissue

Antibody analysis
Low-throughput

Antibodies bind the epitopes of their cognate proteins and similar epitopes on other proteins.

Antibody — Protein A
— Contaminant protein
Barcode — Protein B

1 to 100 proteins per experiment

The barcodes are measured and used to quantify corresponding proteins.

Abundance / Barcodes

Mass spectrometry analysis
High-throughput

Proteins are digested to peptides. The peptides are barcoded and ionized. Individual ion peaks are isolated and fragmented for MS.

MS¹
Abundance / Mass/charge

~10,000 ion peaks per experiment

MS² Barcodes / Peptide fragments
Abundance / Mass/charge

Peptide fragments allow identification of the peptide sequence. MS analysis of the barcodes allows quantification in single cells.

may benefit from single-cell proteomics. Although there has been much progress in developing directed differentiation protocols for certain cell types, these efforts tend to rely on trial-and-error approaches (14). Many of the resulting protocols remain relatively inefficient: Only a fraction of the cells differentiate into the desired cell type, and such cells may not fully recapitulate the desired physiological phenotypes (14).

Next-generation single-cell proteomics analysis offers an alternative to this trial-and-error approach. If the signaling events (usually mediated by protein interactions and PTMs) that guide cell differentiation during normal development can be identified, it should be possible to recapitulate such signaling in induced pluripotent stem cells. This would require identifying the signaling

dent for proteins and their PTMs (1–3). Yet, proteins are key regulators in cells; models that ignore them cannot capture molecular mechanisms involving protein interactions. For example, the absence of direct protein measurements compromises the ability to study signaling networks because most of the key regulatory variables are missing from the data. Currently, when proteins and their PTMs are measured in bulk tissues, they have been analyzed in a few tens to a few hundreds of samples (2, 3). Analyzing so few samples tends to require assumptions about the specific sets of interactions and functional dependencies that occur between interacting proteins and molecules. Such assumptions fundamentally underpin the inferred biological mechanisms and undermine their validity (3).

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

Science

Unpicking the proteome in single cells

Nikolai Slavov

ARTICLE TOOLS	http://science.sciencemag.org/content/367/6477/512
REFERENCES	This article cites 14 articles, 2 of which you can access for free http://science.sciencemag.org/content/367/6477/512#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service