HHS Public Access

Author manuscript

Analyst. Author manuscript; available in PMC 2020 January 28.

Published in final edited form as:

Analyst. 2019 January 28; 144(3): 794–807. doi:10.1039/c8an01574k.

New Mass Spectrometry Technologies Contributing towards Comprehensive and High Throughput Omics Analyses of Single Cells

Sneha P. Couvillion[1], **Ying Zhu**[1], **Gabe Nagy**[1], **Joshua N. Adkins**[1], **Charles Ansong**[1], **Ryan Renslow**[1], **Paul Piehowski**[1], **Yehia Ibrahim**[1], **Ryan Kelly**[1,2], and **Thomas O. Metz**[1,*]

[1]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA, USA

[2]Department of Chemistry and Biochemistry, Brigham Young University, Provo, UT, USA

Abstract

Mass-spectrometry based omics technologies - namely proteomics, metabolomics and lipidomics - have enabled the molecular level systems biology investigation of organisms in unprecedented detail. There has been increasing interest for gaining a thorough, functional understanding of the biological consequences associated with cellular heterogeneity in a wide variety of research areas such as developmental biology, precision medicine, cancer research and microbiome science. Recent advances in mass spectrometry (MS) instrumentation and sample handling strategies are quickly making comprehensive omics analyses of single cells feasible, but key breakthroughs are still required to push through remaining bottlenecks. In this review, we discuss the challenges faced by single cell MS-based omics analyses and highlight recent technological advances that collectively can contribute to comprehensive and high throughput omics analyses in single cells. We provide a vision of the potential of integrating pioneering technologies such as Structures for Lossless Ion Manipulations (SLIM) for improved sensitivity and resolution, novel peptide identification tactics and standards free metabolomics approaches for future applications in single cell analysis.

Why study single cells?

Routine bulk scale omics measurements which involve the analysis of a large number, often millions, of cells are population-averaging techniques that assume that cell populations are uniform. Cell to cell variability or heterogeneity is in fact very common and is a fundamental property of cellular systems.[1] Different cell types in the same multi-cellular organism exhibit differing phenotypes, even though they are genetically identical, due to distinct regulation of gene expression (Figure 1). Cells in the same tissue type may also display differences in phenotypic profiles depending on how they respond to their microenvironment. For example, cells may vary in proximity to blood vessels, other cells or oxygen gradients, which may then trigger a response from only a fraction of the population.[2]

*Corresponding author: Thomas O. Metz, Address: 902 Battelle Blvd., P.O. Box 999, MSIN K8-98, Richland, WA 99352, Phone: 509-371-6581, thomas.metz@pnnl.gov.

The ability to investigate single cells in a high throughput manner can inform us about interesting and important biological phenomena that may be obscured in ensemble measurements. This can be especially valuable in understanding diseases like cancer where heterogeneity can play a pivotal role in disease onset and progression.[3] Studies have shown that the presence of disseminating tumor cells or circulating tumor cells after adjuvant therapy predict a poor prognosis in cancer patients and indicates metastatic progression.[4-6] Although it is possible to detect these rare cell populations, advances in single cell proteomics and metabolomics could provide the much needed insight into the molecular and phenotypic nature of these cells, which is thought to cause marked inter-patient variability with regards to their dormancy or expression.[7]

Single cell analysis methods can also help characterize uncultivable individual microorganisms that influence community stability and resilience in microbiomes inhabiting a variety of ecosystems.[8] However, microbial cells are particularly challenging in that they range in size of 1/1000th - 1/10,000th the size of a eukaryotic cell. Improved methods for analysis of such small samples will allow for important insights, such as in the context of host-pathogen interactions, particularly for intracellular pathogens. Here key differences between uninfected cells, and early-, mid-, and late-infection within a population of cells are typically averaged using conventional analyses such that critical cell-cell signaling and processes of infection are lost. A major hindrance of microbiome proteomics research is the sheer number of different species at different concentrations,[9] which can undermine detection efficiency and control of false-discovery. The microdissection of microbiome and biofilm samples has clear benefits from very small sample processing by greatly improving homogeneity of species studied at one time.

Volume of a single cell & technical challenges

The dimensions and volume of a single cell can differ by orders of magnitude depending on the organism and cell type. Microorganisms such as *Bacillus subtilis, Staphylococcus aureus, Escherichia coli, and Caulobacter crescentus* typically have individual cell volumes between 0.4-3 μm.[3,10] In humans, cell types can vary by several orders of magnitude with erythrocytes, HeLa cells and oocytes having a volume of about 100, 3000 and 4,000,000 μm^3 respectively.

The size of a single cell requires suitable tools and approaches for sample handling and analysis. Depending on the cells of interest, specialized strategies have to be used for separating and isolating individual cells. Ideally these strategies will introduce minimal artifacts during the process. The limited sample volume in a single cell also calls for highly sensitive measurement methods for accurate detection and quantification of the biomolecules of interest.

The development of whole-genome and whole transcriptome amplification (WGA/WTA) has made single cell genomics and transcriptomics possible despite the miniscule amount of genetic material present in a cell.[11] Advancements in WGA technology such as improved performance and the availability of commercial kits and the routine use of unique molecular identifiers (UMIs) to barcode individual RNA molecules has resulted in robust genome and

transcriptome sequence analysis in single cells.[12] In contrast, mass spectrometry-based single cell omics technologies that focus on proteins, metabolites and lipids are less mature, yet are newly emerging fields with great potential. Unlike DNA and RNA, biomolecules such as proteins, metabolites and lipids cannot be amplified, so one has to make do with the quantity actually present in the cell and work on enhancing the intrinsic sensitivity of the method of analysis.

In this paper, rather than providing a comprehensive review of the literature, we highlight key developments in mass spectrometry (MS)-based technologies that can contribute to omics analyses of small samples down to a single cell. Because much of the work on single cell MS-based omics performed to date has focused on proteomics, in part because that field is more mature compared to metabolomics and lipidomics, the majority of our review will in turn focus on recent developments in single cell proteomics analyses. We also focus on approaches for comprehensive, untargeted analyses of small samples, rather than approaches for targeted analyses of more limited sets of biomolecules. Finally, we provide a vision for how the remaining challenges in single cell omics analyses can be overcome using specific technologies as the drivers.

SAMPLE ISOLATION AND MANIPULATION

Only very recently has the prospect of extending MS-based global proteome profiling to single mammalian somatic cells entered the realm of possibility.[13-16] Inefficiencies and sample losses associated with all steps of sample handling, separation, ionization and transmission to the high-vacuum region of the mass analyzer have historically resulted in sample requirements of micrograms or milligrams of protein isolated from millions of cells to achieve a reasonable depth of proteome coverage. These sample size limitations precluded the analysis of rare cells or required averaging over multiple cell and tissue types, greatly limiting the specificity of the measurements and obscuring important underlying heterogeneity. However, consistent improvements in each of these areas has greatly increased the feasibility of in-depth proteome measurements for single cells and other small samples for which only picogram or low-nanogram amounts of protein starting material are available.[17, 18] Below we highlight advances in sample handling, prior to MS analysis.

Isolating single cells

Single cell techniques rely on technologies to enrich cell populations of interest and isolate rare cell types for further characterization. Although most single cell omics studies use a well-characterized cell line such as HeLa for a proof-of-concept demonstration of their method, it is important to keep in mind that actual biological samples will require robust methods to isolate cells of interest. It is imperative that the isolation technique is gentle enough to avoid altering biomolecules in the cell through physical shear forces or by harsh chemicals used during the process. Traditional methods such as filtration and centrifugation for cell isolation have made use of differences in cell density or size for a more gross level purification of samples with low specificity.[19] More selective separation is afforded by affinity-based methods that use antibodies against cell surface antigens that are differentially expressed in different cell types.[20] Flow cytometry has been widely used for rapid sorting of

discrete populations of cells and individual cells based on their cell-surface markers.[21] Adaptations of flow cytometry include fluorescence activated cell sorting (FACS), which is a high-throughput platform for single cell analysis.[22] Affinity substrates and magnetic beads have been used for isolating and capturing intact cells from complex samples.[23] The limitation with affinity-based techniques lies in the fact that they are inherently antibody centered, targeted techniques and the user must have prior knowledge on the target to be measured. Laser capture microdissection (LCM) has been used to harvest specific cells of interest from tissues and is well suited for coupling with techniques that require small numbers of cells.[24] We recently used a LCM approach to gain spatially resolved insights in a proteomic study of lung development in microdissected alveolar tissue.[25] Also, the more traditional approach of manual isolation of cells is used by many research groups. This approach is tedious, but it ensures that rare but morphologically-identifiable cells can be isolated.

Single vessel methods

A typical bottom-up proteomics workflow involves sampling, cell lysis, extraction of proteins from sample, proteolytic digestion of proteins to peptides, separation, fractionation or enrichment of peptide mixtures and finally identification of peptides using MS. Each of these steps contributes to adsorption losses of proteins/peptides to solid supports (tubes, pipette tips etc.), which becomes a problem as the sample size approaches the analytical limits of detection. This can result in limited proteome coverage, with only the most abundant proteins being detected, and given higher analytical variability as one operates closer to the limits of detection, it can be challenging to confidently characterize differences across such samples.[26] To eliminate the contamination and losses typically inevitable in multi-step sample preparation workflows used in traditional MS-based omics, many groups have attempted to carry out all the necessary steps in a single vessel, thereby minimizing the number of surfaces with which the samples comes into contact.

In 2005, Wang et al. reported a ‘single-tube’ protocol that used the organic solvent trifluoroethanol (TFE) for cell lysis and protein denaturation.[27] By evaporating the organic solvent during lyophilization, cleanup steps typically carried out for traditional detergent based protocols and transfer to other tubes were avoided. Using the TFE protocol, 246 peptides and 104 proteins were identified from 5000 MCF-7 human breast cancer cells (~500 ng total protein content). Liang Li’s group used a NP-40 surfactant to lyse the cells, followed by acetone precipitation to profile the proteome of 500-5000 MCF-7 cells.[28] The gradient time for liquid chromatography (LC) separation was optimized for the number of cells being analyzed, resulting in the identification of 167 and 619 proteins from 500 and 5000 cells, respectively. Kulak et al. used a filter-aided sample preparation (FASP)-based method but omitted the use of strong detergents that require the use of molecular weight cut-off filters for subsequent clean-up.[29] By choosing compatible reagents, they were able to carry out lysis, reduction and alkylation in a single step for samples as small as 1000 HeLa cells. Hughes et al. developed a platform (SP3) for proteomic sample preparation of sub-microgram amounts of material in a single tube workflow based on carboxylate-coated paramagnetic beads thus circumventing the need to use a filter membrane.[30] They were able to obtain an increase of ~5000 unique peptide identifications using an equivalent number of

1000 HeLa cells. This improvement was due to the ability of the SP3 approach to use 1% SDS for enhanced lysis and solubilization. More recently Li and co-workers described an approach utilizing acoustics-assisted lysis and ultra-trace LC-MS analysis to achieve zeptomole detection sensitivity and yielding ~4000 protein identifications from 100-200 MCF-7 cells.[31] Instead of directly using the cells of interest, the MCF-7 cells were first spiked into whole blood and then isolated using microfluidic magnetophoretic isolation, thus demonstrating the capability to isolate rare target cells from a complex matrix. Chen et al. developed an integrated device based on a two-position 10-port valve for online digestion, enrichment and analysis of 100 DLD-1 cells.[32] High temperature trypsin digestion was carried out and an average of 635 proteins were identified within 2 hours. Huang and colleagues designed a simple but robust online sample handling platform incorporating immobilized enzyme reactor principles to analyze nanogram (~ 440-500 ng) quantities of protein from mouse blastocysts and identified 348 proteins.[33]

Microfluidic advances

Innovations in the field of microfluidics have enabled researchers to manipulate small quantities of fluids in ways that were not possible before. The ability to integrate multiple functions with the potential for massive parallelization has made microfluidic approaches especially attractive for single cell omics. In 2010, Mellors and coworkers developed an integrated microfluidic device for analysis of individual human erythrocytes using capillary electrophoresis (CE) coupled with MS via electrospray ionization (ESI).[34] The microfluidic structure incorporated continuous cell delivery, detergent-free lysis (using a buffer compatible with CE-ESI-MS), separation of lysate and ionization for detection by MS. Although this early study was used to detect hemoglobin, a highly abundant protein, it did demonstrate the power of combining microfluidics with MS.

Recently, we have utilized microfluidic sample processing in combination with nanoflow LC and a latest-generation MS system to extend label-free proteomic analyses to small cell numbers[14] and single mammalian cells.[15] A robotic nanopipetting platform with microfabricated well plates reduce the total processing volume to just ~200 nL, providing a significant reduction in potentially adsorptive surfaces during preparation relative to standard preparation protocols.[14, 15] Termed nanoPOTS (Nanodroplet Processing in One pot for Trace Samples), the workflow has enabled deep proteome profiling of trace samples that were previously inaccessible. NanoPOTS enables all of the required processing steps for bottom-up proteomics, including cell lysis, protein extraction, reduction, alkylation and digestion to be achieved within a single nanowell (without sample transfer between steps) through a series of 50 nL pipetting steps and incubations.[14, 35, 36] Evaporation is minimized during dispensing by operating the nanopipette within a humidified chamber, and extended incubations take place with a cover over the nanowell plate to minimize the headspace over the array.

Using nanoPOTS, we have reproducibly achieved a proteome coverage of >3000 protein groups from as few as 10 HeLa cells.[14] To our knowledge, this level of coverage was not achieved in past studies for samples comprising ≤5,000 mammalian cells.[37] When coupled with fluorescence-activated cell sorting to dispense single HeLa cells into each nanowell, we

have profiled nearly 700 proteins on average from each cell.[15] We have also utilized nanoPOTS in conjunction with laser capture microdissection to isolate and analyze regions of thin tissue sections with lateral dimensions as small as 50 μm.[38] The nanoPOTS platform has been applied to a variety of biological tissues, including single dissociated lung cells,[15] sections of single pancreatic islets from type 1 diabetic and nondiabetic donors,[14] rat brain,[38] mouse liver,[39] plant tissues[40] and spiked circulating tumor cells.[41] In addition, we have preliminarily isolated and profiled spiked circulating tumor cells from whole blood.[39] Finally, to increase the proteome coverage, we have utilized the robotic platform to fractionate nanoPOTS-prepared samples from a high-pH nanoLC separation into nanowells for a subsequent low-pH separation.[42] Using nanoPOTS in combination with this nanowell-mediated 2D LC workflow, we have increased proteome coverage to ~6000 protein groups for 650 HeLa cells or 10 pooled pancreatic islet thin sections. A plot of proteome coverage achieved using the nanoPOTS platform compared with other nanoscale proteome processing and analysis platforms is shown in Figure 2.

While the microfluidic nanoPOTS platform has reduced the adsorptive losses that have hindered proteomic analysis of trace samples, this is not the only approach that has been utilized to extend MS-based proteomics to the analysis of single mammalian cells. For example, Li et al.[46] recently reported on a microfluidic reaction chamber in an oil layer separated by an air gap minimized evaporation of sample and reagents within a droplet containing a final volume of ~550 nL. When a single HeLa cell was analyzed, 51 proteins were identified as indicated by the blue data point on the left of Figure 2. In addition, Budnik, et al.[16] recently prepared single cells and populations of carrier cells within commercial microtubes and labeled each sample with a tandem mass tag (TMT). The labeled samples were pooled and analyzed by LC/MS, and the carrier cells within the pooled sample provided sufficient MS^1 sensitivity to trigger MS^2 peptide sequencing and identification of peptides from single cell samples based on their respective reporter ions. While the proteome coverage achieved by this method was not reported other than the aggregate number of identified proteins from multiple single cell analyses, it is likely that strategies that combine such multiplexed analyses with reduced losses achieved by microfluidic sample preparation will yield additional insights into the proteomes of single cells and other trace samples.

MOLECULAR IONIZATION, SEPARATION, DETECTION, AND IDENTIFICATION

Because mass spectrometers are mass sensitive detectors and have detection biases towards the most abundant molecules in a sample, the MS and omics communities have placed significant effort into developing methods and technologies that result in maximizing the separations of peptide, metabolite and lipid samples, the efficiency of ion generation, transmission, and utilization inside the MS, and the confident identification of detected molecules. In the below sections, we highlight key advances in these areas as they might be or have been developed for or utilized in the analysis of single cells or small sample sizes.

Advances in molecular ion generation and transmission

One of the most commonly used configurations for the analysis of biomolecules is LC coupled to MS via ESI. Two decades ago, ESI-MS instrumentation involved a high-flow ESI source and highly inefficient ion optics such that very few analyte molecules from solution were ionized and detected at the mass analyzer, resulting in poor analytical sensitivity. Ionization efficiency, the extent of conversion of solution-phase molecules into gas-phase ions, depends on the size of the initial charged droplets emanating from the electrospray source, with smaller droplets being more readily desolvated and producing more ions.[47, 48] By reducing the flow rate of sample delivered to the electrospray source to a few tens of nanoliters, smaller and more readily desolvated initial charged droplets are produced, which greatly improves ionization efficiency.[48] Such nanoESI sources and their coupling with flow-compatible LC or CE separations have been made possible through substantial advances in ESI emitter technology and miniaturization of the accompanying separations. For example, creating an emitter with the appropriate narrow orifice and high aspect ratio to support an electrospray at low nanoliters per minute[48, 49] or even picoliters per minute[50] required the development of heating and pulling techniques or chemically etching of fused silica capillaries.[49-51] Similarly, new methods needed to be developed to prepare LC columns that delivered analyte at the appropriate low flow rates, which has been achieved through either packing of media into narrow-bore capillaries[52-54] or preparing porous layer open tubular LC columns[31, 55, 56]. With these advances and similar developments for capillary electrophoresis (CE)-MS, nanoscale samples could be effectively separated and ionized.[57, 58]

Similar inefficiencies in MS ion optics needed to be overcome to effectively transmit ions from the electrospray ionization source to the mass analyzer that requires high vacuum to operate effectively.[59] A typical ESI-MS instrument interfaced an electrospray emitter with a very narrow orifice or heated capillary inlet such that only a small fraction of the electrospray-generated plume of ions arrived at the first pumping stage of the instrument, which is typically operated at a few Torr. From there, a skimmer orifice again sampled only a tiny portion of the ions for transmission to the high-vacuum region where conventional ion optics such as multi-pole ion guides could efficiently operate. To address the skimmer-related losses, our group developed the electrodynamic ion funnel,[59-61] which could operate in the rough-pumped vacuum regime and captured the entire cloud of ions entering in the first vacuum stage to transmit them to high vacuum. Related advances have also addressed the losses at the atmospheric pressure interface including multicapillary inlets,[62] high-conductance single inlets,[61] and low-pressure electrospray sources.[63, 64] Today, many of these advances have been commercialized and are standard issue on biological mass spectrometers.

SLIM: a new frontier for separations and ion manipulations in the gas-phase

An appealing technique for enhancing single cell analysis methods is the incorporation of ion mobility (IM) spectrometry. [65] IM-MS is a rapid gas-phase technique, whereby ions are

separated on the basis of their mass-to-charge (*m/z*) and their shapes (mobilities) in a given buffer gas.[66] This technique provides a viable, and attractive, alternative to solution-phase methods since separations are performed on the order of milliseconds in IM, while on the order of minutes, to potentially hours, in conventional chromatographic methods.[66] Numerous IM-MS separations approaches have been documented for a wide variety of global, as well as targeted applications.[67-70] Despite their success in many applications, all of these prior approaches fall under the same joint bottleneck: lack of sensitivity and poor resolution. Structures for lossless ion manipulations (SLIM) is a recent traveling wave (TW)-based technology that has successfully been applied to overcome this resolution barrier by greatly extending the separation path length as compared to conventional IM-MS platforms.[71-76] In TWIM-MS-based separations, ions will separate based on their ability to keep up with the motion of the traveling wave, as related to their intrinsic molecular properties, such as *m/z* as well as their shapes (i.e. mobilities). Higher mobility ions (faster/more compact) will move more similarly to the speed of the TW, while lower mobility ions (slower/more elongated) will get passed by the TW and thus fall behind the wave. While much has been discussed in recent literature on the tremendous increases in resolution and resolving power that are enabled by SLIM, here we will discuss its potential in addressing the sensitivity associated with sample-limited applications. In these types of applications, every ion counts and efforts to eliminate ion losses are core to SLIM performance.

This raises the question of: how can more analyte ions be introduced to the mass analyzer. Paradoxically, conventional/commercial IM platforms cannot accommodate a large influx of ions, as it overwhelms the instrument's ability to utilize and quantify them. Thus, the challenge becomes: how can we work with large ion populations? Further, how can we integrate all the required technologies not only into a single instrument platform, but ultimately in a single, standalone, experiment? To begin to address these challenges, we recently reported on the ability to accumulate ions in the SLIM device itself, termed 'in-SLIM ion accumulation' by halting the traveling wave in the second region and thus permitting ions to accumulate at the interface between the first and second regions [71]. This technological advancement has enabled the introduction of over a billion ions, representing a 2-3 orders of magnitude increase in charge capacity as compared to ion introduction via other standard approaches (such as the ion funnel trap).

With this new technology for ion introduction comes the issue of how to work with such large populations. Specifically, when ions are accumulated for several seconds in the SLIM module, the total width of the initial ion packet will be significantly larger, and much broader, than when ions are introduced for a few milliseconds by e.g. an ion funnel trap. Another recent capability for SLIM is the ability to spatially compress ion mobility peaks through compression ratio ion mobility programming (CRIMP).[71, 75] CRIMP successfully permits a broad, diffuse, ion packet (containing multiple unique mobility peaks/features) to be compressed into ones with much higher signal intensity and signal-to-noise (S/N), thereby increasing the sensitivity of measurements. Figure 3 demonstrates an example of how several broad initial peaks can be compressed into much narrower ones. This combination of introducing massive ion populations through 'in-SLIM ion accumulation', followed by a subsequent CRIMP step, enables us to manipulate small sample sizes and greatly increase the sensitivity of measurements. Presently, we have demonstrated limits of

detection (LOD) of 50 pM for phosphoproteomics measurements[77] and as low as 1 pM for the analyses of other peptide mixtures.[71] Future improvements to the in-SLIM ion accumulation process (e.g., variation of traveling wave parameters during accumulation) will enable us to further improve upon our measurement sensitivity and overcome the limitations associated with volume-limited samples.

Even with the combination of in-SLIM ion accumulation and a subsequent CRIMP step, other steps may be necessary to further increase the sensitivity of small sample analyses. Akin to fraction collection and enrichment protocols that are prevalent in conventional chromatography-based applications, our SLIM platform enables similar manipulations to be performed in the gas phase. One such potential manipulation is that of 'ion enrichment', [78] where a specific region of the SLIM module can be used as an ion trapping region to accumulate low abundance ions, and thus build up their population while simultaneously resolving them from unwanted background interferences (Figure 4). This capability permits mobility-selected trapping of desired ions on the order of seconds, in a standalone experiment, as opposed to the tedious, time consuming, and decoupled process of chromatographic fractionation followed by an enrichment step prior to detection.

SLIM has a very promising future as part of the analytical toolbox for sample-limited applications through its ability to: accumulate, and work with, large ion populations [71], increase the sensitivity of measurements through peak compression via CRIMP [75], as well as enriching ions through various mobility-selective trapping regions [78], all in a standalone module. These recent advancements in SLIM present an exciting time for this technology to rise to the forefront of analytical measurements in the search for increased sensitivity without sacrificing resolution.

Peptide identification strategies for single cell proteomics

A particular challenge that arises due to the sample mass limitations of single cells for omics analyses, is the need for MS/MS analysis for confident identification of the analyte of interest. In the case of peptides and proteins, gas-phase fragmentation provides necessary sequence information that is used along with the parent ion mass and the organism's genome to obtain an accurate identification. In a data-dependent acquisition (DDA) experiment, the most commonly applied approach in bottom up proteomics analysis, all ions are injected into the mass spectrometer and a full scan mass spectrum (MS1) is generated. A user-defined number of the most abundant ions from the MS1 scan are then sequentially isolated for fragmentation. As a result, the ion flux required to generate high quality tandem mass spectral (MS/MS) information exceeds that required for MS1 level detection by several fold. Thus, a number of approaches have been developed to achieve confident identification of peptides in high throughput, bottom up proteomics experiments without the need to perform fragmentation.[80-82] Briefly, equivalent samples with adequate sample mass are analyzed using identical LC-MS/MS conditions to identify peptide sequences and their LC elution times in order to establish a comprehensive reference library. To further increase the experimental dynamic range and coverage of the proteome, these samples may also be fractionated prior to LC-MS/MS analysis. The measured LC retention time and exact mass information is then leveraged to identify matching MS1 signals in subsequent analyses,

based on peptide mass and elution time. In the case of sample limited proteomics, the existing MaxQuant MBR algorithm has shown significant utility (any or all nanoPOTS and SNaPP references here.[14, 15, 83-85]

While promising, approaches based on MS1 and elution time information do have limitations at present. First, due to the high sample complexity of bottom up proteomics measurements, accurately assessing and controlling the false discovery rate of identifications based on MS1 and elution time information is challenging.[86] One way that we foresee improving confidence is to add a third dimension of information for matching,[87] as illustrated in Figure 5. The high sensitivity and reproducibility of ion mobility separations,[88] as discussed above, make them the ideal candidate for a third measurement dimension, and the improvements obtainable are just beginning to be realized.

The second limitation to identifying peptides in single cell analyses using MS1 and elution time information and reference libraries, is that one can only identify peptide species that have already been identified in other analyses. The disadvantage here is two-fold. First, the number of analyses required to produce high proteome coverage for an appropriate reference library is increased, placing the burden on platform throughput. Secondly, analyses are limited to what is already known and discovery of novel peptides is not possible. To drive innovation in this area, application of machine learning technologies will play a key role. One potential direction is through the development of models that accurately predict LC elution and IMS drift times. Models, trained on ample existing data, could be used to create matching databases without the need for additional analyses.[90-92] Further, these models could then be extended to predict the behavior of hypothetical peptides or post-translationally modified species not previously identified, expanding the utility of these approaches as ‘discovery’ tools.

Standards-free metabolomics

As described above, MS-based omics analyses of single cells will be challenged by the amount of molecules that can be isolated from the sample, and efficiently ionized and transferred through the instrument to the detector in such a way that allows for their comprehensive, unambiguous identification and accurate quantification. In this section, we focus on unambiguous identification or characterization of the chemical structures associated with detected metabolites and other small molecules. Because peptides and lipids are polymeric or otherwise comprised of well-characterized building blocks that allow for fairly accurate prediction of their tandem mass spectra and therefore algorithms and associated software that enable relatively comprehensive, confident characterizations of the proteome and lipidome,[93-96] we will discuss efforts in the research community that are driving the implementation of standards-free metabolomics analyses and how these might be leveraged for maximizing the amount of information that can be obtained in analysis of single cells.

Metabolites and related chemicals, such as those with anthropogenic origins, are not constrained to polymeric or template-like structures. Of course, there are polymers that can qualify as metabolites or be considered as part of the metabolome, such as carbohydrates

(i.e. the glycome[97, 98]), as well as secondary metabolites that do have template-like structures, such as terpenes[99] and terpenoid;[100] however, the structures of chemicals that largely comprise the metabolome are constrained only by the laws of thermodynamics (i.e. those that determine a chemically stable structure). Because of this chemical diversity, methods have not yet been developed that allow for comprehensive and unambiguous identification of all metabolites detected in MS-based metabolomics studies. This technical gap is due in part to the paradigm typically required for a confident, chemical identification of a molecule,[101-104] which is based on comparison of multiple, orthogonal experimental data (e.g. MS/MS spectra and retention times) to the same data contained in a reference library that was constructed through analysis of authentic reference compounds. This represents the single biggest limitation in MS-based metabolomics studies, because authentic chemical standards are not available for the majority of metabolites or chemicals.[105] When microliter volume samples can yield hundreds of thousands of features,[106] which may represent billions of currently unknown molecules, a paradigm shift is required in order to comprehensively and unambiguously identify all the molecules in complex samples. Therefore, there is growing interest within the metabolomics community, and among funding agencies,[107] to develop methods and tools that accurately predict experimental properties of metabolites encountered in MS-based metabolomics analyses, such as retention times,[108-112] MS/MS spectra,[113-117] collision cross sections (CCS),[118-120] etc. Ultimate success from these efforts will be an analytical approach that provides unambiguous identification of the complete set of detected molecules in a sample.

Because any experimental manipulation of samples, molecules, and ions during MS-based metabolomics analyses will be associated with losses, it is important to minimize such manipulations in analyses of small samples, such as from single cells, in order to maximize the overall analysis efficiency. As such (and as discussed above), approaches that rely on MS1-level information for identification and quantification are preferred, and are characterized by retention time, CCS, molecular ion *m/z*, and isotopic distribution. Of these, *m/z*, isotopic signature, and CCS are those that are most accurately calculated and measured. Calculation of *m/z* and isotopic signatures based on molecular formula is a trivial calculation when the chemical structure is known or proposed. To take advantage of this information, modern MS instrumentation is capable of mass measurement accuracies of a few ppm to sub-ppm,[121, 122] depending on type of mass analyzer.[123] However, until mass analyzer precision can differentiate isomers based on differences in the mass equivalence of their chemical bond binding energies (sub-ppb level), *m/z* measurements will, at best, only ever be able to provide a molecular formula for each detected feature. For a given molecular formula, hundreds of known molecules can be represented (e.g., $C_8H_{16}O_2$, $C_{10}H_{12}O_2$, and $C_8H_8O_3$ all represent over 100 known molecules in PubChem,[124] and theoretically, each formula can represent billions of isomers (i.e., well over 10^{60} possible molecules are possible).[125] The future of small molecule identification must rely upon consistently measurable and accurately calculable attributes that are (semi)-orthogonal to *m/z*. As the number of attributes increases, the effective distance between features increases, enabling unambiguous identification of molecules in small-volume complex biological samples.

As an example semi-orthogonal attribute to *m/z*, CCS can be experimentally measured with high reproducibility and accuracy (condition dependent),[88] since IM separations do not

involve interactions with surfaces, and therefore do not suffer from degradation of chromatographic stationary phases as do GC and LC. CCSs can also be very accurately predicted based on three-dimensional chemical structures using quantum chemistry methods,[126-129] and with lesser accuracy based on two-dimensional structures or chemical properties (e.g. pKa) using machine learning approaches.[130, 131] Accurate prediction of multiple molecular attributes such as *m/z* and CCS will enable more comprehensive identifications of the metabolome through matching to experimental data (Figure 6). Forthcoming applications of IMS separation methods should include consideration of multiple adduct ion types simultaneously, CCS peak shape (as governed by the molecular conformer distribution, see Figure 7), and CCS of adduct ion multimers (e.g., molecular ion hetero- and homo-dimers) to further increase the dimensionality of the IMS data, and thus increase the uniqueness of each molecular identification.

An intriguing option for the identification of molecules from single cells is the incorporation of infrared spectroscopy (IR), or other spectrophotometry techniques, into MS-based measurement approaches. Rizzo and colleagues have demonstrated the coupling of IR with ion traps,[132] and more recently, with IMS in order to generate both CCS and IR spectra for use in characterizing biomolecules (Figure 7).[133] Infrared spectra are almost completely orthogonal to *m/z*, with each molecule having a spectra that is unique (with exception to enantiomers).[134] Critically, IR spectra are also calculable using similar computational pipelines already being developed for quantum chemistry-based CCS calculations.[135] IR combined with ion trap mass spectrometry has been successfully used for small molecule identification by Martens et al.,[136, 137] with examples of acetylhexosamines, glutaric acid, and ethylmalonic acid. By combining a method like IMS, which can separate stable conformers of molecular ions, and cryogenic cooling systems, which reduce the temperature of the molecular ions, uncongested (narrow absorption band and non-convoluted) infrared spectra can be reproducibly obtained for each *m/z*-CCS feature pair. This has successfully been utilized by Masson et al.,[132] demonstrating separation and identification of *cis/trans*-proline versions of the peptide GPGG, with confirmation via density functional theory (quantum chemical electronic structure) calculations. By relying on multiple orthogonal properties such as *m/z*, CCS, and IR spectra, future metabolomics analyses of single cells will likely be completely free from reliance upon authentic reference materials.

SINGLE CELL PROTEOMICS APPLICATION: CELL TYPING

A significant capability offered by single cell transcriptomics is the ability to delineate different cell types present in a tissue sample based on transcript abundance alone. However since transcript levels do not always correlate with protein levels, the utility of the markers identified for specific cell types for enriching or isolating these cells by magnetic-activated cell sorting (MACS) or FACS is usually an unknown. Protein markers for specific cell types by definition should readily allow isolation of specific cell populations of interest more efficiently. We have recently evaluated the feasibility of our nanoPOTS platform to differentiate human cell types from a clinical specimen based on proteome expression.[15] Here primary human lung epithelial and mesenchymal cells were isolated and the cryopreserved lung epithelial and mesenchymal cells were FACS-sorted into nanowells at a count of one cell per well (three single-cell samples were analyzed). 485 proteins were

identified across the single-cell samples, and of the 485 proteins, 328 proteins were quantifiable. The levels of all 328 proteins quantified in single cells were projected onto their principal components. Principal component analysis (PCA) revealed the single-cell proteomes separated by cell type along component 2 (Figure 8a), indicating that our label-free single-cell proteomics platform has the potential to identify cell types based on protein expression alone. To identify features facilitating the distinction of the two cell types, analyses of variance (ANOVA) were performed. ANOVA comparing single-cell epithelial and mesenchymal proteomes revealed 20% of quantified proteins to be significantly differential ($p<0.05$) (Figure 8b). Among the significantly differential proteins were those expected to be cell type specific, such as vimentin (VIME_human) a mesenchymal marker, which was higher in abundance in mesenchymal cells, as well as ezrin (EZRI_human) and keratin 18 (K1C18_human), epithelial cell markers, which were higher in abundance in epithelial cells.

PROSPECT FOR SINGLE-CELL OMICS ANALYSES

Proteomics

Proteins are important because they dictate cell function and therefore changes in the proteome can be a reliable indicator of functional heterogeneity across single cells. Even if the genome of an organism does not typically vary between cells, protein expression can differ amongst cell types and physiological or environmental changes can alter expression patterns even between cells of the same type. Many attempts have been made to achieve deep proteome coverage using increasingly smaller numbers of cells with the goal of making single cell analysis a possibility. While proteome analysis has been preliminarily extended to the single cell level, the coming years will see significant additional gains in terms of proteome coverage, and, critically, measurement throughput. Incremental coverage gains are expected to result from further miniaturization of LC separations and further improvements in ion transmission efficiency and MS sequencing speed. Additionally, multiplexed sample processing and analysis utilizing, e.g., tandem mass tags such as in the recently reported SCOPE-MS [16] workflow will potentially provide further gains in both throughput and proteome coverage for single cells, particularly when implemented within small-volume sample processing workflows. Novel sample acquisition and data analysis workflows such as the recently reported BoxCar MS may additionally enhance single cell analyses.[140] Finally, increasing the separation power of ion mobility-based gas-phase separations through extended pathlengths, as with the SLIM platform,[141, 142] may ultimately supplant liquid-phase separations altogether, resulting in an orders-of-magnitude increase in analysis speed. Given the numerous ongoing developments in this area, we expect to see single cell proteomics rapidly extended from basic proof-of-concept studies to become an important and widely used tool in biological and medical research.

Metabolomics and Lipidomics

Single cell metabolomics and lipidomics is still a budding field when compared to proteomics. Zhang and coworkers recently described a combined droplet extraction and a pulsed direct current ESI-MS method to record over 600 tandem mass spectra from a single human astrocyte cell and identified over 300 phospholipids.[143] Yin et al. extracted picoliter

volumes of material from live Allium cepa cells using electroosmotic extraction and a hydrophobic MS compatible electrolyte followed by nano-ESI analysis to identify more than 50 metabolites.[144] Efficient partitioning of hydrophilic and hydrophobic analytes was observed into the aqueous and hydrophobic electrolyte phase respectively providing additional insight on the molecular properties. The field of single cell metabolomics and lipidomics can greatly benefit by leveraging the recent headway made in single-cell proteomics given that many principles of sample handling, preparation and analysis using MS will be relevant across disciplines. This will open up the exciting prospect of multi-omic analyses of single cells.

Acknowledgments

The authors would like to acknowledge the following funding for support. The National Institutes of Health (NIH), National Institute of Allergy and Infectious Diseases grant U19AI106772 (T.O.M.); National Aeronautics and Space Administration, Exobiology Program (T.O.M.); NIH National Cancer Institute grant R33 CA225248 (R.T.K.); NIH National Institute of Biomedical Imaging and Bioengineering grant R21 EB020976 (R.T.K.); NIH National Heart, Lung, and Blood Institute grant U01 HL122703 (C.A.); National Institute of Diabetes and Digestive and Kidney Diseases grant UC4DK108101 (C.A.); and the Pacific Northwest National Laboratory (PNNL) Laboratory Directed Research and Development Program via the Microbiomes in Transition Initiative. A portion of results presented in this paper was based on data from analyses performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the U.S. Department of Energy (DOE) Office of Biological and Environmental Research, and located at PNNL. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.

References

1. Altschuler SJ and Wu LF, Cell, 2010, 141, 559–563. [PubMed: 20478246]
2. Huang S, Development, 2009, 136, 3853–3862. [PubMed: 19906852]
3. Dagogo-Jack I and Shaw AT, Nature Reviews Clinical Oncology, 2018, 15, 81.
4. Braun S, Vogl FD, Naume B, Janni W, Osborne MP, Coombes RC, Schlimok G, Diel IJ, Gerber B and Gebauer G, New England journal of medicine, 2005, 353, 793–802. [PubMed: 16120859]
5. Braun S, Kentenich C, Janni W, Hepp F, de Waal J, Willgeroth F, Sommer H and Pantel K, Journal of clinical oncology, 2000, 18, 80-80. [PubMed: 10623696]
6. Hayes DF, Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Miller MC, Matera J, Allard WJ, Doyle GV and Terstappen LW, Clinical Cancer Research, 2006, 12, 4218–4224. [PubMed: 16857794]
7. Pantel K, Brakenhoff RH and Brandt B, Nature Reviews Cancer, 2008, 8, 329. [PubMed: 18404148]
8. Nature Biotechnology, 2016, 34, 1077.
9. Lee PY, Chin SF, Neoh HM and Jamal R, J Biomed Sci, 2017, 24, 36. [PubMed: 28606141]
10. Levin PA and Angert ER, Cold Spring Harbor perspectives in biology, 2015, a019216. [PubMed: 26054743]
11. Macaulay IC and Voet T, PLoS genetics, 2014, 10, e1004126. [PubMed: 24497842]
12. Deleye L, Tilleman L, Vander Plaetsen A-S, Cornelis S, Deforce D and Van Nieuwerburgh F, Scientific Reports, 2017, 7, 3422. [PubMed: 28611458]
13. Lombard-Banek C, Reddy S, Moody SA and Nemes P, Molecular & Cellular Proteomics, 2016, mcp. M115. 057760.
14. Zhu Y, Piehowski PD, Zhao R, Chen J, Shen Y, Moore RJ, Shukla AK, Petyuk VA, Campbell-Thompson M and Mathews CE, Nature communications, 2018, 9, 882.
15. Zhu Y, Clair G, Chrisler WB, Shen Y, Zhao R, Shukla AK, Moore RJ, Misra RS, Pryhuber GS, Smith RD, Ansong C and Kelly RT, Angew Chem Int Ed Engl, 2018, 57, 12370–12374. [PubMed: 29797682]
16. Budnik B, Levy E, Harmange G and Slavov N, bioRxiv, 2018, 102681.
17. Smith RD, Shen Y and Tang K, Accounts of chemical research, 2004, 37, 269–278. [PubMed: 15096064]

18. Yi L, Piehowski PD, Shi T, Smith RD and Qian W-J, Journal of Chromatography A, 2017.
19. Tomlinson MJ, Tomlinson S, Yang XB and Kirkham J, Journal of tissue engineering, 2013, 4, 2041731412472690. [PubMed: 23440031]
20. Zhang Y, Lyons V and Pappas D, Electrophoresis, 2018, 39, 732–741. [PubMed: 28960354]
21. Davey HM and Kell DB, Microbiological reviews, 1996, 60, 641–696. [PubMed: 8987359]
22. Spitzer MH and Nolan GP, Cell, 2016, 165, 780–791. [PubMed: 27153492]
23. Sarkar A, Hou HW, Mahan AE, Han J and Alter G, Scientific reports, 2016, 6, 23589. [PubMed: 27026280]
24. Espina V, Wulfkuhle JD, Calvert VS, VanMeter A, Zhou W, Coukos G, Geho DH, Petricoin III EF and Liotta LA, Nature protocols, 2006, 1, 586. [PubMed: 17406286]
25. Clair G, Piehowski PD, Nicola T, Kitzmiller JA, Huang EL, Zink EM, Sontag RL, Orton DJ, Moore RJ, Carson JP, Smith RD, Whitsett JA, Corley RA, Ambalavanan N and Ansong C, Scientific reports, 2016, 6, 39223. [PubMed: 28004771]
26. Gutstein HB, Morris JS, Annangudi SP and Sweedler JV, Mass spectrometry reviews, 2008, 27, 316–330. [PubMed: 18271009]
27. Wang H, Qian W-J, Mottaz HM, Clauss TR, Anderson DJ, Moore RJ, Camp DG, Khan AH, Sforza DM and Pallavicini M, Journal of proteome research, 2005, 4, 2397–2403. [PubMed: 16335993]
28. Wang N, Xu M, Wang P and Li L, Analytical chemistry, 2010, 82, 2262–2271. [PubMed: 20092350]
29. Kulak NA, Pichler G, Paron I, Nagaraj N and Mann M, Nature methods, 2014, 11, 319. [PubMed: 24487582]
30. Hughes CS, Foehr S, Garfield DA, Furlong EE, Steinmetz LM and Krijgsveld J, Molecular systems biology, 2014, 10, 757. [PubMed: 25358341]
31. Li S, Plouffe BD, Belov AM, Ray S, Wang X, Murthy SK, Karger BL and Ivanov AR, Molecular & Cellular Proteomics, 2015, mcp. M114. 045724.
32. Chen Q, Yan G, Gao M and Zhang X, Analytical chemistry, 2015, 87, 6674–6680. [PubMed: 26061007]
33. Huang EL, Piehowski PD, Orton DJ, Moore RJ, Qian W-J, Casey CP, Sun X, Dey SK, Burnum-Johnson KE and Smith RD, Endocrinology, 2016, 157, 1307–1314. [PubMed: 26745641]
34. Mellors JS, Jorabchi K, Smith LM and Ramsey JM, Analytical chemistry, 2010, 82, 967–973. [PubMed: 20058879]
35. Zhu Y, Zhang Y-X, Cai L-F and Fang Q, Analytical chemistry, 2013, 85, 6723–6731. [PubMed: 23763273]
36. Zhu Y, Zhang Y-X, Liu W-W, Ma Y, Fang Q and Yao B, Scientific reports, 2015, 5, 9551. [PubMed: 25828383]
37. Chen W, Wang S, Adhikari S, Deng Z, Wang L, Chen L, Ke M, Yang P and Tian R, Anal Chem, 2016, 88, 4864–4871. [PubMed: 27062885]
38. Zhu Y, Dou M, Piehowski PD, Liang Y, Wang F, Chu RK, Chrisler W, Smith JN, Schwarz KC and Shen Y, Molecular & Cellular Proteomics, 2018, mcp. TIR118. 000686.
39. Xu K, Zhu Y, Liang Y, Piehowski PD, Dou M, Schwarz KC, Zhao R, Sontag RL, Moore R and Kelly RT, Analytical chemistry, 2018, Submitted.
40. Liang Y, Zhu Y, Dou M, Xu K, Chu RK, William B, Zhao R, Hixson KK and Kelly RT, Analytical chemistry, 2018, Submitted.
41. Zhu Y, Podolak J, Zhao R, Shukla AK, Moore RJ, Thomas GV and Kelly RT, Anal Chem, 2018, 90, 11756–11759. [PubMed: 30269481]
42. Dou M, Zhu Y, Liyu A, Liang Y, Chen J, Piehowski P, Xu K, Zhao R, Moore RJ and Atkinson MA, Chemical Science, 2018, DOI: 10.1039/C8SC02680G.
43. Chen Q, Yan G, Gao M and Zhang X, Anal Chem, 2015, 87, 6674–6680. [PubMed: 26061007]
44. Wisniewski JR, Ostasiewicz P and Mann M, J Proteome Res, 2011, 10, 3040–3049. [PubMed: 21526778]
45. Kasuga K, Katoh Y, Nagase K and Igarashi K, Proteomics, 2017, 17. [PubMed: 29275045]

46. Li ZY, Huang M, Wang XK, Zhu Y, Li JS, Wong CCL and Fang Q, Anal Chem, 2018, 90, 5430–5438. [PubMed: 29551058]
47. Wilm MS and Mann M, International Journal of Mass Spectrometry and Ion Processes, 1994, 136, 167–180.
48. Wilm M and Mann M, Analytical chemistry, 1996, 68, 1–8. [PubMed: 8779426]
49. Kelly RT, Page JS, Luo Q, Moore RJ, Orton DJ, Tang K and Smith RD, Analytical chemistry, 2006, 78, 7796–7801. [PubMed: 17105173]
50. Marginean I, Tang K, Smith RD and Kelly RT, Journal of the American Society for Mass Spectrometry, 2014, 25, 30–36. [PubMed: 24122304]
51. Reschke BR and Timperman AT, Journal of the American Society for Mass Spectrometry, 2011, 22, 2115–2124. [PubMed: 21989703]
52. Shen Y, Zhao R, Berger SJ, Anderson GA, Rodriguez N and Smith RD, Analytical chemistry, 2002, 74, 4235–4249. [PubMed: 12199598]
53. Shen Y, Tolić N, Masselon C, Paša-Tolić L, Camp DG, Hixson KK, Zhao R, Anderson GA and Smith RD, Analytical chemistry, 2004, 76, 144–154. [PubMed: 14697044]
54. Shen Y, Moore RJ, Zhao R, Blonder J, Auberry DL, Masselon C, Paša-Tolić L, Hixson KK, Auberry KJ and Smith RD, Analytical chemistry, 2003, 75, 3596–3605. [PubMed: 14570215]
55. Rogeberg M, Vehus T, Grutle L, Greibrokk T, Wilson SR and Lundanes E, Journal of separation science, 2013, 36, 2838–2847. [PubMed: 23813982]
56. Luo Q, Yue G, Valaskovic GA, Gu Y, Wu S-L and Karger BL, Analytical chemistry, 2007, 79, 6174–6181. [PubMed: 17625912]
57. Sun L, Zhu G, Zhao Y, Yan X, Mou S and Dovichi NJ, Angewandte Chemie International Edition, 2013, 52, 13661–13664. [PubMed: 24173663]
58. Lombard-Banek C, Moody SA and Nemes P, Angewandte Chemie International Edition, 2016, 55, 2454–2458. [PubMed: 26756663]
59. Kelly RT, Tolmachev AV, Page JS, Tang K and Smith RD, Mass spectrometry reviews, 2010, 29, 294–312. [PubMed: 19391099]
60. Shaffer SA, Tang K, Anderson GA, Prior DC, Udseth HR and Smith RD, Rapid Commun. Mass Spectrom, 1997, 11, 1813–1817.
61. Shaffer SA, Prior DC, Anderson GA, Udseth HR and Smith RD, Analytical chemistry, 1998, 70, 4111–4119. [PubMed: 9784749]
62. Kim T, Udseth HR and Smith RD, Analytical chemistry, 2000, 72, 5014–5019. [PubMed: 11055723]
63. Page JS, Tang K, Kelly RT and Smith RD, Analytical chemistry, 2008, 80, 1800–1805. [PubMed: 18237189]
64. Marginean I, Page JS, Tolmachev AV, Tang K and Smith RD, Analytical chemistry, 2010, 82, 9344–9349. [PubMed: 21028835]
65. Suhr H, Plasma Chromatography, Plenum Press, New York, 1984.
66. Gabelica V and Marklund E, Curr Opin Chem Biol, 2018, 42, 51–59. [PubMed: 29154177]
67. Zhang X, Quinn K, Cruickshank-Quinn C, Reisdorph R and Reisdorph N, Curr Opin Chem Biol, 2018, 42, 60–66. [PubMed: 29161611]
68. Ben-Nissan G and Sharon M, Current Opinion in Chemical Biology, 2018, 42, 25–33. [PubMed: 29128665]
69. Manz C and Pagel K, Current Opinion in Chemical Biology, 2018, 42, 16–24. [PubMed: 29107930]
70. Hinz C, Liggi S and Griffin JL, Current Opinion in Chemical Biology, 2018, 42, 42–50. [PubMed: 29145156]
71. Deng L, Garimella SVB, Hamid AM, Webb IK, Attah IK, Norheim RV, Prost SA, Zheng X, Sandoval JA, Baker ES, Ibrahim YM and Smith RD, Anal Chem, 2017, 89, 6432–6439. [PubMed: 28497957]
72. Deng L, Ibrahim Yehia M, Baker Erin S, Aly Noor A, Hamid Ahmed M, Zhang X, Zheng X, Garimella Sandilya VB, Webb Ian K, Prost Spencer A, Sandoval Jeremy A, Norheim Randolph V,

Anderson Gordon A, Tolmachev Aleksey V and Smith Richard D, ChemistrySelect, 2016, 1, 2396–2399. [PubMed: 28936476]

73. Ibrahim YM, Hamid AM, Deng L, Garimella SVB, Webb IK, Baker ES and Smith RD, Analyst, 2017, 142, 1010–1021. [PubMed: 28262893]
74. Deng L, Webb IK, Garimella SVB, Hamid AM, Zheng X, Norheim RV, Prost SA, Anderson GA, Sandoval JA, Baker ES, Ibrahim YM and Smith RD, Anal Chem, 2017, 89, 4628–4634. [PubMed: 28332832]
75. Garimella SV, Hamid AM, Deng L, Ibrahim YM, Webb IK, Baker ES, Prost SA, Norheim RV, Anderson GA and Smith RD, Anal Chem, 2016, 88, 11877–11885. [PubMed: 27934097]
76. Deng L, Ibrahim YM, Hamid AM, Garimella SV, Webb IK, Zheng X, Prost SA, Sandoval JA, Norheim RV, Anderson GA, Tolmachev AV, Baker ES and Smith RD, Anal Chem, 2016, 88, 8957–8964. [PubMed: 27531027]
77. Chouinard CD, Nagy G, Webb IK, Shi T, Baker ES, Prost SA, Liu T, Ibrahim YM and Smith RD, Anal Chem, 2018, 90, 10889–10896. [PubMed: 30118596]
78. Chen T-C, Ibrahim YM, Webb IK, Garimella SVB, Zhang X, Hamid AM, Deng L, Karnesky WE, Prost SA, Sandoval JA, Norheim RV, Anderson GA, Tolmachev AV, Baker ES and Smith RD, Analytical Chemistry, 2016, 88, 1728–1733. [PubMed: 26752262]
79. Chen TC, Ibrahim YM, Webb IK, Garimella SV, Zhang X, Hamid AM, Deng L, Karnesky WE, Prost SA, Sandoval JA, Norheim RV, Anderson GA, Tolmachev AV, Baker ES and Smith RD, Anal Chem, 2016, 88, 1728–1733. [PubMed: 26752262]
80. Pasa-Tolic L, Masselon C, Barry RC, Shen Y and Smith RD, Biotechniques, 2004, 37, 621–624, 626-633, 636 passim. [PubMed: 15517975]
81. Cox J and Mann M, Nat Biotechnol, 2008, 26, 1367–1372. [PubMed: 19029910]
82. Shen X, Shen S, Li J, Hu Q, Nie L, Tu C, Wang X, Poulsen DJ, Orsburn BC, Wang J and Qu J, Proc Natl Acad Sci U S A, 2018, 115, E4767–E4776. [PubMed: 29743190]
83. Murgia M, Nagaraj N, Deshmukh AS, Zeiler M, Cancellara P, Moretti I, Reggiani C, Schiaffino S and Mann M, EMBO Rep, 2015, 16, 387–395. [PubMed: 25643707]
84. Hughes CS, Foehr S, Garfield DA, Furlong EE, Steinmetz LM and Krijgsveld J, Mol Syst Biol, 2014, 10, 757. [PubMed: 25358341]
85. Zhu Y, Zhao R, Piehowski PD, Moore RJ, Lim S, Orphan VJ, Paša-Tolić L, Qian W-J, Smith RD and Kelly RT, International journal of mass spectrometry, 2018, 427, 4–10. [PubMed: 29576737]
86. Stanley JR, Adkins JN, Slysz GW, Monroe ME, Purvine SO, Karpievitch YV, Anderson GA, Smith RD and Dabney AR, Anal Chem, 2011, 83, 6135–6140. [PubMed: 21692516]
87. Meier F, Beck S, Grassl N, Lubeck M, Park MA, Raether O and Mann M, J Proteome Res, 2015, 14, 5378–5387. [PubMed: 26538118]
88. Stow SM, Causon TJ, Zheng X, Kurulugama RT, Mairinger T, May JC, Rennie EE, Baker ES, Smith RD, McLean JA, Hann S and Fjeldsted JC, Anal Chem, 2017, 89, 9048–9055. [PubMed: 28763190]
89. Crowell KL, Baker ES, Payne SH, Ibrahim YM, Monroe ME, Slysz GW, LaMarche BL, Petyuk VA, Piehowski PD, Danielson WF, 3rd, Anderson GA and Smith RD, Int J Mass Spectrom, 2013, 354–355, 312-317.
90. Moruz L, Tomazela D and Kall L, J Proteome Res, 2010, 9, 5209–5216. [PubMed: 20735070]
91. Krokhin OV, Ying S, Cortens JP, Ghosh D, Spicer V, Ens W, Standing KG, Beavis RC and Wilkins JA, Anal Chem, 2006, 78, 6265–6269. [PubMed: 16944911]
92. Petritis K, Kangas LJ, Ferguson PL, Anderson GA, Pasa-Tolic L, Lipton MS, Auberry KJ, Strittmatter EF, Shen Y, Zhao R and Smith RD, Anal Chem, 2003, 75, 1039–1048. [PubMed: 12641221]
93. Eng JK, McCormack AL and Yates JR, J Am Soc Mass Spectrom, 1994, 5, 976–989. [PubMed: 24226387]
94. Perkins DN, Pappin DJ, Creasy DM and Cottrell JS, Electrophoresis, 1999, 20, 3551–3567. [PubMed: 10612281]
95. Kyle JE, Crowell KL, Casey CP, Fujimoto GM, Kim S, Dautel SE, Smith RD, Payne SH and Metz TO, Bioinformatics, 2017, 33, 1744–1746. [PubMed: 28158427]

96. Kind T, Liu KH, Lee DY, DeFelice B, Meissen JK and Fiehn O, Nat Methods, 2013, 10, 755–758. [PubMed: 23817071]

97. Rudd P, Karlsson NG, Khoo KH and Packer NH, in Essentials of Glycobiology, eds. rd, Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL and Seeberger PH, Cold Spring Harbor (NY), 2015, DOI: 10.1101/glycobiology.3e.051, pp. 653–666.

98. Gabius HJ, Biosystems, 2018, 164, 102–111. [PubMed: 28709806]

99. Boutanaev AM, Moses T, Zi J, Nelson DR, Mugford ST, Peters RJ and Osbourn A, Proc Natl Acad Sci U S A, 2015, 112, E81–88. [PubMed: 25502595]

100. Lange BM, Annu Rev Plant Biol, 2015, 66, 139–159. [PubMed: 25621517]

101. Neumann S and Bocker S, Anal Bioanal Chem, 2010, 398, 2779–2788. [PubMed: 20936272]

102. Brown M, Dunn WB, Dobson P, Patel Y, Winder CL, Francis-McIntyre S, Begley P, Carroll K, Broadhurst D, Tseng A, Swainston N, Spasic I, Goodacre R and Kell DB, Analyst, 2009, 134, 1322–1332. [PubMed: 19562197]

103. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ and Viant MR, Metabolomics, 2007, 3, 211–221. [PubMed: 24039616]

104. Castle AL, Fiehn O, Kaddurah-Daouk R and Lindon JC, Brief Bioinform, 2006, 7, 159–165. [PubMed: 16772263]

105. Beisken S, Eiden M and Salek RM, Expert Rev Mol Diagn, 2015, 15, 97–109. [PubMed: 25354566]

106. Uppal K, Walker DI, Liu K, Li S, Go YM and Jones DP, Chem Res Toxicol, 2016, 29, 1956–1975. [PubMed: 27629808]

107. N. I. o. Health, RFA-RM-17-013: Compound Identification Development Cores (U2C)).

108. Hou S, Stevenson K and Harynuk JJ, J Sep Sci, 2018, 41, 2544–2552. [PubMed: 29579350]

109. Mollerup CB, Mardal M, Dalsgaard PW, Linnet K and Barron LP, J Chromatogr A, 2018, 1542, 82–88. [PubMed: 29472071]

110. Hu M, Muller E, Schymanski EL, Ruttkies C, Schulze T, Brack W and Krauss M, Anal Bioanal Chem, 2018, 410, 1931–1941. [PubMed: 29380019]

111. Bruderer T, Varesio E and Hopfgartner G, J Chromatogr B Analyt Technol Biomed Life Sci, 2017, 1071, 3–10.

112. Stanstrup J, Neumann S and Vrhovsek U, Anal Chem, 2015, 87, 9421–9428. [PubMed: 26289378]

113. Wolf S, Schmidt S, Muller-Hannemann M and Neumann S, BMC Bioinformatics, 2010, 11, 148. [PubMed: 20307295]

114. Ridder L, van der Hooft JJ, Verhoeven S, de Vos RC, van Schaik R and Vervoort J, Rapid Commun Mass Spectrom, 2012, 26, 2461–2471. [PubMed: 22976213]

115. Allen F, Pon A, Wilson M, Greiner R and Wishart D, Nucleic Acids Res, 2014, 42, W94–99. [PubMed: 24895432]

116. Duhrkop K, Shen H, Meusel M, Rousu J and Bocker S, Proc Natl Acad Sci U S A, 2015, 112, 12580–12585. [PubMed: 26392543]

117. Qiu F, Fine DD, Wherritt DJ, Lei Z and Sumner LW, Anal Chem, 2016, 88, 11373–11383. [PubMed: 27934098]

118. Zhou Z, Shen X, Tu J and Zhu ZJ, Anal Chem, 2016, 88, 11084–11091. [PubMed: 27768289]

119. Zhou Z, Tu J, Xiong X, Shen X and Zhu ZJ, Anal Chem, 2017, 89, 9559–9566. [PubMed: 28764323]

120. Paglia G, Williams JP, Menikarachchi L, Thompson JW, Tyldesley-Worster R, Halldorsson S, Rolfsson O, Moseley A, Grant D, Langridge J, Palsson BO and Astarita G, Anal Chem, 2014, 86, 3985–3993. [PubMed: 24640936]

121. Smith DF, Podgorski DC, Rodgers RP, Blakney GT and Hendrickson CL, Anal Chem, 2018, 90, 2041–2047. [PubMed: 29303558]

122. Shaw JB, Lin TY, Leach FE, 3rd, Tolmachev AV, Tolic N, Robinson EW, Koppenaal DW and Pasa-Tolic L, J Am Soc Mass Spectrom, 2016, 27, 1929–1936. [PubMed: 27734325]
123. Jones DP, Toxicol Rep, 2016, 3, 29–45. [PubMed: 26722641]
124. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, Wang J, Yu B, Zhang J and Bryant SH, Nucleic Acids Res, 2016, 44, D1202–1213. [PubMed: 26400175]
125. Dobson CM, Nature, 2004, 432, 824–828. [PubMed: 15602547]
126. Kaszycki JL and Shvartsburg AA, Journal of the American Society for Mass Spectrometry, 2017, 28, 294–302. [PubMed: 27975328]
127. Shvartsburg AA, Siu KWM and Clemmer DE, Journal of the American Society for Mass Spectrometry, 2001, 12, 885–888. [PubMed: 11506220]
128. Dear GJ, Munoz-Muriedas J, Beaumont C, Roberts A, Kirk J, Williams JP and Campuzano I, Rapid Commun. Mass Spectrom, 2010, 24, 3157–3162. [PubMed: 20941763]
129. Zheng X, Renslow RS, Makola MM, Webb IK, Deng L, Thomas DG, Govind N, Ibrahim YM, Kabanda MM, Dubery IA, Heyman HM, Smith RD, Madala NE and Baker ES, J Phys Chem Lett, 2017, 8, 1381–1388. [PubMed: 28267339]
130. Zhou ZW, Xiong X and Zhu ZJ, Bioinformatics, 2017, 33, 2235–2237. [PubMed: 28334295]
131. Soper-Hopper MT, Petrov AS, Howard JN, Yu SS, Forsythe JG, Grover MA and Fernandez FM, Chem Commun (Camb), 2017, 53, 7624–7627. [PubMed: 28640293]
132. Masson A, Kamrath MZ, Perez MA, Glover MS, Rothlisberger U, Clemmer DE and Rizzo TR, J Am Soc Mass Spectrom, 2015, 26, 1444–1454. [PubMed: 26091889]
133. Kamrath MZ and Rizzo TR, Acc Chem Res, 2018, 51, 1487–1495. [PubMed: 29746100]
134. Coates J, Encyclopedia of analytical chemistry, 2000, 12, 10815–10837.
135. Bouteiller Y, Gillet JC, Gregoire G and Schermann JP, J Phys Chem A, 2008, 112, 11656–11660. [PubMed: 18942809]
136. Martens J, Berden G, Bentlage H, Coene KLM, Engelke UF, Wishart D, van Scherpenzeel M, Kluijtmans LAJ, Wevers RA and Oomens J, J Inherit Metab Dis, 2018, 41, 367–377. [PubMed: 29556837]
137. Martens J, Berden G, van Outersterp RE, Kluijtmans LAJ, Engelke UF, van Karnebeek CDM, Wevers RA and Oomens J, Sci Rep, 2017, 7, 3363. [PubMed: 28611404]
138. Masellis C, Khanal N, Kamrath MZ, Clemmer DE and Rizzo TR, J Am Soc Mass Spectrom, 2017, 28, 2217–2222. [PubMed: 28643189]
139. Zhu Y, Clair G, Chrisler WB, Shen Y, Zhao R, Shukla AK, Moore RJ, Misra RS, Pryhuber GS, Smith RD, Ansong C and Kelly RT, Angew Chem Int Ed Engl, 2018, DOI: 10.1002/anie.201802843.
140. Meier F, Geyer PE, Winter SV, Cox J and Mann M, Nature methods, 2018, 1.
141. Deng L, Ibrahim YM, Baker ES, Aly NA, Hamid AM, Zhang X, Zheng X, Garimella SV, Webb IK and Prost SA, ChemistrySelect, 2016, 1, 2396–2399. [PubMed: 28936476]
142. Deng L, Ibrahim YM, Hamid AM, Garimella SV, Webb IK, Zheng X, Prost SA, Sandoval JA, Norheim RV and Anderson GA, Analytical chemistry, 2016, 88, 8957–8964. [PubMed: 27531027]
143. Zhang X-C, Zang Q, Zhao H, Ma X, Pan X, Feng J, Zhang S, Zhang R, Abliz Z and Zhang X, Analytical Chemistry, 2018.
144. Yin R, Prabhakaran V and Laskin J, Analytical Chemistry, 2018, 90, 7937–7945. [PubMed: 29874047]

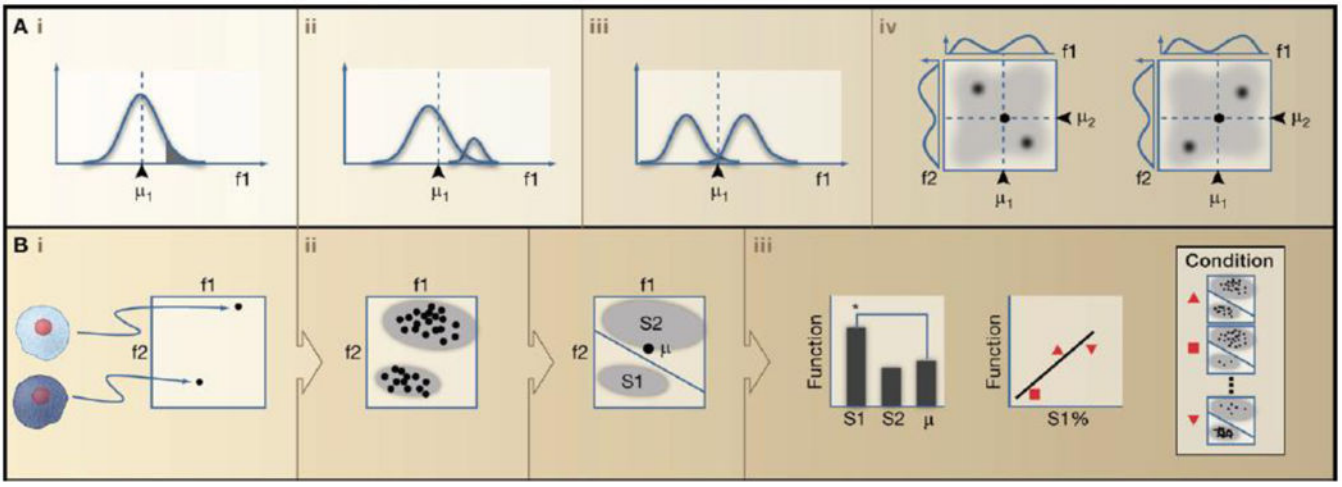

Figure 1. Limitations of bulk cell analysis vs benefits of single cell analysis.

(A) Population averages can mask cell heterogeneity. The mean measurement (indicated by dashed lines) of a population may not capture (i) the shaded tail of the distribution, (ii) a subpopulation or (iii) majority of the cells in case of bimodal behavior. (iv) Univariate analysis of a single measurement from individual cells may not be able to distinguish correlated (left) or anticorrelated (right) expression of cells (f1 and f2 indicate single cell measurements). (B) Interpreting functional significance from heterogeneity. (i) Individual cells can be represented as points in a feature space (ii) Cells can be partitioned into subpopulations (eg: S1 and S2) in regions of the feature space (iii) The presence of significant differences between subpopulations or ensemble averages can be tested. One can assess how informative is an entire decomposition of heterogeneity (middle and right). Reproduced from Altschuler and Wu[1] with permission from Elsevier, Copyright 2010.

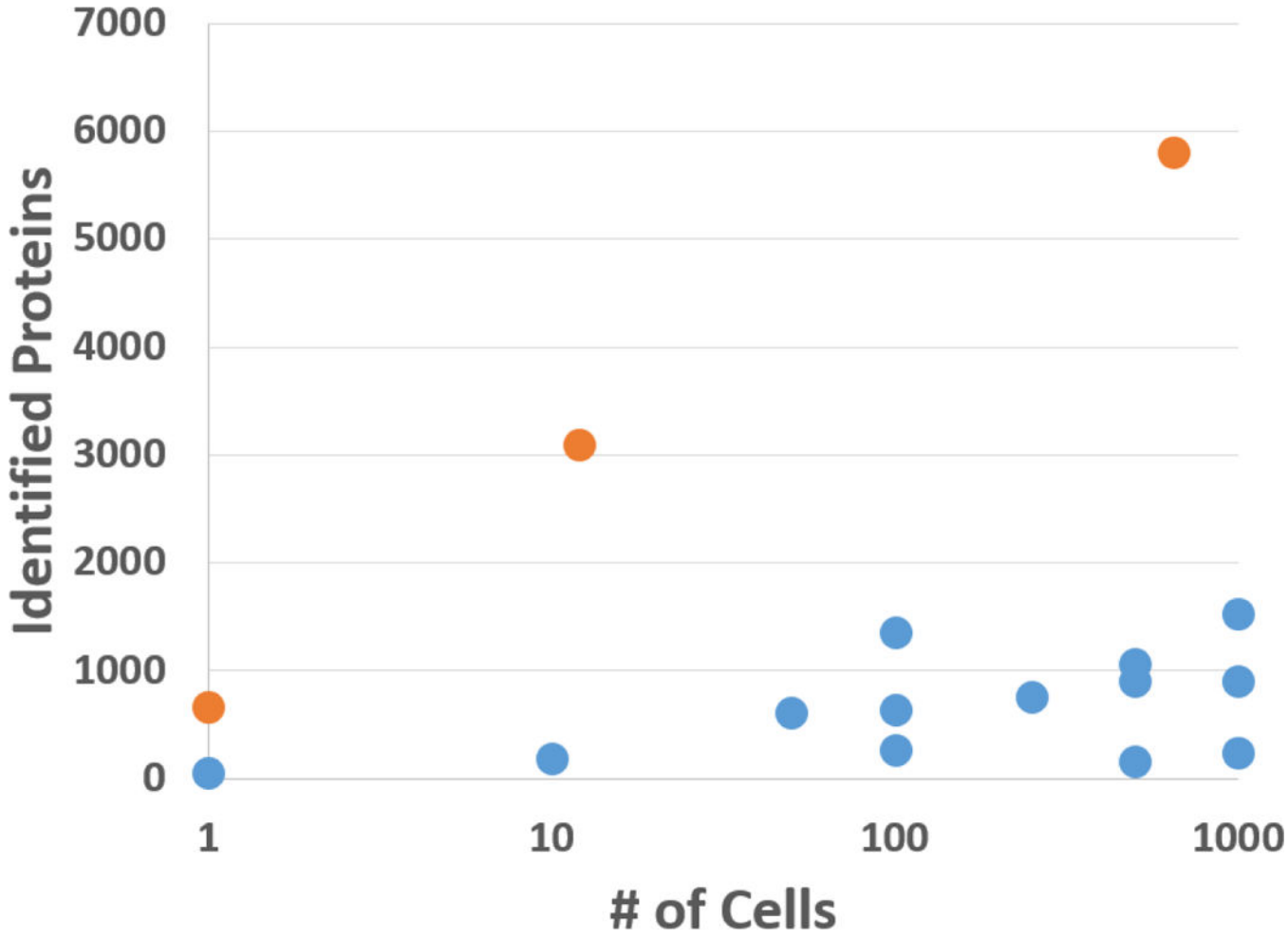

Figure 2. Proteome coverage achieved using nanoPOTS (orange) and other (blue) technologies for samples comprising ≤1000 mammalian somatic cells.

Orange data points, from left to right, are from Zhu et al.,[15] Zhu et al.,[14] and Dou et al.[42] Blue data points represent proteome coverage reported by Wang et al.[28], Zhang et al.,[43] Wiśniewski et al.,[44] Kasuga et al.[45] and Li et al.[46]

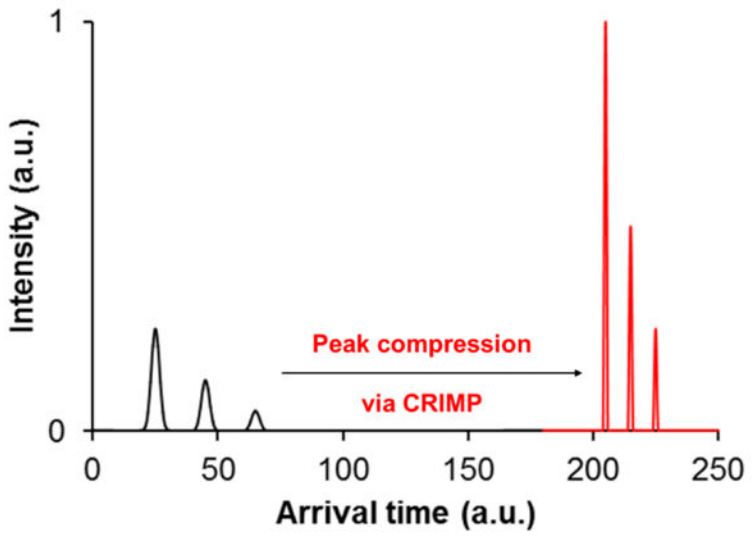

Figure 3.

Depiction of the CRIMP step, where three broad ions are successfully compressed into much narrower ones following CRIMP, thereby increasing signal intensity, and thus S/N.

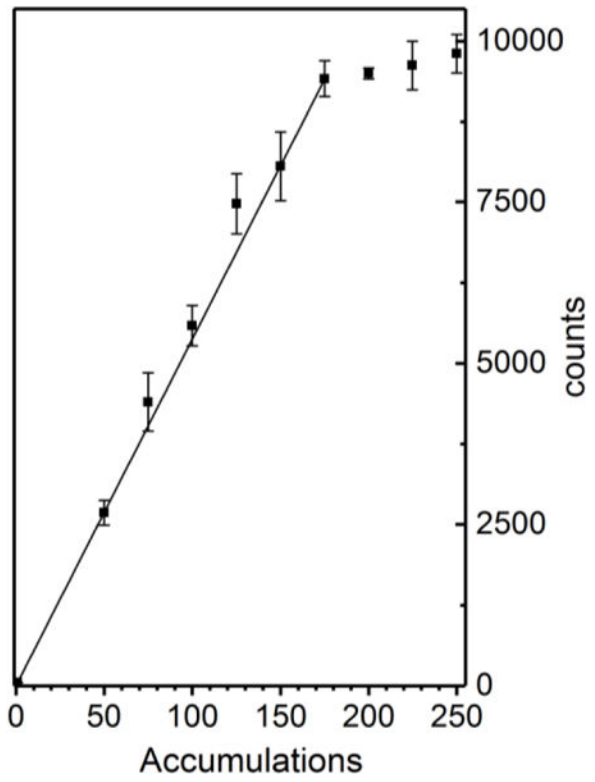

Figure 4. Demonstration of ion enrichment in SLIM.

As an ion is enriched longer (more accumulations), its signal intensity will continue to linearly increase until reaching a horizontal asymptote. Reproduced from Chen et al[79] with permission from American Chemical Society, Copyright 2016.

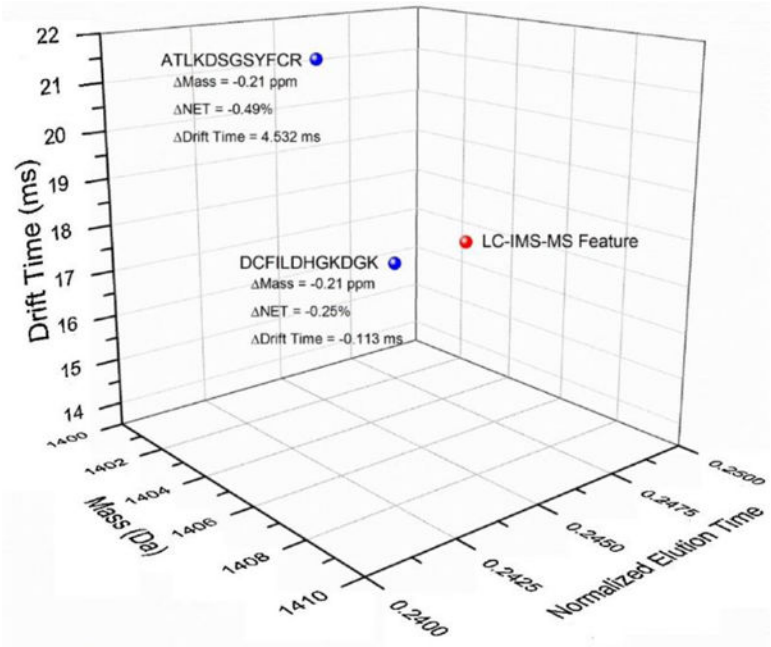

Figure 5. Identification of a peptide feature based on comparison of experimental properties vs those contained in a reference library.

A peptide was detected in a LC-IMS-MS experiment with a measured mass and elution time (red dot) that could match to two different peptides in a reference library (blue dots), when considering only mass and elution time. The mass measurement errors between the experimental mass and library masses are identical, while the normalized elution time errors are different but not sufficiently so to enable a unique identification. Drift time, as measured by drift tube IM, was significantly different between the two library entries to enable the peptide to be identified as DCFILDHGKDGK. Reproduced from Crowell et al.[89] with permission from Elsevier, Copyright 2013.

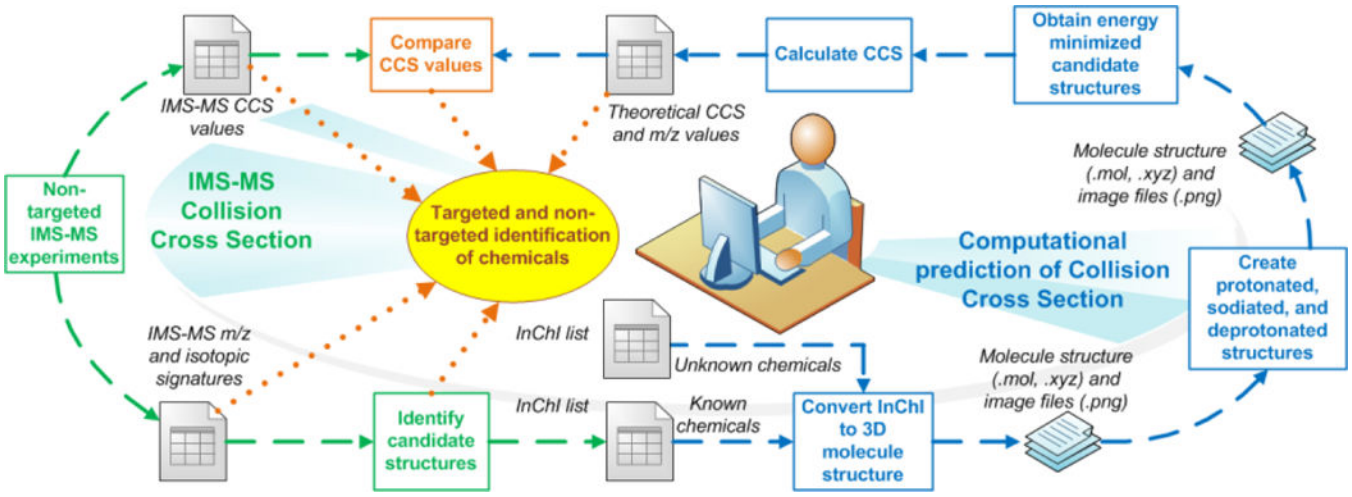

Figure 6.

Representative workflow for identifying molecules based on matching of experimental m/z and CCS to computationally predicted values.

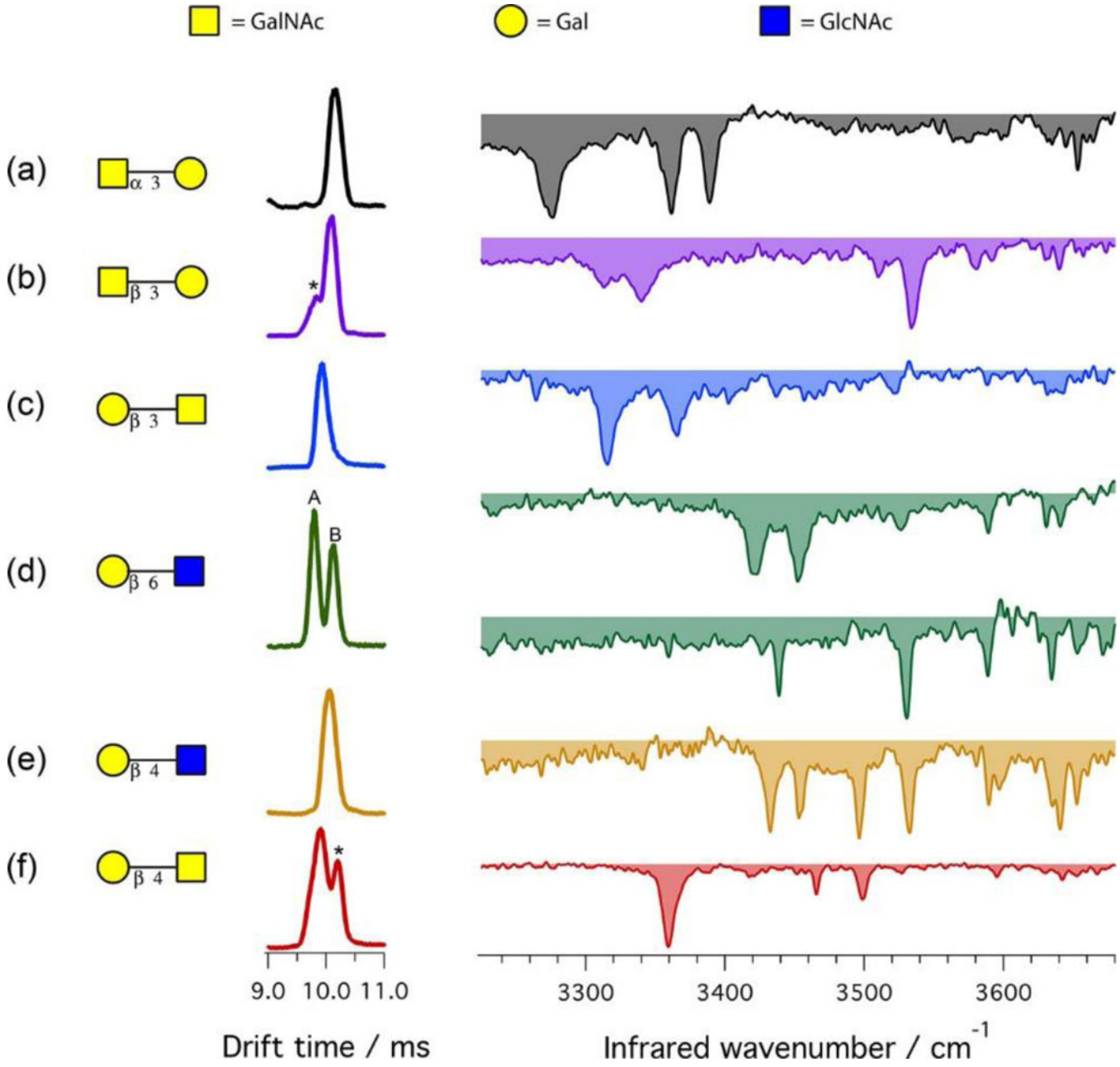

Figure 7. Drift times and infrared spectra for six isomeric disaccharides.

The drift time distributions for the six isobaric disaccharides overlap, and so they would not be distinguishable by IMS alone. However, their vibrational spectra are very different, providing a means for their identification. Reproduced from Masellis et al.[138] with permission from Springer Nature, Copyright 2017.

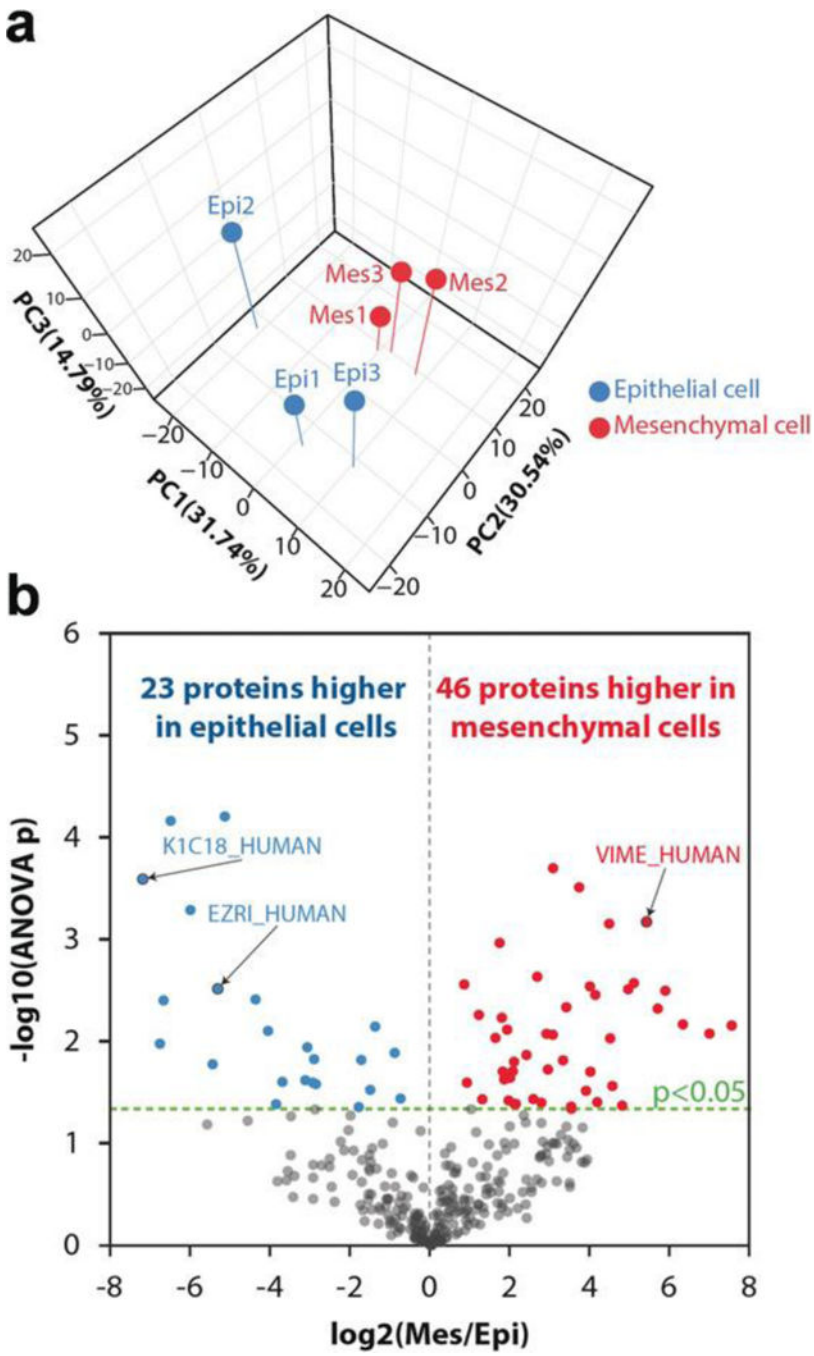

Figure 8. Principal component analysis (PCA) of label-free proteomics data from FACS-sorted cells.

a) Unsupervised PCA based on label-free quantification of proteins expressed in epithelial and mesenchymal cells from human lung. b) Volcano plot of differentially expressed proteins. Epithelial cell Replicate 2 was excluded for this analysis. Reproduced from Zhu et al.[139] with permission from John Wiley and Sons, Copyright 2018.