TECHNOLOGIES AND TECHNIQUES

Spatial proteomics: a powerful discovery tool for cell biology

Emma Lundberg[1,3,4]* *and Georg H. H. Borner*[2]*

Abstract | Protein subcellular localization is tightly controlled and intimately linked to protein function in health and disease. Capturing the spatial proteome — that is, the localizations of proteins and their dynamics at the subcellular level — is therefore essential for a complete understanding of cell biology. Owing to substantial advances in microscopy, mass spectrometry and machine learning applications for data analysis, the field is now mature for proteome-wide investigations of spatial cellular regulation. Studies of the human proteome have begun to reveal a complex architecture, including single-cell variations, dynamic protein translocations, changing interaction networks and proteins localizing to multiple compartments. Furthermore, several studies have successfully harnessed the power of comparative spatial proteomics as a discovery tool to unravel disease mechanisms. We are at the beginning of an era in which spatial proteomics finally integrates with cell biology and medical research, thereby paving the way for unbiased systems-level insights into cellular processes. Here, we discuss current methods for spatial proteomics using imaging or mass spectrometry and specifically highlight global comparative applications. The aim of this Review is to survey the state of the field and also to encourage more cell biologists to apply spatial proteomics approaches.

[1]*Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH – Royal Institute of Technology, Stockholm, Sweden.*

[2]*Max Planck Institute of Biochemistry, Department of Proteomics and Signal Transduction, Martinsried, Germany.*

Present addresses:
[3]*Department of Genetics, Stanford University, Stanford, CA, USA.*

[4]*Chan Zuckerberg Biohub, San Francisco, CA, USA.*

**e-mail: emma.lundberg@scilifelab.se; borner@biochem.mpg.de*

https://doi.org/10.1038/s41580-018-0094-y

Eukaryotic cells are highly compartmentalized to ensure the partitioning of biological processes. Compartments include diverse membrane-bound organelles, such as the Golgi apparatus, the endoplasmic reticulum (ER) and mitochondria, as well as non-membranous structures, such as large protein assemblies (for example, the centrosome and focal adhesions) and phase-separated droplets (for example, RNA granules[1,2]). Protein function is closely linked to subcellular localization, as different compartments provide different chemical environments (such as pH and redox conditions), potential interaction partners or substrates. Tight regulation of protein subcellular localization is hence an important layer of control over cell physiology[3]. Most cell-biological processes involve changes in protein subcellular localization, such as the nucleocytosolic shuttling of transcription factors, the relocalization of mitochondrial proteins during apoptosis and the endocytic uptake of cell-surface cargo receptors and signalling receptors. Conversely, mislocalization of proteins is frequently associated with cellular dysfunction and disease, including neurodegeneration, cancer and metabolic disorders[4–7]. Knowledge of the spatial distribution of proteins at the subcellular level and the ability to capture protein subcellular dynamics are therefore essential for a complete understanding of cell biology.

Recent substantial advances in high-throughput microscopy[8], quantitative mass spectrometry (MS)[9–12] and interactomics mapping[13,14], as well as machine learning applications for data analysis[15–17], have enabled proteome-wide investigations of spatial cellular regulation. A typical human cell expresses more than 10,000 different proteins, spanning an abundance range of seven orders of magnitude[18]. Current large-scale studies of the human spatial proteome suggest that it has a highly complex architecture that includes single-cell variation (in both protein level and localization), dynamic protein translocation, changing interaction networks and the localization of approximately half of all proteins to multiple compartments (implying potential 'moonlighting' activity). The incorporation of global quantification data enables cellular model building and systems analyses that go beyond qualitative descriptions[19–23]. Furthermore, several studies have successfully harnessed the power of global spatial proteomics to investigate diseases, including acute viral infection[24] and liver disease[25], or to pinpoint the cellular defects that underlie monogenic disorders[26,27]. We have now reached the point at which spatial proteomics finally integrates with other 'omics' technologies, cell biology and medical research to provide unbiased systems-level insights into biological processes.

The currently available spatial proteomics approaches are highly complementary, and their individual advantages and limitations make them suitable for different

Dynamic protein translocation
Translocation describes the movement of a protein between cellular compartments. Dynamic translocation refers to a constant change in translocation activity.

types of application. Although spatial proteomics has great potential to benefit most areas of cell biology and is becoming technically amenable, even for non-specialist laboratories, the sheer diversity of approaches can be daunting for newcomers. The aims of this Review are therefore threefold: first, to provide a comprehensive overview of the available concepts and methods in spatial proteomics and how they are applied to understand disease mechanisms, cellular pathways and organellar dynamics; second, to raise awareness of the possibilities provided by spatial proteomics and thereby to stimulate future research; and third, to initiate a discussion of how this still fairly young field could start to combine its data for the benefit of the wider research community. In this Review, we use the term spatial proteomics to refer specifically to subcellular protein mapping; examples of more macroscopic proteomic mapping of organs and tissue regions, which we do not discuss here, include the human tissue atlas[28], mouse brain atlas[29] and human heart atlas[30].

Spatial proteomics methods

In general, three complementary approaches are used for spatial proteomics: MS analysis of fractionated organelles, protein–protein interaction network analysis and proteome-wide imaging of protein localization (FIGS 1,2,3; Supplementary Table 1). This section provides a conceptual overview of these methods, illustrated with recent applications in diverse areas of cell biology.

MS-based organelle profiling

MS can be used to identify proteins and quantify their abundance in highly complex mixtures[9–11]. The traditional biochemical approach to spatial proteomics is to enrich a particular organelle by subcellular fractionation followed by identification of proteins in the peak fraction by MS (FIG. 1a). For homogeneous compartments with a characteristic size, density and shape, substantial enrichment can be achieved (as, for example, for synaptic vesicles[31]). However, although organellar isolation approaches are still commonly used[32], it is now widely recognized that most subcellular compartments are not amenable to genuine 'purification', owing to their intrinsic heterogeneity and overlapping physical properties. Thus, the detection of a protein in a fraction enriched for a target organelle is insufficient evidence of its specific organellar association. The advent of quantitative MS has enabled much progress in this respect, and a suitable experimental design can help to distinguish between compartment constituents and nonspecific background in an unbiased way. These organelle profiling approaches have been applied to single organelles and, importantly, to whole cells, yielding comprehensive high-resolution organellar maps of the cell.

Single-organelle profiling. The basic strategy for organelle profiling is to carry out tailored biochemical fractionation to enrich for a specific organelle and then to quantify proteins across the different steps of the enrichment protocol using MS (FIG. 1b). For every protein, an abundance distribution profile is obtained. Proteins associated with the target organelle or compartment have similar profiles and can thus be distinguished from contaminants, which have different profiles. This powerful approach[33] was first implemented with MS in protein correlation profiling (PCP) of centrosomes purified on a sucrose gradient[34] and has since seen successful application to define several organellar proteomes (including lipid droplets[35], mitochondria[36,37], lysosomes[38] and transport vesicles[39,40]). Conceptually related studies include those featuring the isolation of intact nuclei from *Xenopus laevis* eggs followed by quantification of proteins in the nuclei and cytosol to obtain accurate nucleocytosolic distributions[41] and profiling of the plasma membrane relative to the whole cell proteome following human cytomegalovirus (HCMV) infection in fibroblasts to reveal how the virus reshapes the cell surface to evade the immune response[42]. Furthermore, profiling has been combined with compartment ablation or modification. Thus, analysis of a mitochondria-enriched fraction before and after depletion of the mitochondrial import machinery enabled the identification of the mitochondrial 'importome' in trypanosomes; this approach was also successful for the characterization of peroxisomes[43]. Similarly, profiling of vesicle-enriched fractions from cells that were depleted of different key proteins of the clathrin transport machinery revealed the composition of different types of clathrin-coated vesicle[39]. These studies all achieved high levels of sensitivity and specificity. Owing to the focus on a single organelle, the approach is generally best suited to addressing targeted research questions.

Multi-organelle profiling to create MS-based maps of the cell. The concept of single-organelle profiling can be extended to encompass all subcellular compartments simultaneously. To achieve this, cells are lysed mechanically to release (ideally intact) organelles, which are then partially separated by gradient centrifugation (FIG. 1c). Although different organelles peak in different fractions, they usually have substantially overlapping distributions. Next, fractions from the gradient are analysed by quantitative MS to yield an abundance distribution profile for each protein. Cluster analysis identifies groups of proteins with similar profiles that correspond to the different organelles. Clusters are then overlaid with established organelle markers to reveal cluster identities. Finally, the marker profiles are used to train a classification algorithm (such as support vector machines or neural networks) to predict the compartment association of the non-marker proteins. The result is a detailed organellar map of the cell indicating the subcellular localization of the quantified proteins.

Current implementations of multi-organelle profiling achieve remarkable resolution and coverage of major subcellular compartments in a single experiment and thus provide genuine systems-level analysis tools (see below and BOX 1). Experimentally, multi-organelle profiling is also comparatively rapid and technically simple (Supplementary Box 1). By contrast, the indirect inference of subcellular localization is not always straightforward, as profiles reflect the average steady-state distribution of proteins. The profiles of proteins that are

predominantly associated with a single type of organelle are easily interpreted by comparison with organellar marker profiles. However, for proteins localized to multiple compartments, mixed profiles are obtained, which can lead to uncertain in-between classifications or even erroneous classifications[44]. Analysis of mixed profiles is one of the biggest challenges for profiling approaches, especially because almost no quantitative reference data for multi-organellar distribution are available to guide and validate profile deconvolution attempts. As a result,

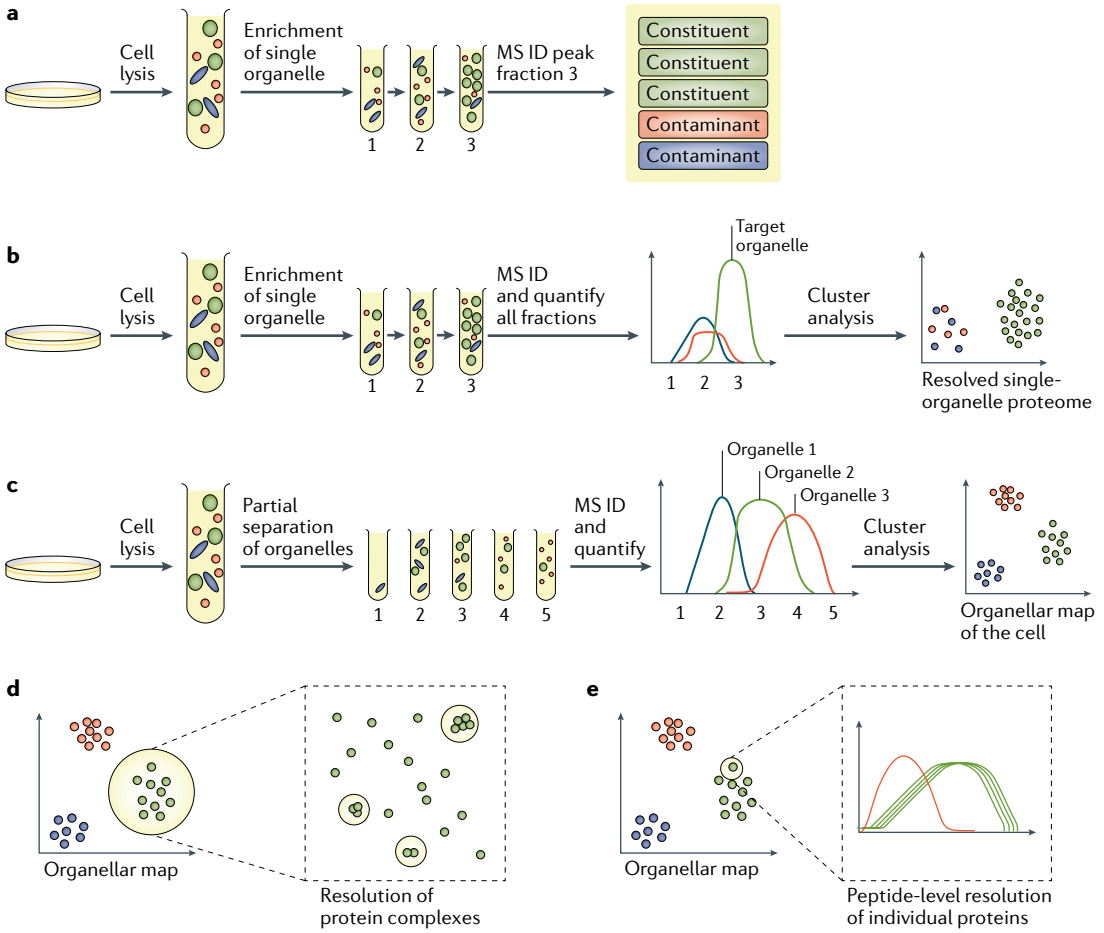
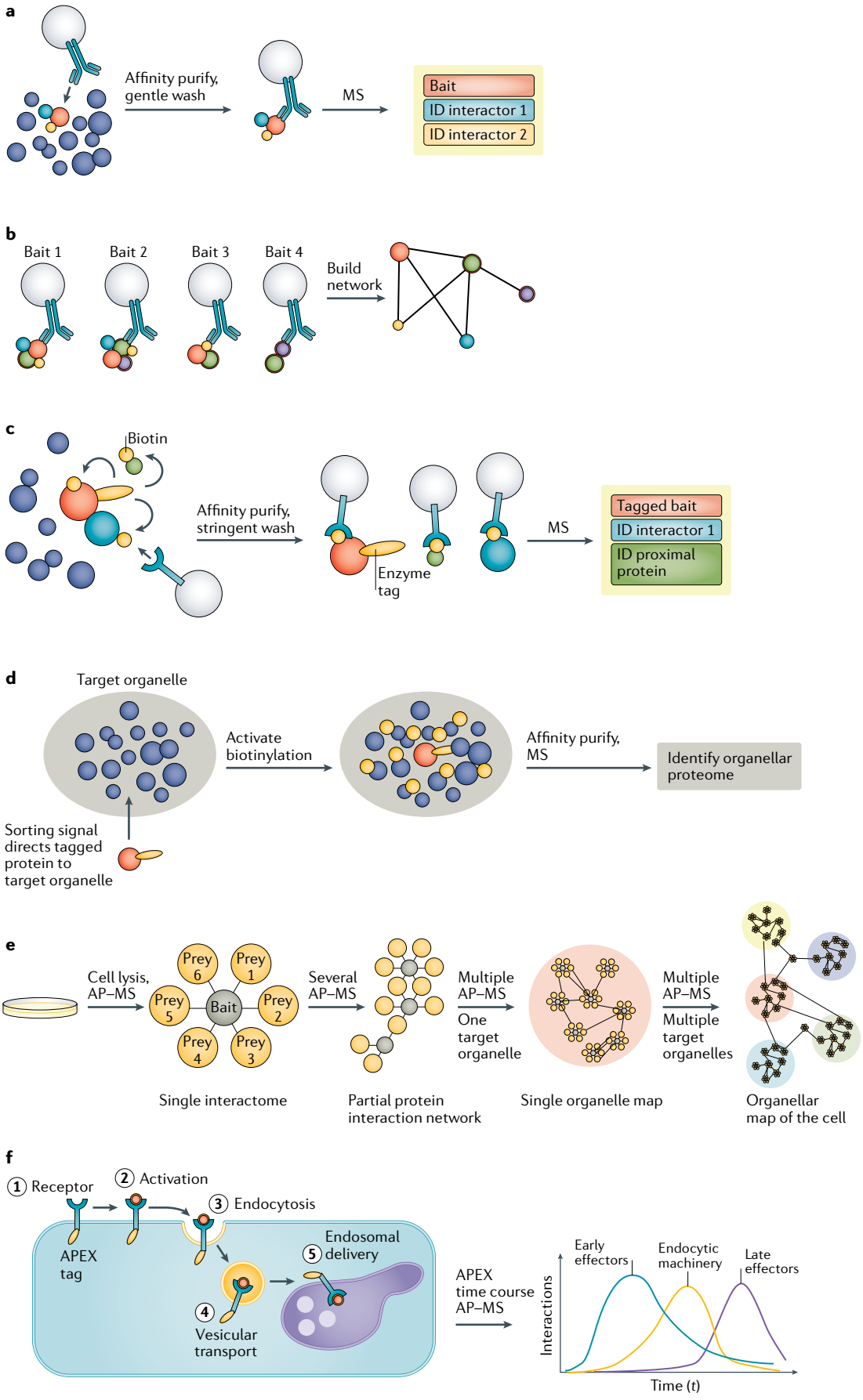


Fig. 1 | **Spatial proteomics by MS analysis of fractionated organelles. a** | In conventional organelle enrichment, cells are lysed, and tailored subcellular fractionation (for example, gradient centrifugation or differential centrifugation) is carried out to enrich a target organelle (green), followed by analysis with mass spectrometry (MS), of the most enriched fraction only. Identified proteins include genuine constituents of the target organelle and co-enriching contaminants (red and blue), which cannot be objectively distinguished; consequently, this approach is not recommended. **b** | In single-organelle profiling, cells are lysed, and tailored subcellular fractionation is carried out (as in part **a**) to enrich a target organelle (green), followed by quantitative MS of the enriched fraction and one or more of the subfractions from the enrichment protocol (such as neighbouring fractions on a gradient or crude fractions). For each protein, an abundance distribution profile is obtained. Proteins associated with the target organelle (green) have similar profiles and can be discriminated from contaminants (red and blue) by statistical analysis. Contaminants are recognized as such but are not necessarily resolved into distinct classes. **c** | In multi-organelle profiling, cells are lysed, and a subfractionation protocol is applied to partially separate all organelles simultaneously (only three are shown to illustrate the principle). No organellar 'purification', as such, is attempted; organelles have largely overlapping distribution. Quantitative MS is then carried out on the subcellular fractions. Each organelle has its own distinct profile, which is shared by all proteins predominantly associated with this organelle. Protein profiles are deconvolved by cluster analysis, and annotation of the plot with established organelle markers reveals cluster identities. In the obtained organellar map, the profile of each protein is represented by a dot; the position indicates the organellar association of the protein. **d** | Organellar maps have high local resolution. Proteins that are part of the same complex have tightly linked fractionation profiles that appear as microclusters within organellar maps, a feature that can be used to identify (ID) novel protein complexes[40]. **e** | Organellar maps can provide peptide-level resolution. Quantification by MS works at the level of peptides. The fractionation profile of a protein (a dot on the organellar map, left panel) is an average (or median) of the profiles from all peptides matching the sequence of that protein. Peptide profiles from the same protein tend to be closely aligned (right panel, green). However, if a post-translational modification (such as phosphorylation) causes a shift in the subcellular localization of a protein, the peptide containing the modification will have a different profile (red). As usually only a fraction of the copies of a protein present within a cell are modified, the profile of the modified peptide is different from that of the bulk of the other peptides.


a
Affinity purify, gentle wash
MS
Bait
ID interactor 1
ID interactor 2
b
Bait 1
Bait 2
Bait 3
Bait 4
Build network
c
Biotin
Affinity purify, stringent wash
Enzyme tag
MS
Tagged bait
ID interactor 1
ID proximal protein
d
Target organelle
Activate biotinylation
Affinity purify, MS
Identify organellar proteome
Sorting signal directs tagged protein to target organelle
e
Cell lysis, AP–MS
Prey 6
Prey 1
Prey 5
Bait
Prey 2
Prey 4
Prey 3
Several AP–MS
Multiple AP–MS
One target organelle
Multiple AP–MS
Multiple target organelles
Single interactome
Partial protein interaction network
Single organelle map
Organellar map of the cell
f
1 Receptor
2 Activation
3 Endocytosis
APEX tag
4 Vesicular transport
5 Endosomal delivery
APEX time course AP–MS
Early effectors
Endocytic machinery
Late effectors
Interactions
Time (t)

◀ Fig. 2 | **Spatial proteomics through interaction networks. a** | Affinity purification–mass spectrometry (AP–MS). Proteins and their binding partners can be affinity purified from complex mixtures, such as whole-cell lysates, using antibodies. Gentle washing conditions are required to preserve protein–protein interactions, and MS analysis is used to identify (ID) the recovered proteins. **b** | AP–MS networks. The interactome of a single protein is a 'local' spatial proteome, as interacting proteins must be in the same subcellular location. Carrying out multiple AP–MS experiments with interacting baits reveals a network of associations that also has spatial information. **c** | Proximity labelling. A bait protein can be tagged with an enzyme (such as engineered ascorbate peroxidase (APEX) or a promiscuous biotin ligase in the biotin identification (BioID) method) that catalyses the biotinylation of proteins in close proximity (<10–20 nm). Targets usually include the tagged protein itself and direct binding partners but can also include transient interactors or even closely juxtaposed (but not directly binding) proteins. As the modification is covalent, biotinylated proteins can be recovered using affinity purification under stringent conditions before identification by MS. **d** | Spatial proteomics through proximity labelling. Proteins tagged with APEX or a promiscuous biotin ligase (BioID) can be chosen or engineered to target a specific organelle (as illustrated here) or subcellular location (such as the synapse). Activation of the enzyme causes widespread biotinylation of nearby proteins, which can be identified using MS. Thus, comprehensive compartment proteomes can be obtained without the need for subcellular fractionation. **e** | Spatial proteomics through expanding interaction networks. When multiple AP–MS or proximity labelling experiments are carried out in the same system, the baits and binding partners begin to overlap (as in part **b**). The use of multiple baits from the same subcellular localization yields detailed maps of individual compartments. Mapping of multiple compartments reveals connections between them, ultimately providing a spatial map of the cell. **f** | Spatiotemporal mapping of protein associations. Proximity labelling can be used to reveal the changing interactions of signalling receptors during their activation (by ligand binding), endocytic uptake and vesicular delivery to the endosomal system (left side of schematic) at subminute temporal resolution. The fluctuating abundance of groups of interacting proteins identified at different time points is illustrated (right side).

most multi-organelle profiling experiments using MS include a relatively high proportion (typically 25–45%) of proteins that cannot be confidently assigned to a compartment. Another limitation is that only compartments for which there are reliable markers can be included in the classification scheme[45], although de novo discovery of compartments is possible in principle[46]. Furthermore, organellar profiling 'averages' the protein distribution from the millions of cells that are required for biochemical fractionation, resulting in the loss of any information on cell-to-cell variability.

Several variations of multi-organelle profiling have been implemented[20,21,25,47–51] that differ in the method used to separate organelles, in the MS quantification strategy and in the bioinformatics analysis. Three approaches have been used in multiple experimental contexts: namely, PCP[25], localization of organelle proteins by isotope tagging (LOPIT)[48] and dynamic organellar maps[21] (see BOX 1 for a detailed comparison of these approaches).

Regardless of the method that was used, all applications of multi-organelle profiling published in the past 3 years[20,21,25,26,47,52] achieved very high levels of organellar resolution (≥10 subcellular compartments), proteomic coverage (>5,000 proteins) and classification accuracy (typically estimated at >90%). The compartment localization predictions obtained using LOPIT and dynamic organellar maps have also been compared directly[21]; the agreement was estimated at >91%, strongly supporting the validity of the results obtained with either method. These studies have provided detailed organellar inventories for mouse liver cells[25], neurons[20] and embryonic stem cells[47], as well as for rat liver cells[50], HeLa cells[21] and USO2 cells[52], enabling the first systematic comparisons of spatial proteomes across different cell types and species[20,52].

Importantly, the resolution of MS profiling maps extends far beyond organelles. Owing to the high precision of current MS quantification, the profiles of individual subunits from a stable multimeric protein complex are often so similar that they can be identified as a tight microcluster within organellar maps[21,47] (FIG. 1d; discussed further elsewhere[40]). Thus, organellar maps intrinsically provide a plethora of potential interactomics data, which have yet to be mined systematically. Furthermore, the protein profiles themselves are averages from multiple individual peptides. Organellar maps can thus, in principle, provide peptide-level resolution (FIG. 1e) and reveal the differential localizations of protein splice isoforms and proteolytically processed forms, as well as localization differences linked to post-translational modifications. For example, a phosphorylated peptide and its unmodified counterpart are clearly distinguishable by MS. If the subcellular localization of a protein is changed in response to phosphorylation (as is the case for many transcription factors), then the modified and unmodified pools of the peptide will have different subcellular localizations, which will be evident from their different positions in a peptide-level organellar map. This remarkable property is only beginning to be exploited, as it pushes existing technology to its limits, with relatively few proof-of-principle demonstrations (for example, splice forms[47] and differentially localized phosphopeptides[25]) published to date. Nevertheless, with further advances in MS technology on the horizon, we expect that peptide-level organellar mapping will soon make substantial contributions to unravelling the intricacies of proteoform subcellular localization.

Spatial protein interaction networks

MS has been used extensively to define protein–protein interactions using antibody-mediated affinity purification–MS (AP–MS) experiments (FIG. 2a). Conceptually, the interactome of a protein is a 'local' spatial proteome, as proteins must be in the same place to interact. After analysis of a sufficiently large number of baits from the same system, the mapped interactions eventually form an interconnected network that provides information about protein subcellular localization (FIG. 2b). The two most extensive large-scale interactomics screens in mammalian cells to date, combined, included a few thousand baits[53,54]; however, complete coverage of the proteome and systematic spatial interpretation of the interactions are still lacking. Nevertheless, the approach has been used successfully to map the proteomes of single organelles. The basic strategy is to identify interaction partners of several proteins that are known to reside in the target organelle. As the overall complexity of organelle proteomes is typically 1–2 orders of magnitude lower than that of the whole cell[20,21,47], a comprehensive network of associations can be obtained, even with only a few baits. An excellent example is the elucidation of

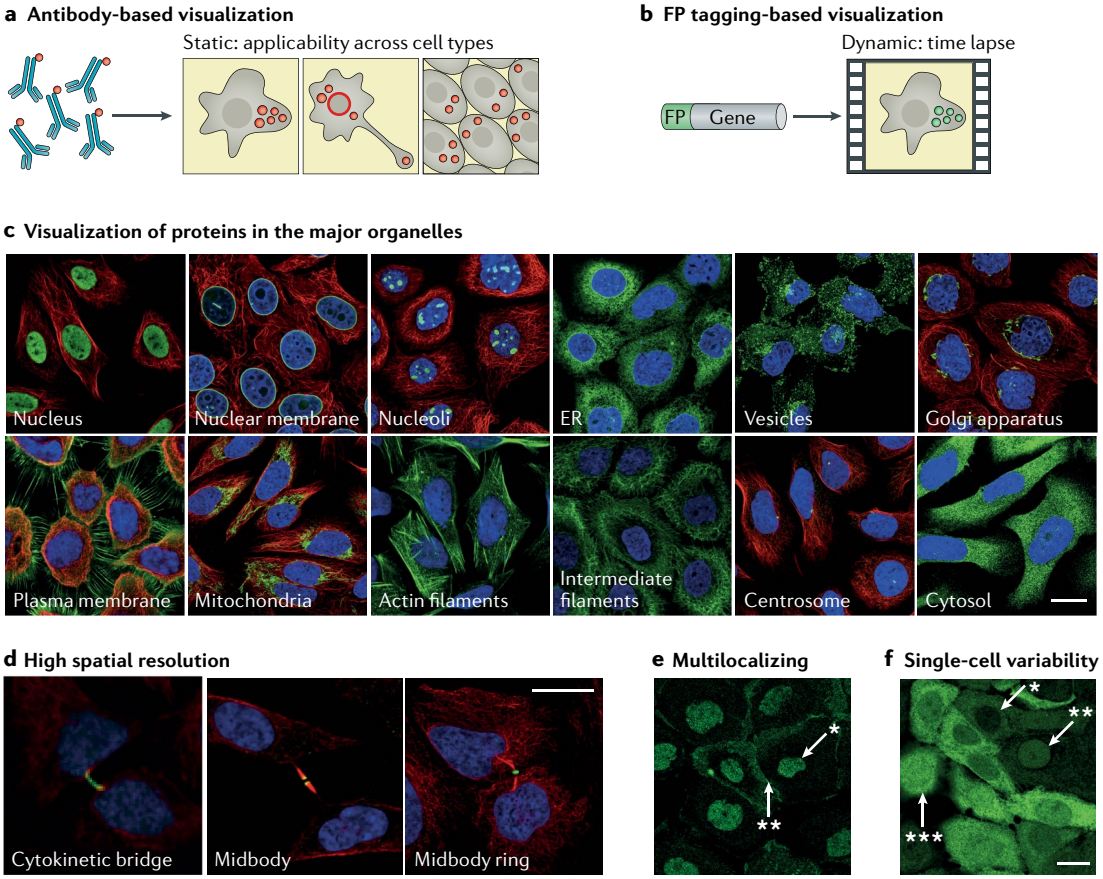


Fig. 3 | **Different approaches to imaging-based spatial proteomics. a** | Antibody-based visualization. Endogenous proteins are visualized using antibodies and immunofluorescence after initial fixation and permeabilization of the cells. For large-scale spatial proteomics studies, the assays are carried out on a massive scale in parallel using robotic pipetting devices. Images of the immunostained cells are acquired using high-throughput fluorescence microscopy. **b** | Fluorescent protein (FP)-tagging-based visualization. Gene-editing technologies are used to tag proteins with an FP reporter, which enables visualization of the protein's localization. Images of live cells are acquired using high-throughput fluorescence microscopy, either in time-lapse series or at static end points. **c** | Cellular organelles and structures. Example images of 12 cellular organelles and structures obtained using antibody-based visualization and fluorescence microscopy are shown. The marker protein for the structure of interest (green) is visualized together with markers for microtubules (red) and the nucleus (blue) to facilitate pattern recognition. **d** | High spatial resolution. Images demonstrate the great spatial detail obtained with confocal microscopy, which can pinpoint protein localization to different substructures of the cytokinetic bridge during cell division. CDCA8 is localized to the entire cytokinetic bridge, whereas APC2 is restricted to the midbody, and CBLN4 is restricted to the midbody ring. **e** | Multilocalizing proteins. HER2 is localized to both the plasma membrane (**) and the nucleus (*) in the same cell. In addition to its canonical function as a receptor tyrosine kinase at the plasma membrane, HER2 has been shown to 'moonlight' as a transcription cofactor in the nucleus. **f** | Single-cell variability. Enolase 1 is a cytosolic metabolic enzyme that moonlights as a DNA-binding protein in the nucleus; single cells with only cytoplasmic localization (*), only nuclear localization (**) or nuclear and cytoplasmic localization (***) are observed. The level of expression (that is, the staining intensity) is also highly variable between cells. Scale bar is 20 μm. ER, endoplasmic reticulum. The images in parts **c**–**f** are from the Cell Atlas[52] of the Human Protein Atlas.

the composition of the algal pyrenoid, a membrane-less compartment in the chloroplast that is crucially important for carbon dioxide concentration and fixation. Following high-throughput GFP tagging and imaging to identify pyrenoid-localized proteins, AP–MS was carried out with only 38 baits[55]. This study revealed many details of the molecular architecture of the pyrenoid and also illustrated the synergy obtained by combining AP–MS with orthogonal imaging.

Proximity labelling methods are especially useful for organellar AP–MS mapping and include, among others, engineered ascorbate peroxidase (APEX) and proximity-dependent biotin identification (BioID)[13,14,56] (FIG. 2c). In both these methods, bait proteins are tagged with an enzyme that biotinylates proteins in close proximity (typically <10–20 nm), and labelled proteins are subsequently recovered using a streptavidin pulldown. As the labelling is not restricted to direct binding partners, even a single bait can identify many proximal proteins, especially in the confined space of an organelle, thereby yielding rich spatial information from a small number of baits (FIG. 2d). Both methods have been used to define

Tandem mass tag multiplexing
A strategy for quantitative proteomic analyses. Peptides from multiple samples are labelled with different mass tags, pooled and analysed as a single sample by mass spectrometry. The tags can be distinguished by their mass and thus enable the simultaneous, relative quantification of peptide and protein abundances across several samples.

organellar proteomes, obtaining excellent coverage and resolution; examples include the analysis of mitochondrial subcompartments using APEX[57,58] and of different types of RNA granule using BioID[59]. Furthermore, the enzyme-tagged constructs can be chosen or engineered to target very specific regions of the cell and thus yield information about compartments that are not amenable to biochemical separation, such as the centrosome–cilia interface[60] and different parts of the synapse (reviewed elsewhere[61]).

Eventually, in-depth mapping of individual organelles will identify links between them and thus begin to generate bottom-up, high-resolution, whole-cell spatial proteomes (FIG. 2e). Progress towards this goal was made by the generation of a database of organellar 'context' BioID proteomes from 18 subcellular localizations using one established marker protein for each compartment[62]. This database, although not exhaustive, can already serve as a reference grid to predict the subcellular localization of a protein from AP–MS data alone based on scoring the known locations of identified interactors.

Proximity labelling is also a powerful tool for spatiotemporal interactomics mapping. Two studies[63,64] combined APEX proximity labelling with time-resolved monitoring of G protein-coupled receptor (GPCR) endocytosis (FIG. 2f). Following activation by the binding of a ligand, plasma membrane GPCRs are rapidly endocytosed by clathrin-coated vesicles and transported to endosomes. APEX labelling at different time points following stimulation enabled the mapping of the changing interactome of tagged GPCRs. Two methodological innovations were key to the success of these studies: first, several additional APEX pulldowns with reference proteins in the various compartments along the endocytic route were included to achieve high specificity[63], and

Box 1 | Current implementations of MS-based multi-organelle profiling

Three major methods of mass spectrometry (MS)-based multi-organelle profiling are currently used, namely, protein correlation profiling (PCP)[142], localization of organelle proteins by isotope tagging (LOPIT)[48] and dynamic organellar maps[21], which all offer comparable levels of compartment resolution and proteome coverage. Although the methods are conceptually similar, differences in organelle separation and MS quantification strategies result in individual-specific and application-specific advantages (see the table for a comparison of these MS-based multi-organelle profiling methods). Of note, all three methods have been applied successfully to generate extensive databases of protein subcellular localization (see below) and have also been used in comparative studies as discovery tools (see Global comparative spatial proteomics).

Protein correlation profiling

Organelles are separated by velocity centrifugation on a sucrose density gradient, and ~20 subfractions are analysed by label-free quantification MS. The large number of subfractions that are quantified in PCP can, to some extent, resolve multimodal organellar distributions, thereby enabling the assignment of proteins to multiple subcellular localizations, which is a key strength of this approach. PCP has been applied to wild-type and steatotic mouse liver cells and provided localization information for >6,000 proteins that were mapped to ten subcellular compartments[25] (see the obesity-induced non-alcoholic fatty liver disease (NAFLD) database).

Localization of organelle proteins by isotope tagging

Organelles are separated by density gradient centrifugation, and ten fractions are selected for further analysis and quantification using tandem mass tag multiplexing (TMT multiplexing). LOPIT achieves good resolution of organelles; first developed to define the plant endomembrane system[141], an updated version (hyper-LOPIT) has been applied to mouse stem cells, resulting in the profiling of >5,000 proteins and mapping of 14 compartments[47] (see the SpatialMap database). Furthermore, LOPIT mapping of U2OS cells enabled the first direct large-scale comparison of MS-based and imaging-derived protein localization data[52], which found a high level of classification agreement between these orthogonal approaches.

Dynamic organellar maps

Organelles are separated by differential centrifugation into eight fractions. In parallel, a total membrane fraction is prepared from metabolically labelled cells, and this fraction is added to every fraction as an internal reference for quantification before MS analysis. This robust workflow yields highly reproducible maps, thereby enabling sensitive comparative analyses (that is, before and after perturbation). The method has been applied to HeLa cells, resulting in the mapping of 8,700 proteins to 15 compartments[21] (see the HeLa Spatial Proteome database). Of note, the workflow has also been implemented with label-free quantification and TMT multiplexing; application of the label-free format provided a detailed organellar map for primary mouse neurons (8,000 proteins and 12 compartments[20]). Furthermore, dynamic organellar maps incorporate absolute protein quantification (that is, protein copy numbers per cell), so that quantitative models of organelle and cell composition can be derived[20,21].

Method	Organelle separation technique	Number of subcellular fractions analysed	Quantification method	Instrument requirements	Approximate measurement time per organellar map (covering >5,000 proteins) (h)	Typical starting material (as published)	Suitable cell types
PCP	Velocity gradient centrifugation	22	Label-free	MS^2	40	One mouse liver	• Tissues (mouse liver) • Cultured cells not yet tested, but most likely suitable
LOPIT	Density gradient centrifugation	10	Tandem mass tags	MS^3	40	10^8 mouse ESCs	• Cultured cells • Tissues (*Arabidopsis* callus)
Dynamic organellar maps	Differential centrifugation	8	• SILAC • Label-free • Tandem mass tags	• MS^2 (SILAC, label-free) • MS^3 (tandem mass tags)	20 (SILAC)	10^7 HeLa cells	• Cultured cells • Tissues not yet tested

ESCs, embryonic stem cells; SILAC, stable isotope labelling with amino acids in cell culture.

second, a fast APEX-labelling protocol was combined with multiplexed quantitative MS to achieve subminute temporal resolution[64].

Imaging-based spatial proteomics

Spatial proteomics using imaging provides a technology for the visualization of proteins in their native cellular environment without the need for cell lysis or the physical separation of compartments or organelles before proteomic analysis (FIG. 3). This in situ localization of proteins makes the detection of proteins with a multimodal organellar distribution tractable. In fact, large-scale imaging studies have revealed that many proteins localize to multiple cellular compartments[52,65]. Furthermore, it has become increasingly clear that populations of genetically identical cells show variability in protein expression levels and protein localization[52,66–68], for example, during cell differentiation[69,70], in response to environmental fluctuations[67,69,71] and following treatment with drugs[19,72–75]. This phenomenon, termed 'bet-hedging', indicates that there is intrinsic variability within a population of cells in the response of individual cells to stress, which serves as a risk-spreading strategy to provide a long-term population fitness advantage. Imaging-based technologies lend themselves to investigating this variability by capturing protein spatial distribution at single-cell resolution.

Imaging requires the visualization of proteins, usually by using affinity reagents such as antibodies, or by the expression of fluorescent protein fusions. Generating antibodies and genetically modified proteins is both expensive and time consuming, which pose limitations for proteome-wide applications. However, once affinity reagents and fluorescent protein fusions are generated, they provide protein-specific handles that can be used for pulldowns, which affords additional opportunities for subsequent analysis of protein interactions and complexes.

Antibody-based protein visualization. Spatial proteomics by imaging has its origin in the visualization of single proteins by immunofluorescence assays[76,77] (FIG. 3a). An advantage of this type of analysis is that affinity reagents enable visualization of the endogenous gene product and can be easily applied to studying the protein in many different cell types and samples. This approach is highly sensitive, as the fluorescent signal can be amplified using, for example, secondary antibodies. The main disadvantage of immunofluorescence is that intracellular staining requires fixation and permeabilization of the cells, limiting applications to static end point measurements. Furthermore, the fixation and permeabilization procedure can introduce artefacts into the cellular morphology and, in the worst case, affect protein localization[78,79]. For targeted studies, a fixation protocol tailored for the organelle of interest can be used, whereas for proteome-wide studies, it is preferable to use nonspecific crosslinking to minimize the risk of losing soluble proteins during cell permeabilization[79].

Historically, antibodies have been the most commonly used affinity reagents, but advances in combinatorial protein chemistry have enabled the engineering of novel affinity reagents from various scaffolds[80]. Independent of the type of affinity reagent, validation of its specificity is essential to avoid false results. A comparative study of 500 proteins that were localized using fluorescent protein tagging and antibodies revealed that cross reactivity with nuclear components is the most common artefact for antibody-based localization[81]. The awareness of a widespread lack of validation and reproducibility in antibody-based research[82] has led to the development of guidelines for application-specific validation[83] and batch citation[84].

Owing to the costly and time-consuming effort required to generate affinity reagents for entire proteomes, the number of published global spatial proteomics studies is still fairly small. The Human Protein Atlas (HPA) has pioneered this field over the past 15 years and aims to map the spatial distribution of all human proteins in all cell types of the human body and to create a knowledge resource for human biology[28,52,85,86]. For this purpose, a near-proteome-wide collection of antibodies has been generated[87,88] and validated[89–91]. Part of this work has been the creation of a reference map detailing the subcellular distribution of the human proteome in a large panel of cell lines, which is termed the HPA Cell Atlas[52] (FIG. 3c). Using 14,000 antibodies, the localizations of 12,003 proteins have been determined at single-cell resolution, of which 5,662 proteins had never been experimentally localized before this study. The spatial distribution of these proteins has been classified into one or more of 30 cellular structures using a combination of manual and computational[15] image analyses. The high resolution of confocal microscopy allowed the localization of proteins to fine structures, such as the fibrillar centre of nucleoli, the ends of microtubules and substructures of the cytokinetic bridge (FIG. 3d), as well as rods and rings, which are functionally uncharacterized cytoplasmic structures. Surprisingly, this study revealed that as many as half of all human proteins localize to multiple compartments. For the first time a map of the patterns of multilocalization (that is, proteome connectivity) across organelles in the cell could be drawn (FIG. 3e). Furthermore, ~16% of the proteome exhibited variation in either protein expression level or spatial distribution at the single-cell level (FIG. 3f). These data emphasize the need for single-cell resolution assays and multiplexed detection to unravel protein covariation patterns, as well as temporal detection to study the dynamics of multimodal protein distribution. Technological developments, such as highly multiplexed imaging using metal-tagged antibodies and mass cytometry[92–94] or DNA barcoded antibodies and fluorescence microscopy[95,96], enable simultaneous analyses of over 50 proteins at subcellular resolution, thus paving the way for analysis of organelle proteome heterogeneity in the context of neighbouring cells.

Although the genome-wide antibody-based procedure is laborious, a great advantage is that once the antibodies are generated, comparative studies of protein localization in different cell or tissue samples are easily carried out.

Multimodal organellar distribution
Refers to the distribution of proteins that simultaneously localize to multiple compartments within a cell.

Affinity reagents
Molecules, such as an antibody, protein, peptide or nucleic acid, that bind specifically to a target protein to enable the identification, visualization, capture or modulation of the target protein or its activity.

Visualization using fluorescent protein fusions. The insertion of genes encoding fluorescent proteins at endogenous genomic loci enables temporal studies of protein dynamics and localization in a near-endogenous cellular context (FIG. 3b). This is best illustrated by studies of the budding yeast *Saccharomyces cerevisiae*. As homologous recombination is very efficient in yeast, the coding sequence of a fluorescent protein can be easily inserted at a gene locus while largely preserving the endogenous expression pattern of the gene. A genome-wide collection of GFP-tagged yeast strains enabled the systematic localization of 75% of all predicted proteins in *S. cerevisiae* to 22 different cellular compartment categories under normal growth conditions in live cells[65]. This research was the first systems-level study of protein localization in a eukaryote, which at the time provided 1,630 novel localizations for proteins in yeast. Further studies of this yeast fusion collection using fluorescence microscopy have refined the spatial map of the yeast proteome and provided independent protein abundance data sets[19,67,73,74] (reviewed elsewhere[75]). Since then, additional genome-wide yeast libraries have been created; for example, the SWAp-Tag approach was used to generate both a carboxy-terminal and an amino-terminal tagged collection[97,98]. Subsequent efforts have been targeted to reveal proteome organization at a higher resolution than was possible in the initial studies. For example, the precise localization of 200 proteins within the yeast endomembrane system was determined using colocalization with seven known markers of endomembrane compartments, resulting in the identification of novel cargoes of the coatomer complex I (COPI)[99]. Owing to these studies, several yeast imaging-based protein localization databases are now available[100–105].

Systematic imaging studies using fluorescent fusion proteins have also been carried out in human cells, although they have so far been limited to a fraction of the proteome. Pioneering spatial localization studies of human proteins were achieved using ectopic expression of tagged cDNAs. The subcellular localization of 1,600 proteins was determined using transient transfection with open reading frames[106]. Similarly, a library of annotated reporter clones was created by exon tagging using retroviral delivery. In this way, 2,180 human proteins fused to YFP and expressed from their endogenous promoter were localized using time-lapse fluorescence microscopy[107–109]. Since then, the development of CRISPR–Cas9-based methods has transformed our ability to tag human genes at their endogenous loci by facilitating homology-directed repair[110,111] and has enabled studies of the subcellular localization dynamics of the encoded fusion protein under endogenous regulatory control[112–115]. Strategies are emerging for systematic tagging in various organisms and systems, such as *Drosophila melanogaster*[116] and different human cell types[117,118], while the development of self-complementing split fluorescent proteins provides less disruptive tagging[119,120] and large-scale protein labelling[121]. Overall, the development of these methods is paving the way for the construction of genome-wide endogenously tagged libraries of proteins in human cells.

Segmentation
Describes the process of partitioning a digital image into segments that represent, for example, a cell or a nucleus.

Sensitivity is still a limiting factor for applications involving imaging of fluorescent proteins. For example, it is estimated that only the top 30% most abundant proteins in human cells are detectable as GFP fusion proteins by direct fluorescence microscopy[121]. To alleviate this limitation, signals from low-abundance proteins can be improved by using brighter and more photostable fluorescent proteins. Schemes for tagging proteins with repeats of fluorescent proteins have also been developed[121], although any engineering of a native protein risks interfering with its function, expression, degradation and localization. Thus, care must be taken about where the fluorescent proteins are inserted. In addition, validation of cells expressing tagged proteins is needed to verify that the observed protein localization represents the endogenous localization. The agreement between fluorescent proteins and antibody-based approaches in terms of protein localization is ~80%, with an equal incidence of localization artefacts for the two methods; mislocalization to the endomembrane system is the most common problem for fluorescent protein fusions[81].

Although the genome-wide tagging procedure is laborious, a great advantage of this approach is that once the cell library is complete, comparative studies of protein localization in live cells after various perturbations are easily carried out.

High-content microscopy. Independent of how proteins are visualized, a crucial part of a spatial proteomics experiment is the image acquisition. The development of automated fluorescence microscopes has expanded the scope of imaging studies to the extent that thousands of samples can now be analysed in multiple conditions[8]. Key considerations for these experiments include the following. First, the need for static or dynamic time-lapse imaging determines which visualization methods are suitable. Second, there is a trade-off between resolution and throughput. A lower resolution at a higher acquisition rate can be preferable when studying live cells or, for example, cytosol-to-nucleus translocations after perturbations, whereas high-resolution technologies, such as oil-immersion confocal microscopy or super-resolution microscopy (reviewed elsewhere[122]), can reveal highly intricate localization patterns at the suborganelle level[22,52,123,124]. Third, cellular reference markers may be needed for segmentation of cells and nuclei and for improved quantitative image analysis. Nuclear[74] and cytosolic[19,67] markers have most commonly been used, but markers of the endomembrane system[52,99] and cytoskeletal structures[52] have also been included for more refined localization assignments.

Image analysis and location classification. Analysis of image data sets is another crucial aspect of spatial proteomic studies. Although preprocessing steps, such as normalization and cell segmentation, are fairly standardized[125], the subsequent pattern classification task can be highly challenging. The aim of this analysis is to recognize the subcellular pattern or patterns of distribution for a protein and, ideally, also to quantify the amount of protein in each compartment.

Manual pattern recognition has been the predominant approach to date; however, for reproducible, scalable and rapid analysis, automated solutions are needed (reviewed elsewhere[125]). Efforts to automate the classification have been based on methods such as *K*-nearest neighbour classifiers, support vector machines, artificial neural networks and decision trees, often utilizing hand-crafted feature sets to capture the protein pattern in relation to the cellular morphology, including hundreds to thousands of parameters describing the shape, positioning and texture of the staining in relation to the cell[125–128]. Currently, deep neural networks[16,129] are gaining popularity as tools for image analysis[130,131]. These networks process images through consecutive layers of compute units that quantify increasingly complex patterns in the data and are trained to predict given labels in the images. One of the strengths of this approach is that image features can be automatically derived. However, the success of deep neural networks for image analysis crucially depends on the availability of training data to learn model parameters. Deep neural networks have been successful in classifying unimodal-localizing proteins in budding yeast[132,133] and human cells[133,134] with an accuracy of up to 91%, but these studies have been largely limited to the classification of single patterns (9–18 locations) in single cell lines. Subsequent studies have addressed the classification of multilocalizing proteins, in both yeast[19] and human cells[15], achieving accuracies of ~70%. Misclassification is more common for low-abundance proteins[19], rare cellular patterns[15] and highly similar cellular patterns[15,132]. Citizen science approaches have also been used for image pattern classification by integration into online multiplayer games[15] or Amazon's 'mechanical turk'[135]. Despite improvements in pattern classification models, manual pattern classification is still substantially more accurate[15]. A pragmatic middle-ground approach would involve unbiased computational image analysis to identify proteins of interest, followed by manual classification of a subset of the images.

For a truly quantitative spatial proteomics analysis, pattern classification needs to be complemented with organelle segmentation. Developments that hold great promise in this context include convolutional networks (such as U-Net[136]), the advent of generative models capable of the segmentation of ambiguous images[137] and in silico synthesis of realistic fluorescent labels from label-free images, such as prediction of labels representing cell state (dead versus alive) and subcellular features such as organelle location[138,139]. Such advancement of computer vision in cell biology points towards a future in which cellular structures are predicted from label-free images, enabling multiparameter spatial dissection of cells.

Pattern classification becomes increasingly challenging in comparative spatial proteomics investigations across cell types or perturbations, which require that changes in morphology can be distinguished from de facto protein relocalizations. With the primary aim of detecting spatial differences between samples, rather than assigning labels, unsupervised computational approaches lend themselves to the task. Developments towards this end include an unsupervised clustering model capable of identifying protein relocalization between six different cellular compartments across all available yeast GFP library screens[68] and a generalizable model capable of classifying patterns across different human cell lines using HPA Cell Atlas images[15]. The fact that these two studies, as well as the majority of the other subcellular image pattern classification work, used the yeast GFP-localization image collection[19,65,67,74] or the human HPA Cell Atlas collection[52] illustrates that these data sets are valuable benchmark resources for the development of improved machine-learning models. To enable the meta-analysis of protein localization and advance the development of image classifiers, a public image repository for biological image data is needed[140].

Global comparative spatial proteomics

Protein subcellular localization is tightly controlled, and many cell-biological processes involve movements of proteins between compartments. Furthermore, numerous diseases are associated with altered subcellular localization of proteins[4–7]. In principle, global comparative spatial proteomics provides an ideal tool to capture these physiological and pathological protein translocations at the systems level. This approach is unbiased and holistic and thus is suited to reveal novel and unexpected aspects of cell biology. However, global comparative spatial proteomics experiments are technically challenging. Below, we discuss key concepts of dynamic spatial mapping and review applications of this approach that illustrate its power as a discovery and systems analysis tool.

Challenges and general considerations

Currently available global spatial proteomic methods generate static 'snapshot' cellular maps rather than continuous information on protein dynamics. Therefore, all localization changes must be inferred by comparing maps obtained before and after a stimulus, genetic alteration or perturbation. This limitation imposes numerous technical and conceptual challenges. Although global spatial proteomics methods have been available for over a decade[48,141,142], successful comparative studies have been published only fairly recently[19–21,24–27,67,73]. Some important considerations for using these methods are discussed below.

Reproducibility. As detection of changes in protein subcellular localization requires a comparison of at least two data sets, reproducibility is crucial.

All MS profiling approaches require the partial separation of organelles by some form of gradient centrifugation. However, it is practically impossible to achieve identical gradient fractions among multiple experiments, and slight differences between replicates are, in fact, the reason why combining several profiling data sets usually improves resolution[20,44]. Furthermore, a perturbation experiment can affect the size or density of organelles and thus severely alter their fractionation patterns. The MS measurements also add experimental noise. Finally, owing to the stochastic element of shotgun MS, the complement of proteins that is identified in each experiment is usually not identical, reducing the overlap

Citizen science
Public participation in scientific research.

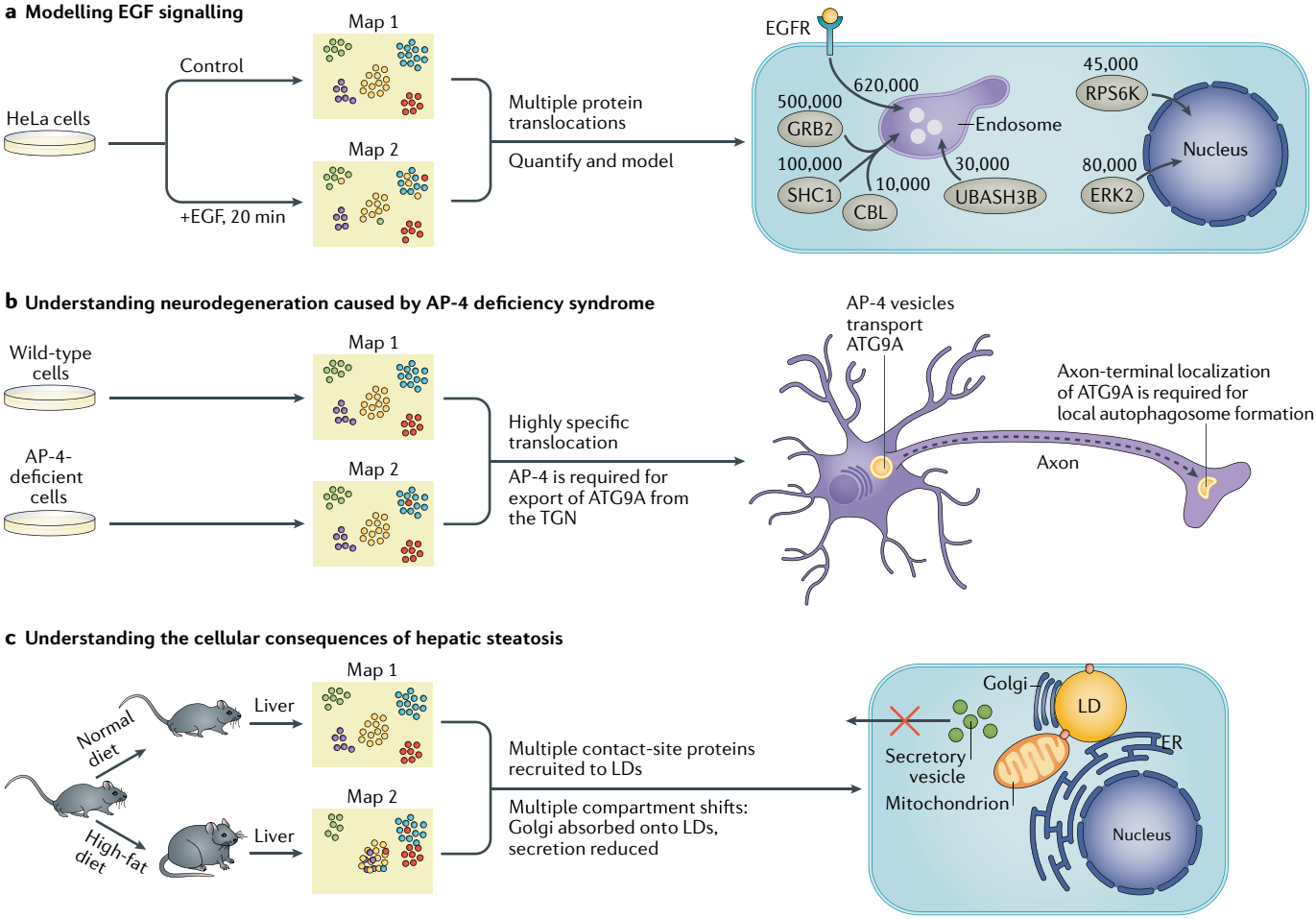
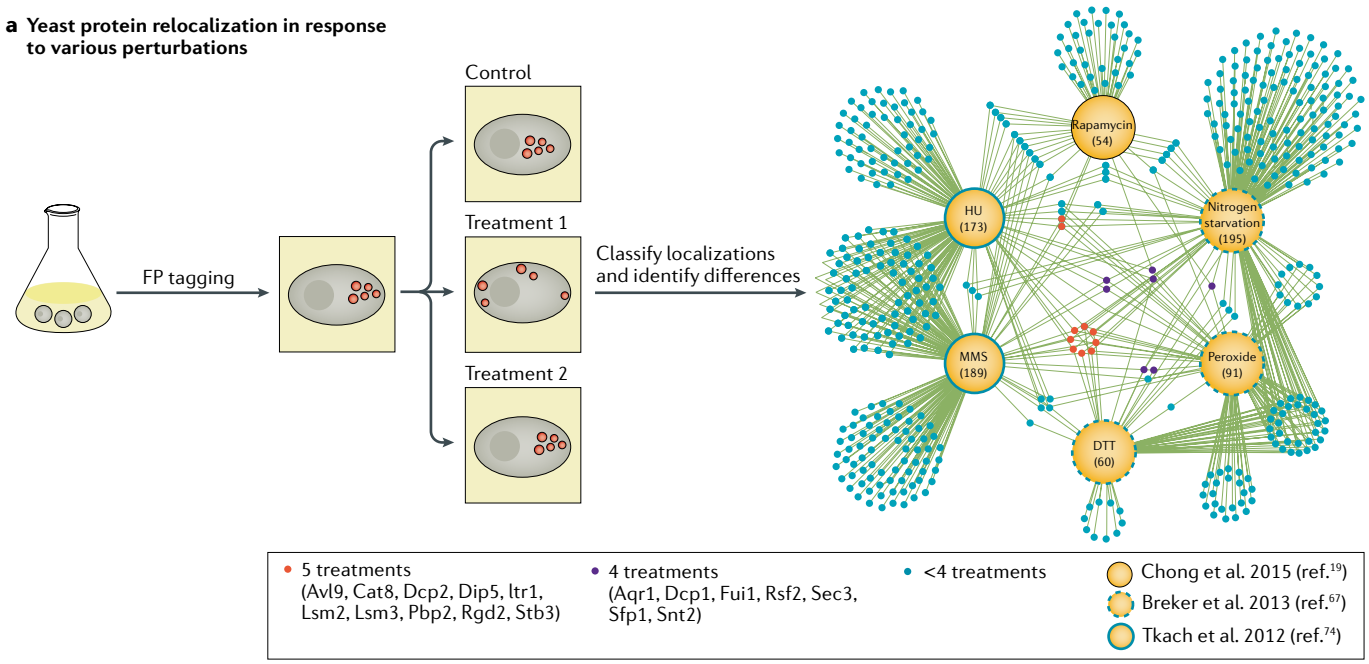
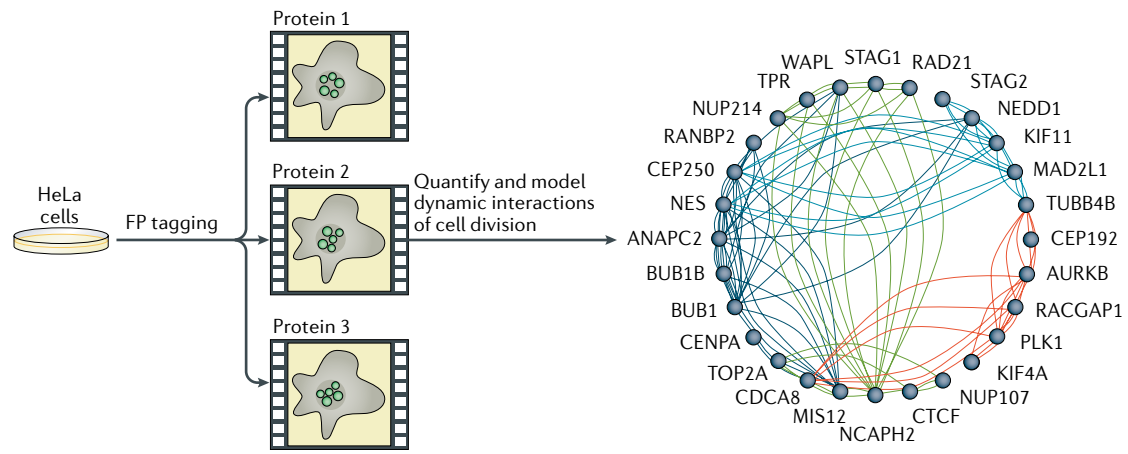
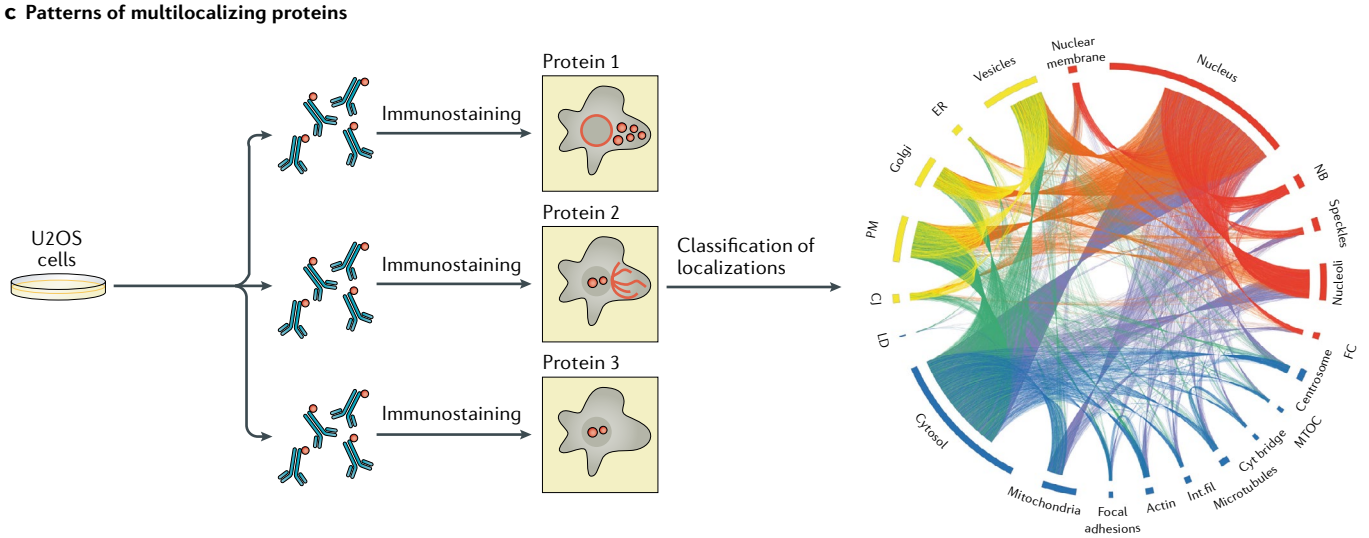


Fig. 4 | **MS-based comparative spatial proteomics: example applications.** **a** | Modelling epidermal growth factor (EGF) signalling[21]. Dynamic organellar maps were generated for HeLa cells that were either untreated or stimulated for 20 min with EGF. Numerous protein translocations were detected, including endocytosis of the EGF receptor (EGFR), recruitment of signalling proteins (GRB2 and SHC1) to EGFR and nuclear shuttling of downstream target transcription factors (ERK2 and RPS6K). Incorporating absolute protein quantification enabled estimation of the copy numbers of proteins undergoing the detected translocations per average cell. Only selected changes are illustrated. **b** | Understanding neurodegeneration caused by adaptor protein 4 (AP-4) complex deficiency syndrome[26]. Mutations in the genes encoding subunits of AP-4 cause severe intellectual disability, axonal degeneration and spastic paraplegia by an unknown mechanism. Application of dynamic organellar maps revealed that the key autophagy protein ATG9A is specifically retained at the *trans*-Golgi network (TGN) in AP-4-deficient cells. An emerging model of the disease is that, in neurons, AP-4 mediates selective transport of ATG9A from the TGN to the distal axon to regulate local autophagosome formation, which is crucial for neuronal homeostasis. **c** | Understanding the cellular consequences of hepatic steatosis[25]. Mice were fed a normal diet or a high-fat diet to induce hepatic steatosis, which is a hallmark of non-alcoholic fatty liver disease (NAFLD). Protein correlation profiling maps were generated to monitor the subcellular translocations of proteins in the liver. Key observations include the relocalization of many contact-site proteins to lipid droplets (LDs), the adsorption of the Golgi onto LDs and ensuing secretion defects, which provide a potential molecular mechanism for the reduced hepatocyte function in NAFLD. **d** | Monitoring cellular responses during human cytomegalovirus (HCMV) infection[24]. Human primary fibroblasts were infected with HCMV. Organellar rearrangements were followed using localization of organelle proteins by isotope tagging (LOPIT) maps every 24 h over a time course of 120 h. This analysis revealed individual protein translocations that are relevant to viral proliferation and organellar rearrangements, such as the merging of the endoplasmic reticulum (ER) and Golgi and a split of lysosomes into two populations. p.i., post-infection; UBASH3B, ubiquitin-associated and SH3 domain-containing protein B. Part **a** adapted with permission from REF.[21], CC-BY-4.0.


a Yeast protein relocalization in response to various perturbations
FP tagging
Control
Treatment 1
Treatment 2
Classify localizations and identify differences
Rapamycin (54)
HU (173)
Nitrogen starvation (195)
MMS (189)
Peroxide (91)
DTT (60)
5 treatments (Avl9, Cat8, Dcp2, Dip5, ltr1, Lsm2, Lsm3, Pbp2, Rgd2, Stb3)
4 treatments (Aqr1, Dcp1, Fui1, Rsf2, Sec3, Sfp1, Snt2)
<4 treatments
Chong et al. 2015 (ref.19)
Breker et al. 2013 (ref.67)
Tkach et al. 2012 (ref.74)


b Spatiotemporal reorganization of human proteins during mitosis
HeLa cells
FP tagging
Protein 1
Protein 2
Protein 3
Quantify and model dynamic interactions of cell division
WAPL
STAG1
RAD21
TPR
STAG2
NUP214
NEDD1
RANBP2
KIF11
CEP250
MAD2L1
NES
TUBB4B
ANAPC2
CEP192
BUB1B
AURKB
BUB1
RACGAP1
CENPA
PLK1
TOP2A
KIF4A
CDCA8
NUP107
MIS12
CTCF
NCAPH2


c Patterns of multilocalizing proteins
U2OS cells
Immunostaining
Immunostaining
Immunostaining
Protein 1
Protein 2
Protein 3
Classification of localizations
Nuclear membrane
Nucleus
Vesicles
ER
Golgi
PM
CL
LD
NB
Speckles
Nucleoli
FC
Centrosome
MTOC
Cyt bridge
Microtubules
Int.fil
Actin
Focal adhesions
Mitochondria
Cytosol

◀ Fig. 5 | **Imaging-based comparative spatial proteomics: example applications.** **a** | Yeast protein relocalization during stress[19,74,100]. Proteome-wide imaging-based studies of GFP-tagged proteins in budding yeast revealed patterns of protein relocalization in response to six environmental stress conditions. The network graph summarizes work from three studies; each edge connects a protein relocalization (cyan, purple and red nodes) to the corresponding perturbation (yellow node). Node colour represents the number of conditions in which the relocalization occurs (proteins indicated in parentheses relocalize in response to all treatments). **b** | Spatiotemporal organization of human proteins during mitosis[144]. 4D imaging integrated with quantitative protein concentration measurements of 28 GFP-tagged human proteins generated an integrated model of spatiotemporal protein reorganization during mitosis in HeLa cells and enabled analysis of dynamic interactions between the proteins. The network graph shows protein colocalizations during cytokinesis (a subphase of mitosis). Each node is a protein, and each edge corresponds to a dynamic localization cluster. The edge colour corresponds to a localization cluster assigned by computational image analysis. **c** | Patterns of multilocalizing proteins in various human cell types[52]. Proteome-wide immunofluorescence analysis of human protein localization across many different cell lines revealed that >50% of proteins localize to multiple cellular compartments. The network graph shows all detected proteins sorted by compartment and organized by meta-compartment (nucleus (red), cytoplasm (blue) and secretory pathway (yellow)). Proteins detected in multiple compartments are connected with an edge, revealing distinct patterns of multilocalization, including between compartments in close proximity (such as nucleoli and nucleoplasm), between compartments that are separated by a membrane (such as the nucleoplasm and cytosol) and between more distant compartments (such as proteins on the plasma membrane and those in the nucleoplasm). CJ, cell junction; Cyt bridge, cytokinetic bridge; DTT, dithiothreitol; ER, endoplasmic reticulum; FC, nucleoli fibrillar centre; FP, fluorescent protein; HU, hydroxyurea; LD, lipid droplet; MMS, methyl methanesulfate; NB, nuclear body; PM, plasma membrane. Part **a** reproduced with permission from REF.[75] (Torres, N. P., Ho, B. & Brown, G. W. High-throughput fluorescence microscopic analysis of protein abundance and localization in budding yeast. *Crit. Rev. Biochem. Mol. Biol.* **51**, 110–119 (2016)), Taylor & Francis Ltd (http://www.tandfonline.com). Part **b** adapted from REF.[144], Springer Nature Limited. Part **c** adapted with permission from REF.[52], AAAS.

between biological repeat experiments and treatment conditions. The only solution for these problems, besides technical optimization, is to carry out multiple replicate comparative experiments to identify consistent changes[21].

Most published global imaging-based comparative spatial proteomics studies to date have used tagged yeast libraries[19,67,68,73,97,98,143], whereas comparative imaging studies of the human proteome have been limited to single compartments of interest[144]. For all imaging-based comparative studies, the greatest challenge is the robust assignment of subcellular localizations, either manually or computationally, and distinguishing between bona fide protein relocalizations and morphological changes that may have resulted from the investigated perturbation.

For both MS-based and imaging-based approaches, another general problem is that two poorly resolved organelles can lead to random and thus inconsistent compartment assignments, resulting in false positives.

Types of translocation and their detection. The simplest type of translocation is the complete transition of a protein from one compartment to another. However, evidence suggests that translocations are often only partial. Furthermore, many proteins have a complex subcellular distribution, before and/or after perturbation, and changes in localization can consist of quantitative rather than qualitative shifts. Finally, a protein can change localization within the same compartment. These scenarios show that the definition of a protein translocation and how it is detected requires careful consideration.

MS-based spatial proteomics produces two types of data for each protein — the raw (numerical) abundance distribution profile and the (qualitative) compartment localization prediction inferred from the profile. Using changes in predicted localization to identify translocating proteins severely limits the scope of the analysis, as only complete transitions can be detected. By contrast, detecting shifts in the raw profiling data is much more inclusive, especially for partial translocations and thus is preferable to using changes in predicted localization[21]. The interpretation of the identified shifts should be a subsequent, uncoupled step.

In imaging-based spatial proteomics, the primary output is the compartment localization assignments, which are also insensitive to detecting partial translocations with hard (that is, single majority location) classifiers. To address this, several studies have used soft classifiers (that is, multiple compartment assignments with likelihood scores) or multiple quantitative feature measurements, which are conceptually similar to the profiles obtained with MS-based spatial proteomics. Thus, imaging data become amenable to statistical detection of localization differences based on profile differences[19,68,73].

Determination of false discovery rates. Comparative spatial proteomics experiments are notoriously prone to producing a high proportion of false positives[44]; estimation of the false discovery rate (FDR) of translocations is therefore essential for a meaningful interpretation of the obtained hits. The best approach is to do a mock experiment (control versus control, with no changes expected) in addition to the actual perturbation experiment and to subject both to the same data analysis pipeline[21,24]. The proportion of hits in the mock versus the perturbation experiment, at a given stringency cut-off, yields a good estimate of the FDR, provided that the experimental noise is similar in both conditions. Published comparative studies all suggest that replicate experiments are essential to achieve low FDR levels[20,21,24].

Time-course experiments. MS-based dynamic mapping and antibody-based imaging are inherently discontinuous methods, as biochemical fractionation destroys the cells and fixation kills the cells. Consequently, to obtain temporal information about spatial changes, a series of comparative mapping experiments is required, much like recording a series of static images to generate a movie[24]. Every time point adds linearly to the overall MS measurement requirements, which are usually the resource bottleneck for MS-based proteomics. Screening a fluorescent protein library can, in principle, provide continuous information on protein dynamics and has produced very impressive results in studies limited to a single compartment[22,145]. For genome-wide fluorescent protein libraries, studies are currently still limited to snapshots, albeit with high temporal resolution[73]. Despite great technical advances in high-throughput image acquisition[8], parallel live-cell imaging on the genomic scale is not yet feasible.

Tagged yeast libraries
Genome-wide libraries of yeast cells, each expressing a protein fused to a fluorescent reporter protein (such as GFP).

Successful applications

Here, we discuss studies that have harnessed the power of global comparative spatial proteomics to illustrate its usefulness in a broad range of biological contexts.

Comparative MS profiling. The dynamic organellar maps approach is currently the most extensively validated method for comparative MS profiling[21] (reviewed elsewhere[146]). It was specifically developed for comparative applications, with special emphasis on the robustness of the workflow and the development of a statistical framework for translocation detection from the primary profiling data. The method was first used to map protein movements in response to epidermal growth factor (EGF) stimulation[21] (FIG. 4a). Numerous novel and known translocations, such as movement of the EGF receptor (EGFR) from the plasma membrane to endosomes, the recruitment of major signalling adaptors and the shuttling of transcription factors into and out of the nucleus, were identified with high specificity. A unique additional feature of this analysis was the incorporation of protein quantification, providing copy number estimates of the observed movements. Since then, the method has been applied twice as a discovery tool to reveal the mechanistic basis of congenital disorders. The adaptor protein 4 (AP-4) complex is required for the formation of transport vesicles with unknown function. Genetic defects in AP-4 cause AP-4 deficiency syndrome, which is characterized by severe intellectual disability and progressive paraplegia. By applying dynamic organellar maps to AP-4-deficient cells[26], it was shown that AP-4 vesicles mediate the cellular distribution of the autophagy protein ATG9A, which is crucial for autophagosome biogenesis (and thus neuronal maintenance), thereby uncovering a novel pathway for spatial control of autophagy and the likely cause of AP-4 deficiency syndrome (FIG. 4b). Remarkably, this mapping approach identified ATG9A as a primary cargo of AP-4 vesicles from among 4,000 profiled proteins with no false-positive identifications and without requiring any prior knowledge about AP-4 function. Application of this approach was also instrumental in elucidating the function of the related AP-5 complex[27].

Comparative PCP has been applied to detect organellar changes in mouse liver cells induced by a high-fat diet[25] to characterize the pathological changes occurring during hepatic steatosis, which is a hallmark of non-alcoholic fatty liver disease (NAFLD) (FIG. 4c). The subcellular localizations of ~4,500 proteins were mapped, and several hundred translocation events were detected, providing evidence of remarkable organellar rearrangements induced by the high-fat diet. For example, the Golgi apparatus was adsorbed onto lipid droplets, and secretion was substantially reduced. Many organellar contact-site proteins were also redirected to lipid droplets, resulting in enhanced contacts between lipid droplets and mitochondria. Furthermore, the authors superimposed the subcellular distribution profiles of phosphorylated peptides onto the corresponding protein profiles. As PCP is capable of assigning each protein to multiple subcellular localizations, organelle-specific phosphorylation events on individual proteins were thus inferred.

Comparative LOPIT has been used to track the cellular rearrangements that occur during HCMV infection[24]. A time course of five LOPIT maps over 120 h tracked the subcellular localization of over 4,000 host and 100 virus proteins (FIG. 4d). This analysis revealed major organellar remodelling, such as the lysosomes splitting into two separate pools, and merging of the ER and Golgi. In addition, the authors identified numerous candidate translocation events, leading to the discovery of the unconventional myosin MYO18A as a new factor required for efficient HCMV replication.

Of note, in the studies described above[24–27], global MS organellar mapping was used as a sensitive and specific screening tool to provide candidates for orthogonal validation by targeted imaging experiments. This generic workflow is emerging as a powerful new strategy for the systematic analysis of cell-biological processes in cells.

Comparative imaging-based spatial proteomics. Gene-tagging approaches are ideally suited for comparative proteomics owing to the simplicity of analysing multiple perturbations once the fluorescent-protein-tagged libraries have been created. Given that these applications rely on libraries of tagged proteins, most comparative studies to date have used the *S. cerevisiae* GFP fusion collection[65] or other yeast collections[98]. Several studies have investigated the localization changes in response to environmental stress conditions, such as DNA replication stress induced by methyl methanesulfonate[73,74,143] and hydroxyurea[19,73,74], or in response to dithiothreitol (DTT), hydrogen peroxide, nitrogen starvation[67] and rapamycin[19]. For example, 254 proteins were found to relocalize after DNA damage[73,74], including the unexpected discovery of 20 metabolic enzymes that entered the nucleus; 235 proteins relocalized after treatment with either DTT or hydrogen peroxide or after nitrogen starvation[67]. Some of these studies relied solely on the manual assignment of protein localizations[67,74], whereas others used a computational approach[19,68,73]. These studies independently concluded that protein relocation is a widely used cellular strategy to cope with environmental stress and that the proportion of the proteome that relocalizes is similar to the proportion that shows changes in abundance (FIG. 5a; reviewed elsewhere[75]). In addition, these two types of response are largely non-overlapping with regard to the affected proteins[19,67,74].

To obtain greater insight into protein relocalization, a computational meta-analysis of 24 yeast screens[104] (comprising 400,000 images in total) was performed. For this purpose, an unsupervised, computational, localization change detection method[147] was used, followed by visual interpretation of the hits. This approach allowed the authors to distinguish responses targeted to specific perturbations from more generalized responses and to identify groups of proteins with similar relocalization patterns, implying functional connectivity. Furthermore, the authors demonstrated that, although some protein localization changes are accompanied

Box 2 | **Complexity of the cellular proteome: future challenges and opportunities for spatial proteomics**

Multiple factors make the cellular proteome much more complex than predicted from a mere gene count. Spatial proteomics is ideally poised to unravel the functional importance of this intriguing cellular complexity.

Multilocalizing and 'moonlighting' proteins

Several studies indicate that ~50% of the human proteome is localized to multiple cellular compartments[52,65]. This observation raises many questions, such as whether the different protein populations are static or dynamic, how their cellular sorting is achieved, whether the different populations correspond to different proteoforms and, most importantly, whether these proteins have context-specific functions (that is, moonlighting activity). Moonlighting is defined as proteins having two or more different cellular functions that do not result from genetic variations, RNA splicing or pleiotropic effects[153,154] but result from differences in subcellular localization, substrates, oligomerization or post-translational modifications (PTMs)[155]. The number of known moonlighting proteins is rapidly increasing. Databases list nearly 700 proteins that have two characterized independent functions[156–158], and it has been estimated that 23% of the human proteome is moonlighting[159]. A remarkable 78% of the known moonlighting proteins are involved in disease, and 48% are current drug targets[156], emphasizing the importance of moonlighting for cellular function. Imaging-based spatial proteomics will be a key technology for the identification of multilocalizing proteins to provide a better understanding of their crucial role in spatial coordination of the complex adaptive systems that underlie cellular functionality.

Proteoforms

Each molecular form of a protein is called a proteoform[160], including variability from the DNA sequence (135,000 coding single nucleotide polymorphisms (SNPs) validated in UniProt[161]), RNA splicing (93% of human genes undergo RNA splicing[162]) and diverse PTMs, such as phosphorylation, ubiquitylation, alkylation and glycosylation. A typical human cell is estimated to contain 6 million coexisting proteoforms[163]; although this number is far below the theoretically possible number of combinations[164] (reviewed elsewhere[165]), it reveals the massive complexity of the proteome. Studies aimed at exploring the functional consequences of alternative splicing have shown that it frequently leads to inclusion or exclusion of targeting sequences, such as membrane anchors or mitochondrial signals, and that the sets of interaction partners of different isoforms of a given protein are often as different as those of proteins encoded by different genes[166]. Furthermore, many PTMs, especially phosphorylation and ubiquitylation, are known to regulate protein subcellular localization. Conceptually, mass spectrometry (MS)-based spatial proteomics, aided by proteogenomics[148], can provide insights about proteoform nature and subcellular localization and will provide a key technology to understand the largely unexplored functional significance of proteoform complexity.

Protein abundance

Cellular protein abundance spans seven orders of magnitude in human cells[18]. To model cellular processes, knowledge of absolute protein copy number per cell and its changes is required. Both MS-based and imaging-based spatial proteomics can provide this information.

Single-cell variability

It is becoming increasingly clear that populations of genetically identical cells show variability in protein expression[52,66], for example, during cell differentiation[69,70] and in response to environmental fluctuations[67,69] or drugs[72]. Our understanding of the consequences of this cellular heterogeneity is still rudimentary. Imaging-based methods will likely have a key role in the elucidation of the causative factors, such as stochasticity, microenvironments, cellular states such as the cell cycle or population bet-hedging strategies. Of note, workflows for single-cell proteomic analysis are currently emerging[167].

Protein dynamics

Protein subcellular localization is tightly controlled, and many proteins change localization in response to a stimulus, a perturbation or in disease. Global comparative spatial proteomics, either imaging-based, interaction-based or MS-based, provides an ideal tool to capture these physiological and pathological protein translocations at the systems level and should become a widespread discovery tool for cell biologists.

by transcriptional or protein abundance changes, subcellular localization is generally an independent layer of regulation.

A seminal study of mitosis in human cells used 4D imaging data for the analysis of protein spatial distribution over time, integrated with 3D concentration data for 28 human proteins that were genomically (homozygously) tagged with enhanced GFP[144]. A canonical model of spatiotemporal protein reorganizations during mitosis in human cells was created using computational image analysis and data integration (FIG. 5b). Although not a global analysis, this study demonstrates a powerful generic approach to the stochiometric analysis of spatiotemporal protein redistribution and paves the way for similar investigations of other cellular processes.

The HPA Cell Atlas mapped the localization of 12,003 human proteins in unperturbed conditions to 30 different cellular structures[52]. This work showed that ~50% (6,163) of the proteins localized to two or more cellular compartments. By mapping the patterns of multilocalization within the cell (FIG. 5c), this study demonstrated that some organelles, such as mitochondria and the ER, mostly contain proteins that are localized exclusively in these organelles, whereas others, such as the plasma membrane and nucleus, contain many multilocalizing proteins; this finding was validated using hyper-LOPIT. The localization of each protein was analysed in three different human cell lines, which revealed that multilocalizing proteins are more likely to show spatial variations between cell lines. In total, 3,546 multilocalizing proteins showed cell line-dependent localization.

Future directions

Recent advances in genomics, transcriptomics, proteomics and other omics technologies have provided unprecedented systems biology tools and drastically reshaped our understanding of cell biology. However, the spatial organization of the cell is not directly assayed by any of these approaches, although evidence suggests that dynamic subcellular reorganization represents an independent layer of regulation as important as changes in abundance at the RNA and protein levels. The systematic application of spatial proteomics will thus be essential for a complete understanding of cellular physiology. Until recently, large-scale proteomics has been the domain of a few specialist laboratories, but this is changing rapidly. The required technology (that is MS or imaging) is becoming more widely available, as are user-friendly software solutions for data analysis. Consequently, spatial proteomics experiments have become fairly

Proteogenomics
Integration of proteomics, transcriptomics and genomics for the discovery and identification of peptides using mass spectrometry. Practically, DNA or RNA sequence information is used to provide an experiment-specific or cell-type-specific tailored database for proteomic protein identification rather than a generic organism-specific database.

accessible, and we encourage readers outside the field to join in (Supplementary Box 1; Supplementary Box 2).

Multiple factors make the cellular proteome much more complex than expected from a gene count (BOX 2). Spatial proteomics is ideally poised to unravel the functional importance of this intriguing cellular complexity, and we expect that it will have a major role in doing so. For example, mapping proteins that localize to multiple compartments will be a key step towards understanding organelle crosstalk, interlinked cellular processes and identifying proteins with moonlighting activity. Mapping proteins with temporal and spatial variability in single cells will add important insights into cellular signalling dynamics. Finally, elucidating the biological roles of proteoforms (that is, splice variants and post-transcriptionally modified forms of proteins) is a central yet largely uncharted aspect of cellular complexity and organization, which can be explored using spatial proteomics.

As we outline here, imaging-based and MS-based approaches offer complementary insights into the spatial proteome (Supplementary Table 1; Supplementary Box 2). Importantly, both approaches are required to fully understand cellular complexity, and several studies have demonstrated the strong synergy obtained by combining MS-based spatial proteomics with imaging[52,55,124].

A further important (and we think imminent) advance will be the combination of spatial proteomics with other omics technologies. As an example, transcriptomics can provide cell-specific expression databases, including cell-type-specific splice variants or single nucleotide polymorphisms, thus empowering proteomics to discover novel proteoforms (proteogenomics[148]). In combination with metabolomics, organellar rearrangements can be functionally linked to changes in metabolism[25]. Similarly, in conjunction with RNA sequencing, spatial proteomics has the potential to connect mRNA biology with subcellular localization[149–151]. In our opinion, spatial proteomics should become an orthogonal and integrated technology in cell mapping efforts such as the Human Cell Atlas, which aims to characterize all human cell types[152].

Finally, a major future challenge will be the collection and integration of the various data sets produced in past and future spatial proteomic studies. At present, there is little crosstalk between major resources such as the HPA Cell Atlas[52] and individual studies that also provide subcellular localization databases[20,21,24,25,47,50]. Given the dynamic nature of protein spatial organization, repositories for both imaging-based and MS-based spatial proteomics data need to be developed 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.

In conclusion, we envisage a new era of cellular modelling in which the spatial dynamics of proteins are integrated with other omics measurements to gain insight into the crosstalk between the different layers of cellular regulation, leading to a greater understanding of cellular phenotype and activity.

Published online 18 January 2019

1. Mitrea, D. M. & Kriwacki, R. W. Phase separation in biology; functional organization of a higher order. *Cell Commun. Signal* **14**, 1 (2016).
2. Wheeler, R. J. & Hyman, A. A. Controlling compartmentalization by non-membrane-bound organelles. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **373**, 20170193 (2018).
3. Bauer, N. C., Doetsch, P. W. & Corbett, A. H. Mechanisms regulating protein localization. *Traffic* **16**, 1039–1061 (2015).
4. Guardia, C. M., De Pace, R., Mattera, R. & Bonifacino, J. S. Neuronal functions of adaptor complexes involved in protein sorting. *Curr. Opin. Neurobiol.* **51**, 103–110 (2018).
5. Banworth, M. J. & Li, G. Consequences of Rab GTPase dysfunction in genetic or acquired human diseases. *Small GTPases* **9**, 158–181 (2018).
6. Bridges, R. J. & Bradbury, N. A. Cystic fibrosis, cystic fibrosis transmembrane conductance regulator and drugs: insights from cellular trafficking. *Handb Exp. Pharmacol.* **245**, 385–425 (2018).
7. Meyer, K. et al. Mutations in disordered regions can cause disease by creating dileucine motifs. *Cell* **175**, 239–253 (2018).
8. Mattiazzi Usaj, M. et al. High-content screening for quantitative cell biology. *Trends Cell Biol.* **26**, 598–611 (2016).
9. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).
10. Hosp, F. & Mann, M. A primer on concepts and applications of proteomics in neuroscience. *Neuron* **96**, 558–571 (2017).
11. Larance, M. & Lamond, A. I. Multidimensional proteomics for cell biology. *Nat. Rev. Mol. Cell Biol.* **16**, 269–280 (2015).
12. Breker, M. & Schuldiner, M. The emergence of proteome-wide technologies: systematic analysis of proteins comes of age. *Nat. Rev. Mol. Cell Biol.* **15**, 453–464 (2014).
13. Kim, D. I. & Roux, K. J. Filling the void: proximity-based labeling of proteins in living cells. *Trends Cell Biol.* **26**, 804–817 (2016).
14. Lonn, P. & Landegren, U. Close encounters - probing proximal proteins in live or fixed cells. *Trends Biochem. Sci.* **42**, 504–515 (2017).
15. Sullivan, D. P. et al. Deep learning is combined with massive-scale citizen science to improve large-scale image classification. *Nat. Biotechnol.* **36**, 820–828 (2018).
16. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
17. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
18. Wilhelm, M. et al. Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587 (2014).
19. Chong, Y. T. et al. Yeast proteome dynamics from single cell imaging and automated analysis. *Cell* **162**, 1413–1424 (2015).
20. Itzhak, D. N. et al. A mass spectrometry-based approach for mapping protein subcellular localization reveals the spatial proteome of mouse primary neurons. *Cell Rep.* **20**, 2706–2718 (2017).
21. Itzhak, D. N., Tyanova, S., Cox, J. & Borner, G. H. Global, quantitative and dynamic mapping of protein subcellular localization. *eLife* **5**, e16950 (2016).
This study presents the first demonstration that MS-based organellar maps of the cell can be used as an unbiased discovery tool, exemplified by EGF signalling. The study also includes a detailed description of how to generate maps and a database of the subcellular localization and copy number information for 8,700 human proteins.
22. Walther, N. et al. A quantitative map of human Condensins provides new insights into mitotic chromosome architecture. *J. Cell Biol.* **217**, 2309–2328 (2018).
23. Ma, J. et al. Using deep learning to model the hierarchical structure and function of a cell. *Nat. Methods* **15**, 290–298 (2018).
24. Jean Beltran, P. M., Mathias, R. A. & Cristea, I. M. A portrait of the human organelle proteome in space and time during cytomegalovirus infection. *Cell Syst.* **3**, 361–373 (2016).
This is the first study to charter the dynamics of a spatial proteome by MS over time during viral infection (with HCMV), revealing new insights into the interplay between the virus and the host cell.
25. Krahmer, N. et al. Organellar proteomics and phospho-proteomics reveal subcellular reorganization in diet-induced hepatic steatosis. *Dev. Cell* **47**, 205–221 (2018).
This is the first application of comparative MS-based organellar mapping to a mammalian tissue, the mouse liver. Organellar rearrangements caused by a high-fat diet were captured, providing a holistic view of the pathological changes during hepatic steatosis.
26. Davies, A. K. et al. AP-4 vesicles contribute to spatial control of autophagy via RUSC-dependent peripheral delivery of ATG9A. *Nat. Commun.* **9**, 3958 (2018).
This is the first example of how MS-based organellar maps of the cell can be used to pinpoint the molecular basis of a genetic disorder, in this case the neurodegenerative AP-4 deficiency syndrome.
27. Hirst, J., Itzhak, D. N., Antrobus, R., Borner, G. H. H. & Robinson, M. S. Role of the AP-5 adaptor protein complex in late endosome-to-Golgi retrieval. *PLOS Biol.* **16**, e2004411 (2018).
28. Uhlen, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
29. Sharma, K. et al. Cell type- and brain region-resolved mouse brain proteome. *Nat. Neurosci.* **18**, 1819–1831 (2015).
30. Doll, S. et al. Region and cell-type resolved quantitative proteomic map of the human heart. *Nat. Commun.* **8**, 1469 (2017).
31. Takamori, S. et al. Molecular anatomy of a trafficking organelle. *Cell* **127**, 831–846 (2006).
32. Tharkeshwar, A. K., Gevaert, K. & Annaert, W. Organellar omics-a reviving strategy to untangle the biomolecular complexity of the cell. *Proteomics* **18**, e1700113 (2018).
33. De Duve, C. Principles of tissue fractionation. *J. Theor. Biol.* **6**, 33–59 (1964).
34. Andersen, J. S. et al. Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* **426**, 570–574 (2003).

35. Krahmer, N. et al. Protein correlation profiles identify lipid droplet proteins with high confidence. *Mol. Cell Proteom.* **12**, 1115–1126 (2013).
36. Harner, M. et al. The mitochondrial contact site complex, a determinant of mitochondrial architecture. *EMBO J.* **30**, 4356–4370 (2011).
37. Morgenstern, M. et al. Definition of a high-confidence mitochondrial proteome at quantitative scale. *Cell Rep.* **19**, 2836–2852 (2017).
38. Wyant, G. A. et al. NUFIP1 is a ribosome receptor for starvation-induced ribophagy. *Science* **360**, 751–758 (2018).
39. Borner, G. H. et al. Multivariate proteomic profiling identifies novel accessory proteins of coated vesicles. *J. Cell Biol.* **197**, 141–160 (2012).
40. Borner, G. H. et al. Fractionation profiling: a fast and versatile approach for mapping vesicle proteomes and protein-protein interactions. *Mol. Biol. Cell* **25**, 3178–3194 (2014).
41. Wuhr, M. et al. The nuclear proteome of a vertebrate. *Curr. Biol.* **25**, 2663–2671 (2015).
42. Weekes, M. P. et al. Quantitative temporal viromics: an approach to investigate host-pathogen interaction. *Cell* **157**, 1460–1472 (2014).
43. Peikert, C. D. et al. Charting organellar importomes by quantitative mass spectrometry. *Nat. Commun.* **8**, 15272 (2017).
44. Gatto, L. et al. A foundation for reliable spatial proteomics data analysis. *Mol. Cell Proteom.* **13**, 1937–1952 (2014).
45. Lund-Johansen, F. et al. MetaMass, a tool for meta-analysis of subcellular proteomics data. *Nat. Methods* **13**, 837–840 (2016).
46. Breckels, L. M. et al. The effect of organelle discovery upon sub-cellular protein localisation. *J. Proteom.* **88**, 129–140 (2013).
47. Christoforou, A. et al. A draft map of the mouse pluripotent stem cell spatial proteome. *Nat. Commun.* **7**, 8992 (2016).
48. Dunkley, T. P., Watson, R., Griffin, J. L., Dupree, P. & Lilley, K. S. Localization of organelle proteins by isotope tagging (LOPIT). *Mol. Cell Proteom.* **3**, 1128–1134 (2004).
49. Gilchrist, A. et al. Quantitative proteomics analysis of the secretory pathway. *Cell* **127**, 1265–1281 (2006).
50. Jadot, M. et al. Accounting for protein subcellular localization: a compartmental map of the rat liver proteome. *Mol. Cell Proteom.* **16**, 194–212 (2017).
51. Mardakheh, F. K. et al. Proteomics profiling of interactome dynamics by colocalisation analysis (COLA). *Mol. Biosyst.* **13**, 92–105 (2016).
52. Thul, P. J. et al. A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).
This study presents the first near proteome-wide spatial map of human cells, at unprecedented resolution in terms of the number of organelles and structures that were mapped, which represents the Cell Atlas component of the HPA. This study provides localization data for more than 12,000 human proteins and revealed a very high number of multilocalizing proteins and a considerable level of single-cell variability.
53. Hein, M. Y. et al. A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* **163**, 712–723 (2015).
54. Huttlin, E. L. et al. Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505–509 (2017).
55. Mackinder, L. C. M. et al. A spatial interactome reveals the protein organization of the algal CO_2-concentrating mechanism. *Cell* **171**, 133–147 (2017).
This is an example of how AP–MS interaction networks can be used in conjunction with imaging to obtain a high-resolution architectural map of a subcellular compartment, the algal pyrenoid.
56. Branon, T. C. et al. Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.* **36**, 880–887 (2018).
This study reports on the latest generation of enzymes for proximity labelling, which can be applied for spatial protein interaction mapping in cultured cells and also in vivo, as demonstrated for flies and worms.
57. Hung, V. et al. Proteomic mapping of the human mitochondrial intermembrane space in live cells via ratiometric APEX tagging. *Mol. Cell* **55**, 332–341 (2014).
58. Rhee, H. W. et al. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **339**, 1328–1331 (2013).
59. Youn, J. Y. et al. High-density proximity mapping reveals the subcellular organization of mRNA-associated granules and bodies. *Mol. Cell* **69**, 517–532 (2018).
60. Gupta, G. D. et al. A dynamic protein interaction landscape of the human centrosome-cilium interface. *Cell* **163**, 1484–1499 (2015).
61. Han, S., Li, J. & Ting, A. Y. Proximity labeling: spatially resolved proteomic mapping for neurobiology. *Curr. Opin. Neurobiol.* **50**, 17–23 (2018).
62. Liu, X. et al. An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations. *Nat. Commun.* **9**, 1188 (2018).
63. Lobingier, B. T. et al. An approach to spatiotemporally resolve protein interaction networks in living cells. *Cell* **169**, 350–360 (2017).
64. Paek, J. et al. Multidimensional tracking of GPCR signaling via peroxidase-catalyzed proximity labeling. *Cell* **169**, 338–349 (2017).
65. Huh, W. K. et al. Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
This study presents the first proteome-wide spatial map of a eukaryotic cell, and revolutionized yeast biology. The yeast library created as part of this work has been used to study, for example, the proteome response to perturbations, single-cell variability, stochasticity and noise in gene expression.
66. Sigal, A. et al. Variability and memory of protein levels in human cells. *Nature* **444**, 643–646 (2006).
67. Breker, M., Gymrek, M. & Schuldiner, M. A novel single-cell screening platform reveals proteome plasticity during yeast stress responses. *J. Cell Biol.* **200**, 839–850 (2013).
68. Lu, A. X. et al. Integrating images from multiple microscopy screens reveals diverse patterns of change in the subcellular localization of proteins. *eLife* **7**, e31872 (2018).
This study reports a meta-analysis of 24 imaging-based yeast screens using unsupervised computational image analysis. Specific and general spatial relocalization responses to various environmental perturbations are identified. Importantly, this study shows that the proportion of the proteome that relocalizes is similar to the proportion showing changes in abundance and that these two responses constitute independent layers of cellular regulation.
69. Balazsi, G., van Oudenaarden, A. & Collins, J. J. Cellular decision making and biological noise: from microbes to mammals. *Cell* **144**, 910–925 (2011).
70. Rubakhin, S. S., Lanni, E. J. & Sweedler, J. V. Progress toward single cell metabolomics. *Curr. Opin. Biotechnol.* **24**, 95–104 (2013).
71. Narayanaswamy, R. et al. Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. *Proc. Natl Acad. Sci. USA* **106**, 10147–10152 (2009).
72. Shaffer, S. M. et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431–435 (2017).
73. Denervaud, N. et al. A chemostat array enables the spatio-temporal analysis of the yeast proteome. *Proc. Natl Acad. Sci. USA* **110**, 15842–15847 (2013).
74. Tkach, J. M. et al. Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. *Nat. Cell Biol.* **14**, 966–976 (2012).
75. Torres, N. P., Ho, B. & Brown, G. W. High-throughput fluorescence microscopic analysis of protein abundance and localization in budding yeast. *Crit. Rev. Biochem. Mol. Biol.* **51**, 110–119 (2016).
76. Coons, A. H., Creech, H. J., Jones, R. N. & Berliner, E. The demonstration of pneumococcal antigen in tissues by the use of fluorescent antibody. *J. Immunol.* **45**, 159–170 (1942).
77. Lazarides, E. & Weber, K. Actin antibody: the specific visualization of actin filaments in non-muscle cells. *Proc. Natl Acad. Sci. USA* **71**, 2268–2272 (1974).
78. Schnell, U., Dijk, F., Sjollema, K. A. & Giepmans, B. N. Immunolabeling artifacts and the need for live-cell imaging. *Nat. Methods* **9**, 152–158 (2012).
79. Stadler, C., Skogs, M., Brismar, H., Uhlen, M. & Lundberg, E. A single fixation protocol for proteome-wide immunofluorescence localization studies. *J. Proteom.* **73**, 1067–1078 (2010).
80. Simeon, R. & Chen, Z. In vitro-engineered non-antibody protein therapeutics. *Protein Cell* **9**, 3–14 (2018).
81. Stadler, C. et al. Immunofluorescence and fluorescent-protein tagging show high correlation for protein localization in mammalian cells. *Nat. Methods* **10**, 315–323 (2013).
82. Baker, M. Reproducibility crisis: blame it on the antibodies. *Nature* **521**, 274–276 (2015).
83. Uhlen, M. et al. A proposal for validation of antibodies. *Nat. Methods* **13**, 823–827 (2016).
84. Bandrowski, A. et al. The Resource Identification Initiative: a cultural shift in publishing. *F1000Res* **4**, 134 (2015).
85. Uhlen, M. et al. Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28**, 1248–1250 (2010).
86. Uhlen, M. et al. A pathology atlas of the human cancer transcriptome. *Science* **357**, eaan2507 (2017).
87. Nilsson, P. et al. Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. *Proteomics* **5**, 4327–4337 (2005).
88. Uhlen, M. et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell Proteom.* **4**, 1920–1932 (2005).
89. Algenas, C. et al. Antibody performance in western blot applications is context-dependent. *Biotechnol. J.* **9**, 435–445 (2014).
90. Skogs, M. et al. Antibody validation in bioimaging applications based on endogenous expression of tagged proteins. *J. Proteome Res.* **16**, 147–155 (2017).
91. Stadler, C. et al. Systematic validation of antibody binding and protein subcellular localization using siRNA and confocal microscopy. *J. Proteom.* **75**, 2236–2251 (2012).
92. Angelo, M. et al. Multiplexed ion beam imaging of human breast tumors. *Nat. Med.* **20**, 436–442 (2014).
93. Bodenmiller, B. Multiplexed epitope-based tissue imaging for discovery and healthcare applications. *Cell Syst.* **2**, 225–238 (2016).
94. Giesen, C. et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* **11**, 417–422 (2014).
95. Goltsev, Y. et al. Deep profiling of mouse splenic architecture with CODEX multiplex imaging. *Cell* **174**, 968–981 (2018).
96. Gut, G., Herrmann, M. D. & Pelkmans, L. Multiplexed protein maps link subcellular organization to cellular states. *Science* **361**, eaar7042 (2018).
97. Meurer, M. et al. Genome-wide C-SWAT library for high-throughput yeast genome tagging. *Nat. Methods* **15**, 598–600 (2018).
98. Weill, U. et al. Genome-wide SWAp-Tag yeast libraries for proteome exploration. *Nat. Methods* **15**, 617–622 (2018).
99. Weill, U. et al. Toolbox: creating a systematic database of secretory pathway proteins uncovers new cargo for COPI. *Traffic* **19**, 370–379 (2018).
100. Breker, M., Gymrek, M., Moldavski, O. & Schuldiner, M. LoQAtE — Localization and Quantitation ATlas of the yeast proteomE. A new tool for multiparametric dissection of single-protein behavior in response to biological perturbations in yeast. *Nucleic Acids Res.* **42**, D726–D730 (2014).
101. Riffle, M. & Davis, T. N. The Yeast Resource Center Public Image Repository: a large database of fluorescence microscopy images. *BMC Bioinformatics* **11**, 263 (2010).
102. Chuartzman, S. G. & Schuldiner, M. Database for High Throughput Screening Hits (dHITS): a simple tool to retrieve gene specific phenotypes from systematic screens done in yeast. *Yeast* **35**, 477–483 (2018).
103. Cherry, J. M. et al. Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* **40**, D700–D705 (2012).
104. Koh, J. L. et al. CYCLoPs: a comprehensive database constructed from automated analysis of protein abundance and subcellular localization patterns in *Saccharomyces cerevisiae*. *G3 (Bethesda)* **5**, 1223–1232 (2015).
105. Dubreuil, B. et al. YeastRGB: comparing the abundance and localization of yeast proteins across cells and libraries. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gky941 (2018).
106. Simpson, J. C., Wellenreuther, R., Poustka, A., Pepperkok, R. & Wiemann, S. Systematic subcellular

localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep.* **1**, 287–292 (2000).
107. Frenkel-Morgenstern, M. et al. Dynamic proteomics: a database for dynamics and localizations of endogenous fluorescently-tagged proteins in living human cells. *Nucleic Acids Res.* **38**, D508–D512 (2010).
108. Sigal, A. et al. Generation of a fluorescently labeled endogenous protein library in living human cells. *Nat. Protoc.* **2**, 1515–1527 (2007).
109. Sigal, A. et al. Dynamic proteomics in individual human cells uncovers widespread cell-cycle dependence of nuclear proteins. *Nat. Methods* **3**, 525–531 (2006).
110. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
111. Jinek, M. et al. RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013).
112. Cho, W. K. et al. Super-resolution imaging of fluorescently labeled, endogenous RNA polymerase II in living cells with CRISPR/Cas9-mediated gene editing. *Sci. Rep.* **6**, 35949 (2016).
113. Dambournet, D., Hong, S. H., Grassart, A. & Drubin, D. G. Tagging endogenous loci for live-cell fluorescence imaging and molecule counting using ZFNs, TALENs, and Cas9. *Methods Enzymol.* **546**, 139–160 (2014).
114. Doyon, J. B. et al. Rapid and efficient clathrin-mediated endocytosis revealed in genome-edited mammalian cells. *Nat. Cell Biol.* **13**, 331–337 (2011).
115. Ratz, M., Testa, I., Hell, S. W. & Jakobs, S. CRISPR/Cas9-mediated endogenous protein tagging for RESOLFT super-resolution microscopy of living human cells. *Sci. Rep.* **5**, 9592 (2015).
116. Li-Kroeger, D. et al. An expanded toolkit for gene tagging based on MiMIC and scarless CRISPR tagging in *Drosophila*. *eLife* **7**, e38709 (2018).
117. Merkle, F. T. et al. Efficient CRISPR-Cas9-mediated generation of knockin human pluripotent stem cells lacking undesired mutations at the targeted locus. *Cell Rep.* **11**, 875–883 (2015).
118. Roberts, B. et al. Systematic gene tagging using CRISPR/Cas9 in human stem cells to illuminate cell organization. *Mol. Biol. Cell* **28**, 2854–2874 (2017).
119. Feng, S. et al. Improved split fluorescent proteins for endogenous protein labeling. *Nat. Commun.* **8**, 370 (2017).
120. Kamiyama, D. et al. Versatile protein tagging in cells with split fluorescent protein. *Nat. Commun.* **7**, 11046 (2016).
121. Leonetti, M. D., Sekine, S., Kamiyama, D., Weissman, J. S. & Huang, B. A scalable strategy for high-throughput GFP tagging of endogenous human proteins. *Proc. Natl Acad. Sci. USA* **113**, E3501–E3508 (2016).
122. Sahl, S. J., Hell, S. W. & Jakobs, S. Fluorescence nanoscopy in cell biology. *Nat. Rev. Mol. Cell Biol.* **18**, 685–701 (2017).
123. Monnich, M. et al. CEP128 localizes to the subdistal appendages of the mother centriole and regulates TGFβ/BMP signaling at the primary cilium. *Cell Rep.* **22**, 2584–2592 (2018).
124. Jakobsen, L. et al. Novel asymmetrically localizing components of human centrosomes identified by complementary proteomics methods. *EMBO J.* **30**, 1520–1535 (2011).
125. Caicedo, J. C. et al. Data-analysis strategies for image-based cell profiling. *Nat. Methods* **14**, 849–863 (2017).
126. Coelho, L. P. et al. Determining the subcellular location of new proteins from microscope images using local features. *Bioinformatics* **29**, 2343–2349 (2013).
127. Li, J., Newberg, J. Y., Uhlen, M., Lundberg, E. & Murphy, R. F. Automated analysis and reannotation of subcellular locations in confocal images from the Human Protein Atlas. *PLOS ONE* **7**, e50514 (2012).
128. Li, J., Xiong, L., Schneider, J. & Murphy, R. F. Protein subcellular location pattern classification in cellular images using latent discriminative models. *Bioinformatics* **28**, i32–i39 (2012).
129. Schmidhuber, J. Deep learning in neural networks: an overview. *Neural Netw.* **61**, 85–117 (2015).
130. Girshick, R., Donahue, J., Darrell, T. & Malik, J. in *2014 IEEE Conference on Computer Vision and Pattern Recognition* 580–587 (IEEE, 2014).
131. He, K. M., Zhang, X. Y., Ren, S. Q. & Sun, J. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (IEEE, 2016).
132. Parnamaa, T. & Parts, L. Accurate classification of protein subcellular localization from high-throughput microscopy images using deep learning. *G3 (Bethesda)* **7**, 1385–1392 (2017).
133. Kraus, O. Z. et al. Automated analysis of high-content microscopy data with deep learning. *Mol. Syst. Biol.* **13**, 924 (2017).
134. Kraus, O. Z., Ba, J. L. & Frey, B. J. Classifying and segmenting microscopy images with deep multiple instance learning. *Bioinformatics* **32**, i52–i59 (2016).
135. Hughes, A. et al. Quanti.us: a tool for rapid, flexible, crowd-based annotation of images. *Nat. Methods* **15**, 587–590 (2018).
136. Ronneberger, O., Fischer, P. & Brox, T. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Lecture Notes in Computer Science Book Series) Vol. 9351 234–241 (Springer, 2015).
137. Simon, A. A. et al. A probabilistic U-Net for segmentation of ambiguous images. Preprint at *arXiv* https://arxiv.org/abs/1806.05034 (2018).
138. Christiansen, E. M. et al. In silico labeling: predicting fluorescent labels in unlabeled images. *Cell* **173**, 792–803 (2018).
139. Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F. & Johnson, G. R. Label-free prediction of three-dimensional fluorescence images from transmitted light microscopy. *Nat. Methods* **15**, 917–920 (2018).
140. Ellenberg, J. et al. Public archives for biological image data. Preprint at *arXiv* https://arxiv.org/abs/1801.10189 (2018).
141. Dunkley, T. P. et al. Mapping the *Arabidopsis* organelle proteome. *Proc. Natl Acad. Sci. USA* **103**, 6518–6523 (2006).
142. Foster, L. J. et al. A mammalian organelle map by protein correlation profiling. *Cell* **125**, 187–199 (2006).
143. Mazumder, A., Pesudo, L. Q., McRee, S., Bathe, M. & Samson, L. D. Genome-wide single-cell-level screen for protein abundance and localization changes in response to DNA damage in *S. cerevisiae*. *Nucleic Acids Res.* **41**, 9310–9324 (2013).
144. Cai, Y. et al. Experimental and computational framework for a dynamic protein atlas of human cell division. *Nature* **561**, 411–415 (2018).
This is a seminal study that integrates imaging-based spatiotemporal protein mapping with absolute quantification to enable stoichiometric modelling of the spatial and temporal dynamics of cell division — this paper shows where the field is heading.
145. Politi, A. Z. et al. Quantitative mapping of fluorescently tagged cellular proteins using FCS-calibrated four-dimensional imaging. *Nat. Protoc.* **13**, 1445–1464 (2018).
146. Larochelle, S. Tracking the proteome. *Nat. Methods* **13**, 821–821 (2016).
147. Lu, A. X. & Moses, A. M. An unsupervised kNN method to systematically detect changes in protein localization in high-throughput microscopy images. *PLOS ONE* **11**, e0158712 (2016).
148. Nesvizhskii, A. I. Proteogenomics: concepts, applications and computational strategies. *Nat. Methods* **11**, 1114–1125 (2014).
149. Queiroz, R. M. L. et al. Unbiased dynamic characterization of RNA-protein interactions by OOPS. Preprint at *bioRxiv* https://doi.org/10.1101/333336 (2018).
150. Kaewsapsak, P., Shechner, D. M., Mallard, W., Rinn, J. L. & Ting, A. Y. Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking. *eLife* **6**, e29224 (2017).
151. Fazal, F. M. et al. Atlas of subcellular RNA localization revealed by APEX-seq. Preprint at *bioRxiv* https://doi.org/10.1101/454470 (2018).
152. Regev, A. et al. The Human Cell Atlas. *eLife* **6**, e27041 (2017).
153. Jeffery, C. J. Moonlighting proteins. *Trends Biochem. Sci.* **24**, 8–11 (1999).
154. Jeffery, C. J. Why study moonlighting proteins? *Front. Genet.* **6**, 211 (2015).
155. Jeffery, C. J. Protein species and moonlighting proteins: very small changes in a protein's covalent structure can change its biochemical function. *J. Proteom.* **134**, 19–24 (2016).
156. Franco-Serrano, L. et al. MultitaskProtDB-II: an update of a database of multitasking/moonlighting proteins. *Nucleic Acids Res.* **46**, D645–D648 (2018).
157. Chen, C., Zabad, S., Liu, H. P., Wang, F. & Jeffery, C. MoonProt 2.0: an expansion and update of the moonlighting proteins database. *Nucleic Acids Res.* **46**, D640–D644 (2018).
158. Chapple, C. E. et al. Extreme multifunctional proteins identified from a human protein interaction network. *Nat. Commun.* **6**, 7412 (2015).
159. Khan, I. K., Bhuiyan, M. & Kihara, D. DextMP: deep dive into text for predicting moonlighting proteins. *Bioinformatics* **33**, i83–i91 (2017).
160. Smith, L. M. & Kelleher, N. L. & The Consortium for Top Down Proteomics. Proteoform: a single term describing protein complexity. *Nat. Methods* **10**, 186–187 (2013).
161. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **46**, 2699 (2018).
162. Wang, E. T. et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
163. Ponomarenko, E. A. et al. The size of the human proteome: the width and depth. *Int. J. Anal. Chem.* **2016**, 7436849 (2016).
164. Skinner, O. S. et al. Top-down characterization of endogenous protein complexes with native proteomics. *Nat. Chem. Biol.* **14**, 36–41 (2018).
165. Aebersold, R. et al. How many human proteoforms are there? *Nat. Chem. Biol.* **14**, 206–214 (2018).
166. Yang, X. et al. Widespread expansion of protein interaction capabilities by alternative splicing. *Cell* **164**, 805–817 (2016).
167. Zhu, Y. et al. Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nat. Commun.* **9**, 882 (2018).

Acknowledgements
E.L. acknowledges funding from the Knut and Alice Wallenberg foundation (KAW 2016.0204), the Swedish Research Council (2017-05327) and the Chan Zuckerberg Initiative (173965 (5022)). G.H.H.B. received funding from the German Research Foundation (DFG/Gottfried Wilhelm Leibniz Prize MA 1764/2-1) and the Max Planck Society for the Advancement of Science.

Author contributions
Both authors contributed equally to all aspects of preparing the article (researching data for the article, substantial contributions to the discussion of the content and writing, reviewing and editing of the manuscript before submission).

Competing interests
The authors declare no competing interests.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reviewer information
Nature Reviews Molecular Cell Biology thanks G. W. Brown and other anonymous reviewer(s), for their contribution to the peer review of this work.

Supplementary information
Supplementary information is available for this paper at https://doi.org/10.1038/s41580-018-0094-y.

RELATED LINKS
HeLa Spatial Proteome: www.mapofthecell.org
Human Protein Atlas (HPA): www.proteinatlas.org
Obesity-induced non-alcoholic fatty liver disease (NAFLD): http://nafld-organellemap.org/
SpatialMap: http://spatialmap.org/