Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

J Proteome Res. Author manuscript; available in PMC 2024 August 03.

Published in final edited form as:

J Proteome Res. 2024 August 02; 23(8): 2700–2722. doi:10.1021/acs.jproteome.3c00839.

Recent Advancements in Subcellular Proteomics: Growing Impact of Organellar Protein Niches on the Understanding of Cell Biology

Vanya Bhushan,

Functional Cellular Networks Section, Laboratory of Immune System Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, United States

Aleksandra Nita-Lazar

Functional Cellular Networks Section, Laboratory of Immune System Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, United States

Abstract

The mammalian cell is a complex entity, with membrane-bound and membrane-less organelles playing vital roles in regulating cellular homeostasis. Organellar protein niches drive discrete biological processes and cell functions, thus maintaining cell equilibrium. Cellular processes such as signaling, growth, proliferation, motility, and programmed cell death require dynamic protein movements between cell compartments. Aberrant protein localization is associated with a wide range of diseases. Therefore, analyzing the subcellular proteome of the cell can provide a comprehensive overview of cellular biology. With recent advancements in mass spectrometry, imaging technology, computational tools, and deep machine learning algorithms, studies pertaining to subcellular protein localization and their dynamic distributions are gaining momentum. These studies reveal changing interaction networks because of “moonlighting proteins” and serve as a discovery tool for disease network mechanisms. Consequently, this review aims to provide a comprehensive repository for recent advancements in subcellular proteomics subcontexting methods, challenges, and future perspectives for method developers. In summary, subcellular proteomics is crucial to the understanding of the fundamental cellular mechanisms and the associated diseases.

Graphical Abstract

Corresponding Author: Aleksandra Nita-Lazar – Functional Cellular Networks Section, Laboratory of Immune System Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, United States; nitalazarau@niaid.nih.gov.

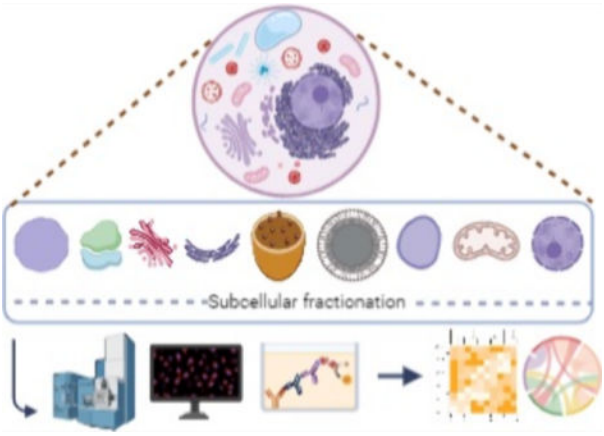
Author Contributions

Conceptualization: V.B. and A.N.L. and writing: V.B (initial draft, editing, and final draft) and A.N.L. (editing and final draft). All authors read and commented on the manuscript.

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.3c00839

The authors declare no competing financial interest.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Keywords

subcellular localization; mass spectrometry; imaging technology; proximity protein labeling; dynamic organellar maps; BioID; APEX; computational tools; machine learning algorithms

INTRODUCTION

Cellular compartmentalization and dynamic distribution of cellular macromolecules, such as proteins, are essential for maintaining cell function and homeostasis.[1] Proteins are vital in bridging the gap between genetic information and disease phenotypes.[2,3] With the advancement in proteomic approaches and machine learning tools, proteomes can be studied at the subcellular level with ever-increasing depth. The subcellular organization of proteins directly impacts their function, as the compartmentalization of cells changes the microenvironment with distinct molecular compositions, chemical properties, and physical attributes. Subcellular proteomics is an emerging dimension of proteomics used to study proteomes at the organellar level.[4] Eukaryotic cellular compartmentalization includes many membrane-bound organelles, such as Golgi apparatus, mitochondria, lysosome, endoplasmic reticulum, and nucleus, as well as membrane-less structures like cytosol, centrosomes, polysomes, and ribonucleoprotein granules.[5,6] Proteins often localize to distinct subcellular niches according to their functions. Their movement is tightly regulated through cellular processes such as signaling mechanisms, cell motility, cell progression, macromolecule interactome, and apoptosis.[7] Distinct localization of proteins into cellular organelles is vital since incorrect protein localization has been reportedly associated with cellular dysfunction and diseases such as neurodegenerative diseases,[8,9] cancer,[10] obesity,[11] laminopathies,[12] liver disease,[13] and metabolic disorders.[14] Therefore, understanding the subcellular distribution of proteins and their dynamics (in homeostatic and external or internal perturbation conditions) is crucial for comprehending vital cell biochemical functions.

Interestingly, most cellular proteins are present in multiple subcellular locations, and their distribution varies in a context-specific manner.[15] Studies have also shown that around 50% of the total cell proteins are multiorganelle localized, reflecting their “moonlighting” activities, which could be entirely independent or organelle modulated.[16,17]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Human subcellular proteome studies suggest that there are single-cell variations at protein location, expression levels, protein–protein interaction networks, and changes in their dynamic translocation, indicating a high level of cellular regulation of protein dynamics.[18] Additionally, with the help of state-of-the-art mass spectrometry advances, it is now possible to integrate subcellular proteomics with quantitative “omics” techniques to explore unbiased system-level insights into biological processes.[19,20] These findings provide us with many opportunities to explore and study the complex world of cellular proteins and their dynamic behavior more constructively.

Recent technological advancements in high-throughput quantitative mass spectrometry,[21,22] imaging-based microscopy,[23] interactome mapping,[24] and computational machine-learning methods for data analysis[25] have revolutionized how we investigate organellar protein functions. The synergistic subcellular proteomic approaches and refined experimental techniques have enabled significant progress in understanding and quantifying subcellular protein domains at single-cell levels,[26] intact tissue specimens,[27,28] and even at subcellular phospho-proteome levels.[29,30]

Subcellular proteomics methods, such as mass spectrometry-based cell-wide dynamic organellar protein mapping and deep fluorescent-based imaging techniques combined with artificial intelligence, are powerful tools for quantifying subcellular proteomes. Our literature review has a broad perspective and aims to provide a comprehensive overview of the methods of subcellular proteomics. We will discuss recent advances and their application in different fields. Additionally, we will cover the latest advancements in artificial intelligence and machine learning tools for subcellular proteomics.

SUBCELLULAR PROTEOMICS METHODS

Subcellular proteomic data acquisition methods for localization and protein abundance detection within intricate subcellular compartments can be divided into three broad categories: organelle or cell-wide fractionation and quantitative mass spectrometry-based methods,[31] imaging-based approaches using tagged/fusion recombinant proteins,[32] and protein–protein interaction studies to identify the local subcellular distribution of proteins through enzyme-linked proximity labeling and affinity purification.[33] Appropriate data acquisition approaches coupled with rigorous data analysis following data acquisition are paramount.

Mass Spectrometry-Based Methods

Mass spectrometry-based acquisition methods provide valuable insights into the proteome maps of cells, along with their subcellular abundance profiles. These methods often rely on cellular fractionation approaches or proximity labeling techniques as prerequisite steps.[34] A diverse range of subcellular fractionation techniques, such as detergent-based fractionation,[35] centrifugation-sedimentation-based,[36] density centrifugation-based methods,[37] electrophoresis,[33] and affinity purification,[38] are available to support MS-based workflows. Furthermore, organellar profiling methods have been extensively applied to cell-wide and single organelles. These techniques can be highly effective in generating

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

accurate data, and further advancements in this area can help improve our understanding of cell biochemical functions (Figure 1).

Single Organelle Protein Profiling.—Single-organelle proteins are a less complex subset of the proteome than the proteome of whole cells or tissues.[39] This makes organellar proteins attractive targets for in-depth biological and functional analysis. By enriching subcellular organelles through subcellular fractionation methods and coupling those methods with mass spectrometry identification, the abundance distribution profile of each protein distinct to that cell compartment can be identified.[40] It should be noted that proteins associated with a given compartment often have similar abundance profiles to outliers or contaminants.[41] However, recent advances in mass spectrometry have made single-organelle protein profiling a feasible approach for identifying proteins in the nucleus,[42] mitochondria,[43] Golgi apparatus,[44] and plasma membrane.[45] This approach generates highly specific data sets, which makes it suitable for addressing targeted research questions such as drug target discoveries. Overall, using single-organelle protein profiling is a promising avenue for deepening our understanding of organelles' biological and functional properties.

Cell-Wide Organellar Protein Profiling.—Multiorganellar protein profiling encompasses all subcellular compartments simultaneously. A prerequisite for multiorganelle protein mapping is physical cell disruption to release intact organelles (Table 1) for subcellular compartment enrichment. Further, intact organelles are enriched by density centrifugation[37] or differential centrifugation[21] methods. Different organelles are enriched in density-dependent methods, but their distribution is not absolute and can significantly overlap. Each fraction protein is identified through quantitative MS, producing an abundance distribution profile for each protein. Machine learning methods, such as cluster analysis, group similar profile proteins together based on organelle marker proteins. Multiorganelle profiling provides genuine systems-level analysis tools that achieve remarkable resolution and coverage of major subcellular compartments in a single experiment.[46]

Subcellular Fractionation and Enrichment Methods

Centrifugation-based Methods.—Biochemical centrifugation fractionation methods enrich specific organelles based on their physical parameters, such as size, mass, and density. These methods provide high sensitivity and coverage, as revealed in MS data analysis. Differential and density centrifugation techniques can isolate organelles from a crude preparation. The differential centrifugation fractionation method exploits the sedimentation property of organelles varying in their densities.[47] Density gradient centrifugation separates cell compartments based on their buoyancy density or sedimentation rate.[21] The selection of gradient media relies on the sample type (Table 2).

These techniques use sucrose[48] Percoll,[49] and Iodixanol[50] as density mediums to separate organelles based on density. Equilibrium density centrifugation (gradient centrifugation) is another method to separate cellular organelles with closely related densities. A nonionic medium of low density, low osmolarity, and viscosity is ideal for separation. Synthetic medium such as Nycodenz[51] is used with great success and better-resolving ability. Natural media (such as sucrose) are limited by changes in osmolarity at higher concentrations.[52] To

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ensure that only target organelle components are enriched, subtractive proteomics[53] filters out contaminants during data analysis. Multiple subcellular fractions collected during target organelle enrichment can be analyzed using the protein correlation profiling technique, based on bona fide organellar marker protein strategy, to distinguish the target niche from the control. Enriching the target organelle adequately is essential, as density centrifugation fractions do not ensure absolute enrichment.[54]

Electrophoresis and Affinity Purification-Based Methods.—Electrophoresis and affinity purification-based methods recently gained momentum for organelle enrichment and proteomic mapping studies coupled with MS data analysis. For example, ER in *Arabidopsis* was enriched using the free-flow electrophoresis (FFE) technique to study the translocation of ER vesicles.[33] The working principle for zone electrophoresis (ZE-FFE) is that the organelles are separated according to their specific surface charge. This technique of narrowing down the organelles through respective surface charges can be used for applications such as target identification during drug discovery. Similarly, flow field-flow fractionation (Fl-FFF) is a size-based separation technique that is highly versatile and capable of separating macromolecules such as DNA, whole cells, subcellular organelles, extracellular species, and protein/protein aggregates. Studies reviewing Fl-FFF application in organelle enrichment studies are available, citing their far reach in proteomic and lipidomic analysis coupled with mass spectrometry.[55] Affinity purification strategies for biochemical enrichment are highly efficient and reliable for isolating/enriching intracellular organelles, such as lysosomes, mitochondria, and peroxisomes from mammalian cells using twin tag and streptavidin variants.[38] The study workflow enriches organelles in just 3 min. This method can help study transient and fast cellular activities like organelle content of small molecular metabolites. The reproducibility of these methods is the major limiting factor in cell subcellular enrichment and purification.

Protein Correlation Profiling

Protein correlation profiling (PCP) allows the profiling of multiprotein complexes enriched by fractionation but not purified to homogeneity.[39] PCP employs distribution profiles of unique marker proteins in cell compartments to determine the subcellular location of uncharacterized proteins.[56] Recent studies using PCP to study organellar protein fraction to identify Golgi membrane remodeling pathway during nutrient stress[57] and subcellular interrogation of the centrosome proteome in hiPSCs coupled with mass spectrometry[58] provide deeper insights into organellar proteomics. PCP has a distinct advantage over other methods as it combines low-resolution differential centrifugation with high-throughput mass spectrometry to generate accurate quantitative protein maps.[59] PCP methods such as localization of organelle proteins using isotope tagging (LOPIT)[60] rely on the comparison of the abundance distribution of each protein with known organelle markers to identify organelle protein distribution patterns[61] and protein dynamics information in the cell.[62] LOPIT and HyperLOPIT are widely used and are the most reproducible methods used in subcellular proteomics. LOPIT is a method used to generate cell-wide protein subcellular map structures from complex biological mixtures in a single experiment.[2] Unlike traditional methods, it does not require absolute organelle purification. Instead, it measures protein distribution across multiple density gradient fractions.[63] Subcellular localization is assigned

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

by comparing protein profiles to those of well-curated organelle markers using multivariate statistical analysis and machine-learning approaches.[64] An improved version of this method, called hyperLOPIT, was developed recently. This upgraded approach integrates novel methods for sample preparation, MS data acquisition, and protein localization classification, and it was used to create a high-resolution map of protein subcellular localization in E14TG2a mouse embryonic stem cells.[16] Other methods for profiling subcellular global proteins, such as Prolocate,[36] COLA,[65] and SubCellBarCode,[66] rely on separating cell organelles using centrifugation gradients and analyzing them with MS. These methods use multivariate statistics along with machine learning algorithms to obtain results. We have compiled a list of recent global subcellular protein profiling methods (Table 3). PCP can even identify changes in protein distribution caused by structural alterations, which are often missed by traditional fractionation methods focused on enriching specific subcellular niches.[67]

Dynamic Organellar Maps

Dynamic organellar maps (DOMs) are a powerful global subcellular protein profiling tool. This method combines cell fractionation and shotgun proteomics to generate protein abundance profiles. The approach is comparative and highly robust due to its fractionation protocol. To generate the profiles, cultured cells are mechanically lysed and separated by differential centrifugation, and the resulting fractions are analyzed by mass spectrometry. The abundance profiles are then used to predict protein localization using supervised machine learning algorithms.[31] DOMs are highly reproducible and allow for the analysis of changes in protein localization. A recent study used DIA-DOMs to analyze subcellular protein translocation in response to lysosomal stress in HeLa cells. The study identified a large number of translocating Golgi proteins with high confidence.[68]

MS Subcellular Proteomics: Database tools

Mass spectrometry-based protein identification of organellar fractionation uses database searching tools such as X! Tandem,[69] Sequest,[70] and Mascot[71] to compare experimental peptide spectra and theoretical in-silico spectra to generate peptide spectral match scores. High PSM scores correspond to a higher probability of a particular peptide in the sample. Common MS data identification software such as Proteome Discoverer, MaxQuant works in data-dependent acquisition mode (DDA).[72] In contrast, applications such as OpenSWATH,[73] Skyline, and DIA-NN[74] are suitable for data-independent acquisition (DIA) mode.

Tools and Software for Analyzing Fractionation Experiments.—Proteins are present in various subcellular locations due to their dynamic nature. Therefore, it is crucial to validate their presence using subcellular markers.[75] To ensure the accuracy of subtractive proteomics data analysis methods, it is necessary to include a reference set of known protein markers of the target organelle and reference lists of protein contaminants from other subcellular compartments. One can choose reference protein lists using Gene Ontology,[76] COMPARTMENTS,[77] and UniProt.[78] OpenCell[79] and Human Protein Atlas[18] tools use cutting-edge technology such as genome engineering, confocal live-cell imaging, mass spectrometry, and data protein engineering to systematically map the localization

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and interaction of human proteins, providing a comprehensive understanding of the human cell subcellular proteome. Correlation profiling data do not require a contaminant reference and can be visualized using several organelle profile plots. Principal component analysis (PCA) is commonly used to simplify complex data by reducing dimensions. PCA can assess subcellular resolution as proteins associated with the same subcellular niche cluster together.[80] Machine learning algorithms, such as supervised, semisupervised, or unsupervised, are also used for correlation profiling. Supervised and semisupervised algorithms require the spatial profile of known organelle marker proteins to assign unknown protein localizations to subcellular niches.[81] These algorithms include support vector machines, neural networks, and random forest models. BANDLE (Bayesian analysis of differential localization experiments) is a semisupervised functional mixture model used to obtain the probability of a protein being differentially localized upon cellular perturbation.[82] Unsupervised clustering algorithms like k-means clustering or DBSCAN are helpful when training nonmodel organisms with limited marker proteins. These methods are best suited for static and single-locale localization of proteins. The aforementioned methods are challenging for the dynamic classification of proteins and MLPs (multilocalizing proteins) but have been implemented through T-augmented Gaussian mixture model approaches and TRANSPIRE (Translocation analysis of subcellular proteomics).[81] Dimension reduction is another approach to simplifying complex multidimensional data to make it easier to interpret. One of the most used linear methods for this is PCA. Still, nonlinear methods like t-distributed stochastic neighbor embedding (t-SNE) help detect more complex patterns, such as polynomial relationships.[83] In addition, unsupervised clustering techniques, such as DBSCAN and hierarchical clustering, can be helpful for marker curation in challenging or poorly researched systems.[84] The choice of machine learning algorithm depends on available computation power, processing timeframes, experience, and appropriate data assumptions.

Proximity Labeling

Proximity labeling-based methods, in combination with mass spectrometry (MS), offer a high-throughput approach for systematically analyzing subcellularly restricted proteomes (Figure 2).

A protein–protein interaction network is used to identify the local subcellular distribution of proteins through enzyme-linked proximity labeling.[85] Proximity labeling involves fusing “bait” protein with labeling enzymes to covalently label neighboring “prey” proteins. Four standard proximity labeling approaches are used in subcellular proteomics. The first approach is BioID and BioID2, which rely on mutated bacterial biotin ligase (*E. coli* proteins BirA,[86] TurboID,[87] miniTurbo variants, and BASU[88]) that can covalently attach biotinylated lysine amines to proximal proteins. Then, biotinylated proteins are extracted or recovered from the complex lysate using streptavidin-agarose or magnetic beads, followed by MS data analysis. Recent publications include a BioID-based map of HEK293 cells[89] and mitochondria.[90] The second approach is APEX and APEX2,[91,92] which biotinylates proteins’ tyrosine residues within a 20 nm radius after peroxide stimulation. This approach does not require specific antibodies, but simultaneous analysis of multiple organelle proteins is the only limiting factor. The third approach is HRP (horseradish peroxidase).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

HRP, a peroxidase, activates upon exposure to H_2O_2 and converts substrates into highly reactive radicals that covalently tag neighboring proteins on electron-rich amino acids.[93] HRP is inactive in reducing environments like the cytosol, limiting its use in intracellular interactomes[94] (thus motivating the development of APEX). However, HRP remains active in oxidizing environments like the ER, Golgi, and extracellular region.[95] Although HRP can catalyze a variety of substrates, for proximity labeling, two have been used extensively: (1) EMARS (enzyme-mediated activation of radical source)[96] and (2) SPPLAT (selective proteomic proximity labeling assay using tyramide), also known as biotin-phenol.[97] The fourth approach is PUP-IT, which stands for pupylation-based interaction tagging. PUP-IT is a novel system for proximity labeling that uses the bacterial enzyme PafA to attach a small protein called Pup to lysine residues on target proteins. This process, known as pupylation, marks the proteins for degradation. PafA is particularly useful for labeling promiscuous proteins since it does not require a consensus binding motif.[98]

Proximity labeling is a powerful technique that provides detailed information about cell subcellular maps. Unlike membrane-less and membrane-bound organelles with a more extensive protein labeling range, proximity labeling has a far shorter range (diameter ranging from nano- to micrometers). This makes it ideal for obtaining highly localized information about proteins and their interactions within cells. With proximity labeling, researchers can gain a deeper understanding of complex cellular processes and pave the way for discoveries in the field of cell biology. Gaining a deeper understanding of disease pathologies requires understanding localized biomolecular networks or microenvironments. For instance, flavoprotein fusion is an exceptional technique that accurately labels surrounding proteins. With the correct probe, it enables precise determination of labeling pathways.[99] Similarly, MicroMap (μMap) is a protein–protein interaction identification technique that specifically labels antibody binding targets and their microenvironment protein neighbors with higher precision. This has been demonstrated in the PD-L1 microenvironment in live lymphocytes.[100]

Data Analysis Tools for Proximity Labeling Approaches.—Labeling strategies face a significant challenge distinguishing candidate protein from background noise in MS data. However, experimental procedures have been developed to generate high-confidence MS data through (i) proximity labeling coupled with quantitative MS using metabolic labeling such as SILAC (stable isotope labeling by amino acid in cell culture)[101] or by chemical labeling tags, such as iTRAQ[102] (isobaric tags for relative and absolute quantification) and TMT[103] (tandem mass tags); (ii) use of multiple negative controls to filter out background labeled proteins;[104] and (iii) reduction of background due to nonspecific labeling by using endogenous CRISPR/Cas9 tagging of bait proteins to maintain physiological levels of the labeling enzyme.[105]

Proximity Labeling: In Vivo Systems.—Proximity labeling is often performed in cultured cells due to its technical advantages, such as easy delivery of labeling reagents and efficient cell lysis of large cell quantities. However, *in vivo*, protein labeling offers specific benefits, such as identifying organelle components or protein interactions from cells in a homeostatic physiological environment, including difficult to grow cells like neurons.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Protein labeling can also be restricted to specific cell types or developmental stages by expressing labeling enzymes from transgenes. These cells can also be transplanted into wild-type host organisms.[106] Proximity labeling approaches have been used in developmental systems like *Arabidopsis*,[107] *C. elegans*,[108] *Drosophila*,[109] and mouse[110] to discover new components of developmental processes. However, penetrating labeling substrate into target tissues and cells is a significant technical challenge of using proximity labeling tools in vivo.

Imaging Methods

Imaging techniques give a microscopic view of protein distribution in the cell and across the organelles. The two distinct ways to study imaging-based cellular macromolecular environments are immunocytochemistry and live cell imaging. Immunocytochemistry works by adding a fluorescent tag to the antibodies/aptamers/nanobodies[111–113] against the targeted protein/s.[114] Live-cell imaging is a technique that involves observing living cells in real time using time-lapse microscopy. Live-cell imaging reagents comprise targeted fluorescent proteins and small, membrane-permeant fluorescent dyes. These reagents are specifically designed to facilitate live-cell imaging experiments. Some reagents are intended for time-lapse imaging over several hours or days, while others are better suited for end-point assays, where cells are stained, imaged, and analyzed immediately.[115] Whether static or dynamic, the method chosen for studying cellular protein maps depends on sample availability, handling expertise, and the necessary instruments and processing techniques. To observe organellar protein movement, it is crucial to maintain cell viability by ensuring optimal temperature and CO_2 levels. Time-lapse microscopy is the most effective approach.[116,117] In contrast, fixed cells labeled with fluorescent antibodies are preferred for mapping proteins difficult to tag or express or for analyzing the static localization of multiple proteins in the cell.[57] However, both methods can be used for comparative studies to detect one protein in different cell types under varying test conditions.[118–120] Staining concentration, incubation time, and imaging interval/frequency should be determined empirically to minimize cytotoxicity and maintain cellular function (Figure 3).

Prerequisites for Imaging-Sample Preparation.—Imaging-based subcellular localization of proteins can be applied to many samples, such as tissue sections, cultured cells, cell lines, and organoids. The general working principle of visual-subcellular proteomics depends on the starting material, i.e., (a) live cell imaging and (b) fixed cell imaging. Paraformaldehyde (PFA) is the most widely used fixing agent to fix the cell and preserve the subcellular organelles.[121] Alternatively, organic solvents or alcohols such as acetone and methanol can also be used to fix and permeabilize cells in one step, which could be helpful when visualizing more rigid structures of the cells, such as cytoskeletal components or nuclear structures.[122,123] Tween-20, saponin glycosides, Triton X-100, and NP-40 are commonly used for cell permeabilization.[124,125] Optimizing fixation and permeabilization protocols is crucial for specific target protein and affinity reagents.[126,127]

Mass Spectrometry Imaging (MSI).—Mass spectrometry imaging (MSI) is a highly advanced analytical technique that allows visualization of precise spatial distribution of biomolecules, including proteins, metabolites, and biomarkers, in samples with great accuracy. This is done by analyzing the mass spectrum of different spots on tissue samples,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

which generates pictures of the spatially resolved distribution of proteins pixel by pixel. The signal generated by MSI is proportional to the relative abundance of the analyte. The operation principle of MSI depends on the technique used to obtain the spatial information, which can be microprobe (ionization beam-based) or microscope-based (discussed further in detail). Overall, MSI is a qualitative technique that relies on the spatial distribution of the sample to generate useful data. The most common ionization technologies in MSI are DESI imaging, MALDI imaging, secondary ion mass spectrometry imaging (SIMS imaging), nanoscale SIMS (NanoSIMS), and laser ablation electrospray ionization (LAESI).[128–130] SPUTNIK, an R package, is widely used to analyze MSI data. SPUTNIK offers a wide range of filters to remove peaks with an uncorrelated spatial distribution from the data.[131] More detailed information on MSI technique, application, and challenges can be found in the review.[128]

Microscopy Advancements in Subcellular Proteomics.—Fluorescence imaging provides detailed spatiotemporal information in living cells and tissue sections. Obtaining subcellular data from various organelles and cells through imaging can often lead to inaccurate data.[132] Therefore, employing highly multiplexed imaging techniques such as imaging mass cytometry (IMC)[133] and multiplexed ion beam imaging (MIBI)[134] is essential to gather subcellular protein information within the same cell. These approaches can detect 36 proteins with low resolution (0.5–1 μm). However, new high-dimensional fluorescence imaging techniques such as DNA-barcoded codetection by indexing (CODEX)[135] imaging, multiplexed immunofluorescence microscopy (MxIF),[136] and cyclic and sequential IF[137] can map up to 50 proteins. Simultaneous detection of signals from multiple proteins is possible with different fluorophores. However, spectral overlap can limit simultaneous protein detection with multiple fluorophores. New cyclic detection methods allow more proteins to be detected in a single sample without additional fluorophores.[138,139] The quality of captured images depends on resolution, sample size, and speed. High-throughput screening techniques prioritize throughput over resolution and are widely used in the pharmaceutical industry. These approaches require many replicates to gain statistical power. Advanced super-resolution microscopy techniques, such as photoactivated localization microscopy (PALM),[140] stimulated emission depletion microscopy (STED),[141] and stochastic optical reconstruction microscopy (STORM),[142] have been developed to overcome the confocal microscopy diffraction limit. Similarly, coupling cryoelectron microscopy or tomography (cryo-EM or cryo-ET) with mass spectrometry allows subnanometer visualization of biological samples in near-native and purified states.[143] Rapid multiplexed immunofluorescence (RapMIF) provides multiple rounds of immunostaining and fluorophore inactivation to enable high throughput *in situ* proteomic analysis using conventional microscopes.[144]

Processing Imaging Data.—Imaging data provides subcellular information about target proteins within intact cells. Using reference proteins or dyes to differentiate between organelles with near-similar localization is necessary. Proper controls and replicates are crucial to distinguish accurate localization from false staining patterns. Negative controls include samples without the affinity reagent or protein target expression to subtract background and ensure accurate results.[145] To ensure reproducibility, fixation,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and proper permeabilization, it is vital to use affinity reagents targeting specific cellular structures. These reagents can be validated by assessing signal loss following the knockdown of the target protein, comparing staining patterns, or examining correlations with RNA sequencing data. Manual pattern recognition is currently the primary approach for assigning subcellular localization. Popular open-source software for image analysis includes ImageJ,[146] CellProfiler,[147] QuPath,[148] ilastik,[149] and Orbit[150] (Table 4). Proteins localized to multiple cellular structures[151] and cells with varying protein expression levels require quantitative analysis techniques to reveal subtle cell-to-cell variations and partial translocations.[152] This involves segmenting cells and subcellular structures using structure-specific markers and fluorescence intensities. Scalable, high-performance machine and deep learning techniques are advantageous for image analysis, particularly when analyzing large data sets, which makes them appealing for predicting protein localization patterns.[153,154] Uniform Manifold Approximation and Projection (uMAP) is another dimension reduction method for complex subcellular imaging data analysis. uMAP is a technique that reduces the complexity of a data set by analyzing the subcellular location of human proteins using confocal microscopy images. After extracting features from each image using a machine learning model, the uMAP method is employed to reduce the dimensionality of the data set. The result is displayed in a scatter plot where each data point represents an image, enabling researchers to visualize and explore the highly dimensional protein localization data. Coloring each data point according to subcellular localization makes identifying images of proteins localizing to the same compartment possible. Integrating diverse data sets with the uMAP projection can facilitate the identification of unknown staining patterns and the recognition of distinct gene clusters in a large and intricate data set.[155]

Comparative Subcellular Proteomics

Comparative MS Subcellular Profiling.—The Dynamic Organellar Map (DOM) is a well-established method for comparing MS profiles. This method requires a reliable statistical framework to detect protein translocations accurately. Initially, the DOM method was employed to track protein movements after EGF stimulation, revealing various translocations such as the movement of EGFR from the plasma membrane to endosomes, the recruitment of necessary signaling adaptors, and the transportation of transcription factors to and from the nucleus.[31] The method also permits protein quantification, providing estimates of copy numbers. The DOM method has also been used to understand the mechanism behind congenital disorders such as AP-4 deficiency syndrome. The DOM application discovered that AP-4 vesicles regulate the cellular localization of autophagy protein ATG9A, revealing a new pathway for controlling autophagy.[19] To understand the pathological changes that occur during hepatic steatosis, a hallmark of nonalcoholic fatty liver disease (NAFLD), comparative PCP has been used to detect changes in subcellular protein distribution in mouse liver cells after exposure to a high-fat diet.[156] The high-fat diet resulted in significant organellar rearrangements, with translocation events detected in several hundred of the ~4,500 mapped protein subcellular localizations. Cellular changes, such as Golgi apparatus adsorption onto lipid droplets and secretion, were substantially reduced. Many proteins that mediate contact between organelles were redirected to lipid droplets, forming more robust connections between the droplets and mitochondria. Similarly, comparative LOPIT has been utilized to track cellular changes during HCMV infection.[61] The study used five LOPIT

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

maps over 120 h to track 4000 hosts and 100 viral protein rearrangements. This analysis revealed organellar remodeling events, such as merging endoplasmic reticulum into Golgi body and lysosomal splitting. In addition, candidate translocation events led to the discovery of a new factor, MYO18A, required for efficient HCMV replication.

Comparative Imaging-Based Subcellular Proteomics.—Comparative proteomics can benefit from gene tagging methods. These methods enable the analysis of multiple perturbations by creating fluorescent-protein-tagged libraries. One of the most widely used yeast collections in comparative studies is the *S. cerevisiae* GFP fusion collection.[157] These studies have independently concluded that cells often relocate proteins in response to environmental stress conditions, such as DNA replication inhibitors by methylmethanesulfonate, hydroxyurea, DTT, hydrogen peroxide, and rapamycin.[158] This suggests that protein relocation is a common cellular strategy for environmental stress. Gene tagging methods and comparative proteomics are powerful tools for understanding cellular responses to cellular stress conditions. A meta-analysis of 24 yeast screens was performed to gain more insights into protein translocation under stress conditions.[159] The authors used a specific approach to identify groups of proteins with similar relocation patterns, implying functional connectivity. To study mitosis in human cells, 4D imaging data was integrated with 3D concentration data for twenty-eight human proteins tagged with genomic tags. This integration created a model for protein reorganizations during mitosis using computational image analysis.[160] This study offers a powerful approach for analyzing protein redistribution in cellular processes. Similarly, the Human Protein Atlas (HPA) mapped 12,003 human proteins to 30 cellular organelles. Fifty percent of the proteins were seen to localize to two or more cellular compartments.[18] HPA also revealed that multilocalizing proteins are more likely to show spatial variation between the cell lines; 3546 multilocalizing proteins showed cell linedependent localization.

DATA RESOURCES IN SUBCELLULAR PROTEOMICS

Data Analysis: Importance and Methods

A thorough analysis of raw data is just as crucial as acquiring experimental data. Computational methods and models are required to gain biological insights from complex subcellular proteomic data.[161] Protein subcellular localization was initially predicted using amino acid sequences.[162] However, this method was later found inaccurate in identifying protein clusters due to data identification issues.[163] With the advancement of MS techniques, it is now possible to map proteins to subcellular locations based on organelle enrichment/fractionation.[164] Statistical tools such as Student's *t* test[165] and chi-squared test[166] can map single organellar proteins in a nondynamic context. When mapping multiple organelle proteins in a dynamic context, statistical tools combined with deep machine-learning methods are required to generate high-confidence output. Statistical techniques validate and refine machine learning models, helping to quantify their performance and avoid problems like overfitting. With increasing advancements in mass spectrometry and image-based subcellular protein localization methods, new machine-learning algorithms are constantly emerging (Figure 4).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Computational Tools Based on Protein Features

Sequence-based approaches for localization prediction include (i) using the public database Uniprot[78] or (ii) Chou's Pseudo amino acid composition (PseAAC), which considers amino acid composition and potential interactions among adjacent residues.[167] Homology-based methods predict protein function based on conserved homologous sequences of known function.[168] Homology-based methods, such as BLAST,[169] PSI-BLAST,[170] and hidden Markov models (HMM),[171] are commonly used to search for homologous sequences. However, these methods may not be effective when no homology is found between query and annotated protein sequences. Position-Specific Scoring Matrices (PSSMs) are utilized for predicting protein localization. PSSMs represent the frequency of each amino acid in a protein multiple sequence alignment. The PSI-BLAST algorithm can generate PSSMs by searching for protein sequences similar to a given query sequence. BLOSUM62 is a quick and efficient matrix for protein BLAST that can detect weak protein similarities. It can be used as an alternative to PSSM when the acquisition process is slow or unsuccessful.[172] Functional motifs-based methods include PROSITE,[173] MEME,[174] and simple modular architecture research tool (SMART).[175] K-mer sequence can be used to extract sequence patterns with a sliding window. The motif length, k, is generally based on prior knowledge. Recently published approaches, such as TetraMito,[176] use over-represented tetrapeptides as critical features to predict submitochondrial protein localization. Similarly, SubGolgi 2.0,[177] a sub-Golgi protein location predictor, identifies g-gap dipeptide compositions (two amino acids with g residues between them). LOCALIZER[178] predicts protein localization to chloroplasts, mitochondria, and nuclei using k-mer motifs of varying lengths. The signal peptide-based method identifies oligos-amino acid sequences at the N-terminal of newly synthesized proteins. Signal peptide prediction tools such as SignalP 6.0,[179] TargetP 2.0,[180] SigUNet,[181] and DeepSig[182] are extensively used. Not only sequence-based protein identification features but also various nonsequence-derived features have been used to predict subcellular localization. LOC3D is one such tool that utilizes structural information to predict proteins' subcellular locations.[183] Combining multiple approaches that consider a variety of characteristics, such as protein sequence, structure, physicochemical properties, and function, can enhance the accuracy of protein subcellular localization prediction.

The pretrained sequence embedding method uses the pretrain model adopted from Natural Language Processing (NLP). SeqVec[184] and ProtTrans[185] are pretrained models on UniRef[186] for pretrain embedding and transfer learning for subcellular localization prediction. In the case of interacting proteins, they are considered neighbors in a protein–protein interaction network (PPI). The location of the neighbors in a PPI network can provide information about the location of unannotated proteins. Protein interaction data can be retrieved from databases such as BioGRID[187] and STRING.[188] Gene/protein expression can also be considered a protein feature as genes/proteins found in the same compartment at the organelle or suborganelle level tend to be coexpressed to carry out related functions. Gene/protein expression data can establish interactions and create features like k-NN scores in the MU-LOC[172] method or standalone features in the SLocX[189] method.

Classification Algorithms Used in Organellar Protein Prediction

Machine learning (ML) techniques have proven effective in predicting the subcellular localization of proteins. These techniques use sequence data from all known proteins with

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

subcellular localization information to create a meaningful model of biological data. The parameters required to assign an ML tool for prediction are based on (i) support vector machine (SVM), a supervised machine learning algorithm that annotates classification and regression in complex biological problems. SVMs aim to locate a hyperplane in an *n*-dimensional space that can classify data into various classes with the maximum margin. SVM-based methods include SubMitoPred,[190] which uses Pfam domain information to predict mitochondrial proteins and their submitochondrial localization. ERPred[191] predicts ER-resident proteins by training an SVM with a combination of amino acid compositions from different parts of proteins. SubNucPred[192] predicts protein localization for ten subnuclear locations sequentially by combining the presence or absence of a unique Pfam domain and an amino acid composition based SVM model. VacPred[193] predicts vacuole-resident proteins through SVM training. Combining an SVM-based localization prediction method with homology search, CELLO2GO[194] uses GO terms of homologous proteins as possible functional annotations for a queried protein when available. Otherwise, the SVM classifier provides localization prediction. (ii) Random Forest (RF) is a deep learning classification and regression analysis method. The main idea of Random Forest is to construct multiple decision trees and combine them to obtain a more accurate and stable prediction.[195] Apart from predicting subcellular localization, RF is used in various applications such as prediction of cancer drug,[196] gene expression analysis,[197] biomarker discovery,[198] and protein–protein interaction study.[199] (iii) Neural Network and Deep Learning is a feed-forward neural network connecting all neurons in hidden and output layers.[200] An artificial neural network (ANN) is a collection of connected nodes called artificial neurons that model the neurons in a biological framework. Each neuron receives a signal, and output is calculated using a nonlinear function of the sum of its inputs. Deep learning uses larger networks such as deep neural networks (DNNs),[201] convolutional neural networks (CNNs),[202] recurrent neural networks (RNNs),[203] and attention mechanisms.[204] These deep learning methods, as well as traditional ANNs, have been applied in protein localization prediction. Several neural network-based methods are available to predict subcellular protein localization. SCLpred[205] is a neural network for protein localization prediction that can map a sequence into fixed-length properties without needing predefined features. DeepLoc[206] uses CNN and attention mechanisms to predict protein localization and detect contributing regions. MULocDeep[172] extends target localization coverage to 10 main subcellular compartments and their suborganellar compartments with 44 localization classes in total. DeepMito[207] is a deep-learning method that predicts submitochondrial localization using CNNs. It uses physical-chemical properties, PSSM, and raw sequence one-hot encoding as features. Some methods predict targeting peptides and their cleavage sites, which help infer corresponding proteins' localization. DeepSig[208] and SignalP 5.0[209] use deep learning to predict signal peptides (Table 5).

DATA REPOSITORIES

In MS-based techniques, the raw output contains mass spectral data that requires analysis using quantification measurement tools. MS-based subcellular proteomic methods yield high-resolution protein subcellular maps (Figure 5).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The SubCellBarCode database has generated an organellar map of 12,418 proteins in five human cell lines.[66] The PSL-LCCL database provides information on subcellular localization and abundance of six membranous organelles in human cancer cell line SK_HEP1.[210] The PROLOCATE database assigned 6000 proteins to eight subcellular locations in rat liver cells.[36] Meanwhile, the HeLa subcellular proteome database quantified 8700 proteins in HeLa cells[31] for absolute protein quantification. Other available databases are BioPlex[211] and Human Cell Map[34] and incorporate cell line-specific protein–protein interaction information and subcellular niche information.

High-throughput imaging subcellular proteomic methods readily picture large-scale cellular images of living cells for analysis. DeepCell, integrated with the LiveCellNet data set, generates clear cellular maps by amalgamating images obtained by fluorescence microscopy.[212] The yeast GFP Fusion Localization Database is widely used for generating *S. cerevisiae* organellar map with 22 distinct locations.[213] Imaging studies using fluorescent proteins describing not only protein location but also its abundance in *S. cerevisiae* in different conditions (mutation, environment, stimulus) are available in collections such as Cellbase,[214] Collection of Yeast Cells Localization Patterns (CYCLoPs),[215] YeastRGB,[216] Database of High Throughput Screening Hits (dHITS),[217] and Localization and Quantitation Atlas of yeast proteomE (LoQAtE).[218] The fluorescent protein fusion method has been applied to study the subcellular organization of human cells using the Dynamic Proteomics database.[155] The results from multiple experiments across multiple species confirm the high reliability of fusion protein integration for imaging ease.

Data sets that combine MS-based LOPIT fractionation with imaging-based immune fluorescence methods are proving more effective for gaining insights into the subcellular proteome domain. Such an integration offers higher coverage resolution, reliability, and accuracy. For instance, the Cell Atlas map is a part of the Human Protein Atlas, which has localized 12003 proteins across 30 subcellular compartments.[18] The integration strategy of these two approaches enables simultaneous and precise subcellular and temporal resolution of the subcellular proteome.

Raw Data Repositories

Raw data files generated from various subcellular proteomic methods are invaluable data resources. In MS-based subcellular proteomics studies, raw output files can be submitted to standard protein repositories:[219] PRIDE,[220] Panorama,[221] PeptideAtlas,[222] and Mass Spectrometry Interactive Virtual Environment (MassIVE).[223]

For imaging-based spatial subcellular proteomics, the raw image data files can be submitted to public repositories like Image Data Resource (IDR),[224] Cell Image Library (CIL),[225] and Broad Bioimage Benchmark Collection (BC).[226] Data repositories ensure public data accessibility and reusability for ML-based subcellular proteomic data.

CHALLENGES IN THE FIELD OF SUBCELLULAR PROTEOMICS

Subcellular proteomics has emerged as a rapidly growing field, with many publications shedding light on subcellularly resolved protein information from subcellular compartments

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

to single cells.[4,62,63,81,82] From a technology development perspective, it is a highly interdisciplinary field that integrates imaging and image analysis, subcellular analyses, mass spectrometry, and computational machine learning methods. Published global subcellular proteomic approaches focused more on static cellular maps[56,60] rather than information on protein dynamics.[21,22,31] Subcellular proteomics employs multiple translational techniques to study cellular protein dynamics, protein–protein interaction, and protein relocalization following perturbations. Besides the numerous advantages, subcellular proteomics studies post specific technical and conceptual challenges for method developers. One of the main challenges is reproducibility. Detection of protein subcellular dynamics requires comparative analysis of at least two data sets; thus, reproducibility of the methods is crucial.

Mass spectrometry-based subcellular protein profiling studies typically rely on the fractionation of the subcellular content of the cell based on gradient centrifugation or cell homogenization techniques. However, it is impossible to attain identical gradient fractions and homogenization steps across experiments, and minor differences between the replicates can impact the data resolution, fractionation patterns, and subcellular niche enrichment.[64] Furthermore, the mass spectrometry sample run adds significant experimental noise and some degree of nonoverlapping proteins in the same set of biological replicate conditions. Robust and meticulously scrutinized technical optimization of experiments to cover all substantial changes across MS runs is the most reliable way to increase reproducibility.[31] Similarly, developing protocols beyond conventional fluorescent staining imaging and probing is essential for imaging-based comparative studies. Using alternative tags for primary antibody conjugation (like DNA or metals) has facilitated the accessibility of different kit-based protocols.[57] Also, poorly resolved organelles can further dilute the comparative assignment of a protein and can lead to false positive compartment localizations. Limitations in imaging resolution and MS sensitivity still constitute significant barriers. Integration of super-resolution[227] and expansion microscopy approaches[114] has shown significant improvements in sensitivity and resolution for imaging-based subcellular methods.[228]

Another challenge is the detection of protein subcellular dynamics. Translating proteins from one organelle to another is essential for maintaining cellular homeostasis. However, studies have reported that protein translocation between organelles is not absolute but somewhat relative. The mass spectrometry analysis of a protein subcellular location and dynamics results in (a) a qualitative interpretation of the data: compartment localization prediction from protein profile, and (b) a quantitative interpretation of the data: intensity abundance profile. Qualitative interpretation of protein translocation is biased as only a complete transition between organelles can be detected.

However, the shift in protein abundance profiles between organelles is more inclusive.[229] Similarly, in image-based subcellular proteomics studies, detecting partial protein translocation between organelles is hard to address as only absolute translocations are imaged using complex classifiers (one fit score). Recent advances have addressed the partial translocations between organelles using soft classifiers (multiple compartment localization with likelihood scores).[230]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Subcellular-Temporal Analysis

Comparative mapping experiments are conducted with various cell fractionation methods to determine the temporal movement of proteins in a subcellular setup. However, imaging-based and MS-based studies are inherently discontinuous because cell fixation and fractionation lead to cell destruction.[21] To perform a global subcellular-temporal analysis of proteins in any cell compartment, it is necessary to assemble a series of static images to see the whole picture in one frame.[61] The main limitation in imaging-based subcellular research is parallel live cell imaging, which is used to observe real-time protein dynamics.

False Discovery Rate in Comparative Analysis

Comparative spatial proteomics experiments frequently lead to many false positives, making it imperative to estimate the false discovery rate (FDR) of translocations to ensure a meaningful interpretation of the results. The best approach is to set up a mock experiment (control versus control, with no expected changes) in addition to the actual perturbation experiment, using the same data analysis pipeline. A precise estimate of the FDR can be obtained by comparing the number of hits in the mock and perturbation experiments at a specific stringency cutoff, provided that the experimental noise is comparable in both conditions. Replicate experiments are indispensable to achieve low FDR levels.[61]

Challenges of Imaging-Based Methods

Choosing the most appropriate imaging method for cell biology studies can be challenging with the availability of many advanced microscopy techniques. Specimen type, type of fluorophores being used, and experiment duration all affect the suitable microscopy technique. Data management is also a growing concern, with 4D data generating terabytes of information that must be organized and analyzed. However, with affordable memory storage and cloud technology, it is becoming easier to make these exciting new techniques accessible to the scientific community.[231]

Imaging Artifacts.—Fluorescence microscopy is a valuable tool for investigating protein behavior in living cells, but it has certain limitations regarding spatial and temporal resolution. To overcome these limitations, researchers often turn to fixation, a technique that immobilizes cells to enable high-resolution imaging of protein localization. However, this approach can sometimes result in protein redistribution that does not accurately reflect their behavior in living cells. The extent of such artifacts varies depending on the specific fixation method and the protein or cellular structure under investigation. While there are ways to minimize these effects, no single method can guarantee perfect preservation of protein localization. As a result, it is recommended that fixed-cell imaging be complemented with live-cell imaging to ensure accuracy.[232]

False Positives in Imaging Acquisition.—Imaging-based methods may result in false positives/negatives due to contamination/failures in image analysis. Proteins may show bimodal expression due to mixed genetic strains. Image transformations like segmentation can introduce errors. In Cluster Q analysis, the segmentation algorithm may incorrectly identify bud tips as bud cells. This can lead to incorrect attribution of protein localization changes to the bud cells instead of the mother cell. Similarly, clustering methods can also

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

introduce false positives into clusters, particularly when some protein change profiles have weaker signals. Data analysis uses distance-based clustering methods, which gives less efficient output as data dimensionality increases. Manual evaluation is also recommended to validate results and reduce false positives. The unsupervised method focuses on strong signals, reducing the search space to hundreds of images. This saves time and effort, as human evaluation can be laborious. Predictive approaches based on protein–protein interactions and mRNA expression patterns may miss many localization dynamics. Some transcription factors observed to change when using manual analysis were not predicted to have any localization changes by a previous predictive method.[233]

Effect of Laser on Imaging Samples.—Laser systems with high power consumption can cause damage to tissues due to the rise in temperature generated by laser irradiation. This can lead to cellular metabolism and electrical membrane capacitance changes, ultimately resulting in cell death. The temperature known to induce cell injury is 10 °C when studied *in vitro*. Tailoring laser parameters to the optical characteristics of the target tissue is crucial for effective optical imaging. Careful selection of laser parameters can optimize imaging efficacy while minimizing undesirable tissue damage. Determining the energy delivered to the tissues is crucial to ensure safety standards in optical imaging. Testing the safety of optical imaging lasers is essential, especially in the specific, sensitive neonatal population.[234]

Challenges of Mass Spectrometry-Based Methods

Labeling-based proteomics methods, such as stable isotope labeling by amino acids in culture (SILAC) and tandem mass tagging (TMT), offer numerous advantages in identifying protein interactions, resolving proteome-wide temporal kinetics and spatial distributions in a single experiment. They also allow observations of the spatial distribution of new and old protein pools. Under baseline conditions, the spatial profiles of light and heavy proteins are highly similar, providing additional assurance of spatial assignments. However, the number of dynamic SILAC labeling time points that can be investigated must be optimized for different cell types with distinct intrinsic protein turnover rates. Additionally, including earlier time points to investigate early translocation events may be technically challenging due to low-intensity heavy SILAC-labeled peptides. The double labeling design also requires independent MS2 acquisition of light and heavy peptides, which can decrease the depth and data completeness of mass spectrometry-based proteomics analysis. Future work may alleviate this limitation by modifying the mass spectrometry acquisition methods to trigger heavy peptide acquisition and reduce incomplete light-heavy pairs automatically. However, protein correlation profiling-based techniques face challenges in recognizing proteins with multiple localizations or partial translocations, such as p97/VCP, which has multiple subcellular localizations, and its precise subcellular translocation profile is complex to interpret from TMT data. Translocated proteins can have lower confidence in location classification as translocation may be substoichiometric.[235]

Data Reproducibility.—Spatial proteomics faces challenges in achieving consistent and accurate results due to minor differences in sample handling, staining protocols, and instrumentation.[236] Standardizing protocols is difficult due to the many parameters

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

involved in optimizing protocols. Automation, microfluidics, barcoding technologies, and imaging mass spectrometry have been improving the scalability and reproducibility of the methods. Reliable automation systems are needed to support different sample formats and deliver high reproducibility. Second, each centrifugation technique has advantages and disadvantages, depending on the type and quality of the sample and the desired outcome. Differential centrifugation is simpler, faster, and cheaper than density equilibrium centrifugation, as it does not require a density gradient or any unique rotor. However, it is less precise and is more prone to contamination, as the organelles may not be separated or damaged by the high speed and force. On the other hand, density centrifugation is more accurate, sensitive, and gentle than differential centrifugation, allowing for better resolution and preservation of the particles. However, samples must be washed several times after spinning in density ultracentrifugation to ensure no cross-contamination between fractions. Samples for preparative centrifugation are usually limited in size (e.g., tissues) or volume (e.g., cell suspensions or blood). There is a material loss in every wash step a sample is subjected to, thus the yield can be meager after an ultracentrifugation protocol. Ultracentrifugation is still time-consuming, and it can take several hours to fractionate a single mixture's components. Ultracentrifuges are costly devices that require constant maintenance. The cell disruption/homogenization method is another essential step while considering protocol reproducibility. Disruption of the cellular cell membrane is a crucial preparatory step for fractionation protocols. Lysing cells/breaking cellular membranes can pose additional problems downstream of homogenization, such as sample contamination between fractions.[237]

Protein Turnover Rate and Localization

Protein turnover is an essential cellular process that regulates protein synthesis and degradation rates. Disruption of protein turnover and cellular homeostasis often leads to a diseased state. Cellular stress upon external perturbation leads to misfolded proteins, triggering unfolded protein response (UPR). Despite continued efforts, we do not fully understand how the cellular proteome reorganizes in response to proteostatic stress. Proteins utilize the ER vesicular transport and secretory pathway for trafficking, a crucial step in producing membrane and extracellular proteins. A mismatch in protein synthesis and localization can cause ER stress and mislocalization. Proteins must fold correctly and be directed to the right location in the cell, which requires coordination between protein synthesis and localization.[235]

Data Handling and Computational Challenges

Data handling and computational model generation is a prerequisite in spatiotemporal analysis studies. The development and application of high-accuracy image segmentation methods, as well as the adaptation of analysis tools utilized initially in single-cell omics for subcellular omics data, have been instrumental during the early stages of field development. However, subcellular omics data pose distinct computational challenges[238,239] due to the added dimensions and increased data size and the unique nature of the data, which is often different from the bulk omics concerning coverage, sensitivity, level of noise, and the overall amount of represented information. Single-cell-focused tools also offer limited utility for leveraging new opportunities provided by the subcellular content of the data. The

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

large subcellular data sets impose heavy requirements on data handling, infrastructure, and computational performance.

Moreover, the unique aspects of subcellular data, including the subcellular information, need for custom visualization, and often custom data formats require novel data frameworks and repositories focused on subcellular omics data. Previously generated data sets and atlases, such as human protein atlas[240] or Allen brain atlases,[132] have become very useful as references and frameworks for building subcellular omics data. Creating new repositories supporting raw and processed data deposition will provide reference platforms and enable benchmarking data reprocessing and analysis with future computational tools. Unlike other omics approaches, subcellular proteomics methods and protocols are relatively nascent and expected to improve significantly over time. The progress made in automation, artificial intelligence, and machine learning has resulted in significant advancements. Streamlining advanced technical workflows, imaging, mass spectrometry techniques, microfluidics, and barcoding technologies has improved the reproducibility and scalability of subcellular recognition methods and will positively impact the field.

CONCLUSIONS AND PERSPECTIVES

The complexity of a cell's proteome cannot be explained solely by the number of genes present in it. Other factors contribute to this complexity. Subcellular proteomics is a field that helps us understand this complexity and its functional significance. For example, mapping proteins in multiple compartments can tell us more about how organelles interact and how different cellular processes are linked. It can also help identify proteins that have multiple functions. Additionally, mapping proteins with temporal and subcellular variability in single cells can provide important insights into cellular signaling. Exploring different protein isoforms can help us understand the complexity and organization of cells. There are two main approaches to studying subcellular proteomics: imaging-based and MS-based. Both approaches are necessary to comprehend cellular complexity fully. Several studies have shown the synergy of MS-based subcellular proteomics with imaging. Deep machine learning-based methods will remain important for image and MS-based sequence analyses. As subcellular proteomics-based localization studies become more popular, predicting unknown protein localization computationally becomes less critical. Instead, computational approaches can help improve proteome-based experiments, contribute to understanding molecular mechanisms of protein sorting, characterize dynamic translocation processes, and contribute to synthetic biology. By integrating subcellular proteomics with other omics technologies, such as transcriptomics and metabolomics, an integrated orthogonal field can be developed for subcellular omics studies. Repositories must be developed for imaging-based and MS-based subcellular proteomics data to enable meta-analysis of studies using different cell types, perturbations, and growth conditions. Existing repositories such as UniProt should consider how to incorporate and cross-reference these data sets for the benefit of all cell biologists. We envision a new era of cellular modeling, where the subcellular dynamics of proteins are integrated with other omics measurements to gain insight into the crosstalk between the different layers of cellular regulation. This will lead to a greater understanding of cellular phenotype and activity.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ACKNOWLEDGMENTS

This research was supported by the Intramural Research Program of NIAID, NIH.

REFERENCES

(1). Regev A; Teichmann SA; Lander ES; Amit I; Benoist C; Birney E; Bodenmiller B; Campbell P; Carninci P; Clatworthy M; Clevers H; Deplancke B; Dunham I; Eberwine J; Eils R; Enard W; Farmer A; Fugger L; Göttgens B; Hacohen N; Haniffa M; Hemberg M; Kim S; Klenerman P; Kriegstein A; Lein E; Linnarsson S; Lundberg E; Lundeberg J; Majumder P; Marioni JC; Merad M; Mhlanga M; Nawijn M; Netea M; Nolan G; Pe'er D; Phillipakis A; Ponting CP; Quake S; Reik W; Rozenblatt-Rosen O; Sanes J; Satija R; Schumacher TN; Shalek A; Shapiro E; Sharma P; Shin JW; Stegle O; Stratton M; Stubbington MJT; Theis FJ; Uhlen M; van Oudenaarden A; Wagner A; Watt F; Weissman J; Wold B; Xavier R; Yosef N Human Cell Atlas Meeting Participants. The Human Cell Atlas. Elife 2017, 6, No. e27041.

(2). Jimenez CR; Verheul HM W. Mass Spectrometry-Based Proteomics: From Cancer Biology to Protein Biomarkers, Drug Targets, and Clinical Applications. Am. Soc. Clin Oncol Educ Book 2014, 34, No. e504.

(3). Hosp F; Mann M A Primer on Concepts and Applications of Proteomics in Neuroscience. Neuron 2017, 96 (3), 558–571. [PubMed: 29096073]

(4). Thul PJ; Åkesson L; Wiking M; Mahdessian D; Geladaki A; Ait Blal H; Alm T; Asplund A; Björk L; Breckels LM; Bäckström A; Danielsson F; Fagerberg L; Fall J; Gatto L; Gnann C; Hober S; Hjelmare M; Johansson F; Lee S; Lindskog C; Mulder J; Mulvey CM; Nilsson P; Oksvold P; Rockberg J; Schutten R; Schwenk JM; Sivertsson C5; Sjöstedt E; Skogs M; Stadler C; Sullivan DP; Tegel H; Winsnes C; Zhang C; Zwahlen M; Mardinoglu A; Pontén F; von Feilitzen K; Lilley KS; Uhlén M; Lundberg E A Subcellular Map of the Human Proteome. Science 2017, 356 (6340), No. eaal3321.

(5). Boeynaems S; Alberti S; Fawzi NL; Mittag T; Polymenidou M; Rousseau F; Schymkowitz J; Shorter J; Wolozin B; Van Den Bosch L; Tompa P; Fuxreiter M Protein Phase Separation: A New Phase in Cell Biology. Trends Cell Biol. 2018, 28 (6), 420–435. [PubMed: 29602697]

(6). Schueder F; Lara-Gutiérrez J; Haas D; Beckwith KS; Yin P; Ellenberg J; Jungmann R Super-Resolution Spatial Proximity Detection with Proximity-PAINT. Angew. Chem., Int. Ed. Engl 2021, 60 (2), 716–720. [PubMed: 32936507]

(7). Hung M-C; Link W Protein Localization in Disease and Therapy. J. Cell Sci 2011, 124 (20), 3381–3392. [PubMed: 22010196]

(8). Guardia CM; De Pace R; Mattera R; Bonifacino JS Neuronal Functions of Adaptor Complexes Involved in Protein Sorting. Curr. Opin Neurobiol 2018, 51, 103–110. [PubMed: 29558740]

(9). Thelen MP; Kye MJ The Role of RNA Binding Proteins for Local mRNA Translation: Implications in Neurological Disorders. Front Mol. Biosci 2020, 6, 161. [PubMed: 32010708]

(10). Neel DS; Allegakoen DV; Olivas V; Mayekar MK; Hemmati G; Chatterjee N; Blakely CM; McCoach CE; Rotow JK; Le A; Karachaliou N; Rosell R; Riess JW; Nichols R; Doebele RC; Bivona TG Differential Subcellular Localization Regulates Oncogenic Signaling by ROS1 Kinase Fusion Proteins. Cancer Res. 2019, 79 (3), 546–556. [PubMed: 30538120]

(11). Siljee JE; Wang Y; Bernard AA; Ersoy BA; Zhang S; Marley A; Von Zastrow M; Reiter JF; Vaisse C Subcellular Localization of MC4R with ADCY3 at Neuronal Primary Cilia Underlies a Common Pathway for Genetic Predisposition to Obesity. Nat. Genet 2018, 50 (2), 180–185. [PubMed: 29311635]

(12). Hübner S; Eam JE; Hübner A; Jans DA Laminopathy-Inducing Lamin A Mutants Can Induce Redistribution of Lamin Binding Proteins into Nuclear Aggregates. Exp. Cell Res 2006, 312 (2), 171–183. [PubMed: 16289535]

(13). Mechanisms of protein-folding diseases at a glance | Disease Models & Mechanisms | The Company of Biologists. https://journals.biologists.com/dmm/article/7/1/9/19965/Mechanisms-of-protein-folding-diseases-at-a-glance (accessed 2024-02-11).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(14). Meyer K; Kirchner M; Uyar B; Cheng JY; Russo G; Hernandez-Miranda LR; Szymborska A; Zauber H; Rudolph IM; Willnow TE; Akalin A; Haucke V; Gerhardt H; Birchmeier C; Kühn R; Krauss M; Diecke S; Pascual JM; Selbach M Mutations in Disordered Regions Can Cause Disease by Creating Dileucine Motifs. Cell 2018, 175 (1), 239–253. [PubMed: 30197081]

(15). Bauer NC; Doetsch PW; Corbett AH Mechanisms Regulating Protein Localization. Traffic 2015, 16 (10), 1039–1061. [PubMed: 26172624]

(16). Christoforou A; Mulvey CM; Breckels LM; Geladaki A; Hurrell T; Hayward PC; Naake T; Gatto L; Viner R; Arias AM; Lilley KS A Draft Map of the Mouse Pluripotent Stem Cell Spatial Proteome. Nat. Commun 2016, 7, 8992. [PubMed: 26754106]

(17). Jeffery CJ Protein Moonlighting: What Is It, and Why Is It Important? Philos. Trans R Soc. Lond B Biol. Sci 2018, 373 (1738), 20160523. [PubMed: 29203708]

(18). Thul PJ; Lindskog C The Human Protein Atlas: A Spatial Map of the Human Proteome. Protein Sci. 2018, 27 (1), 233–244. [PubMed: 28940711]

(19). Davies AK; Itzhak DN; Edgar JR; Archuleta TL; Hirst J; Jackson LP; Robinson MS; Borner GHH AP-4 Vesicles Contribute to Spatial Control of Autophagy via RUSC-Dependent Peripheral Delivery of ATG9A. Nat. Commun 2018, 9 (1), 3958. [PubMed: 30262884]

(20). Mund A; Brunner A-D; Mann M Unbiased Spatial Proteomics with Single-Cell Resolution in Tissues. Mol. Cell 2022, 82 (12), 2335–2349. [PubMed: 35714588]

(21). Mulvey CM; Breckels LM; Crook OM; Sanders DJ; Ribeiro ALR; Geladaki A; Christoforou A; Britovsek NK; Hurrell T; Deery MJ; Gatto L; Smith AM; Lilley KS Spatiotemporal Proteomic Profiling of the Pro-Inflammatory Response to Lipopolysaccharide in the THP-1 Human Leukaemia Cell Line. Nat. Commun 2021, 12 (1), 5773. [PubMed: 34599159]

(22). Moloney NM; Barylyuk K; Tromer E; Crook OM; Breckels LM; Lilley KS; Waller RF; MacGregor P Mapping Diversity in African Trypanosomes Using High Resolution Spatial Proteomics. Nat. Commun 2023, 14, 4401. [PubMed: 37479728]

(23). Paul I; White C; Turcinovic I; Emili A Imaging the Future: The Emerging Era of Single-Cell Spatial Proteomics. FEBS J 2021, 288 (24), 6990–7001. [PubMed: 33351222]

(24). Pankow S; Martínez-Bartolomé S; Bamberger C; Yates JR Understanding Molecular Mechanisms of Disease through Spatial Proteomics. Curr. Opin Chem. Biol 2019, 48, 19–25. [PubMed: 30308467]

(25). Mou M; Pan Z; Lu M; Sun H; Wang Y; Luo Y; Zhu F Application of Machine Learning in Spatial Proteomics. J. Chem. Inf Model 2022, 62 (23), 5875–5895. [PubMed: 36378082]

(26). Lohani V; Akhiya AR; Kundu S; Akhter MQ; Bag S Single-Cell Proteomics with Spatial Attributes: Tools and Techniques. ACS Omega 2023, 8 (20), 17499–17510. [PubMed: 37251119]

(27). Bhatia HS; Brunner A-D; öztürk F; Kapoor S; Rong Z; Mai H; Thielert M; Ali M; Al-Maskari R; Paetzold JC; Kofler F; Todorov MI; Molbay M; Kolabas ZI; Negwer M; Hoeher L; Steinke H; Dima A; Gupta B; Kaltenecker D; Caliskan ö. S.; Brandt D; Krahmer N; Müller S; Lichtenthaler SF; Hellal F; Bechmann I; Menze B; Theis F; Mann M; Ertürk A Spatial Proteomics in Three-Dimensional Intact Specimens. Cell 2022, 185 (26), 5040–5058. [PubMed: 36563667]

(28). Brožová K; Hantusch B; Kenner L; Kratochwill K Spatial Proteomics for the Molecular Characterization of Breast Cancer. Proteomes 2023, 11 (2), 17. [PubMed: 37218922]

(29). Liu Y; Zeng R; Wang R; Weng Y; Wang R; Zou P; Chen PR Spatiotemporally Resolved Subcellular Phosphoproteomics. Proc. Natl. Acad. Sci. U. S. A 2021, 118 (25), No. e2025299118.

(30). Martinez-Val A; Bekker-Jensen DB; Steigerwald S; Koenig C; Østergaard O; Mehta A; Tran T; Sikorski K; Torres-Vega E; Kwasniewicz E; Brynjólfsdóttir SH; Frankel LB; Kjøbsted R; Krogh N; Lundby A; Bekker-Jensen S; Lund-Johansen F; Olsen JV Spatial-Proteomics Reveals Phospho-Signaling Dynamics at Subcellular Resolution. Nat. Commun 2021, 12 (1), 7113. [PubMed: 34876567]

(31). Itzhak DN; Tyanova S; Cox J; Borner GH Global, Quantitative and Dynamic Mapping of Protein Subcellular Localization. Elife 2016, 5, No. e16950.

(32). Minoshima M; Umeno T; Kadooka K; Roux M; Yamada N; Kikuchi K Development of a Versatile Protein Labeling Tool for Live-Cell Imaging Using Fluorescent β-Lactamase Inhibitors. Angew. Chem., Int. Ed. Engl 2023, 62 (18), No. e202301704.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(33). Parsons HT Preparation of Highly Enriched ER Membranes Using Free-Flow Electrophoresis. Methods Mol. Biol 2018, 1691, 103–115. [PubMed: 29043672]

(34). Go CD; Knight JDR; Rajasekharan A; Rathod B; Hesketh GG; Abe KT; Youn J-Y; Samavarchi-Tehrani P; Zhang H; Zhu LY; Popiel E; Lambert J-P; Coyaud C9; Cheung SWT; Rajendran D; Wong CJ; Antonicka H; Pelletier L; Palazzo AF; Shoubridge EA; Raught B; Gingras A-C A Proximity-Dependent Biotinylation Map of a Human Cell. Nature 2021, 595 (7865), 120–124. [PubMed: 34079125]

(35). Senichkin VV; Prokhorova EA; Zhivotovsky B; Kopeina GS Simple and Efficient Protocol for Subcellular Fractionation of Normal and Apoptotic Cells. Cells 2021, 10 (4), 852. [PubMed: 33918601]

(36). Jadot M; Boonen M; Thirion J; Wang N; Xing J; Zhao C; Tannous A; Qian M; Zheng H; Everett JK; Moore DF; Sleat DE; Lobel P Accounting for Protein Subcellular Localization: A Compartmental Map of the Rat Liver Proteome. Mol. Cell Proteomics 2017, 16 (2), 194–212. [PubMed: 27923875]

(37). Villanueva E; Smith T; Pizzinga M; Elzek M; Queiroz RML; Harvey RF; Breckels LM; Crook OM; Monti M; Dezi V; Willis AE; Lilley KS System-Wide Analysis of RNA and Protein Subcellular Localization Dynamics. Nat. Methods 2024, 21 (1), 60–71. [PubMed: 38036857]

(38). Xiong J; He J; Xie WP; Hinojosa E; Ambati CSR; Putluri N; Kim H-E; Zhu MX; Du G Rapid Affinity Purification of Intracellular Organelles Using a Twin Strep Tag. J. Cell Sci 2019, 132 (24), jcs235390.

(39). Andersen JS; Wilkinson CJ; Mayor T; Mortensen P; Nigg EA; Mann M Proteomic Characterization of the Human Centrosome by Protein Correlation Profiling. Nature 2003, 426 (6966), 570–574. [PubMed: 14654843]

(40). Gingras A-C; Abe KT; Raught B Getting to Know the Neighborhood: Using Proximity-Dependent Biotinylation to Characterize Protein Complexes and Map Organelles. Curr. Opin Chem. Biol 2019, 48, 44–54. [PubMed: 30458335]

(41). Pandya NJ; Koopmans F; Slotman JA; Paliukhovich I; Houtsmuller AB; Smit AB; Li KW Correlation Profiling of Brain Sub-Cellular Proteomes Reveals Co-Assembly of Synaptic Proteins and Subcellular Distribution. Sci. Rep 2017, 7 (1), 12107. [PubMed: 28935861]

(42). Hendriks IA; Larsen SC; Nielsen ML An Advanced Strategy for Comprehensive Profiling of ADP-Ribosylation Sites Using Mass Spectrometry-Based Proteomics*. *Molecular &* Cellular Proteomics 2019, 18 (5), 1010a.

(43). Gonzalez-Franquesa A; Stocks B; Chubanava S; Hattel HB; Moreno-Justicia R; Peijs L; Treebak JT; Zierath JR; Deshmukh AS Mass-Spectrometry-Based Proteomics Reveals Mitochondrial Supercomplexome Plasticity. Cell Rep 2021, 35 (8), 109180. [PubMed: 34038727]

(44). Hargett AA; Renfrow MB Glycosylation of Viral Surface Proteins Probed by Mass Spectrometry. Curr. Opin Virol 2019, 36, 56–66. [PubMed: 31202133]

(45). Sun S; Hu Y; Ao M; Shah P; Chen J; Yang W; Jia X; Tian Y; Thomas S; Zhang H N-GlycositeAtlas: A Database Resource for Mass Spectrometry-Based Human N-Linked Glycoprotein and Glycosylation Site Mapping. Clin Proteomics 2019, 16, 35. [PubMed: 31516400]

(46). Lund-Johansen F; de la Rosa Carrillo D; Mehta A; Sikorski K; Inngjerdingen M; Kalina T; Røysland K; de Souza GA; Bradbury ARM; Lecrevisse Q; Stuchly J MetaMass, a Tool for Meta-Analysis of Subcellular Proteomics Data. Nat. Methods 2016, 13 (10), 837–840. [PubMed: 27571551]

(47). Geladaki A; Kočevar Britovšek N; Breckels LM; Smith TS; Vennard OL; Mulvey CM; Crook OM; Gatto L; Lilley KS Combining LOPIT with Differential Ultracentrifugation for High-Resolution Spatial Proteomics. Nat. Commun 2019, 10 (1), 331. [PubMed: 30659192]

(48). Taguchi Y; Schätzl HM Small-Scale Subcellular Fractionation with Sucrose Step Gradient. Bio Protoc 2014, 4 (11), No. e1138.

(49). Bouchnak I; Brugiere S; Moyet L; Le Gall S; Salvi D; Kuntz M; Tardif M; Rolland N Unraveling Hidden Components of the Chloroplast Envelope Proteome: Opportunities and Limits of Better MS Sensitivity. Mol. Cell Proteomics 2019, 18 (7), 1285–1306. [PubMed: 30962257]

(50). Chapel A; Kieffer-Jaquinod S; Sagné C; Verdon Q; Ivaldi C; Mellal M; Thirion J; Jadot M; Bruley C; Garin J; Gasnier B; Journet A An Extended Proteome Map of the Lysosomal

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Membrane Reveals Novel Potential Transporters. Mol. Cell Proteomics 2013, 12 (6), 1572–1588. [PubMed: 23436907]

(51). Manner A; Islinger M Isolation of Mammalian Peroxisomes by Density Gradient Centrifugation. Methods Mol. Biol 2023, 2643, 1–12. [PubMed: 36952174]

(52). Hirohata K; Yamaguchi Y; Maruno T; Shibuya R; Torisu T; Onishi T; Chono H; Mineno J; Yuzhe Y; Higashiyama K; Masumi-Koizumi K; Uchida K; Yamamoto T; Uchida E; Okada T; Uchiyama S Applications and Limitations of Equilibrium Density Gradient Analytical Ultracentrifugation for the Quantitative Characterization of Adeno-Associated Virus Vectors. Anal. Chem 2024, 96 (2), 642–651. [PubMed: 38165078]

(53). Schirmer EC; Florens L; Guan T; Yates JR; Gerace L Nuclear Membrane Proteins with Potential Disease Links Found by Subtractive Proteomics. Science 2003, 301 (5638), 1380–1382. [PubMed: 12958361]

(54). Morgenstern M; Stiller SB; Lübbert P; Peikert CD; Dannenmaier S; Drepper F; Weill U; Höß P; Feuerstein R; Gebert M; Bohnert M; van der Laan M; Schuldiner; Schütze C; Oeljeklaus S; Pfanner N; Wiedemann N; Warscheid B Definition of a High-Confidence Mitochondrial Proteome at Quantitative Scale. Cell Rep 2017, 19 (13), 2836–2852. [PubMed: 28658629]

(55). Kim KH; Moon MH Chip-Type Asymmetrical Flow Field-Flow Fractionation Channel Coupled with Mass Spectrometry for Top-down Protein Identification. Anal. Chem 2011, 83 (22), 8652–8658. [PubMed: 21981549]

(56). Foster LJ; de Hoog CL; Zhang Y; Zhang Y; Xie X; Mootha VK; Mann M A Mammalian Organelle Map by Protein Correlation Profiling. Cell 2006, 125 (1), 187–199. [PubMed: 16615899]

(57). Hickey JW; Neumann EK; Radtke AJ; Camarillo JM; Beuschel RT; Albanese A; McDonough E; Hatler J; Wiblin AE; Fisher J; Croteau J; Small EC; Sood A; Caprioli RM; Angelo RM; Nolan GP; Chung K; Hewitt SM; Germain RN; Spraggins JM; Lundberg E; Snyder MP; Kelleher NL; Saka SK Spatial Mapping of Protein Composition and Tissue Organization: A Primer for Multiplexed Antibody-Based Imaging. Nat. Methods 2022, 19 (3), 284–295. [PubMed: 34811556]

(58). Uzbas F; O'Neill AC Spatial Centrosome Proteomic Profiling of Human iPSC-Derived Neural Cells. Bio Protoc 2023, 13 (17), No. e4812.

(59). Harner M; Korner C; Walther D; Mokranjac D; Kaesmacher J; Welsch U; Griffith J; Mann M; Reggiori F; Neupert W The Mitochondrial Contact Site Complex, a Determinant of Mitochondrial Architecture. EMBO Journal 2011, 30 (21), 4356–4370. [PubMed: 22009199]

(60). Dunkley TPJ; Hester S; Shadforth IP; Runions J; Weimar T; Hanton SL; Griffin JL; Bessant C; Brandizzi F; Hawes C; Watson RB; Dupree P; Lilley KS Mapping the Arabidopsis Organelle Proteome. Proc. Natl. Acad. Sci. U. S. A 2006, 103 (17), 6518–6523. [PubMed: 16618929]

(61). Jean Beltran PM; Mathias RA; Cristea IM A Portrait of the Human Organelle Proteome In Space and Time during Cytomegalovirus Infection. Cell Syst 2016, 3 (4), 361–373. [PubMed: 27641956]

(62). Shin JJH; Crook OM; Borgeaud AC; Cattin-Ortolá J; Peak-Chew SY; Breckels LM; Gillingham AK; Chadwick J; Lilley KS; Munro S Spatial Proteomics Defines the Content of Trafficking Vesicles Captured by Golgin Tethers. Nat. Commun 2020, 11 (1), 5987. [PubMed: 33239640]

(63). Nikolovski N; Shliaha PV; Gatto L; Dupree P; Lilley KS Label-Free Protein Quantification for Plant Golgi Protein Localization and Abundance1[W]. Plant Physiol 2014, 166 (2), 1033–1043. [PubMed: 25122472]

(64). Gatto L; Breckels LM; Wieczorek S; Burger T; Lilley KS Mass-Spectrometry-Based Spatial Proteomics Data Analysis Using pRoloc and pRolocdata. Bioinformatics 2014, 30 (9), 1322–1324. [PubMed: 24413670]

(65). Mardakheh FK; Sailem HZ; Kümper S; Tape CJ; McCully RR; Paul A; Anjomani-Virmouni S; Jørgensen C; Poulogiannis G; Marshall CJ; Bakal C Proteomics Profiling of Interactome Dynamics by Colocalisation Analysis (COLA). Mol. Biosyst 2017, 13 (1), 92–105.

(66). Orre LM; Vesterlund M; Pan Y; Arslan T; Zhu Y; Fernandez Woodbridge A; Frings O; Fredlund E; Lehtiö J SubCellBarCode: Proteome-Wide Mapping of Protein Localization and Relocalization. Mol. Cell 2019, 73 (1), 166–182. [PubMed: 30609389]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(67). Jean Beltran PM; Cook KC; Cristea IM Exploring and Exploiting Proteome Organization during Viral Infection. J. Virol 2017, 91 (18), No. e00268–17. [PubMed: 28679763]

(68). Schessner JP; Albrecht V; Davies AK; Sinitcyn P; Borner GHH Deep and Fast Label-Free Dynamic Organellar Mapping. Nat. Commun 2023, 14 (1), 5252. [PubMed: 37644046]

(69). Weston LA; Bauer KM; Hummon AB Comparison of Bottom-up Proteomic Approaches for LC-MS Analysis of Complex Proteomes. Anal Methods 2013, 5 (18), 4615.

(70). Lambert J-P; Picaud S; Fujisawa T; Hou H; Savitsky P; Uusküla-Reimand L; Gupta GD; Abdouni H; Lin Z-Y; Tucholska M; Knight JDR; Gonzalez-Badillo B; St-Denis N; Newman JA; Stucki M; Pelletier L; Bandeira N; Wilson MD; Filippakopoulos P; Gingras A-C Interactome Rewiring Following Pharmacological Targeting of BET Bromodomains. Mol. Cell 2019, 73 (3), 621–638. [PubMed: 30554943]

(71). Ludwig C; Gillet L; Rosenberger G; Amon S; Collins BC; Aebersold R Data-Independent Acquisition-Based SWATH-MS for Quantitative Proteomics: A Tutorial. Mol. Syst. Biol 2018, 14 (8), No. e8126.

(72). Palomba A; Abbondio M; Fiorito G; Uzzau S; Pagnozzi D; Tanca A Comparative Evaluation of MaxQuant and Proteome Discoverer MS1-Based Protein Quantification Tools. J. Proteome Res 2021, 20 (7), 3497–3507. [PubMed: 34038140]

(73). Rauniyar N; Yates JR Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics. J. Proteome Res 2014, 13 (12), 5293–5309. [PubMed: 25337643]

(74). Boersema PJ; Raijmakers R; Lemeer S; Mohammed S; Heck AJR Multiplex Peptide Stable Isotope Dimethyl Labeling for Quantitative Proteomics. Nat. Protoc 2009, 4 (4), 484–494. [PubMed: 19300442]

(75). Mendik P; Dobronyi L; Hári F; Kerepesi C; Maia-Moço L; Buszlai D; Csermely P; Veres DV Translocatome: A Novel Resource for the Analysis of Protein Translocation between Cellular Organelles. Nucleic Acids Res. 2019, 47 (D1), D495–D505. [PubMed: 30380112]

(76). Ashburner M; Ball CA; Blake JA; Botstein D; Butler H; Cherry JM; Davis AP; Dolinski K; Dwight SS; Eppig JT; Harris MA; Hill DP; Issel-Tarver L; Kasarskis A; Lewis S; Matese JC; Richardson JE; Ringwald M; Rubin GM; Sherlock G Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium. Nat. Genet 2000, 25 (1), 25–29. [PubMed: 10802651]

(77). Binder JX; Pletscher-Frankild S; Tsafou K; Stolte C; O'Donoghue SI; Schneider R; Jensen LJ COMPARTMENTS: Unification and Visualization of Protein Subcellular Localization Evidence. Database (Oxford) 2014, 2014, bau012. [PubMed: 24573882]

(78). UniProt Consortium. UniProt: A Worldwide Hub of Protein Knowledge. Nucleic Acids Res. 2019, 47 (D1), D506–D515, . [PubMed: 30395287]

(79). Cho NH; Cheveralls KC; Brunner A-D; Kim K; Michaelis AC; Raghavan P; Kobayashi H; Savy L; Li JY; Canaj H; Kim JYS; Stewart EM; Gnann C; McCarthy F; Cabrera JP; Brunetti RM; Chhun BB; Dingle G; Hein MY; Huang B; Mehta SB; Weissman JS; Gómez-Sjöberg R; Itzhak DN; Royer LA; Mann M; Leonetti MD OpenCell: Endogenous Tagging for the Cartography of Human Cellular Organization. Science 2022, 375 (6585), No. eabi6983.

(80). Guo J; Pu X; Lin Y; Leung H Protein Subcellular Localization Based on PSI-BLAST and Machine Learning. J. Bioinform Comput. Biol 2006, 04 (06), 1181–1195.

(81). Crook OM; Breckels LM; Lilley KS; Kirk PDW; Gatto L A Bioconductor Workflow for the Bayesian Analysis of Spatial Proteomics. F1000Res. 2019, 8, 446. [PubMed: 31119032]

(82). Crook OM; Davies CTR; Breckels LM; Christopher JA; Gatto L; Kirk PDW; Lilley KS Inferring Differential Subcellular Localisation in Comparative Spatial Proteomics Using BANDLE. Nat. Commun 2022, 13 (1), 5948. [PubMed: 36216816]

(83). Pan X; Li H; Zeng T; Li Z; Chen L; Huang T; Cai Y-D Identification of Protein Subcellular Localization With Network and Functional Embeddings. Front Genet 2021, 11, 626500. [PubMed: 33584818]

(84). Ester M; Kriegel H-P; Sander J; Xu X A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining; KDD'96; AAAI Press: Portland, OR, 1996; pp 226–231.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(85). Pino L; Schilling B Proximity Labeling and Other Novel Mass Spectrometric Approaches for Spatiotemporal Protein Dynamics. Expert Rev. Proteomics 2021, 18 (9), 757–765. [PubMed: 34496693]

(86). Chalkley RJ Modification Site Localization in Peptides. Adv. Exp. Med. Biol 2016, 919, 243–247. [PubMed: 27975222]

(87). Mittal RD Tandem Mass Spectroscopy in Diagnosis and Clinical Research. Indian J. Clin Biochem 2015, 30 (2), 121–123. [PubMed: 25883417]

(88). Balgley BM; Laudeman T; Yang L; Song T; Lee CS Comparative Evaluation of Tandem MS Search Algorithms Using a Target-Decoy Search Strategy. Mol. Cell Proteomics 2007, 6 (9), 1599–1608. [PubMed: 17533222]

(89). Roux KJ; Kim DI; Burke B; May DG BioID: A Screen for Protein-Protein Interactions. Curr. Protoc Protein Sci 2018, 91, 19.23.1.

(90). Sears RM; May DG; Roux KJ BioID as a Tool for Protein-Proximity Labeling in Living Cells. Methods Mol. Biol 2019, 2012, 299–313. [PubMed: 31161514]

(91). Fazal FM; Han S; Parker KR; Kaewsapsak P; Xu J; Boettiger AN; Chang HY; Ting AY Atlas of Subcellular RNA Localization Revealed by APEX-Seq. Cell 2019, 178 (2), 473–490. [PubMed: 31230715]

(92). Lam SS; Martell JD; Kamer KJ; Deerinck TJ; Ellisman MH; Mootha VK; Ting AY Directed Evolution of APEX2 for Electron Microscopy and Proximity Labeling. Nat. Methods 2015, 12 (1), 51–54. [PubMed: 25419960]

(93). Li X-W; Rees JS; Xue P; Zhang H; Hamaia SW; Sanderson B; Funk PE; Farndale RW; Lilley KS; Perrett S; Jackson AP New Insights into the DT40 B Cell Receptor Cluster Using a Proteomic Proximity Labeling Assay. J. Biol. Chem 2014, 289 (21), 14434–14447. [PubMed: 24706754]

(94). Hopkins C; Gibson A; Stinchcombe J; Futter C Chimeric Molecules Employing Horseradish Peroxidase as Reporter Enzyme for Protein Localization in the Electron Microscope. Methods Enzymol 2000, 327, 35–45. [PubMed: 11044972]

(95). Li J; Han S; Li H; Udeshi ND; Svinkina T; Mani DR; Xu C; Guajardo R; Xie Q; Li T; Luginbuhl DJ; Wu B; McLaughlin CN; Xie A; Kaewsapsak P; Quake SR; Carr SA; Ting AY; Luo L Cell-Surface Proteomic Profiling in the Fly Brain Uncovers Wiring Regulators. Cell 2020, 180 (2), 373–386. [PubMed: 31955847]

(96). Miyagawa-Yamaguchi A; Kotani N; Honke K Expressed Glycosylphosphatidylinositol-Anchored Horseradish Peroxidase Identifies Co-Clustering Molecules in Individual Lipid Raft Domains. PLoS One 2014, 9 (3), No. e93054.

(97). Rees JS; Li X-W; Perrett S; Lilley KS; Jackson AP Selective Proteomic Proximity Labeling Assay Using Tyramide (SPPLAT): A Quantitative Method for the Proteomic Analysis of Localized Membrane-Bound Protein Clusters. Curr. Protoc Protein Sci 2015, 80, 19.27.1.

(98). Liu Q; Zheng J; Sun W; Huo Y; Zhang L; Hao P; Wang H; Zhuang M A Proximity-Tagging System to Identify Membrane Protein-Protein Interactions. Nat. Methods 2018, 15 (9), 715–722. [PubMed: 30104635]

(99). Hananya N; Ye X; Koren S; Muir TW A Genetically Encoded Photoproximity Labeling Approach for Mapping Protein Territories. Proc. Natl. Acad. Sci. U. S. A 2023, 120 (16), No. e2219339120.

(100). Geri JB; Oakley JV; Reyes-Robles T; Wang T; McCarver SJ; White CH; Rodriguez-Rivera FP; Parker DL; Hett EC; Fadeyi OO; Oslund RC; MacMillan DWC Microenvironment Mapping via Dexter Energy Transfer on Immune Cells. Science 2020, 367 (6482), 1091–1097. [PubMed: 32139536]

(101). Ong S-E; Blagoev B; Kratchmarova I; Kristensen DB; Steen H; Pandey A; Mann M Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. Mol. Cell Proteomics 2002, 1 (5), 376–386. [PubMed: 12118079]

(102). Ross PL; Huang YN; Marchese JN; Williamson B; Parker K; Hattan S; Khainovski N; Pillai S; Dey S; Daniels S; Purkayastha S; Juhasz P; Martin S; Bartlet-Jones M; He F; Jacobson A; Pappin DJ Multiplexed Protein Quantitation in Saccharomyces Cerevisiae Using Amine-Reactive Isobaric Tagging Reagents. Mol. Cell Proteomics 2004, 3 (12), 1154–1169. [PubMed: 15385600]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(103). Thompson A; Schafer J; Kuhn K; Kienle S; Schwarz J; Schmidt G; Neumann T; Hamon C Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. Anal. Chem 2003, 75 (8), 1895–1904. [PubMed: 12713048]

(104). Hung V; Zou P; Rhee H-W; Udeshi ND; Cracan V; Svinkina T; Carr SA; Mootha VK; Ting AY Proteomic Mapping of the Human Mitochondrial Intermembrane Space in Live Cells via Ratiometric APEX Tagging. Mol. Cell 2014, 55 (2), 332–341. [PubMed: 25002142]

(105). Long S; Brown KM; Sibley LD CRISPR-Mediated Tagging with BirA Allows Proximity Labeling in Toxoplasma Gondii. Bio Protoc 2018, 8 (6), No. e2768.

(106). Han S; Li J; Ting AY Proximity Labeling: Spatially-Resolved Proteomic Mapping for Neurobiology. Curr. Opin Neurobiol 2018, 50, 17–23. [PubMed: 29125959]

(107). Mair A; Xu S-L; Branon TC; Ting AY; Bergmann DC Proximity Labeling of Protein Complexes and Cell-Type-Specific Organellar Proteomes in Arabidopsis Enabled by TurboID. Elife 2019, 8, No. e47864.

(108). Branon TC; Bosch JA; Sanchez AD; Udeshi ND; Svinkina T; Carr SA; Feldman JL; Perrimon N; Ting AY Efficient Proximity Labeling in Living Cells and Organisms with TurboID. Nat. Biotechnol 2018, 36 (9), 880–887. [PubMed: 30125270]

(109). Shinoda N; Hanawa N; Chihara T; Koto A; Miura M Dronc-Independent Basal Executioner Caspase Activity Sustains Drosophila Imaginal Tissue Growth. Proc. Natl. Acad. Sci. U. S. A 2019, 116 (41), 20539–20544. [PubMed: 31548372]

(110). Brudvig JJ; Cain JT; Schmidt-Grimminger GG; Stumpo DJ; Roux KJ; Blackshear PJ; Weimer JM MARCKS Is Necessary for Netrin-DCC Signaling and Corpus Callosum Formation. Mol. Neurobiol 2018, 55 (11), 8388–8402. [PubMed: 29546593]

(111). Marx V Calling the next Generation of Affinity Reagents. Nat. Methods 2013, 10 (9), 829–833. [PubMed: 23985727]

(112). Tiede C; Bedford R; Heseltine SJ; Smith G; Wijetunga I; Ross R; AlQallaf D; Roberts AP; Balls A; Curd A; Hughes RE; Martin H; Needham SR; Zanetti-Domingues LC; Sadigh Y; Peacock TP; Tang AA; Gibson N; Kyle H; Platt GW; Ingram N; Taylor T; Coletta LP; Manfield I; Knowles M; Bell S; Esteves F; Maqbool A; Prasad RK; Drinkhill M; Bon RS; Patel V; Goodchild SA; Martin-Fernandez M; Owens RJ; Nettleship JE; Webb ME; Harrison M; Lippiat JD; Ponnambalam S; Peckham M; Smith A; Ferrigno PK; Johnson M; McPherson MJ; Tomlinson DC Affimer Proteins Are Versatile and Renewable Affinity Reagents. Elife 2017, 6, No. e24903.

(113). Wang L; Frei MS; Salim A; Johnsson K Small-Molecule Fluorescent Probes for Live-Cell Super-Resolution Microscopy. J. Am. Chem. Soc 2019, 141 (7), 2770–2781. [PubMed: 30550714]

(114). Wen L; Fan Z; Mikulski Z; Ley K Imaging of the Immune System - towards a Subcellular and Molecular Understanding. J. Cell Sci 2020, 133 (5), jcs234922.

(115). Chen SJT; Samuelson MI; KD AR A Reassessment of the Impact and Significance of Social Media to Pathology. Archives of Pathology & Laboratory Medicine 2023, DOI: 10.5858/arpa.2022-0463-RA.

(116). Vijayaraghavan A; Jadhav C; Pradeep B; Bindu H; Kumaran S A Histopathological Study of Endometrial Biopsy Samples in Abnormal Uterine Bleeding. Cureus 2022, 14 (11), No. e31264.

(117). Wong C; Chen AA; Behr B; Shen S Time-Lapse Microscopy and Image Analysis in Basic and Clinical Embryo Development Research. Reprod Biomed Online 2013, 26 (2), 120–129. [PubMed: 23273754]

(118). Neumann B; Walter T; Hériché J-K; Bulkescher J; Erfle H; Conrad C; Rogers P; Poser I; Held M; Liebel U; Cetin C; Sieckmann F; Pau G; Kabbe R; Wünsche A; Satagopam V; Schmitz MHA; Chapuis C; Gerlich DW; Schneider R; Eils R; Huber W; Peters J-M; Hyman AA; Durbin R; Pepperkok R; Ellenberg J Phenotypic Profiling of the Human Genome by Time-Lapse Microscopy Reveals Cell Division Genes. Nature 2010, 464 (7289), 721–727. [PubMed: 20360735]

(119). Sarov M; Murray JI; Schanze K; Pozniakovski A; Niu W; Angermann K; Hasse S; Rupprecht M; Vinis E; Tinney M; Preston E; Zinke A; Enst S; Teichgraber T; Janette J; Reis K; Janosch S; Schloissnig S; Ejsmont RK; Slightam C; Xu X; Kim SK; Reinke V; Stewart AF; Snyder M;

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Waterston RH; Hyman AA A Genome-Scale Resource for In Vivo Tag-Based Protein Function Exploration in C. Elegans. Cell 2012, 150 (4), 855–866. [PubMed: 22901814]

(120). Chong YT; Koh JLY; Friesen H; Kaluarachchi Duffy S; Cox MJ; Moses A; Moffat J; Boone C; Andrews BJ Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis. Cell 2015, 161 (6), 1413–1424. [PubMed: 26046442]

(121). Cheng R; Zhang F; Li M; Wo X; Su Y-W; Wang W Influence of Fixation and Permeabilization on the Mass Density of Single Cells: A Surface Plasmon Resonance Imaging Study. Frontiers in Chemistry 2019, 7, 1. [PubMed: 30778383]

(122). Maria Nowakowska A; Borek-Dorosz A; Leszczenko P; Adamczyk A; Pieczara A; Jakubowska J; Pastorczak A; Ostrowska K; Marzec KM; Majzner K Reliable Cell Preparation Protocol for Raman Imaging to Effectively Differentiate Normal Leukocytes and Leukemic Blasts. Spectrochim Acta A Mol. Biomol Spectrosc 2023, 292, 122408. [PubMed: 36812801]

(123). Stadler C; Skogs M; Brismar H; Uhlén M; Lundberg E A Single Fixation Protocol for Proteome-Wide Immunofluorescence Localization Studies. J. Proteomics 2010, 73 (6), 1067–1078. [PubMed: 19896565]

(124). Amidzadeh Z; Behbahani AB; Erfani N; Sharifzadeh S; Ranjbaran R; Moezi L; Aboualizadeh F; Okhovat MA; Alavi P; Azarpira N Assessment of Different Permeabilization Methods of Minimizing Damage to the Adherent Cells for Detection of Intracellular RNA by Flow Cytometry. Avicenna J. Med. Biotechnol 2014, 6 (1), 38–46. [PubMed: 24523954]

(125). Oliver C; Jamur MC Immunocytochemical Methods and Protocols. Methods Mol. Biol 2010, 588, iv–v. [PubMed: 20025087]

(126). Syrbu SI; Cohen MB An Enhanced Antigen-Retrieval Protocol for Immunohistochemical Staining of Formalin-Fixed, Paraffin-Embedded Tissues. Methods Mol. Biol 2011, 717, 101–110. [PubMed: 21370027]

(127). Scheffler JM; Schiefermeier N; Huber LA Mild Fixation and Permeabilization Protocol for Preserving Structures of Endosomes, Focal Adhesions, and Actin Filaments during Immunofluorescence Analysis. Methods Enzymol 2014, 535, 93–102. [PubMed: 24377919]

(128). Buchberger AR; DeLaney K; Johnson J; Li L Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights. Anal. Chem 2018, 90 (1), 240–265. [PubMed: 29155564]

(129). Nemes P; Vertes A Laser Ablation Electrospray Ionization for Atmospheric Pressure Molecular Imaging Mass Spectrometry. Methods Mol. Biol 2010, 656, 159–171. [PubMed: 20680590]

(130). Siuzdak G Subcellular Quantitative Imaging of Metabolites at the Organelle Level. Nat. Metab 2023, 5 (9), 1446–1448. [PubMed: 37679555]

(131). Inglese P; Correia G; Takats Z; Nicholson JK; Glen RC SPUTNIK: An R Package for Filtering of Spatially Related Peaks in Mass Spectrometry Imaging Data. Bioinformatics 2019, 35 (1), 178–180. [PubMed: 30010780]

(132). Viana MP; Chen J; Knijnenburg TA; Vasan R; Yan C; Arakaki JE; Bailey M; Berry B; Borensztejn A; Brown EM; Carlson S; Cass JA; Chaudhuri B; Cordes Metzler KR; Coston ME; Crabtree ZJ; Davidson S; DeLizo CM; Dhaka S; Dinh SQ; Do TP; Domingus J; Donovan-Maiye RM; Ferrante AJ; Foster TJ; Frick CL; Fujioka G; Fuqua MA; Gehring JL; Gerbin KA; Grancharova T; Gregor BW; Harrylock LJ; Haupt A; Hendershott MC; Hookway C; Horwitz AR; Hughes HC; Isaac EJ; Johnson GR; Kim B; Leonard AN; Leung WW; Lucas JJ; Ludmann SA; Lyons BM; Malik H; McGregor R; Medrash GE; Meharry SL; Mitcham K; Mueller IA; Murphy-Stevens TL; Nath A; Nelson AM; Oluoch SA; Paleologu L; Popiel TA; Riel-Mehan MM; Roberts B; Schaefbauer LM; Schwarzl M; Sherman J; Slaton S; Sluzewski MF; Smith JE; Sul Y; Swain-Bowden MJ; Tang WJ; Thirstrup DJ; Toloudis DM; Tucker AP; Valencia V; Wiegraebe W; Wijeratna T; Yang R; Zaunbrecher RJ; Labitigan RLD; Sanborn AL; Johnson GT; Gunawardane RN; Gaudreault N; Theriot JA; Rafelski SM Integrated Intracellular Organization and Its Variations in Human iPS Cells. Nature 2023, 613 (7943), 345–354. [PubMed: 36599983]

(133). Giesen C; Wang HAO; Schapiro D; Zivanovic N; Jacobs A; Hattendorf B; Schüffler PJ; Grolimund D; Buhmann JM; Brandt S; Varga Z; Wild PJ; Günther D; Bodenmiller B Highly Multiplexed Imaging of Tumor Tissues with Subcellular Resolution by Mass Cytometry. Nat. Methods 2014, 11 (4), 417–422. [PubMed: 24584193]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(134). Keren L; Bosse M; Thompson S; Risom T; Vijayaragavan K; McCaffrey E; Marquez D; Angoshtari R; Greenwald NF; Fienberg H; Wang J; Kambham N; Kirkwood D; Nolan G; Montine TJ; Galli SJ; West R; Bendall SC; Angelo M MIBI-TOF: A Multiplexed Imaging Platform Relates Cellular Phenotypes and Tissue Structure. Sci. Adv 2019, 5 (10), No. eaax5851.

(135). Goltsev Y; Samusik N; Kennedy-Darling J; Bhate S; Hale M; Vazquez G; Black S; Nolan GP Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. Cell 2018, 174 (4), 968–981. [PubMed: 30078711]

(136). Gerdes MJ; Sevinsky CJ; Sood A; Adak S; Bello MO; Bordwell A; Can A; Corwin A; Dinn S; Filkins RJ; Hollman D; Kamath V; Kaanumalle S; Kenny K; Larsen M; Lazare M; Li Q; Lowes C; McCulloch CC; McDonough E; Montalto MC; Pang Z; Rittscher J; Santamaria-Pang A; Sarachan BD; Seel ML; Seppo A; Shaikh K; Sui Y; Zhang J; Ginty F Highly Multiplexed Single-Cell Analysis of Formalin-Fixed, Paraffin-Embedded Cancer Tissue. Proc. Natl. Acad. Sci. U. S. A 2013, 110 (29), 11982–11987. [PubMed: 23818604]

(137). Ko J; Oh J; Ahmed MS; Carlson JCT; Weissleder R Ultra-Fast Cycling for Multiplexed Cellular Fluorescence Imaging. Angew. Chem., Int. Ed. Engl 2020, 59 (17), 6839–6846. [PubMed: 32004403]

(138). Nakagawa T; Ohnishi K; Kosaki Y; Saito Y; Horlad H; Fujiwara Y; Takeya M; Komohara Y Optimum Immunohistochemical Procedures for Analysis of Macrophages in Human and Mouse Formalin Fixed Paraffin-Embedded Tissue Samples. J. Clin Exp Hematop 2017, 57 (1), 31–36. [PubMed: 28679964]

(139). Cohen M; Varki NM; Jankowski MD; Gagneux P Using Unfixed, Frozen Tissues to Study Natural Mucin Distribution. J. Vis Exp 2012, No. 67, No. e3928.

(140). Shroff H; White H; Betzig E Photoactivated Localization Microscopy (PALM) of Adhesion Complexes. Current Protocols in Cell Biology; Wiley, 2013; Chapter 4, pp 4.21.1–4.21.28, .

(141). Vicidomini G; Bianchini P; Diaspro A STED Super-Resolved Microscopy. Nat. Methods 2018, 15 (3), 173–182. [PubMed: 29377014]

(142). Rust MJ; Bates M; Zhuang X Sub-Diffraction-Limit Imaging by Stochastic Optical Reconstruction Microscopy (STORM). Nat. Methods 2006, 3 (10), 793–795. [PubMed: 16896339]

(143). Klykov O; Kopylov M; Carragher B; Heck AJR; Noble AJ; Scheltema RA Label-Free Visual Proteomics: Coupling MS-and EM-Based Approaches in Structural Biology. Mol. Cell 2022, 82 (2), 285–303. [PubMed: 35063097]

(144). Venkatesan M; Zhang N; Marteau B; Yajima Y; De Zarate Garcia NO; Fang Z; Hu T; Cai S; Ford A; Olszewski H; Borst A; Coskun AF Spatial Subcellular Organelle Networks in Single Cells. Sci. Rep 2023, 13 (1), 5374. [PubMed: 37005468]

(145). Bhattacharya N; Nagornov K; Verheggen K; Verhaert M; Sciot R; Verhaert P MS1-Based Data Analysis Approaches for FFPE Tissue Imaging of Endogenous Peptide Ions by Mass Spectrometry Histochemistry (MSHC). Methods Mol. Biol 2023, 2688, 187–202. [PubMed: 37410294]

(146). Schneider CA; Rasband WS; Eliceiri KW NIH Image to ImageJ: 25 Years of Image Analysis. Nat. Methods 2012, 9 (7), 671–675. [PubMed: 22930834]

(147). Carpenter AE; Jones TR; Lamprecht MR; Clarke C; Kang IH; Friman O; Guertin DA; Chang JH; Lindquist RA; Moffat J; Golland P; Sabatini DM CellProfiler: Image Analysis Software for Identifying and Quantifying Cell Phenotypes. Genome Biol. 2006, 7 (10), R100. [PubMed: 17076895]

(148). Bankhead P; Loughrey MB; Fernández JA; Dombrowski Y; McArt DG; Dunne PD; McQuaid S; Gray RT; Murray LJ; Coleman HG; James JA; Salto-Tellez M; Hamilton PW QuPath: Open Source Software for Digital Pathology Image Analysis. Sci. Rep 2017, 7 (1), 16878. [PubMed: 29203879]

(149). Berg S; Kutra D; Kroeger T; Straehle CN; Kausler BX; Haubold C; Schiegg M; Ales J; Beier T; Rudy M; Eren K; Cervantes JI; Xu B; Beuttenmueller F; Wolny A; Zhang C; Koethe U; Hamprecht FA; Kreshuk A Ilastik: Interactive Machine Learning for (Bio)Image Analysis. Nat. Methods 2019, 16 (12), 1226–1232. [PubMed: 31570887]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(150). Stritt M; Stalder AK; Vezzali E Orbit Image Analysis: An Open-Source Whole Slide Image Analysis Tool. PLOS Computational Biology 2020, 16 (2), No. e1007313.

(151). Lin J-R; Fallahi-Sichani M; Sorger PK Highly Multiplexed Imaging of Single Cells Using a High-Throughput Cyclic Immunofluorescence Method. Nat. Commun 2015, 6, 8390. [PubMed: 26399630]

(152). Gut G; Herrmann MD; Pelkmans L Multiplexed Protein Maps Link Subcellular Organization to Cellular States. Science 2018, 361 (6401), No. eaar7042.

(153). Rathbun LI; Colicino EG; Manikas J; O'Connell J; Krishnan N; Reilly NS; Coyne S; Erdemci-Tandogan G; Garrastegui A; Freshour J; Santra P; Manning ML; Amack JD; Hehnly H Cytokinetic Bridge Triggers de Novo Lumen Formation in Vivo. Nat. Commun 2020, 11 (1), 1269. [PubMed: 32152267]

(154). Chen C; Henderson JN; Ruchkin DA; Kirsh JM; Baranov MS; Bogdanov AM; Mills JH; Boxer SG; Fang C Structural Characterization of Fluorescent Proteins Using Tunable Femtosecond Stimulated Raman Spectroscopy. Int. J. Mol. Sci 2023, 24 (15), 11991. [PubMed: 37569365]

(155). Becht E; McInnes L; Healy J; Dutertre C-A; Kwok IWH; Ng LG; Ginhoux F; Newell EW Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP. Nat. Biotechnol 2019, 37, 38.

(156). Krahmer N; Najafi B; Schueder F; Quagliarini F; Steger M; Seitz S; Kasper R; Salinas F; Cox J; Uhlenhaut NH; Walther TC; Jungmann R; Zeigerer A; Borner GHH; Mann M Organellar Proteomics and Phospho-Proteomics Reveal Sub-cellular Reorganization in Diet-Induced Hepatic Steatosis. Dev Cell 2018, 47 (2), 205–221. [PubMed: 30352176]

(157). Huh W-K; Falvo JV; Gerke LC; Carroll AS; Howson RW; Weissman JS; O'Shea EK Global Analysis of Protein Localization in Budding Yeast. Nature 2003, 425 (6959), 686–691. [PubMed: 14562095]

(158). Tkach JM; Yimit A; Lee AY; Riffle M; Costanzo M; Jaschob D; Hendry JA; Ou J; Moffat J; Boone C; Davis TN; Nislow C; Brown GW Dissecting DNA Damage Response Pathways by Analysing Protein Localization and Abundance Changes during DNA Replication Stress. Nat. Cell Biol 2012, 14 (9), 966–976. [PubMed: 22842922]

(159). Koh JLY; Chong YT; Friesen H; Moses A; Boone C; Andrews BJ; Moffat J CYCLoPs: A Comprehensive Database Constructed from Automated Analysis of Protein Abundance and Subcellular Localization Patterns in Saccharomyces Cerevisiae. G3 (Bethesda) 2015, 5 (6), 1223–1232. [PubMed: 26048563]

(160). Cai Y; Hossain MJ; Hériché J-K; Politi AZ; Walther N; Koch B; Wachsmuth M; Nijmeijer B; Kueblbeck M; Martinic-Kavur M; Ladurner R; Alexander S; Peters J-M; Ellenberg J Experimental and Computational Framework for a Dynamic Protein Atlas of Human Cell Division. Nature 2018, 561 (7723), 411–415. [PubMed: 30202089]

(161). Rahmatbakhsh M; Gagarinova A; Babu M Bioinformatic Analysis of Temporal and Spatial Proteome Alternations During Infections. Front Genet 2021, 12, 667936. [PubMed: 34276775]

(162). Scholz M; Selbig J Visualization and Analysis of Molecular Data. Methods Mol. Biol 2007, 358, 87–104. [PubMed: 17035682]

(163). Unadkat HV; Hulsman M; Cornelissen K; Papenburg BJ; Truckenmüller RK; Carpenter AE; Wessling M; Post GF; Uetz M; Reinders MJT; Stamatialis D; van Blitterswijk CA; de Boer J An Algorithm-Based Topographical Biomaterials Library to Instruct Cell Fate. Proc. Natl. Acad. Sci. U. S. A 2011, 108 (40), 16565–16570. [PubMed: 21949368]

(164). Tang X; Pu Y; Peng H; Li K; Faouzi S; Lu T; Pu D; Cerezo M; Xu J; Li L; Robert C; Shen S Spatial Patterns of the Cap-Binding Complex eIF4F in Human Melanoma Cells. Comput. Struct Biotechnol J 2023, 21, 1157–1168. [PubMed: 36789267]

(165). Kim H-Y Statistical Notes for Clinical Researchers: The Independent Samples t-Test. Restor Dent Endod 2019, 44 (3), No. e26.

(166). McHugh ML The Chi-Square Test of Independence. Biochem Med. (Zagreb) 2013, 23 (2), 143–149. [PubMed: 23894860]

(167). Chou KC Prediction of Protein Cellular Attributes Using Pseudo-Amino Acid Composition. Proteins 2001, 43 (3), 246–255. [PubMed: 11288174]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(168). Cozzetto D; Jones DT Computational Methods for Annotation Transfers from Sequence. Methods Mol. Biol 2017, 1446, 55–67. [PubMed: 27812935]

(169). Boratyn GM; Camacho C; Cooper PS; Coulouris G; Fong A; Ma N; Madden TL; Matten WT; McGinnis SD; Merezhuk Y; Raytselis Y; Sayers EW; Tao T; Ye J; Zaretskaya I BLAST: A More Efficient Report with Usability Improvements. Nucleic Acids Res. 2013, 41, W29–W33. [PubMed: 23609542]

(170). Jones DT; Swindells MB Getting the Most from PSIBLAST. Trends Biochem. Sci 2002, 27 (3), 161–164. [PubMed: 11893514]

(171). Yoon B-J Hidden Markov Models and Their Applications in Biological Sequence Analysis. Curr. Genomics 2009, 10 (6), 402–415. [PubMed: 20190955]

(172). Jiang Y; Wang D; Yao Y; Eubel H; Künzler P; Møller IM; Xu D MULocDeep: A Deep-Learning Framework for Protein Subcellular and Suborganellar Localization Prediction with Residue-Level Interpretation. Comput. Struct Biotechnol J 2021, 19, 4825–4839. [PubMed: 34522290]

(173). Bailey TL; Williams N; Misleh C; Li WW MEME: Discovering and Analyzing DNA and Protein Sequence Motifs. Nucleic Acids Res. 2006, 34, W369–W373. [PubMed: 16845028]

(174). Bailey TL; Johnson J; Grant CE; Noble WS The MEME Suite. Nucleic Acids Res. 2015, 43 (W1), W39–49. [PubMed: 25953851]

(175). Mott R; Schultz J; Bork P; Ponting CP Predicting Protein Cellular Localization Using a Domain Projection Method. Genome Res. 2002, 12 (8), 1168–1174. [PubMed: 12176924]

(176). Lin H; Chen W; Yuan L-F; Li Z-Q; Ding H Using Over-Represented Tetrapeptides to Predict Protein Submitochondria Locations. Acta Biotheor 2013, 61 (2), 259–268. [PubMed: 23475502]

(177). Jiao Y-S; Du P-F Prediction of Golgi-Resident Protein Types Using General Form of Chou's Pseudo-Amino Acid Compositions: Approaches with Minimal Redundancy Maximal Relevance Feature Selection. J. Theor. Biol 2016, 402, 38–44. [PubMed: 27155042]

(178). Sperschneider J; Catanzariti A-M; DeBoer K; Petre B; Gardiner DM; Singh KB; Dodds PN; Taylor JM LOCALIZER: Subcellular Localization Prediction of Both Plant and Effector Proteins in the Plant Cell. Sci. Rep 2017, 7, 44598. [PubMed: 28300209]

(179). Teufel F; Almagro Armenteros JJ; Johansen AR; Gíslason MH; Pihl SI; Tsirigos KD; Winther O; Brunak S; von Heijne G; Nielsen H SignalP 6.0 Predicts All Five Types of Signal Peptides Using Protein Language Models. Nat. Biotechnol 2022, 40 (7), 1023–1025. [PubMed: 34980915]

(180). Almagro Armenteros JJ; Salvatore M; Emanuelsson O; Winther O; von Heijne G; Elofsson A; Nielsen H Detecting Sequence Signals in Targeting Peptides Using Deep Learning. Life Sci. Alliance 2019, 2 (5), No. e201900429.

(181). Wu J-M; Liu Y-C; Chang DT-H SigUNet: Signal Peptide Recognition Based on Semantic Segmentation. BMC Bioinformatics 2019, 20 (24), 677. [PubMed: 31861981]

(182). Savojardo C; Martelli PL; Fariselli P; Casadio R DeepSig: Deep Learning Improves Signal Peptide Detection in Proteins. Bioinformatics 2018, 34 (10), 1690–1696. [PubMed: 29280997]

(183). Nair R; Rost B LOC3D: Annotate Sub-Cellular Localization for Protein Structures. Nucleic Acids Res. 2003, 31 (13), 3337–3340. [PubMed: 12824321]

(184). Heinzinger M; Elnaggar A; Wang Y; Dallago C; Nechaev D; Matthes F; Rost B Modeling Aspects of the Language of Life through Transfer-Learning Protein Sequences. BMC Bioinformatics 2019, 20 (1), 723. [PubMed: 31847804]

(185). Elnaggar A; Heinzinger M; Dallago C; Rehawi G; Wang Y; Jones L; Gibbs T; Feher T; Angerer C; Steinegger M; Bhowmik D; Rost B ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. IEEE Trans Pattern Anal Mach Intell 2022, 44 (10), 7112–7127. [PubMed: 34232869]

(186). Jaakkola T; Diekhans M; Haussler D A Discriminative Framework for Detecting Remote Protein Homologies. J. Comput. Biol 2000, 7 (1–2), 95–114. [PubMed: 10890390]

(187). Oughtred R; Stark C; Breitkreutz B-J; Rust J; Boucher L; Chang C; Kolas N; O'Donnell L; Leung G; McAdam R; Zhang F; Dolma S; Willems A; Coulombe-Huntington J; Chatr-Aryamontri A; Dolinski K; Tyers M The BioGRID Interaction Database: 2019 Update. Nucleic Acids Res. 2019, 47 (D1), D529–D541. [PubMed: 30476227]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(188). Szklarczyk D; Gable AL; Lyon D; Junge A; Wyder S; Huerta-Cepas J; Simonovic M; Doncheva NT; Morris JH; Bork P; Jensen LJ; von Mering C STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. Nucleic Acids Res. 2019, 47 (D1), D607–D613. [PubMed: 30476243]

(189). Ryngajllo M; Childs L; Lohse M; Giorgi FM; Lude A; Selbig J; Usadel B SLocX: Predicting Subcellular Localization of Arabidopsis Proteins Leveraging Gene Expression Data. Front Plant Sci. 2011, 2, 43. [PubMed: 22639594]

(190). Kumar R; Kumari B; Kumar M Proteome-Wide Prediction and Annotation of Mitochondrial and Sub-Mitochondrial Proteins by Incorporating Domain Information. Mitochondrion 2018, 42, 11–22. [PubMed: 29032233]

(191). Kumar R; Kumari B; Kumar M Prediction of Endoplasmic Reticulum Resident Proteins Using Fragmented Amino Acid Composition and Support Vector Machine. PeerJ 2017, 5, No. e3561.

(192). Kumar R; Jain S; Kumari B; Kumar M Protein SubNuclear Localization Prediction Using SVM and Pfam Domain Information. PLoS One 2014, 9 (6), No. e98345.

(193). Yadav AK; Singla D VacPred: Sequence-Based Prediction of Plant Vacuole Proteins Using Machine-Learning Techniques. J. Biosci 2020, 45, 106. [PubMed: 32975233]

(194). Yu C-S; Cheng C-W; Su W-C; Chang K-C; Huang S-W; Hwang J-K; Lu C-H CELLO2GO: A Web Server for Protein subCELlular LOcalization Prediction with Functional Gene Ontology Annotation. PLoS One 2014, 9 (6), No. e99368.

(195). Zhao L; Wang J; Nabil MM; Zhang J Deep Forest-Based Prediction of Protein Subcellular Localization. Curr. Gene Ther 2018, 18 (5), 268–274. [PubMed: 30209998]

(196). Kim Y; Kim D; Cao B; Carvajal R; Kim M PDXGEM: Patient-Derived Tumor Xenograft-Based Gene Expression Model for Predicting Clinical Response to Anticancer Therapy in Cancer Patients. BMC Bioinformatics 2020, 21 (1), 288. [PubMed: 32631229]

(197). Clayton EA; Pujol TA; McDonald JF; Qiu P Leveraging TCGA Gene Expression Data to Build Predictive Models for Cancer Drug Response. BMC Bioinformatics 2020, 21 (14), 364. [PubMed: 32998700]

(198). Guan X; Runger G; Liu L Dynamic Incorporation of Prior Knowledge from Multiple Domains in Biomarker Discovery. BMC Bioinformatics 2020, 21 (S2), 77. [PubMed: 32164534]

(199). Yu X; Lai S; Chen H; Chen M Protein-Protein Interaction Network with Machine Learning Models and Multiomics Data Reveal Potential Neurodegenerative Disease-Related Proteins. Hum. Mol. Genet 2020, 29 (8), 1378–1387. [PubMed: 32277755]

(200). Almagro Armenteros JJ; Sønderby CK; Sønderby SK; Nielsen H; Winther O DeepLoc: Prediction of Protein Subcellular Localization Using Deep Learning. Bioinformatics 2017, 33 (24), 4049. [PubMed: 29028934]

(201). Bengio Y Learning Deep Architectures for AI. Found. Trends Mach. Learn 2009, 2 (1), 1–127.

(202). Yamashita R; Nishio M; Do RKG; Togashi K Convolutional Neural Networks: An Overview and Application in Radiology. Insights Imaging 2018, 9 (4), 611–629. [PubMed: 29934920]

(203). Liao Z; Pan G; Sun C; Tang J Predicting Subcellular Location of Protein with Evolution Information and Sequence-Based Deep Learning. BMC Bioinformatics 2021, 22 (S10), 515. [PubMed: 34686152]

(204). Niu Z; Zhong G; Yu H A Review on the Attention Mechanism of Deep Learning. Neurocomputing 2021, 452, 48–62.

(205). Kaleel M; Zheng Y; Chen J; Feng X; Simpson JC; Pollastri G; Mooney C SCLpred-EMS: Subcellular Localization Prediction of Endomembrane System and Secretory Pathway Proteins by Deep N-to-1 Convolutional Neural Networks. Bioinformatics 2020, 36 (11), 3343–3349. [PubMed: 32142105]

(206). Almagro Armenteros JJ; Sønderby CK; Sønderby SK; Nielsen H; Winther O DeepLoc: Prediction of Protein Subcellular Localization Using Deep Learning. Bioinformatics 2017, 33 (21), 3387–3395. [PubMed: 29036616]

(207). Savojardo C; Bruciaferri N; Tartari G; Martelli PL; Casadio R DeepMito: Accurate Prediction of Protein Sub-Mitochondrial Localization Using Convolutional Neural Networks. Bioinformatics 2020, 36 (1), 56–64. [PubMed: 31218353]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(208). Qiu K; Zheng S; Zhang L; Lou C; Yang X DeepSIG: A Hybrid Heterogeneous Deep Learning Framework for Radio Signal Classification. IEEE Transactions on Wireless Communications 2024, 23 (1), 775–788.

(209). Almagro Armenteros JJ; Tsirigos KD; Sønderby CK; Petersen TN; Winther O; Brunak S; von Heijne G; Nielsen H SignalP 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks. Nat. Biotechnol 2019, 37 (4), 420–423. [PubMed: 30778233]

(210). Huang F; Tang X; Ye B; Wu S; Ding K PSL-LCCL: A Resource for Subcellular Protein Localization in Liver Cancer Cell Line SK_HEP1. Database (Oxford) 2022, 2022, baab087. [PubMed: 35134877]

(211). Huttlin EL; Ting L; Bruckner RJ; Gebreab F; Gygi MP; Szpyt J; Tam S; Zarraga G; Colby G; Baltier K; Dong R; Guarani V; Vaites LP; Ordureau A; Rad R; Erickson BK; Wühr M; Chick J; Zhai B; Kolippakkam D; Mintseris J; Obar RA; Harris T; Artavanis-Tsakonas S; Sowa ME; De Camilli P; Paulo JA; Harper JW; Gygi SP The BioPlex Network: A Systematic Exploration of the Human Interactome. Cell 2015, 162 (2), 425–440. [PubMed: 26186194]

(212). Kobayashi H; Cheveralls KC; Leonetti MD; Royer LA Self-Supervised Deep Learning Encodes High-Resolution Features of Protein Subcellular Localization. Nat. Methods 2022, 19 (8), 995–1003. [PubMed: 35879608]

(213). Stringer C; Wang T; Michaelos M; Pachitariu M Cellpose: A Generalist Algorithm for Cellular Segmentation. Nat. Methods 2021, 18 (1), 100–106. [PubMed: 33318659]

(214). Nadif M; Role F Unsupervised and Self-Supervised Deep Learning Approaches for Biomedical Text Mining. Brief Bioinform 2021, 22 (2), 1592–1603. [PubMed: 33569575]

(215). McKinley ET; Shao J; Ellis ST; Heiser CN; Roland JT; Macedonia MC; Vega PN; Shin S; Coffey RJ; Lau KS MIRIAM: A Machine and Deep Learning Single-Cell Segmentation and Quantification Pipeline for Multi-Dimensional Tissue Images. Cytometry A 2022, 101 (6), 521–528. [PubMed: 35084791]

(216). Borner GHH Organellar Maps Through Proteomic Profiling - A Conceptual Guide. Mol. Cell Proteomics 2020, 19 (7), 1076–1087. [PubMed: 32345598]

(217). Devarajan K; Cheung VCK On Nonnegative Matrix Factorization Algorithms for Signal-Dependent Noise with Application to Electromyography Data. Neural Comput 2014, 26 (6), 1128–1168. [PubMed: 24684448]

(218). Giannuzzi L; Lombardo T; Juárez I; Aguilera A; Blanco G A Stochastic Characterization of Hydrogen Peroxide-Induced Regulated Cell Death in Microcystis Aeruginosa. Front Microbiol 2021, 12, 636157. [PubMed: 34394016]

(219). Bar N; Nikparvar B; Jayavelu ND; Roessler FK Constrained Fourier Estimation of Short-Term Time-Series Gene Expression Data Reduces Noise and Improves Clustering and Gene Regulatory Network Predictions. BMC Bioinformatics 2022, 23 (1), 330. [PubMed: 35945515]

(220). Csordas A; Ovelleiro D; Wang R; Foster JM; Ríos D; Vizcaíno JA; Hermjakob H PRIDE: Quality Control in a Proteomics Data Repository. Database (Oxford) 2012, 2012, bas004. [PubMed: 22434838]

(221). Wong DSV; Wong FK; Wood GR A Multi-Stage Approach to Clustering and Imputation of Gene Expression Profiles. Bioinformatics 2007, 23 (8), 998–1005. [PubMed: 17308340]

(222). Wang Y; Yu Z; Li S; Bian C; Liang Y; Wong K-C; Li X scBGEDA: Deep Single-Cell Clustering Analysis via a Dual Denoising Autoencoder with Bipartite Graph Ensemble Clustering. Bioinformatics 2023, 39 (2), btad075. [PubMed: 36734596]

(223). Pang S; Orgun MA; Yu Z A Novel Biomedical Image Indexing and Retrieval System via Deep Preference Learning. Comput. Methods Programs Biomed 2018, 158, 53–69. [PubMed: 29544790]

(224). Spitzer H; Berry S; Donoghoe M; Pelkmans L; Theis FJ Learning Consistent Subcellular Landmarks to Quantify Changes in Multiplexed Protein Maps. Nat. Methods 2023, 20 (7), 1058–1069. [PubMed: 37248388]

(225). Seton-Rogers S A Multiplexed View of Immune Infiltration. Nat. Rev. Cancer 2018, 18 (11), 666–667.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(226). Kim Y; Wang T; Xiong D; Wang X; Park S Multiple Instance Neural Networks Based on Sparse Attention for Cancer Detection Using T-Cell Receptor Sequences. BMC Bioinformatics 2022, 23 (1), 469. [PubMed: 36348271]

(227). Jacquemet G; Carisey AF; Hamidi H; Henriques R; Leterrier C The Cell Biologist's Guide to Super-Resolution Microscopy. J. Cell Sci 2020, 133 (11), jcs240713.

(228). Alon S; Goodwin DR; Sinha A; Wassie AT; Chen F; Daugharthy ER; Bando Y; Kajita A; Xue AG; Marrett K; Prior R; Cui Y; Payne AC; Yao C-C; Suk H-J; Wang R; Yu C-CJ; Tillberg P; Reginato P; Pak N; Liu S; Punthambaker S; Iyer EPR; Kohman RE; Miller JA; Lein ES; Lako A; Cullen N; Rodig S; Helvie K; Abravanel DL; Wagle N; Johnson BE; Klughammer J; Slyper M; Waldman J; Jané-Valbuena J; Rozenblatt-Rosen O; Regev A; IMAXT Consortium; Church GM; Marblestone AH; Boyden ES Expansion Sequencing: Spatially Precise in Situ Transcriptomics in Intact Biological Systems. Science 2021, 371 (6528), No. eaax2656.

(229). Watson J; Ferguson HR; Brady RM; Ferguson J; Fullwood P; Mo H; Bexley KH; Knight D; Howell G; Schwartz J-M; Smith MP; Francavilla C Spatially Resolved Phosphoproteomics Reveals Fibroblast Growth Factor Receptor Recycling-Driven Regulation of Autophagy and Survival. Nat. Commun 2022, 13 (1), 6589. [PubMed: 36329028]

(230). Lu Y; Phillips CA; Langston MA A Robustness Metric for Biological Data Clustering Algorithms. BMC Bioinformatics 2019, 20 (S15), 503. [PubMed: 31874625]

(231). Myers KA; Janetopoulos C Recent Advances in Imaging Subcellular Processes. F1000Res. 2016, 5, 1553.

(232). Yoshida SR; Maity BK; Chong S Visualizing Protein Localizations in Fixed Cells: Caveats and the Underlying Mechanisms. J. Phys. Chem. B 2023, 127 (19), 4165–4173. [PubMed: 37161904]

(233). Lu AX; Chong YT; Hsu IS; Strome B; Handfield L-F; Kraus O; Andrews BJ; Moses AM Integrating Images from Multiple Microscopy Screens Reveals Diverse Patterns of Change in the Subcellular Localization of Proteins. Elife 2018, 7, No. e31872.

(234). Nourhashemi M; Mahmoudzadeh M; Wallois F Thermal Impact of Near-Infrared Laser in Advanced Noninvasive Optical Brain Imaging. Neurophotonics 2016, 3 (1), 015001. [PubMed: 27115020]

(235). Currie J; Manda V; Robinson SK; Lai C; Agnihotri V; Hidalgo V; Ludwig RW; Zhang K; Pavelka J; Wang ZV; Rhee J-W; Lam MPY; Lau E Simultaneous Proteome Localization and Turnover Analysis Reveals Spatiotemporal Features of Protein Homeostasis Disruptions. bioRxiv 2024, 2023.01.04.521821.

(236). Liao Y-C; Fulcher JM; Degnan DJ; Williams SM; Bramer LM; Veličković D; Zemaitis KJ; Veličković M; Sontag RL; Moore RJ; Paša-Tolić L; Zhu Y; Zhou M Spatially Resolved Top-Down Proteomics of Tissue Sections Based on a Microfluidic Nanodroplet Sample Preparation Platform. Mol. Cell Proteomics 2023, 22 (2), 100491. [PubMed: 36603806]

(237). Alexandrov T; Saez-Rodriguez J; Saka SK Enablers and Challenges of Spatial Omics, a Melting Pot of Technologies. Mol. Syst. Biol 2023, 19 (11), No. e10571.

(238). Alexandrov LB; Kim J; Haradhvala NJ; Huang MN; Tian Ng AW; Wu Y; Boot A; Covington KR; Gordenin DA; Bergstrom. The Repertoire of Mutational Signatures in Human Cancer. Nature 2020, 578 (7793), 94–101. [PubMed: 32025018]

(239). Lähnemann D; Köster J; Szczurek E; McCarthy DJ; Hicks SC; Robinson MD; Vallejos CA; Campbell KR; Beerenwinkel N; Mahfouz A; Schonhuth A Eleven Grand Challenges in Single-Cell Data Science. Genome Biol. 2020, 21 (1), 31. [PubMed: 32033589]

(240). Uhlén M; Fagerberg L; Hallström BM; Lindskog C; Oksvold P; Mardinoglu A; Sivertsson A; Kampf C; Sjöstedt E; Asplund A; Ponten F Proteomics. Tissue- Based Map of the Human Proteome. Science 2015, 347 (6220), 1260419. [PubMed: 25613900]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

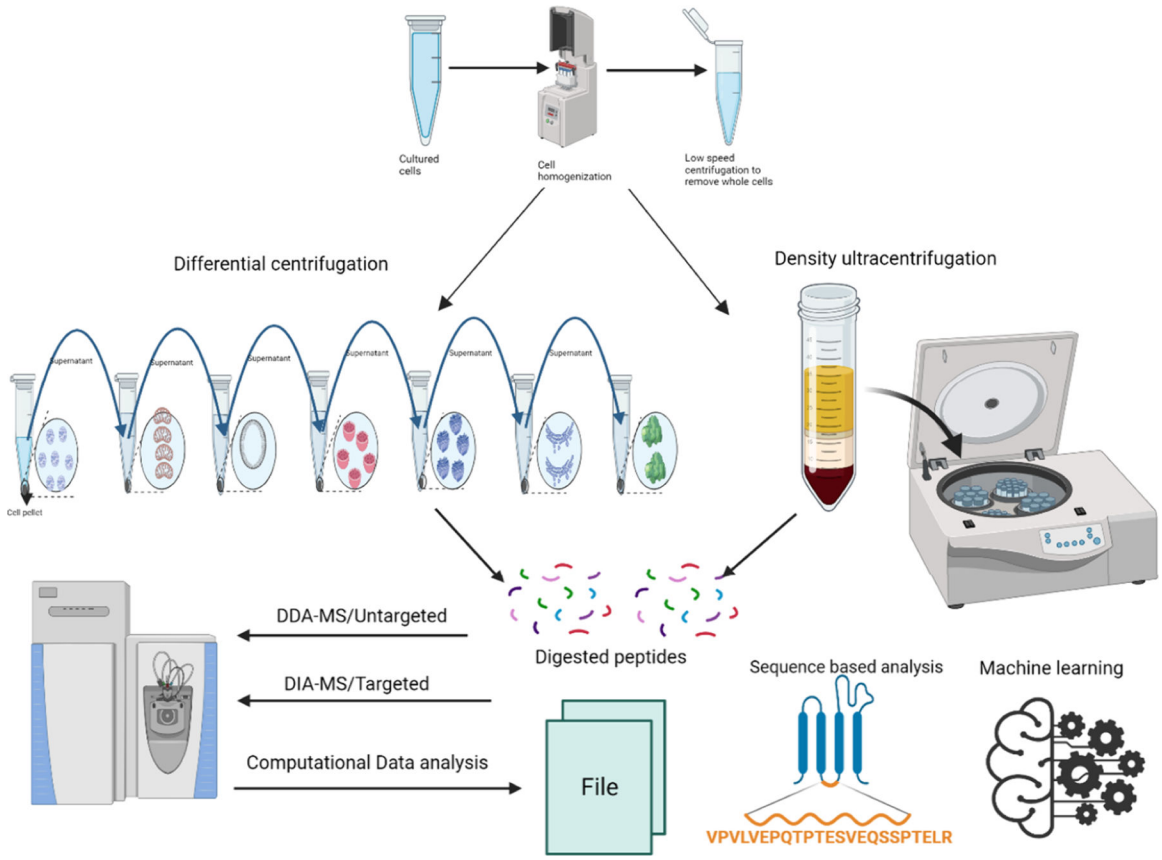


Figure 1.

A schematic workflow of mass spectrometry-based subcellular protein niche identification involves the following steps: (1) Cell lysis and homogenization to obtain membrane-bound and membrane-less organelles, (2) differential and density gradient centrifugation to separate organelles, and (3) data analysis using deep machine learning tools for distinct subcellular niche identification and protein dynamic translocation. Machine learning (ML) tools help to simplify complex data sets.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

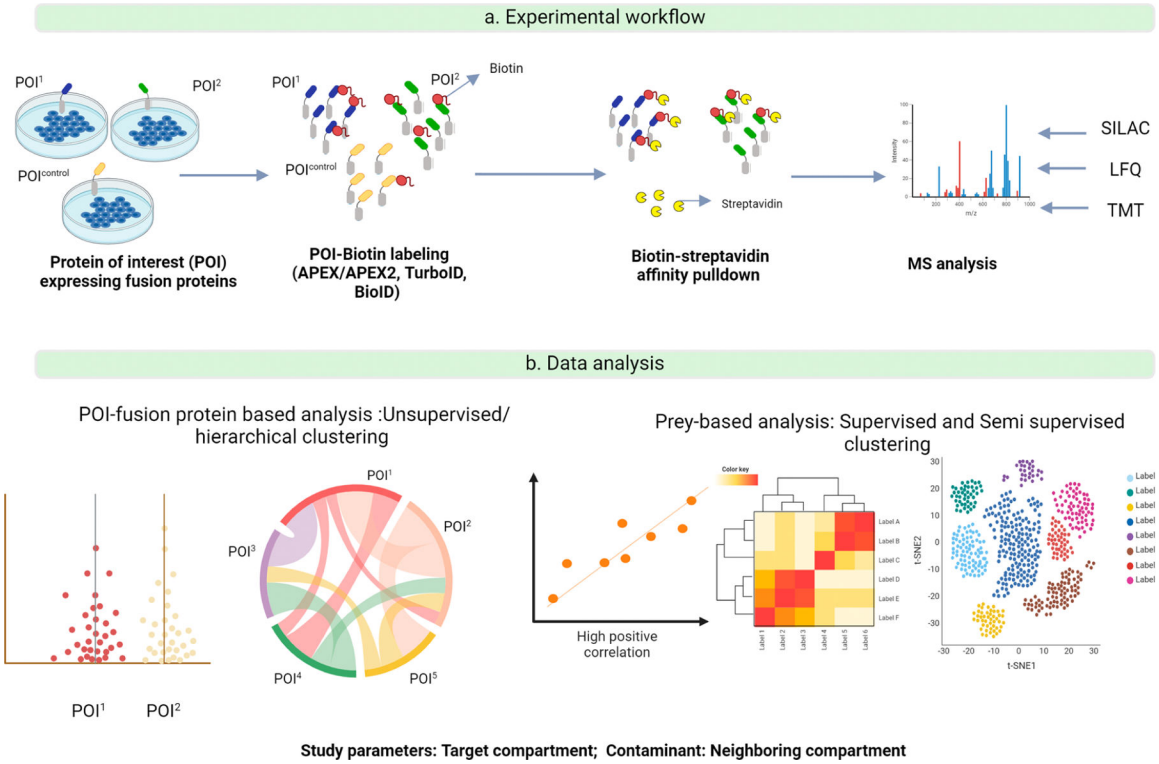


Figure 2.

Schematic workflow of proximity labeling strategies. (a) Proteins of interest (Bait) are genetically fused with an enzyme such as APEX/APEX2/BirA, BioID2, miniTurbo, or TurboID (1) that biotinylates adjacent proteins upon incubation with biotin (2). Control lines can indicate the labeling enzyme, which is fused to a control bait that is nonspecifically localized, such as GFP. (3). After labeling, proteins are enriched through a streptavidin pull-down, followed by identification through mass spectrometry. These labeled proteins are termed “Prey.” The prey is compared with proteins isolated from control lines to identify high-confidence proximity interactors. (b) Two types of analyses can be used to study subcellular components through proximity labeling, namely bait-centric and prey-centric analyses. Isotopic labeling and bait quantification techniques are used to identify proteins in organelles. Clustering baits and prey-centric studies can reveal proximity interaction networks.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

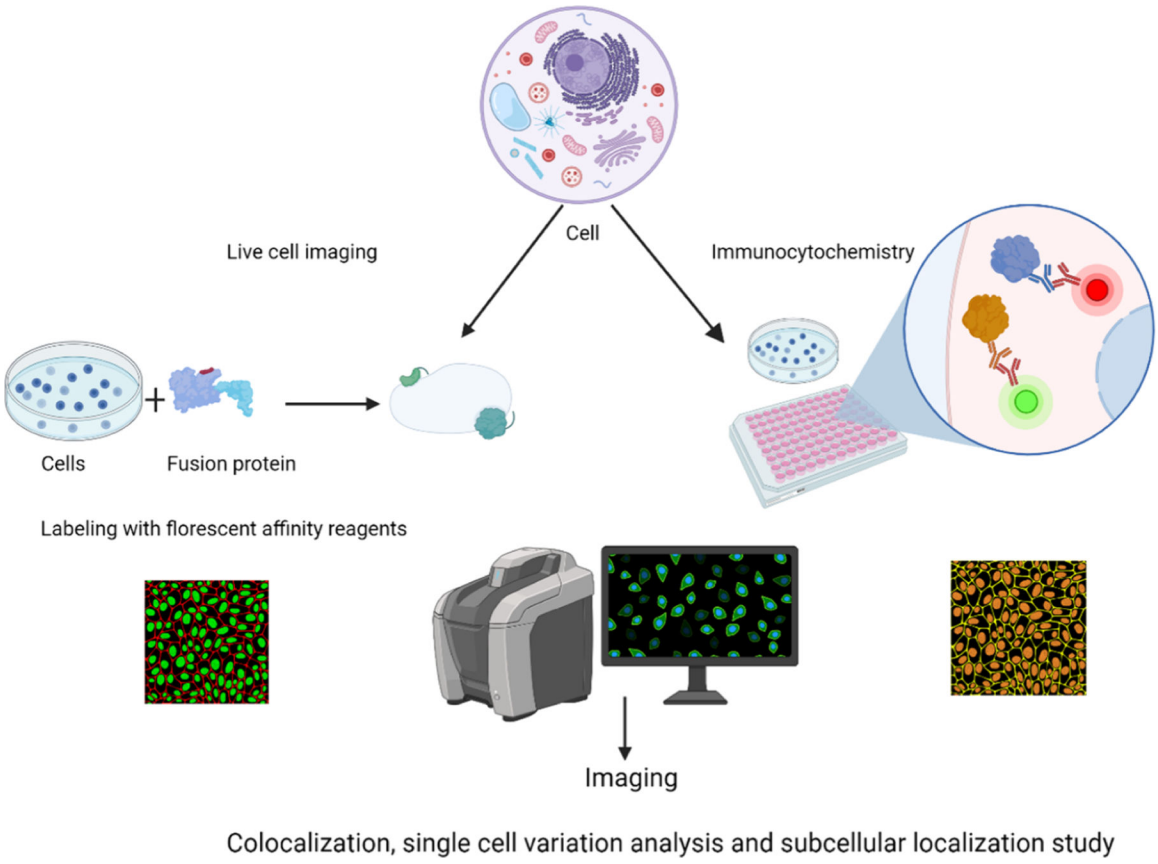


Figure 3.

Schematic workflow of imaging-based subcellular protein niche identification. Imaging analysis can be done through live cell imaging using membrane-permeable dyes and fusion proteins. Time-lapse microscopy studies the protein translocation between subcellular compartments. Alternatively, the immunochemistry-based method uses antibodies, aptamers, and nanobodies against target proteins to study the cell compartment proteome.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

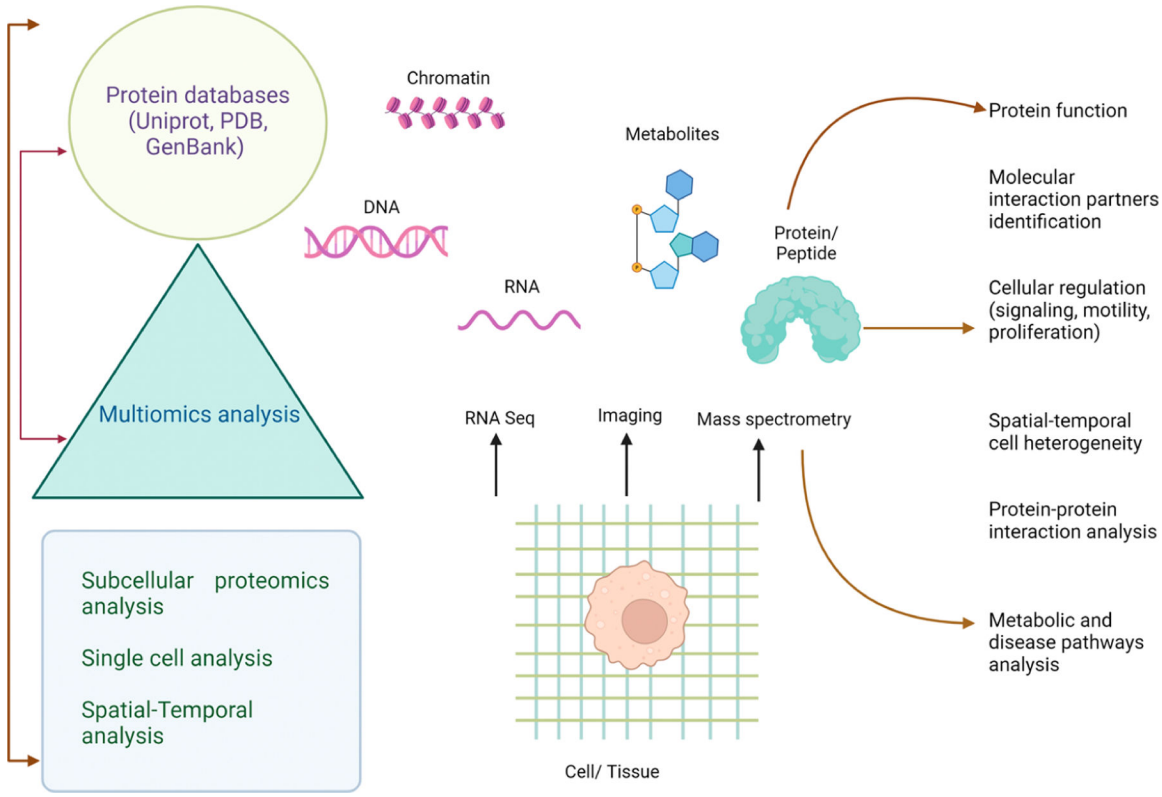


Figure 4.
Summary of machine learning-aided spatial proteomics applications in cell biology.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

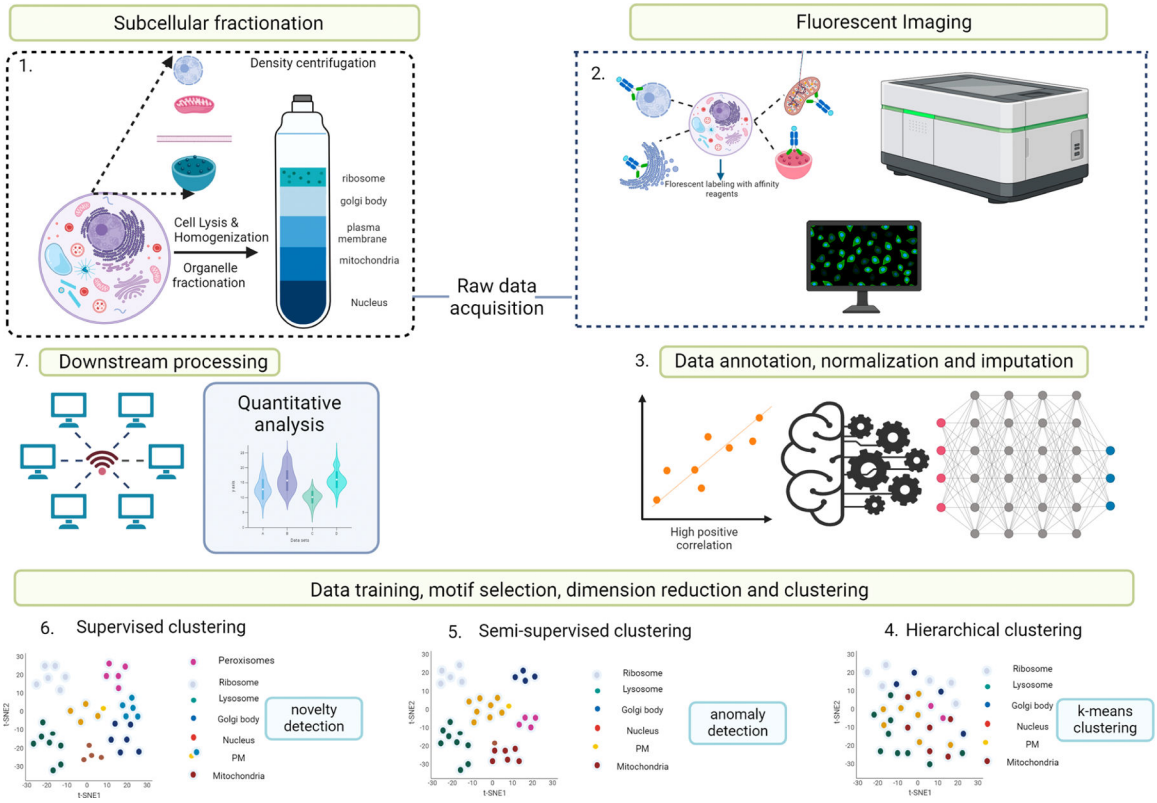


Figure 5.

Following is a schematic workflow of machine learning tools that can be used with MS-based spatial proteome data analysis: (1) Cell homogenization and subcellular fractionation are performed to determine the enrichment of the organelle of interest. (2) Mass spectrometry is used to identify each subcellular component proteome, which provides a large amount of raw data to analyze. (3) The data processing step is crucial, as missing values are imputed, and the data is normalized against the database. (4) Processing large data sets can be challenging, but machine learning tools can assist in data reduction and clustering. (5) Semisupervised clustering is generally used for novelty detection of the cellular compartments. (6) Similarly, supervised clustering can predict the subcellular niche of the protein of interest. (7) Downstream quantitative analysis methods such as cluster analysis algorithms are then used to visualize the data.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1.

Cell Lysis and Disruption Methods Used for Subcellular Fractionation and Subsequent MS Analyses

Lysis method	Homogenization	Description	Application	Advantages	Disadvantages
Mechanical breakage	Waring Blender Polytron	Rotating blades grind and disperse cells and tissues	Complex tissues (Liver, muscle)	Large scale possible	Large scale possible
				Can be used to disperse cells from tissue samples	Can be used to disperse cells from tissue samples
				Works well with complex tissues (e.g., liver, muscle)	Works well with complex tissues (e.g., liver, muscle)
Tissue/cell homogenization	Dounce and Potter-Elvehjem homogenizer	Cell or tissue suspensions are sheared by forcing them through a narrow space	Dounce and Potter-Elvehjem homogenizers: small volumes, soft tissues, cultured cells	Easy to use with smaller volumes	Low throughput
	French Press		French press: bacterial cells		Dounce: reproducibility may vary
Sonication	Sonicator	High-frequency sound waves shear cells	Eukaryotic, bacterial cells, tissue	Independent of cell type	Produces heat that needs to be mitigated to prevent damage to sensitive proteins
				High lysing efficiency	Produces cellular debris
				Shears chromosomes, eliminating the need for nuclease treatment	Loud process and requires sound mitigation
Freeze—thaw	Freezer or dry ice with ethanol	Repeated cycles of freezing and thawing disrupt cells through ice crystal formation	Mammalian cells	Inexpensive technique	Cannot be used for extracting Cellular components sensitive to temperature
					Time-consuming
Nitrogen cavitation	Cavitation machine	High-pressure nitrogen effervescence	Mammalian cells	Fast, uniform, and controllable technique	Not well suited to bacterial, yeast cells
					Expensive instrumentation
					Best suited for cell membrane study
Manual grinding	Mortar and pestle	Grinding tissue, frozen in liquid nitrogen	Hard tissues (e.g., plants), large tissue samples	Works well with challenging tissue samples (e.g., plants); reconstitution in a buffer of choice	Time-consuming
					Reproducibility may vary

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Standard Gradient Media Used in the Density Centrifugation Method

gradient medium type	uses
Sucrose	Organelles, proteins, vesicles, polysomes
Glycerol	Mammalian protein separation
Ficoll, polysucrose and dextrans	Mammalian cells (sometimes in combination with iodinated density gradient media), mammalian subcellular particles (infrequent)
CsCl	DNA, proteins
Cs_2SO_4	DNA, RNA
Diatrizoate	WBC separation and enrichment
Nycodenz, Histodenz	Mammalian cells, organelles, membrane vesicles, viruses
Iodixanol	Mammalian cells, organelles, membrane vesicles, viruses, plasma lipoproteins, proteins, DNA
Colloidal silica media	
Percoll	cell compartments and membrane fractions

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3.

Recent Publications Indicating Spatial Proteomics Methods and Integration of Machine Learning Algorithms

Profiling method	Cell fractionation method	MS technique	ML algorithm	Sample	Study title	DDA/DIA acquisition
Protein Correlation Profiling (PCP)	differential centrifugation	LFQ	Neural network algorithm implemented in pRoloc	HaCaT skin cells	Spatial proteomics reveals subcellular reorganization in human keratinocytes exposed to UVA light	DDA
HyperLOPIT	Nitrogen cavitation	LFQ	Semisupervised and unsupervised clustering	*Trypanosoma brucei Trypanosoma congolense*	Mapping diversity in African trypanosomes using high-resolution spatial proteomics	DDA
LOPIT	Differential centrifugation	LFQ	Support vector machine-based supervised clustering	U2OS cells	Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics	DDA
	Differential centrifugation and Nycodenz density gradient centrifugation	LFQ	Clustering tools	Yeast cells	Analysis of Yeast peroxisomes via spatial proteomics	DDA
Dynamic Organellar Maps (DOMs)	Differential centrifugation	LFQ	SVM supervised clustering	HeLa cells	Deep and fast label-free dynamic organellar mapping	DIA

J Proteome Res. Author manuscript; available in PMC 2024 August 03.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 4.

Open-Source Image Analysis Tools[a]

Name	Based on	Source	Distinct feature	Skills required
ImageJ	Java-based image processing program	NIH-LOCI Collaboration	Not suitable for whole slide image analysis	Basic programming skills
CELLPROFILER	MATLAB	Broad Institute of MIT	To analyze and batch-process cells in biological images	For researchers without programming skills
Fiji	Based on ImageJ	ImageJ plugins	Ready-to-use ImageJ plugins for use in life sciences	Minimal computational skills
QUPATH	Comprehensive desktop software	Queen's University Belfast	Specifically, to analyze whole slide imaging	Researchers without computer programming skills
ILASTIK	Based on random forest classifier	–	For classification of 2, 3, and 4D images in a unified way	Researchers without computer programming skills
Orbit	Tile-based processing pipeline	–	For quantifying large-format images such as whole slide images of tissue	Software development skills required
CYTOMINE	Web-based annotation, image analysis tool	–	For large-scale image-based studies	Researchers without computer skills
ICY	Photoshop tool dedicated to image analysis	Institute Pasteur and France-BioImaging	Whole slide imaging capability	Researchers with software development skills
PATHML	Python 3-based image framework generator	Dana-Farber Cancer Institute in Boston, MA, USA.	Whole slide imaging capability	Researchers with software development skills

[a]Abbreviations: National Institutes of Health (NIH) and Laboratory for Optical and Computational Instrumentation (LOCI) at the University of Wisconsin.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 5.

Summary of Subcellular Protein Localization Prediction Tools[a]

Tool	Species	Computational algorithm	Compartments identified	Web server
BUSCA (2018)	Eukaryote, Prokaryote	Integrated method	16 (for plant proteins); 9 (for animal/fungi proteins); 4 (for Gram-negative bacterial proteins); 3 (for Gram-positive bacterial proteins)	https://busca.biocomp.unibo.it/
CELL02GO (2014)	Eukaryote, Prokaryote, virus	SVM and homology search	4 subcellular localizations in archaea and Gram-positive bacteria, 5 subcellular localizations in Gram-negative bacteria, and 12 subcellular localizations in eukaryotes	http://cello.life.nctu.edu.tw/cello2go/
MULocDeep (2021)	Eukaryote	LSTM + attention	10 major subcellular compartments-	www.mu-loc.org
DeepLoc (2017)	Eukaryote	CNN + LSTM + attention	10 different localizations	https://services.healthtech.dtu.dk/services/DeepLoc-2.0/
TargetP 2.0 (2019)	Eukaryote, Prokaryote	LSTM + attention	TargetP-2.0 server predicts the presence of N-terminal sequences: signal peptide (SP), mitochondrial transit peptide (mTP)	https://services.healthtech.dtu.dk/services/TargetP-2.0/
MU-LOC (2018)	Plant	SVM and neural network	5	https://www.mu-loc.org/
LocTree3 (2014)	Eukaryote, Prokaryote	SVM and homology search	18 EUK, 6 BAC	https://rostlab.org/services/loctree3/
MitoFates (2015)	Eukaryote	SVM	MitoFates predicts mitochondrial sequences, a cleavable localization signal located in the N-terminal, and its cleaved position	https://bioweb.pasteur.fr/packages/pack@MitoFates@1.2
LOCALIZER (2017)	Plant	SVM	5	http://localizer.csiro.au/
SignalP 5.0 (2019)	Eukaryote, Prokaryote	CNN, bidirectional LSTM, and CRF	signal peptides and the location of their cleavage sites in proteins	https://services.healthtech.dtu.dk/services/SignalP-5.0/
DeepSig (2018)	Eukaryote, Bacteria	Deep Convolutional Neural Networks (DNNs) and Conditional Random Fields	signal peptides and the location of their cleavage sites in proteins	https://deepsig.biocomp.unibo.it/welcome/default/index
PSORTb 3.0 (2010)	Bacteria	SVM and homology search	5 (for gram-negative bac) 4 (for gram positive bac)	https://www.psort.org/psortb/
WoLF PSORT (2007)	Eukaryote	k-NN classifier	10 cellular compartments	https://wolfpsort.hgc.jp/
SubCons (2017)	HUMAN	Random Forest classifier	9 cellular compartments	http://subcons.bioinfo.se/
TPpred 3.0 (2015)	Eukaryote	Integrated method	Peptide prediction	https://tppred3.biocomp.unibo.it/welcome/default/index
ERPred (2017)	Eukaryote	SVM	ER resident proteins	http://proteininformatics.org/mkumar/erpred/indexhtml
SCLpred-EMS (2020)	Eukaryote	DEEP neural network	Endomembrane system and secretory pathway (EMS)	http://distilldeep.ucd.ie/SCLpred2/
SeqVec (2019)	Eukaryote	Language Model + FNN	–	https://github.com/mldlproject/2022-iPromoter-Seqvec

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Tool	Species	Computational algorithm	Compartments identified	Web server
LA (2021)	Eukaryote	Language Model + attention	–	https://embed.protein.properties/
ProtTrans (2020)	Eukaryote	Language Model + FNN	–	https://github.com/agemagician/ProtTrans

[a]Abbreviation: supervised learning algorithm (SVM), long short-term memory recurrent neural networks (LSTM), convolutional neural networks (CNN), conditional random fields (CRF), and feedforward neural network (FNN).

J Proteome Res. Author manuscript; available in PMC 2024 August 03.