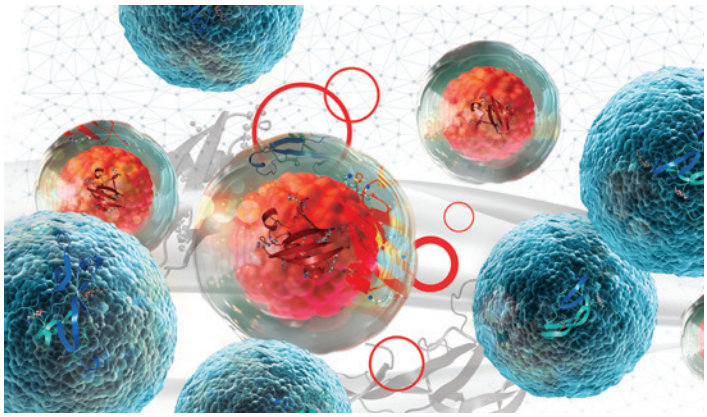
WHITE PAPER 65730

Challenges and emerging directions in single-cell proteomics

Will It go mainstream like genomics?

The Human Genome Project was a landmark achievement and signaled the beginning of a new way to approach the understanding of biology. Until then, biologists explained the secrets of life using a reductionist approach where a given biological system was decomposed into its parts and those parts were connected back to explain the chemical basis of the different molecular processes. The arrival of the -omics technologies, together with the growth of bioinformatics and computational biology, enables biologists to produce diverse and abundant data with adequate precision and accuracy to explain them using the laws of physics, chemistry and mathematics as a common language. Altogether, this development allows biologists to study biological systems in a holistic manner in which networks of genes, proteins or metabolites interact synergistically with each other, and the emergent properties of those interactions can be explained and predicted.



Introduction

Impact of omics sciences in biology and medicine

-Omics technologies represent a critical breakthrough in biology that marks the end of the 20th century and the beginning of the 21st. Researchers have moved from studying individual genes, proteins and metabolites toward studying entire populations of these molecules designated with the suffix “ome,” such as the genome, proteome, and so on. Since their inception in the early ’90s, these approaches have become entire fields of their own, each with different technologies and methodologies that have enabled big transitions in the way biological systems are investigated. The most famous accomplishment of -omics remains the Human Genome Project,[1] a multi-pronged effort to catalogue all of humankind’s genes using the then-emergent tools of genomics. Proteomics achieved similar visibility when two groups published drafts of the human proteome, a catalogue of all the proteins detectable in human cell lines, tissues and body fluids.[2,3]

Since then, -omics fields have transitioned from enabler technologies for basic research to more industrialized approaches, taking on endeavors such as tailoring medical treatments to each patient according to their individual molecular profile, an approach called personalized medicine. Genomic, transcriptomic, proteomic and metabolomic data create the possibility of deep and

expansive data sets about individual people's biology, which guide physicians to tailor treatments to specific bodies to facilitate recovery from diseases. Personalized medicine has proven especially effective in oncology, as it is increasingly clear that every cancer is its own disease with its own vulnerabilities and patterns.[4] Other conditions that benefit from personalized medicine include autoimmune disorders[5] and mental health,[6] both of which often involve long "trial and error" periods when these tools are not available to guide health care practitioners. -Omics sciences provide the possibility of both early diagnosis and precision treatment, leading to better outcomes[7] or even better, the creation of profiles of wellness with personalized details. However, the most important outcome will be enabling the study and understanding of life using engineering design principles that explain why biological systems are built the way they are.

Why single-cell analysis?

When talking about advances in the -omics fields, it's important to keep in mind that as analytical techniques become more sensitive, accurate and precise, the methodology used to measure biomolecules becomes less challenging and enables us to interrogate biological systems with much more detail. Biology is a science of complex systems; biological systems must be modular from a functional and a physical perspective, but the modularity doesn't prevent the different parts of the system from communicating with each other. Human body tissues and other study systems consist of numerous types of cells that are often at different stages of differentiation. Bulk-sample studies may provide detailed information about the average state of a study system, but they will miss the emergent properties derived from fine-scale differences, which can only be observed by looking into the state of individual cells.[8] Bulk -omics studies obscure variability between cells and provide a misleading sense of uniformity in their study systems. In highly heterogeneous systems, such as brains, such studies can produce average readings that are not representative of any single cell found within the system, misdirecting research conclusions.[9] Discovering microbes that cannot be cultured with current cell-culture methods, an important frontier for cataloguing the world's microbial biodiversity, depends critically on being able to distinguish individual cells from one another.[10] Even within seemingly homogeneous populations, hidden cell-to-cell variations can create differences that are impossible to fully investigate using bulk methods, and many biochemical phenomena occur at the cellular level and so are difficult or impossible to examine in bulk tissues.[11] Therefore, advances in the methodology will greatly help as we interrogate the granularity of biological systems.

Single-cell -omics techniques also create the possibility of fine-grained time and space series in -omics measurements. Cells in multicellular organisms exist on a continuum from their progenitor stem cells to end-of-life cells that will soon die and be recycled, and each point in this lifecycle comes with different molecular activity. Bulk tissues cannot provide insight into these differences because they flatten multiple cell types and stages into a single measurement. This is important for rapidly renewing tissues such as skin, blood and digestive epithelium, whose states are short-lived and thus require special care to isolate. Isolated cells also create the possibility of thoroughly describing individual cell–cell interactions, providing a more complete understanding of an organ's activities than bulk tissues can. At the single-cell level, categories that seem obvious in bulk tissues become porous or irrelevant, and this insight can rewrite the whole story of molecular biology.[12]

All of this is especially important in oncology research. Tumors are highly heterogeneous in composition, as the loss of DNA repair mechanisms leads to the proliferation of more and more differences between tumor cell generations. A bulk biopsy measurement provides a useful average view of a tumor at the moment it is taken but cannot account for this heterogeneity.[13] Additionally, advanced tumors discharge small populations of circulating tumor cells into the bloodstream, which are the seeds for future metastatic tumors elsewhere in the body. These seed populations are too rare to examine with -omics technologies that are not designed for use with single cells, but they provide critical insight into their parent tumor and information on how to prevent metastatic tumors from forming in that patient.[14]

Recent technological advances provide new opportunities to investigate complex biological systems at the level of single cells. Both high-throughput methods for reading many cells individually and careful isolation protocols for smaller numbers of cells enable researchers to study the genomes, transcriptomes, proteomes and metabolomes of more individual cells at more time points than ever before. Now, labs can generate massive single-cell data sets that facilitate the discoveries of new cell types, variations within cell types, and accurate measures of the rarity of these types and variations.[8] As this technology continues to

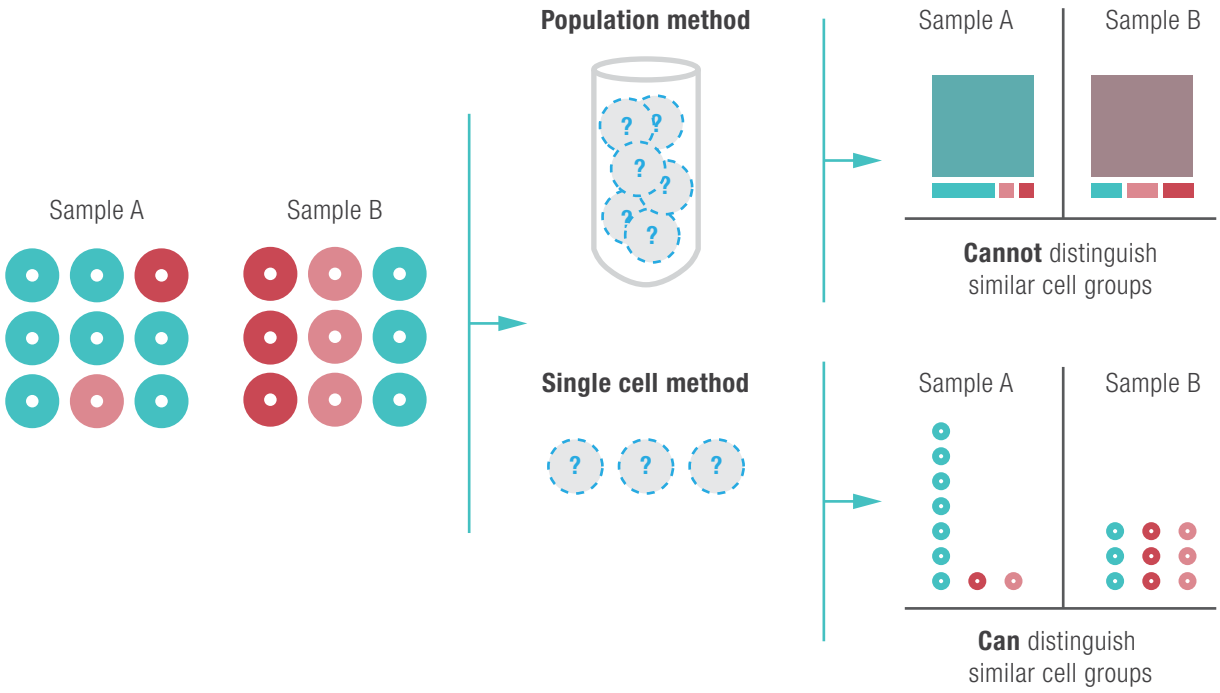


Figure 1. Population vs. single cell omics

develop, single-cell measurement possibilities will become more and more dynamic, able to show more transient cellular states and deliver more insight into how cells function from moment to moment.[11]

Current state of single-cell -omics

A variety of new technologies to extract -omic-level information from single cells have reached widespread use in recent years. These tools provide unprecedented investigative power to researchers examining cellular heterogeneity, whether at the level of DNA, RNA, proteins or metabolites. A summary of each of these technologies follows.

Single-cell genomics

Genomics is the oldest of the -omics, and the Human Genome Project of 1984–2003 is its most famous accomplishment.[1] Until recently, however, genomics was not a single-cell field. Single-cell genomics effectively began with experiments that allowed the detection of gene expression in single cells using microarrays[15] and came into its own once next-generation sequencing (NGS) entered the mainstream.[16] This method is sensitive enough to read genomes from single cells, especially with tools such as polymerase chain reaction (PCR) available to amplify the available DNA, and improvements in sensitivity and throughput continue to make this technology more accessible for new applications.[17] Most often, NGS helps determine the number of relevant single-nucleotide variants (SNVs) present in a sample, which occur at an estimated rate of approximately 1,500 per human cell and which are often associated with disease states.[18]

Single-cell DNA sequencing has been particularly useful for cancer biology. Tumors are heterogeneous tissues that arise from multiple clones and change over time as DNA repair mechanisms fail, so neither large biopsies nor study of the originating tissue can offer a complete picture of a tumor's genetics and anticipated behavior.[19] Single-cell genomics provides a powerful tool for following the progression of individual clones within a tumor and the shifts in the balance between them.[20,21] Bulk analyses are unsuited to a number of specific tasks in cancer research that are proving increasingly important, such as studying circulating tumor cells and cancer stem cells. These cells are exceedingly rare compared to ordinary tumor cells and to healthy cells, so they disappear into margins of error in bulk analyses. However, they play critical roles in tumorigenesis and metastasis and are thus important to track, understand and sequence when determining cancer prognosis and recommending treatment.[14,17] Further improvements in the sensitivity and accessibility of single-cell genomics technology will make these analyses more and more routine.

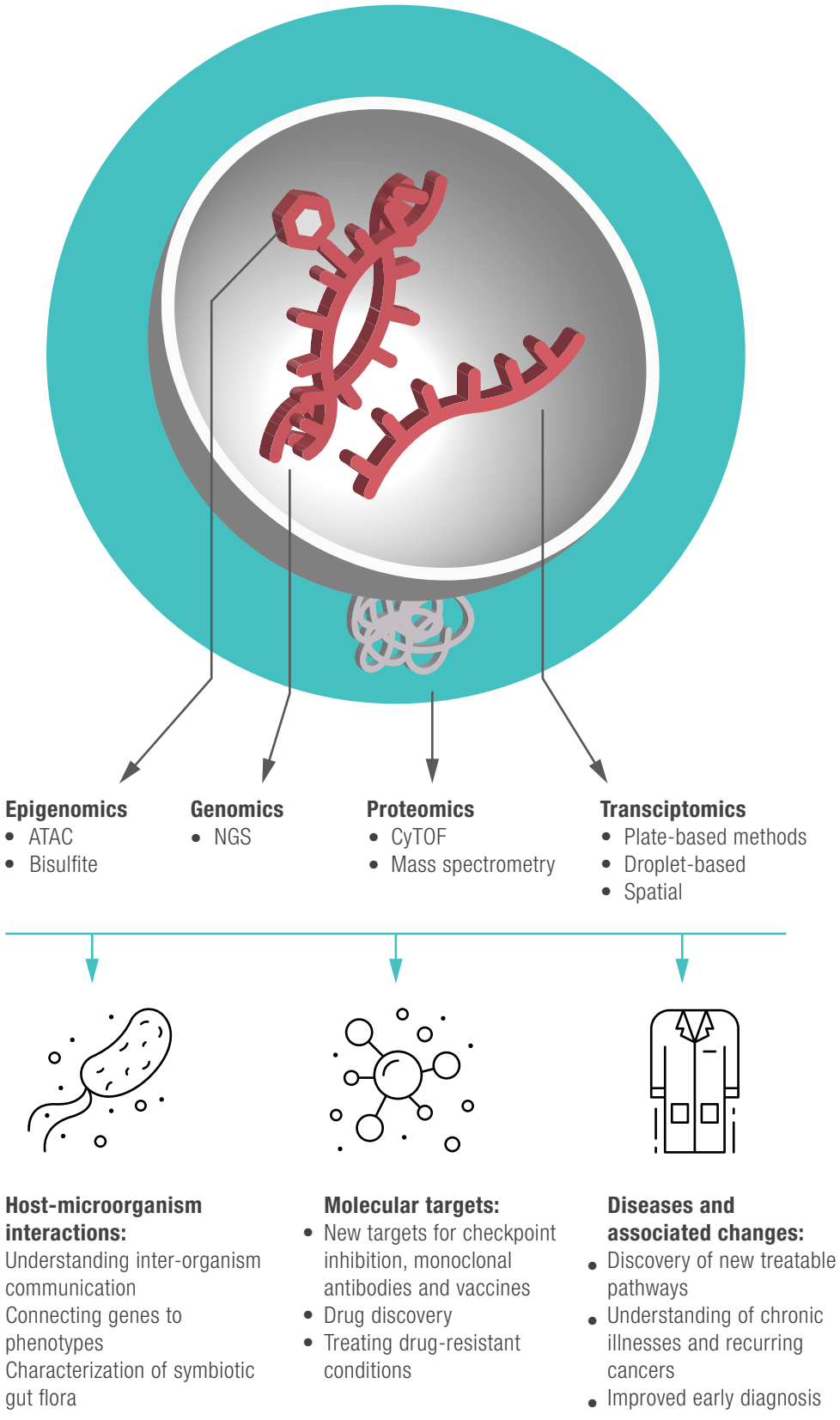


Figure 2. Single cell technologies and applications

Single-cell epigenomics

Epigenetics, referring to heritable changes to DNA that are not part of the nucleotide base sequence, entered the scientific consensus in the 1990s and has since become increasingly important to understanding a variety of heritable phenomena.[22] Epigenetic modifications regulate gene expression, and reading them provides necessary insight into the effects of a genome on an organism. In effect, epigenetics is a critical part of the story that the genome tells. These changes can occur during a cell's or organism's life without directly affecting the genome and while remaining heritable. Single-cell approaches have been extended to provide data on DNA accessibility,[23–25] methylation[26] and chromosome conformation,[27] all of which affect which genomic DNA can be expressed and at what levels. Epigenetic changes are cell-specific and difficult to observe in bulk samples, where they are only theoretically accessible if a large population shows the same modifications.

Compared to genomics and metabolomics, single-cell epigenomics remains a new and difficult field. Epigenetic signatures take many forms and their effects are difficult to determine, preventing the kind of rapid advancement that other -omics fields have enjoyed. Nevertheless, the importance of epigenetic changes to cancer[19] and embryonic development[28] ensures the growth of this field and the continuous improvement of tools for collecting epigenomic data.

Single-cell transcriptomics

The transcriptome, or sum of all of a cell's RNA transcripts, occupies an intermediate position between the genome and the proteome, showing which portions of the genome are being actively translated into proteins. It provides a dynamic picture of a cell's current functioning, rather than the more static data provided by its genome and proteome. RNA is much more fragile and transient than DNA and occurs at much lower masses in a cell, but single-cell RNA sequencing has nevertheless rapidly advanced in recent years. These technologies typically rely on converting RNA into complementary DNA and then amplifying that DNA to make it accessible to DNA sequencing tools. Like DNA sequencing, transcriptomics can be performed with specific targets in mind or in an untargeted, exploratory way, with the former method offering greater speed and the latter greater coverage.[29] Sample multiplexing allows the analysis of hundreds of cells with up to 4 million reads per cell, and droplet- or nanowell-based methods can analyze thousands of cells with reduced read numbers closer to 200,000.[29] Modern transcriptomic methods offer scientists a more accurate and complete picture of cellular activity than genomics and proteomics can offer alone, and these techniques measure thousands of cells separately to provide a picture of activity patterns in an entire cell population. Future improvements will drive down costs for more and more complete transcriptomics platforms.

Single-cell proteomics

Proteomics has been one of the most challenging of the -omics sciences to extend to single-cell applications. Unlike genomics and transcriptomics, there is no amplification process in proteomics because proteins cannot be amplified, so there is no intrinsic workaround for the minute amounts of protein available as research material in a single cell. Proteins provide necessary detail about the cell's current activity and structure that nucleic acids cannot, making proteomics and its adaptation to single-cell systems a top priority for many research laboratories. High-throughput methods like those available in genomics have not yet arrived, but improvements are ongoing.

The main methods currently available for single-cell proteomics are antibody-based, cytometry-based or mass spectrometry (MS)–based. Cytometry-based approaches depend or are based on fluorescence-activated cell sorting (FACS) and depend on antibodies to tag proteins of interest. These approaches are thus limited by available antibodies and by the ability of their device to read multiple tags. Single-cell mass spectrometry (such as CyTOF) approaches have successfully detected and identified up to 450 proteins in single oocytes, some of the largest human cells.[29,30] Improving sample preparation techniques to make sure more of every sample arrives at its analysis device and improving analysis devices so that more of every sample turns into data are the challenges ahead of single-cell proteomics.[30]

Single-cell metabolomics

Metabolites are the endpoints and intermediate steps of most biological processes, and their sum, the metabolome, represents a thorough catalogue of a cell's biochemical activities. The metabolome represents the most immediate way to identify and begin to study phenotypic differences at the cellular level, and it can provide the information required to inform proteomic and genomic studies. In synthetic biology, metabolomics provides a critical check to determine whether an alteration had the desired effect

on a cell line,[31] because the metabolome can dynamically react to the environment on a very short time scale, while changes at the protein expression level take longer and are not acute to the cell's actual behavior.

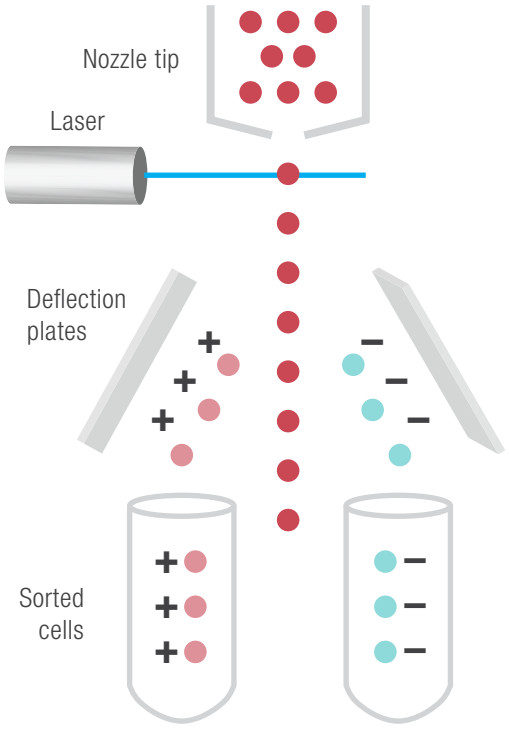


Figure 3. Fluorescence activated cell sorting

Like proteins, metabolites cannot be amplified. But unlike proteins, metabolites are highly heterogeneous, differing extensively in size, polarity, solubility and more. This combination makes single-cell metabolomics particularly difficult.[7,32] A cell's metabolome is particularly responsive to environmental cues compared to its genome, transcriptome or proteome, making metabolomes critical for studying cells as they exist in particular moments in time, and cells with identical genomes can still have different metabolomes if their circumstances differ enough.[33] This makes metabolomics an increasingly common tool for understanding the effects of both environmental changes and genomic changes on health and disease.[7,32] However, one of the major limitations is that tools used for isolating and selecting specific cells, such as flow cytometry, have been shown to change the metabolome of the cell during the sorting process and therefore bias the results.[34,35] Improvements increasing the sensitivity and speed of mass spectrometers in conjunction with new sample introduction techniques will lead to a more complete coverage of the metabolome, making single-cell metabolomics easier, more accessible and more effective.

Single-cell multiomics

No single -omics science can provide a truly complete picture of a cell's activities. The most comprehensive studies combine different -omics approaches to simultaneously collect data at different levels. This is particularly difficult in single-cell applications because of the extremely limited amount of research material available, making "multiomics" perhaps the most difficult of all the -omics approaches. Single-cell multiomics is ideal for directly linking a cell's phenotype to its genotype, and for providing the basis of later studies that could not be combined into the multiomic approach. A common single-cell multiomics approach combines genome and transcriptome sequencing, generating data from both RNA and DNA.[36] The other -omics methods are different enough from one another that combining them continues to be an elusive goal. With the increasing sensitivity of proteomic and metabolomic assays, multiomic single-cell approaches involving these methods will become more and more possible.

Challenges of single-cell -omics

Bringing single cells into a field designed for chemistry is not a simple matter. Biochemistry has advanced as far as it has in part because tissue homogenates can be treated like any other complex mixture, with organic extractions and similar tools. Conventional genomics can begin with homogenized tissue in special buffers from which DNA or RNA are precipitated, but working with single cells involves much smaller quantities of material. Single cells cannot be treated this way for two main reasons: they typically must be recovered intact to be useful, and they represent much less material than tissue homogenates.

For most single-cell -omics applications, individual cells must be laboriously isolated and retrieved intact so that no material is lost. Cell lysis greatly complicates single-cell omics measurements and must be carefully avoided, and the various preparation methods coming into prominence all work by minimizing these losses. For many applications, cells must also be viable rather than dead or frozen, which creates even more time and handling constraints. This level of sample preparation takes up a disproportionate amount of single-cell -omics workflows and represents one of the biggest challenges to more widespread adoption of single-cell -omics practices.

For MS-based proteomics and metabolomics, these barriers are particularly acute. Single-cell genomics and transcriptomics are more mature fields that have had more time to solve their practical concerns,[37] and they benefit

from PCR, for which proteomics and metabolomics have no equivalent. This tool allows researchers to analyze a tiny amount of material by amplifying it until less sensitive instruments can detect it. Single-cell proteomics and metabolomics assays must be sensitive enough to collect data from single cells without such amplification. Designing instruments and assays with sufficient sensitivity has been a practical barrier to more widespread use of MS-based single-cell proteomics, metabolomics and multiomics.

Profiling the proteome and metabolome of individual cells at the single-cell level also remains a serious challenge due to the high diversity and large dynamic range of the cellular proteome and metabolome. Proteins and especially metabolites are much more different from one another than nucleic acids are, including in their size, charge state and three-dimensional complexity, and fully cataloguing them requires a more complex array of extraction and assay conditions than in equivalent genomics or proteomics applications. Proteins are also "sticky," making them difficult to deliver from sample preparation to MS, and the chemicals used to digest or isolate them before reading should be meticulously removed (along with a noteworthy fraction of the proteins themselves) before the analysis to make the samples MS-compatible. Single-cell metabolomics is an even more daunting challenge, because cells are dynamic and need to be captured or quenched and kept in that native state for as long as possible to get measurements of the specific time point of interest, without disturbing that state. Cell sorting techniques like FACS have been shown to change the state of the cells.[35]

Advances for both single-cell proteomics and metabolomics are consistently based on maximizing the efficiency of the tools involved, both by reducing sample loss and by designing more sensitive MS systems that can detect these molecules from ever-smaller amounts of materials. These improvements involve increasing absolute sensitivity, dynamic range and multiplexing capacity, and result in improved throughput as well as improved data quality.

These challenges have limited the reach of both single-cell proteomics and single-cell metabolomics. As a field that is both difficult and more expensive compared to genomics, it is for now limited to relatively well-funded and well-equipped laboratories. But the necessary technologies become less expensive and more accessible every year, bringing this field into wider and wider use. A tremendous level of insight remains to be acquired through the widespread adoption of single-cell -omics technologies, and especially single-cell proteomics.

Why single-cell protein/proteomics analysis

Single-cell proteomics is a nascent field that is already delivering change. Until now, protein levels in single cells often had to be inferred from bulk samples or from cellular mRNA levels, but proteomic methods allow such data to be directly tested, removing layers of abstraction and enabling deeper studies. There is great potential in the future as science moves past simple profiling and abundance measurements to dynamic examinations of cells as systems that change through time. As single-cell proteomics continues to grow, it will enable the direct study of protein interactions and modifications such as phosphorylation, providing a more complete understanding of cellular activities.

Significant innovations in the protein-level analysis of single cells have emerged in recent years. Most of these new methods are based on antibodies and/or fluorescence, and they work by tagging proteins of interest with detectable antibodies or other sorts of markers. However, this strategy limits their ability to generate global, untargeted insights about a cell's protein environment. These approaches usually face multiplexing limits, as it is rare that more than 100 proteins can be studied simultaneously, and their quantification accuracy is limited by the available antibodies. These methods are useful, but a truly global approach requires MS-based methods. A selection of antibody- and fluorescence-based methods is presented here.

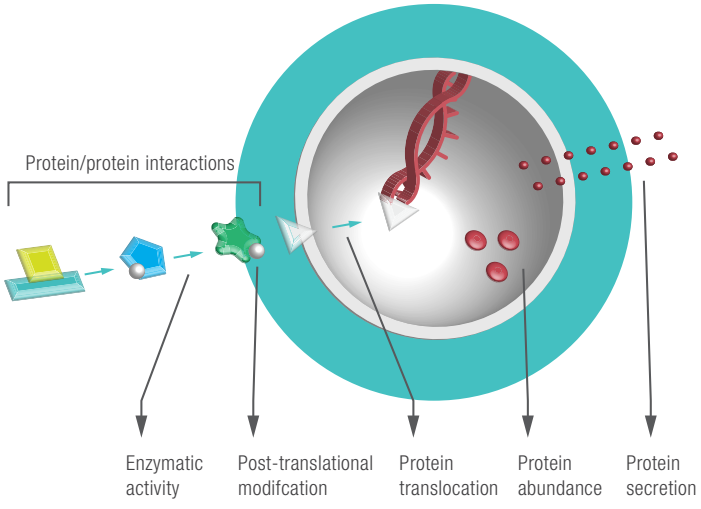


Figure 4. Protein analysis of single cells can move beyond simple profiling and abundance changes

Single-cell western blotting

Western blotting was one of the first technologies for observing and isolating proteins, and innovations in polyacrylamide gel technology have brought it into the single-cell age. Functional proteomic studies of thousands of single cells can be achieved on a single microscope slide using single-cell western blots. Hughes et al. conducted approximately 10^3 concurrent single-cell western blots using a microscope slide with photoactive polyacrylamide gel with single-cell microwells and in-situ lysis.[38] This four-hour experiment monitored the differentiation of single rat neural stem cells and their response to mitogen stimulation using 11 multiplexed protein targets. Detection thresholds were as low as <30,000 molecules and, with integrated fluorescence-activated cell sorting, starting cell numbers as low as 200 could be analyzed. Western blotting offers higher protein specificity compared to pure antibody-based assays, as the method reports both the target molecular mass and probe binding.

Although this is the state-of-the-art level for this technology and it offers the convenience of being an at-the-bench operation, western blotting is still quite limited compared to other proteomics methods. The fluorescence signals used in western blotting are noisy and diffuse on a surface as large as a polyacrylamide gel, and accurate quantification (as opposed to identification) is difficult with this method. Like other antibody-based methods, western blotting has a very limited ability to yield data on unknown proteins and it is best suited to studies with pre-identified targets for which antibodies already exist.

Single-cell flow cytometry

The most established method for single-cell protein analysis is flow cytometry, which was invented in the 1960s. Its effectiveness derives from the fact that, although the actual protein amounts in single cells are exceedingly small, they can be very concentrated. When the cells are kept intact throughout measurement, as in flow cytometry, these high concentrations become measurable via fluorescent antibody-based tags. At first, flow cytometry was limited to measuring one or two fluorescent species at a time, but modern versions can measure up to 15, allowing the profiling of entire pathways.[39,40] The ability to perform correlated measurements of multiple proteins in single cells has allowed flow cytometry to become a powerful tool for quantitatively analyzing pathways and understanding diseases associated with them.[41,42] Improvements in both instrumentation and the availability of highly specific antibodies has brought flow cytometry this far, and the advent of barcoding methodology, improved tags and dyes, and microfluidic technologies for sample handling will continue to improve this technology, keeping it relevant for future studies.

Mass cytometry

Connecting a flow cytometer machine to a mass spectrometer rather than the usual fluorometer[43] created mass cytometry. The most relevant instrumentation for mass cytometry is CyTOF, cytometry by time of flight, which uses inductively coupled plasma ionization to get the ions into the gas phase and inside the mass spectrometer. The ions' time of flight is used to detect and distinguish them from one another. Like other flow cytometry workflows, CyTOF uses antibodies to tag proteins of interest in cells. These antibodies are conjugated with metal isotopes, which have the same elemental composition but different masses. These metal tags can be detected and quantified by the mass spectrometer to decode which protein was quantified for each signal. This concept is very similar to conventional flow cytometry tags that are quantified when passed through a fluorometer. Palii et al. used this approach to study the temporal dynamics of transcription factors during human hematopoiesis over multiple time points.[44] They measured 27 proteins simultaneously and found that quantitative changes in lineage-specific transcription factor abundance can determine cell fates. These findings are far from the limits of this technology, given that the precision of MS removes the issue of spectral overlap complicating fluorescence measurements in conventional flow cytometry. Like all antibody-based methods, CyTOF's reliance on flagging known protein targets prevents this method from being easily applied to exploratory studies in which discovering unknown proteins is important.

Reverse transcription and proximity extension assays

Antibody-based methods are extremely sensitive and powerful to detect and quantify their targets. However, one of the major limitations for protein multiplex assays is the spectral overlap among the different targets, as well as the types of protein targets that can be assayed, which are often extracellular proteins. To overcome this limitation, nucleic acid proximity-based methods with dual reporters have been shown to be particularly useful.[45] These methods are based on an assay that uses pairs of antibodies equipped with DNA reporter molecules. When these antibodies bind to their targets, they create DNA amplicons that bar code those targets. These amplicons can then be quantified with real-time

quantitative PCR, combining the data of proteomics with the precision, sensitivity and speed of the genomics tools. By combining genetic material with protein analysis, these proximity extension assays allow the well-developed tools of genomics to become part of proteomics. Using DNA tags means that the assays are limited by the multiplexing capacities of quantitative PCR devices rather than those of fluorometers, enabling much greater multiplexing capacity than many other methods. These methods can even be combined to measure protein and RNA targets in the same sample at pictogram-per-milliliter resolution. The fact that proximity assays can also simultaneously detect RNA and protein expression at the single-cell level enables the identification of biologically meaningful differences between cells and their molecular markers, and generates insights into the processes of protein synthesis that may yield the drivers to understand cellular heterogeneity. This is important, as more and more research shows that mRNA is an imperfect proxy at best for protein levels and cannot be relied upon to provide solid insights into how much protein is being produced.[46]

Global single-cell proteomics by mass spectrometry

Mass spectrometry is the gold-standard technology for proteomics and remains a strong choice in single-cell applications where the goal is to study global protein profiles. With MS, the goal of analyzing proteins at scale with sensitivities down to the low nanogram level is an everyday task. These technologies have been built on decades of work with faster and more sensitive chromatographic separations and higher-resolution MS systems. To address the most complex proteomics questions, specially designed workflows include reagents such as tandem mass tagging (TMT), and special software has been developed to process this data. Adapting an MS workflow for single-cell proteomics requires extremely high sensitivity, minimal sample loss, and the ability to multiplex both for throughput and sensitivity gains.

Sample handling and preparation

Before any kind of single-cell proteomics measurement with MS, the cells must be isolated. A variety of techniques are available for cell separation,[47] including immunomagnetic cell sorting,[48] fluorescence-activated cell sorting (FACS),[49] density gradient centrifugation[50] and microfluidic cell sorting.[51] These methods are not covered in this white paper, which instead focuses on later steps in the sample preparation process.

In single-cell measurements, maintaining precision and care in sample handling is critical. Small losses of analytes that would be negligible in bulk tissue or cell population-based methods may lead to dramatic fluctuations in a single cell, overwhelming the sample signal. Single-cell data also require *many* single cells to overcome naturally occurring noise and to improve data quality, much as any other experiment benefits from increasing sample size. The extreme nature of single-cell systems means that simply scaling down methods designed for larger samples is not sufficient, and sample handling methods specific to single-cell systems must be devised. Microfluidics represents a particularly promising source of these innovations, enabling cell isolation, lysing, culturing and transporting for large numbers of individual cells without losing analytes.[52] Microfluidic sample preparation techniques come with significant improvements in throughput performance, cost-effectiveness, workflow complexity and assay consistency compared to alternatives.

The proteomics community has not yet agreed upon one single method as the best suited for single-cell analysis. Nevertheless, a few groups have successfully demonstrated single-cell proteomics analysis. This section describes their methods.

Nanodroplet processing in one pot for trace samples (NanoPOTS)

Proteomics sample preparation typically includes protein extraction, proteolytic digestion, cleanup and delivery to the analytical platform. As sample amounts decrease without a concomitant reduction in reaction volume (often limited by evaporation and the ~microliter volumes addressable by pipette), the nonspecific adsorption of proteins and peptides to the surfaces of reaction vessels, along with inefficient digestion kinetics, become increasingly problematic. Efforts to improve this aspect of sample preparation have included the use of low-binding sample tubes and the advent of “one-pot” digestion protocols that reduce losses by removing the need to move the sample through multiple vessels. Nanodroplet processing in one pot for trace samples (NanoPOTS) is one such protocol.

NanoPOTS addresses the issues of miniaturizing protein digestion and cleanup by reducing the processing volume to less than 200 nL, which significantly accelerates reaction kinetics. However, by reducing the volume more than 200 times compared with conventional methods, it significantly reduces sample losses due to nonspecific adsorption of the proteins to surfaces. The method consists of the

use of a liquid handler capable of dispensing nanoliter volumes into wells etched in a glass slide with a volume of 200 nL each. The system is typically integrated into a flow cytometry system or to a laser-capture microdissection system. Cells are sorted and deposited into the wells, and then the sample is prepared by adding all necessary reagents. Because of its architecture, this system allows multiple digestion and extraction steps to take place without changing containers. The digested peptides are then retrieved and delivered to the mass spectrometer via glass capillary tube or are directly placed into an autosampler plate.[53] When combined with ultrasensitive liquid chromatography-MS, nanoPOTS allows the identification of ~1,500 to ~3,000 proteins from ~10 to ~140 cells, respectively, with efforts to decrease that number even further.[54] The team has also demonstrated the method's compatibility with tandem mass tags (TMTs), which allows for the analysis of several cells at a time when all the proteins for each given cell have been labeled with a specific mass tag.

The capability of nanoPOTS to be combined with other cell isolation techniques and its high sensitivity mean it promises to be one of the most important technologies in this field.

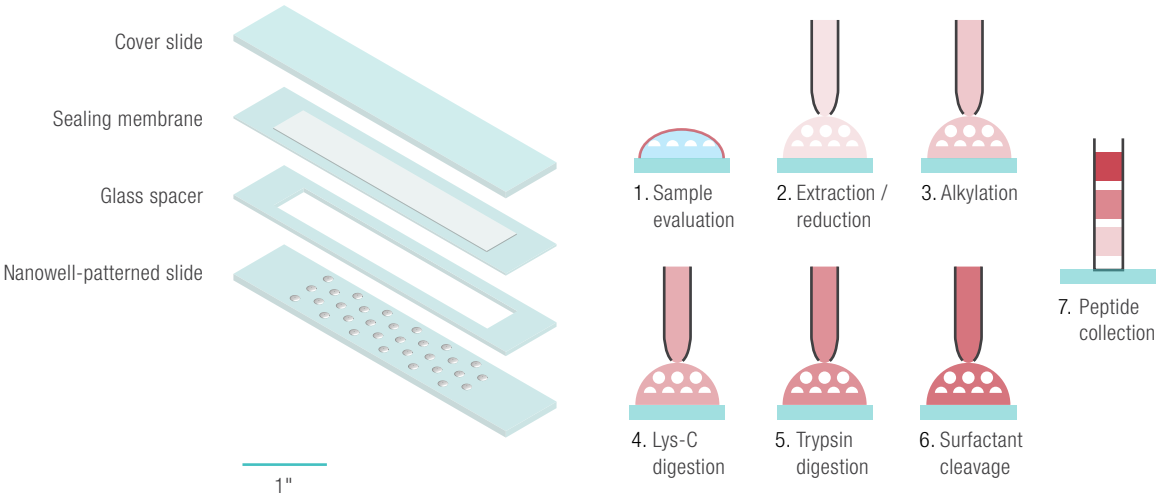


Figure 5. Nanodroplet processing for proteomics applications

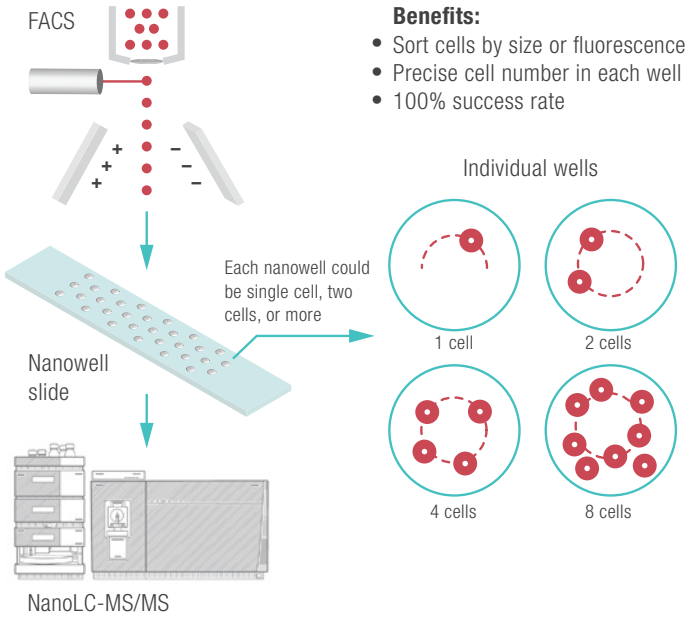


Figure 6. Single cell proteomics by mass spectrometry

Single-cell proteomics by mass spectrometry (SCoPE2-MS)

SCoPE-MS is a mass spectrometry workflow optimized for single-cell proteomics developed by Budnik and Slavov, currently in its second version (hence, SCoPE2).[55] It is designed to address two major issues with conventional MS approaches when applied to single cells: minimizing losses during sample preparation and achieving the simultaneous identification and quantification of peptides from multiple samples.[56] In Budnik and Slavov's test, SCoPE-MS enabled the quantification of over 2,000 proteins in 356 single monocytes and macrophages in about 85 hours of instrument time, and the quantified proteins were used to discern single cells by cell type. With such an abundance of highly precise and complete data, they were able to analyze the emergence of cellular heterogeneity as homogeneous monocytes differentiated into macrophage-like cells in the absence of polarizing cytokines. This workflow shows great promise and future

developments will increase its throughput, speed and ease of use, eventually enabling it to quantify and identify thousands of proteins and peptides in single cells.

SCoPE-MS features several key innovations over other MS approaches. To minimize losses, live cells are lysed via sonication or freeze-thaw cycles rather than using chemical detergents, which are generally incompatible with MS measurements. Since these chemicals are not used, they do not then need to be cleaned out of the sample, which removes the danger of sample losses during cleaning steps.

To aid with simultaneous identification and quantification and signal enhancement, SCoPE-MS uses tandem mass tags (TMTs). These isobaric labeling reagents allow for the quantification of each tagged peptide and connect them across samples, providing enough material to generate a complete sequence when all of the tagged peptides are pooled together. The SCoPE-MS method improves identification by also including with each single-cell set a sample composed of more than one cell, typically between 10 to 200 cells. This sample is what scientists have named the boost sample or the boost channel, because it includes sufficient peptide ions to provide enough signal to yield a peptide sequence identification from the mass spectrum without the sensitivity limitations of single-cell samples. Meanwhile, the TMT provides the precision required for quantitative analysis of the identified peptides.

Boost channels for throughput, increased sensitivity and quantification

Building on Budnik et al.'s success,[55] Maowei Dou and his team combined nanoPOTS sample preparation with TMT to improve both proteomic sample processing efficiency and analysis throughput for single cells. Their boost-channel experiment achieved multiplex analysis of single-cell–level protein quantities to a depth of 1,600 proteins with a median CV of 10.9% and a correlation coefficient of 0.98.[57] They also measured protein expression in 72 single murine epithelial, immune and endothelial cells. In this study, they were able to identify 2,300 proteins with less than two days of instrument time. Erwin Schoof's team used a similar approach to derive quantitative information about 10 single cells per MS injection. Schoof's team studied a leukemia culture system containing functionally defined leukemic stem cells, progenitors and terminally differentiated cells, and the boost-channel approach helped them gain information about this aberrant developmental hierarchy.[58] The new TMT 16-plex reagents currently on the market and further improvements in platform automation will continue to improve the throughput of these approaches.

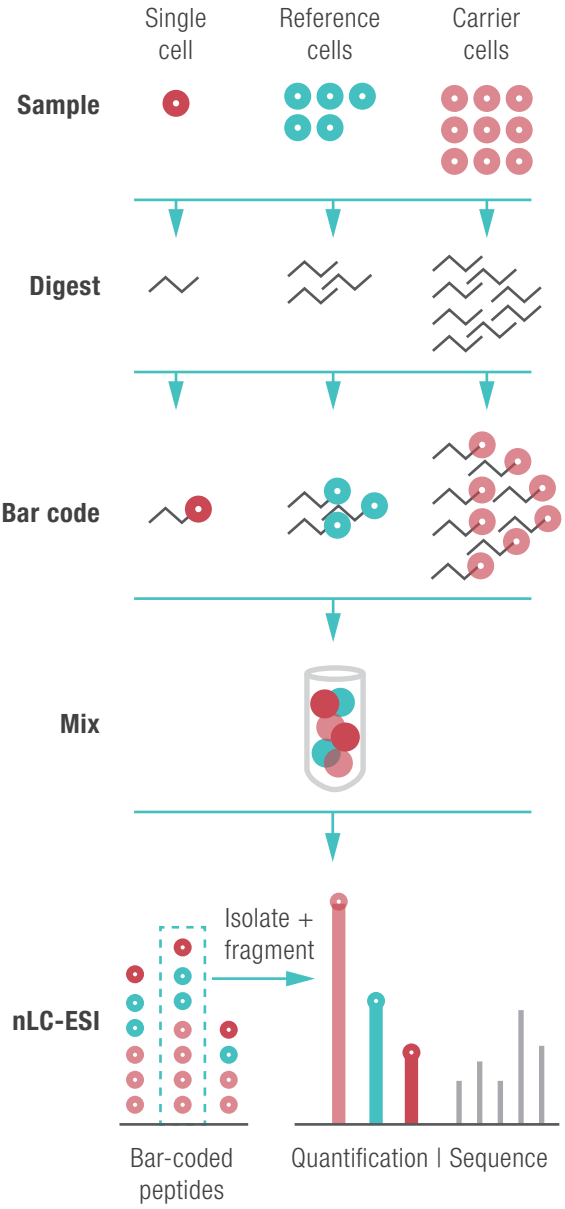


Figure 7. SCoPE2 workflow

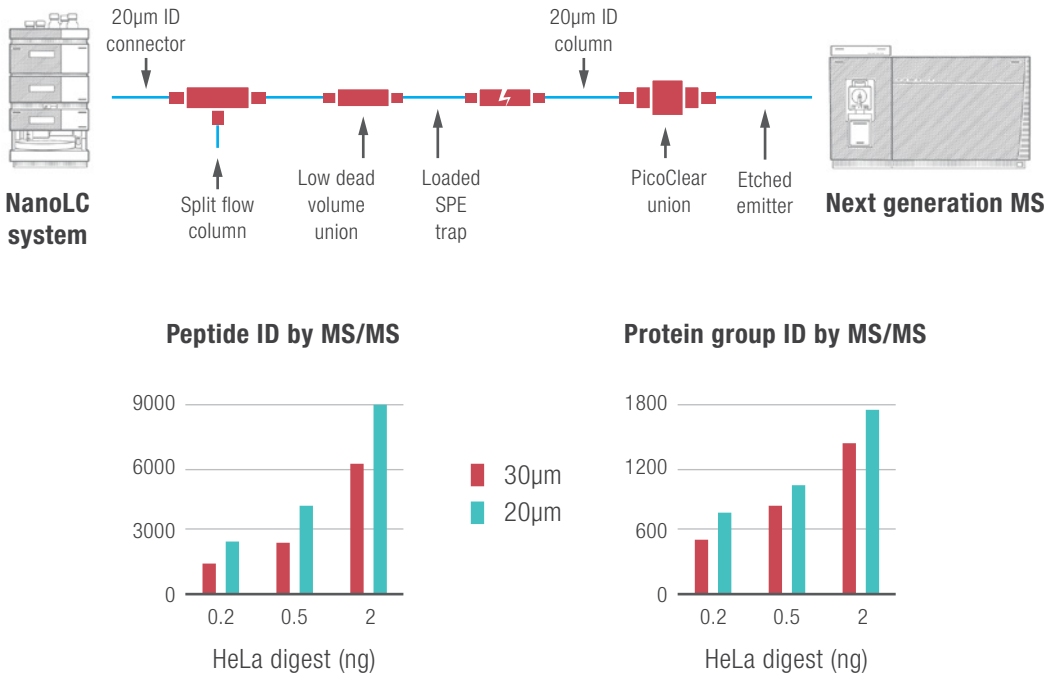


Figure 8. Next generation ultra high sensitivity LC/MS platform for single-cell proteomics

Liquid chromatography separations for low-level samples

High-performance liquid chromatography (HPLC) has a long history of use as part of MS workflows, to the point that they are often combined as LC-MS systems. Modern proteomics separations employ low-flow or nanoflow HPLC with small internal diameter (ID) chromatography columns coupled to electrospray ionization (ESI) to get the peptides into the gas phase and ready to be analyzed by the MS. The ESI process is critical because the signal in the mass spectrometer depends on how well the setup is able to transition the peptides from the liquid phase into the gas phase. To minimize losses, it's important to increase the analyte concentration and reduce the size of the droplet during the ESI process. This is why it is key to reduce the chromatography to low nL/min flow rates and to reduce the internal diameter of the chromatography columns to maximize peak capacity. These settings enable the generation of high-quality data from complex, tiny samples such as single cells.

Our team recently demonstrated that switching to 30-µm-ID nanoLC columns rather than the conventional 75-µm-ID columns can substantially enhance sensitivity due to increased ionization efficiency at the nanoelectrospray ion source and increased concentration of each component eluting from the narrow-bore columns. ESI emitter technology that accommodates the resulting lower flow rates could be employed to improve the detection sensitivity of the LC-MS system. However, practical issues still remain, such as the challenge in interfacing single-cell samples with the MS instrument.

Gas phase separations

One of the major gaps in single-cell MS has always been the lack of sensitivity. The technologies described above have been focused on how to reduce losses and how to get the most signal from the sample. However, in these types of experiments, one of the major limitations is the trace chemical contaminants intrinsic to the process or to the sample. These chemical contaminants tend to be small molecules that compete for ionization and signal during the MS analysis. Recent studies show that the most recent generation of Thermo Scientific™ Orbitrap™ mass analzyer-based mass spectrometers are capable of analyzing single cells despite this hazard.[30,44] Further enhancements on proteome coverage are mandatory for the field to become more useful. New technologies such as FAIMS (field asymmetric ion mobility spectrometry) can be used to remove impurities, in this case +1 ions that are not peptides, increasing the signal-to-noise ratio and improving the sensitivity of all measurements significantly.

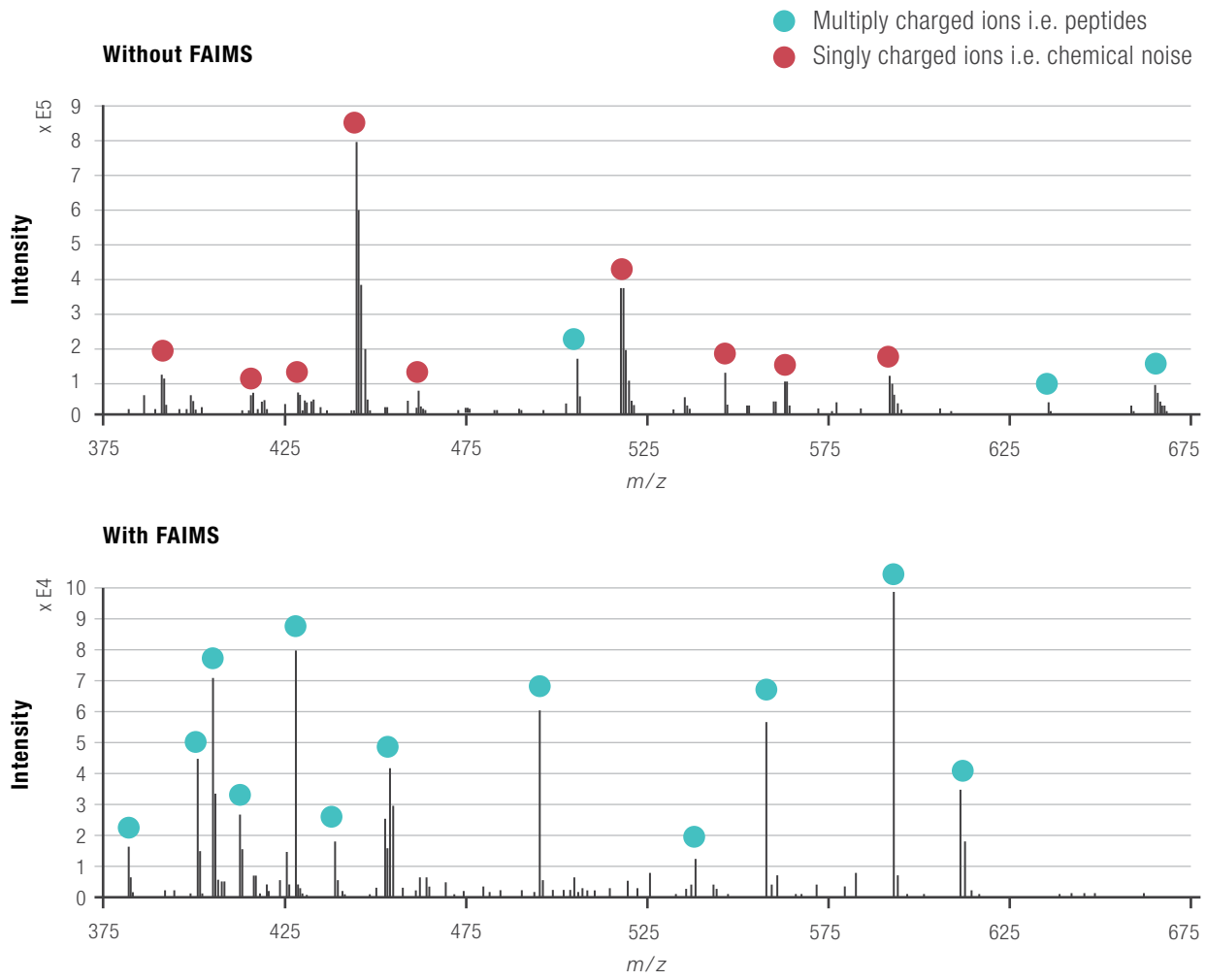


Figure 9. Effect of gas phase separations to remove singly charged ions (interferences) to maximize detection of multiply charged peptides

Future outlook

Single-cell proteomics is increasingly necessary to answer the most pressing questions in numerous fields of biology. As it becomes more accessible to more researchers, its reach can only grow. Challenges remain before single-cell methods can become as ubiquitous as mass spectrometry or flow cytometry have already become. Leading researchers in single-cell proteomics are all striving to make these techniques accessible to the community by developing experimental protocols for all steps of the workflow: cell sorting, sample preparation, LC-MS and data analysis.

For single-cell proteomics to be a viable alternative to single-cell RNA sequencing, it needs to match the former for throughput capacity, cover the same order of magnitude in terms of the number of unique proteins detected and identified, and be easily implementable as part of a wide range of cellular assays. The new technologies and methods redefining single-cell proteomics possibilities are all working to fill these gaps. Throughput improvements are being achieved via the automation of cell sorting processes, parallelization, multiplexing with TMT reagents, and fast, robust LC-MS protocols. Sample losses can be minimized with procedures that allow smaller volumes during sample prep, the use of microfluidics, and the incorporation of carrier cells and other boosting practices to protect precious samples and separate identification from quantification. TMT, FAIMS, Thermo Scientific™ Tribrid™ approaches and boost channels also improve the sensitivity of analysis platforms, allowing ever more complete data to come from tiny samples. All of these improvements will continue to make single-cell proteomics more accessible to more researchers.

Intriguingly, one approach that may yet have a major impact on single-cell proteomics is not based on mass spectrometry at all. This method is called nanopore sequencing, and it is based on driving a protein or peptide molecule through a nanopore, typically using an electric field, and measuring the ionic current in the nanopore as it changes depending on the amino acid that is passing at a given time.[59] Still in its infancy, this new technology may yet become a rival for mass spectrometry–based methods for protein sequencing due to how easily it scales and its high sensitivity. Until then, mass spectrometry remains the gold standard for single-cell proteomics.

References

1. The Human Genome Project. Genome.gov. https://www.genome.gov/human-genome-project.
2. Kim, M.-S., et al. (2014). A draft map of the human proteome. Nature 509, 575–581.
3. Wilhelm, M., et al. (2014). Mass-spectrometry-based draft of the human proteome. Nature 509, 582–587.
4. Ramaswamy, S., & Perou, C.M. (2003). DNA microarrays in breast cancer: The promise of personalised medicine. Lancet Lond. Engl. 361, 1576–1577.
5. Tavakolpour, S. (2017). Towards personalized medicine for patients with autoimmune diseases: Opportunities and challenges. Immunol. Lett. 190, 130–138.
6. Fernandes, B.S., et al. (2017). The new field of 'precision psychiatry'. BMC Med. 15, 80.
7. Putri, S.P., et al. (2013). Current metabolomics: practical applications. J. Biosci. Bioeng. 115, 579–589.
8. Strzelecka, P.M., Ranzoni, A.M., & Cvejic, A. (2018). Dissecting human disease with single-cell omics: Application in model systems and in the clinic. Dis. Model. Mech. 11.
9. Altschuler, S.J., & Wu, L.F. (2010). Cellular heterogeneity: do differences make a difference? Cell 141, 559–563.
10. Nesme, J., et al. (2016). Back to the Future of Soil Metagenomics. Front. Microbiol. 7.
11. Stuart, T., & Satija, R. (2019). Integrative single-cell analysis. Nat. Rev. Genet. 20, 257–272.
12. Yuan, G.-C., et al. (2017). Challenges and emerging directions in single-cell analysis. Genome Biol. 18, 84.
13. Conrads, T.P., Hood, B.L., Petricoin, E.F., Liotta, L.A., & Veenstra, T.D. (2005). Cancer proteomics: Many technologies, one goal. Expert Rev. Proteomics 2, 693–703.
14. Jan, Y.J., et al. (2018). NanoVelcro rare-cell assays for detection and characterization of circulating tumor cells. Adv. Drug Deliv. Rev. 125, 78–93.
15. Lee, H.H., & Collins, J.J. (2012). Microbial environments confound antibiotic efficacy. Nat. Chem. Biol. 8, 6–9.
16. Tang, F. et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods 6, 377–382 (2009).
17. Ramsköld, D., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol. 30, 777–782.
18. Lodato, M.A. et al. (2015). Somatic mutation in single human neurons tracks developmental and transcriptional history. Science 350, 94–98.
19. Meerzaman, D., et al. (2016). The promise of omics-based approaches to cancer prevention. Semin. Oncol. 43, 36–48.
20. Gawad, C., Koh, W., & Quake, S.R. (2016). Single-cell genome sequencing: current state of the science. Nat. Rev. Genet. 17, 175–188.
21. Navin, N., et al. (2011). Tumour evolution inferred by single-cell sequencing. Nature 472, 90–94.
22. Moore, D.S. (2015). The Developing Genome: An Introduction to Behavioral Epigenetics. Oxford University Press..
23. Buenrostro, J.D., et al. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 523, 486–490.
24. Cusanovich, D.A., et al. (2015). Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 348, 910–914.
25. Jin, W., et al. (2015). Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. Nature 528, 142–146.
26. Smallwood, S.A., et al. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat. Methods 11, 817–820.
27. Nagano, T., et al. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature 502, 59–64.
28. Clark, S.J., Lee, H.J., Smallwood, S.A., Kelsey, G., & Reik, W. (2016). Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. Genome Biol. 17, 72.
29. Mincarelli, L., Lister, A., Lipscombe, J., & Macaulay, I.C. (2018). Defining cell identity with single-cell omics. Proteomics 18, 1700312.
30. Virant-Klun, I., Leicht, S., Hughes, C., & Krijgsveld, J. (2016). Identification of maturation-specific proteins by single-cell proteomics of human oocytes. Mol. Cell. Proteomics 15, 2616–2627.
31. Chong, W.P.K., et al. (2012). LC-MS-based metabolic characterization of high monoclonal antibody-producing Chinese hamster ovary cells. Biotechnol. Bioeng. 109, 3103–3111.
32. Minakshi, P., et al. (2019). Chapter 15 - Single-Cell Metabolomics: Technology and Applications. In Single-Cell Omics (eds. Barh, D., & Azevedo, V.) 319–353 (Academic Press). doi:10.1016/B978-0-12-814919-5.00015-4.
33. Duncan, K.D., Fyrestam, J., & Lanekoff, I. (2019). Advances in mass spectrometry based single-cell metabolomics. Analyst 144, 782–793.
34. Llufrio, E.M., Wang, L., Naser, F.J. & Patti, G.J. (2018). Sorting cells alters their redox state and cellular metabolome. Redox Biol. 16, 381–387.
35. Binek, A. et al. (2019). Flow cytometry has a significant impact on the cellular metabolome. J. Proteome Res. 18, 169–181.
36. Macaulay, I.C., Ponting, C.P., & Voet, T. (2017). Single-cell multiomics: Multiple measurements from single cells. Trends Genet. TIG 33, 155–168.
37. Linnarsson, S., & Teichmann, S.A. (2016). Single-cell genomics: Coming of age. Genome Biol. 17, 97.
38. Hughes, A.J., et al. (2014). Single-cell western blotting. Nat. Methods 11, 749–755.
39. Perfetto, S.P., Chattopadhyay, P.K., & Roederer, M. (2004). Seventeen-colour flow cytometry: Unravelling the immune system. Nat. Rev. Immunol. 4, 648–655.
40. Perez, O.D., & Nolan, G.P. (2002). Simultaneous measurement of multiple active kinase states using polychromatic flow cytometry. Nat. Biotechnol. 20, 155–162.
41. Irish, J.M., et al. (2004). Single cell profiling of potentiated phospho-protein networks in cancer cells. Cell 118, 217–228.
42. Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D.A., & Nolan, G.P. (2005). Causal protein-signaling networks derived from multiparameter single-cell data. Science 308, 523–529.
43. Gadalla, R., et al. (2019). Validation of CyTOF against flow cytometry for immunological studies and monitoring of human cancer clinical trials. Front. Oncol. 9, 415.
44. Palii, C.G., et al. (2019). Single-cell proteomics reveal that quantitative changes in co-expressed lineage-specific transcription factors determine cell fate. Cell Stem Cell 24, 812-820.e5.
45. Assarsson, E. et al. (2014). Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLOS ONE* 9, e95192.
46. Darmanis, S., et al. (2016). Simultaneous multiplexed measurement of RNA and proteins in single cells. Cell Rep. 14, 380–389.
47. Selby, W.S., Janossy, G., Bofill, M., & Jewell, D.P. (1984). Intestinal lymphocyte subpopulations in inflammatory bowel disease: an analysis by immunohistological and cell isolation techniques. Gut 25, 32–40.
48. Thiel, A., Scheffold, A., & Radbruch, A. (1998). Immunomagnetic cell sorting—pushing the limits. Immunotechnology 4, 89–96.
49. Brown, M., & Wittwer, C. (2000). Flow cytometry: Principles and clinical applications in hematology. Clin. Chem. 46, 1221–1229.
50. Brakke, M.K. (1951). Density gradient centrifugation: A new separation technique. J. Am. Chem. Soc. 73, 1847–1848.

thermo scientific

51. Shen, Y., Yalikun, Y., & Tanaka, Y. (2019). Recent advances in microfluidic cell sorting systems. Sens. Actuators B Chem. 282, 268–281.
52. Liu, Y., Chen, X., Zhang, Y., & Liu, J. (2019). Advancing single-cell proteomics and metabolomics with microfluidic technologies. Analyst 144, 846–858.
53. Piehowski, P.D., et al. (2020). Automated mass spectrometry imaging of over 2,000 proteins from tissue sections at 100-µm spatial resolution. Nat. Commun. 11, 8.
54. Zhu, Y., et al. (2018). Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. Nat. Commun. 9, 882.
55. Budnik, B., Levy, E., Harmange, G., & Slavov, N. (2018). SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. Genome Biol. 19, 161.
56. Marx, V. (2019). A dream of single-cell proteomics. Nat. Methods 16, 809–812.
57. Dou, M., et al. (2019). High-throughput single cell proteomics enabled by multiplex isobaric labeling in a nanodroplet sample preparation platform. Anal. Chem. 91, 13119–13127.
58. Schoof, E.M., et al. (2019). A quantitative single-cell proteomics approach to characterize an acute myeloid leukemia hierarchy. bioRxiv 745679. doi:10.1101/745679.
59. Ouldali, H., et al. (2019). Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. Nat. Biotechnol. 1–6. doi:10.1038/s41587-019-0345-2.

Find out more at **thermofisher.com/singlecellproteomics**

ThermoFisher SCIENTIFIC