ANNUAL REVIEWS

Annual Review of Biomedical Data Science

Mapping the Multiscale Proteomic Organization of Cellular and Disease Phenotypes

Anthony Cesnik,[1,*] Leah V. Schaffer,[2,*] Ishan Gaur,[1] Mayank Jain,[2] Trey Ideker,[2,3] and Emma Lundberg[1,4,5,6]

[1]Department of Bioengineering, Stanford University, Stanford, California, USA; email: emmalu@stanford.edu

[2]Department of Medicine, University of California San Diego, La Jolla, California, USA; email: tideker@health.ucsd.edu

[3]Departments of Computer Science and Engineering and Bioengineering, University of California San Diego, La Jolla, California, USA

[4]Department of Pathology, Stanford University, Palo Alto, California, USA

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

[6]Chan Zuckerberg Biohub, San Francisco, California, USA

Annu. Rev. Biomed. Data Sci. 2024. 7:369–89

The *Annual Review of Biomedical Data Science* is online at biodatasci.annualreviews.org

https://doi.org/10.1146/annurev-biodatasci-102423-113534

*These authors contributed equally to this article

Keywords

spatial, proteomics, interactomics, multiscale, integration, machine learning

Abstract

While the primary sequences of human proteins have been cataloged for over a decade, determining how these are organized into a dynamic collection of multiprotein assemblies, with structures and functions spanning biological scales, is an ongoing venture. Systematic and data-driven analyses of these higher-order structures are emerging, facilitating the discovery and understanding of cellular phenotypes. At present, knowledge of protein localization and function has been primarily derived from manual annotation and curation in resources such as the Gene Ontology, which are biased toward richly annotated genes in the literature. Here, we envision a future powered by data-driven mapping of protein assemblies. These maps can capture and decode cellular functions through the integration of protein expression, localization, and interaction data across length scales and timescales.

In this review, we focus on progress toward constructing integrated cell maps that accelerate the life sciences and translational research.

1. INTRODUCTION

Every organism is composed of cells with tightly coordinated subcellular organization. This organization spans length scales of many orders of magnitude, from small protein complexes to phase-separated systems and entire organelles. Translocation, moonlighting, and phase separation are all classic examples of how protein function is regulated in space and time. These processes allow the dynamic formation and reformation of critical cellular subsystems that are necessary for the life cycle of all organisms from bacteria to human cells (1). The gram-negative bacterium *Caulobacter crescentus*, for example, has inspired many systems biologists due to its fascinating asymmetric division and the intriguing way its cell division cycle is regulated by localizing certain proteins to poles of the cell, including CtrA at the start of replication (2). In human cells, the glycolytic enzyme Eno1 moonlights with additional functions of DNA binding (3) and as a plasminogen receptor (4), depending on its location. Phase separation is another type of compartmentalization that is critical for cellular function, such as the role disordered proteins like Ddx4 play in forming membraneless organelles to help stabilize RNAs (5, 6). While knowledge of individual genes or complexes, like these examples, contributes to our understanding of protein functions, we generally need more information about the context of subcellular structure as a whole to fully understand the emergence of complex cellular phenotypes (7). Mapping cellular organization improves our ability to decode cellular functions and predict phenotypic changes (**Figure 1**).

The proteome is arranged into cellular subsystems at various spatial scales that can be addressed with a spectrum of current technologies. Mass spectrometry (MS)-based proteomics has provided a means not only to measure entire proteomes in a high-throughput fashion, but also to measure the interactions between individual proteins (11). At larger physical scales (i.e., length, area, volume), spatial proteomics technologies can capture the localization of proteins within organelles (12). Techniques for spatial proteomics include immunofluorescence- and fluorescent protein (FP)-based imaging, which produce highly correlated data for protein localization (13). Such biological organization extends downward in scale to atomic structures, including conformational states, which can be measured by cryogenic electron microscopy (cryo-EM) and tomography (14, 15), and post-translational modifications (PTMs), which are frequently measured by MS (16). At scales larger than a cell, cellular communities in tissues and whole organisms can be characterized with imaging technologies (17), single-cell RNA sequencing, and metagenomics in the case of microbial communities (18). Many current cell mapping efforts that use these technologies have matured and are continuing to be refined (**Table 1**).

Because these technologies capture proteomic organization at different scales, combining their data into integrated cell maps can capture emergent, multiscale processes by which protein function translates to phenotype (36). Initial efforts to represent the cell as a hierarchy of cellular subsystems have already begun to provide a deeper understanding of mutational landscapes (37) and DNA damage response (38). By creating additional scalable methods that intelligently combine diverse data modalities, we will be able to describe more specific spatial relationships, add temporal logic to proteomic organization, and expand these models to include diverse biomolecules of interest. We envision that the future of genetics and translational research will require integrated mapping of the multiscale organization of cells in order to more fully understand complex cellular phenotypes with genomic and other biological data. Here, we review

Figure 1

Changes at the various levels of cellular and organismal organization inform our understanding of cellular pathways and disease. We can decode cellular and disease phenotypes using data-driven methods to create multiscale maps of protein assemblies through integration of data modalities that capture information at different scales. Expanded data modalities and length scales of interactions will reveal more about the functional significance of modifications and conformational differences on the atomic side and about tissue architecture and interactions with other organisms on the larger end. Abbreviations: cryo-EM, cryogenic electron microscopy; IF, immunofluorescence; MS, mass spectrometry. Whole zebrafish embryo image reproduced with permission from Reference 8, whole mouse brain image reproduced from RIKEN National Science Institute (CC BY 4.0). Human duodenum tissue sample IF image taken by Anna Bäckström (CC BY 4.0). Human cell diagram and IF images from The Human Protein Atlas, Subcellular Atlas (CC BY-SA 3.0). Protein structures from Protein Data Bank (PDB ID: 1YCR). Interactome image reproduced with permission from Reference 9. Cell map hierarchy image adapted from Reference 10.

Table 1 Data and techniques for subcellular mapping

Dataset	Data type	Description	URL (sample)	Reference(s)
HPA Subcellular Atlas	Subcellular protein location	Aims to measure human protein subcellular locations in single cells using the HPA antibody resource and IF confocal microscopy; contains ~80,000 images, corresponding to one to five antibodies for each of over 13,000 proteins	**https://www.proteinatlas.org** (various cell lines)	19–21
OpenCell	Subcellular protein location and interactome	1,310 endogenously epitope-tagged proteins from HEK293-derived cells; PPIs extracted from protein immunoprecipitations, covering ~30,000 interactions; also subcellular locations in single cells using fluorescent proteins and microscopy (~6,000 images)	**https://opencell.czbiohub.org/** (HEK293)	22
hyperLOPIT	Subcellular protein location	Centrifugation-based techniques for determining protein subcellular location; proteins are identified by MS in fractions that contain proteins belonging to one or several organelles, which are identified by organellar markers	**https://proteome.shinyapps.io/hyperlopit-u2os2018/** (U-2 OS)	23
LOPIT-DC	Subcellular protein location		**https://proteome.shinyapps.io/lopitdc-u2os2018/** (U-2 OS)	23
SubCellBarcode	Subcellular protein location		**https://www.subcellbarcode.org/** (A431, H322, HCC827, MCF7, U251)	24
JUMP Cell Painting Dataset	Cell imaging and single-cell RNA-seq	Cell morphological images and single-cell RNA sequencing profiles for over 100,000 unique compounds and over 15,000 genetic perturbations	**https://jump-cellpainting.broadinstitute.org/** (U-2 OS)	25, 26
BioPlex Interactome	Interactome	The biophysical interactions of the ORFeome-based (BioPlex) network resulting in the creation of thousands of cell lines, each expressing a tagged version of a protein from the ORFeome collection; currently includes ~200,000 interactions identified by affinity purification MS for over 10,000 affinity purified proteins	**https://bioplex.hms.harvard.edu/index.php** (HEK293T, HCT116)	11
Human Cell Map	Interactome	A PPI map using BioID, a proximity-dependent biotinylation technique that provides an alternative approach to affinity purification (i.e., BioPlex above) to define the composition of cellular compartments in living cells; currently covers ~5,000 proteins in HEK293-derived cells	**https://humancellmap.org** (HEK293)	27

(*Continued*)

Table 1 (*Continued*)

Dataset	Data type	Description	URL (sample)	Reference(s)
Coelution MS	Interactome	Protein complexes separated by size-exclusion chromatography, and interacting proteins identified by MS by their co-occurrence in fractions	**http://human.med.utoronto.ca/** (HEK293, HeLa S3)	28
			https://sec-explorer.ethz.ch/ (HEK293)	29
Crosslinking MS	Interactome	Protein interactions determined using chemical cross-linking combined with MS analysis	N/A (HEK293)	30
Surface protein interactome	Interactome	Catalogs over 100 immune surface PPIs using HEK293 cells, which is noteworthy given that cell surface proteins are underrepresented in most interatomic datasets	**https://github.com/jshilts/shilts-et-al-2022-immunoreceptors** (HEK293)	31
HPA Tissue Atlas	Tissue atlas	Normal and tumor tissue–based maps of protein expression in normal tissues with cellular resolution; in some cases, these data can also provide rough subcellular protein localization information, albeit at much lower resolution than fluorescent protein– and IF-based imaging	**https://www.proteinatlas.org/humanproteome/tissue** (tissues)	32
HPA Pathology Atlas	Tissue atlas		**https://www.proteinatlas.org/humanproteome/pathology** (tissues)	32
HPA Cell Cycle Protein Expression Dataset	Cell types and dynamics	Single-cell temporal proteomics of the human cell cycle; monitors single-cell abundance levels for 1,166 proteins; in addition, the quantitative cell cycle phase of each individual cell has been characterized using endogenously tagged cell cycle markers; data are based on fluorescent microscopy images that show cell-to-cell variability and were selected for in-depth analysis of protein localization and abundance using immunofluorescence in U-2 OS FUCCI cells over the cell cycle	**https://www.proteinatlas.org/humanproteome/cell/cell+cycle+dependent** (U-2 OS) **https://www.ebi.ac.uk/biostudies/BioImages/studies/S-B** IAD34 (U-2 OS)	21
HPA Cell Cycle mRNA Expression Dataset	Cell types and dynamics	Single-cell temporal transcriptomics of the human cell cycle; monitors single-cell abundance levels for 13,450 mRNA species in U-2 OS FUCCI cells; in addition, the quantitative cell cycle phase of each individual cell has been characterized using endogenously tagged cell cycle markers	**https://www.proteinatlas.org/humanproteome/cell/cell+cycle+dependent** (U-2 OS)	21
Human Cell Atlas	Cell types and dynamics	Contains over 100 human single-cell transcriptomic studies that could inform multiscale model construction or be used for dynamic modeling	**https://data.humancellatlas.org/** (various)	33

(*Continued*)

Table 1 (*Continued*)

Dataset	Data type	Description	URL (sample)	Reference(s)
HPA Single Cell Type Dataset	Cell types and dynamics	Single-cell RNA sequencing dataset, aggregated from various publicly available sequencing data, for 444 individual human cell type clusters across 25 tissues	**https://www.protein%20atlas.org/humanproteome/single+cell+type** (various)	34
Human Developmental Atlas	Whole-organism mapping	Can map whole embryos for cell types over various stages of development	**https://developmental.cellatlas.io/** (embryo)	35

Abbreviations: FUCCI, fluorescent ubiquitination-based cell cycle indicator; HPA, Human Protein Atlas; IF, immunofluorescence; MS, mass spectrometry; PPI, protein–protein interaction.

key developments and future directions in determining the multiscale proteomic organization of cells. We discuss the main approaches for mapping protein assemblies, recent efforts to integrate these data, and the role that multiscale maps can play in advancing systems biology. Finally, we present relevant data modalities and integration approaches that we envision will expand future multiscale maps to help us better understand biological systems.

2. MAPPING PROTEIN ASSEMBLIES

The endeavor of creating maps of protein assemblies is built upon previous work that constructed parts lists and distinguished cell types. The Human Genome Project, completed in 2003, determined both the primary DNA sequence and a list of elemental biological objects (39), consisting of genes and their corresponding putative proteins (40). Since that time, these sequence maps and parts lists have been refined, including the recently completed telomere-to-telomere genome (41), the creation of inventories of proteomes (42, 43) and transcriptomes (33, 44, 45) across diverse cell types, and ongoing efforts to identify all human proteoforms in many cell types (46, 47). Presently, there is a rise in systematic efforts to map how the proteome is organized within cells in higher-order structures, including the interactions (48–50) and locations (12, 51) of proteins within cells and tissues.

Multiscale map: data-driven hierarchy of cellular subsystems represented by protein assemblies (i.e., proximal sets of proteins)

Proteoform: a unique molecular form of a protein defined by its primary amino acid sequence and modifications at specific residues

Protein interactome: the set of all physical interactions between proteins in a cell

2.1. Protein Interactomes

Protein–protein interaction (PPI) maps represent subcellular structure by determining the physical interactions of proteins to capture protein complexes and other higher-order structures of proteins. Key examples of efforts to map the protein interactome include BioPlex (11), which uses systemic affinity purification mass spectrometry (AP-MS) analysis, and Human Cell Map (27), which uses biotinylation probes to pull down and analyze interacting proteins with MS. These approaches have revealed global interaction networks, but they are low-throughput approaches when expanded to other cell contexts. Recently, coelution MS (29) has emerged as a higher-throughput means of globally determining protein interactions. In this technique, native protein complexes are separated, often by size-exclusion chromatography (SEC), into dozens of fractions prior to MS analysis. Proteins identified as coeluting in the same fractions are considered likely interactors. This approach enables high-throughput collection of interactomic data in different cellular contexts, although filtering false positive interactions from coeluting proteins can be challenging.

The endogenous epitope tagging of proteins with an FP has enabled both live-cell imaging and PPI mapping using immunoprecipitation (IP) of the FP, followed by MS-based proteomic analysis

(IP-MS) (22, 52). The use of endogenous protein levels in SEC-MS and endogenous tagging potentially allow for the calculation of stoichiometries, which are obscured by the overexpression used in most AP-MS workflows (29, 52, 53). Both techniques have trade-offs. The use of IP-MS following endogenous epitope tagging is less robust for detecting protein interactors with lower abundance or stoichiometry, but the overexpression of bait proteins that allows AP-MS to have a lower limit of detection may lead to more false positive interactions. Another emerging technique is cross-linking mass spectrometry (XL-MS), which is increasingly able to produce global maps of protein interactions and may at the same time be able to aid in determining subunit organization and conformational heterogeneity within protein complexes and cellular subsystems (30, 54–56). Cross-linking can reveal transient interactions, such as those between disordered proteins or proteins that are less solubilized, which may be missed by other methods (56). In summary, these protein interaction mapping efforts are producing global interactome datasets that are expanding across different cell lines (9) and cellular contexts (29).

Subcellular protein localization: the physical locations of a protein within the cell, including organelles, condensates, and membranes

2.2. Subcellular Protein Localizations

Resolving subcellular protein localizations within cellular compartments, including organelles and condensates, can shed light on a larger physical scale of cellular organization than PPIs. One such mapping effort is the Human Protein Atlas (HPA) (19), which has used classic antibody-based imaging to study the localizations of proteins within cells, tissues, and tumors. The HPA Subcellular Atlas (20) has cataloged the locations of nearly all nonsecreted proteins across many cell lines including nearly complete coverage of the U-2 OS osteosarcoma cell line; the HPA Tissue Atlas (32) has information about the protein expression and subcellular location of proteins in normal human tissues; and the HPA Pathology Atlas has information about protein expression in a large cohort of human tumors with subcellular resolution (32). Beyond subcellular localization, these data have also revealed important information about the heterogeneity of protein expression, including new cell cycle associations for hundreds of proteins (21). Overall, these data provide extensive mapping of proteins in larger structures like condensates, organelles, cells, and even cell types within tissue contexts.

Proteins can also be localized using endogenously tagged FPs (57–59); early work in this area determined how best to tag and image the proteins for localization. Since then, the advent of CRISPR-based technologies has facilitated subcellular localization of proteins using gene editing to endogenously tag FPs prior to imaging (22, 60). One limitation of using these techniques is the need to ensure that the FP tags do not alter the proteins' functions, localizations, or interactions (22).

Subcellular localizations of proteins can also be determined with centrifugation fractionation techniques, in which larger organelles like the nucleus can be separated from smaller ones like vesicles. In centrifugation fractionation techniques such as LOPIT (61) and SubCellBarCode (24), maps of proteins are created through MS-based identification of proteins in the same centrifugation fractions as proteins from one or several organelles. Fractions are assigned to organelles using a training set of proteins with known locations. While this assignment method may produce some bias, it is a high-throughput option for identifying the contents of higher-order structures within the cell.

2.3. Creating Multiscale Maps of Protein Assemblies with Data Integration

Integrating more than one proteomic data modality can extract complementary information and bridge gaps in physical scales and proteome coverage. Adding new data modalities introduces new degrees of freedom that allow biologically meaningful separation of previously indistinguishable

Multiscale community detection: a hierarchical clustering of nodes in a network, defining relationships within communities and between different communities, across multiple levels of detail

subsets of data, such as the subcellular location of subsystems or the structure of proteins within complexes. This multimodal data integration can also strengthen the true signal in an existing dataset, as many sources of noise are unlikely to be correlated across assays. We see this effect in integrative structural biology, where multiple data types (e.g., X-ray crystallography, XL-MS, nuclear magnetic resonance spectroscopy) are combined in order to improve the accuracy and completeness of structural models (62). This approach has been used to create accurate models of protein subsystems that are difficult to model with one approach alone, such as the ∼50-MDa nuclear pore complex (63).

Recent efforts are bridging the gap across larger scales by integrating proteomic interaction data with immunofluorescent images that capture subcellular location. In our recent work, a multiscale integrated cell map (36) systematically integrated an imaging proteomic dataset from the HPA Subcellular Atlas (20) for HEK293T with the global BioPlex AP-MS interaction network (11). An embedding was created for each protein for each data modality separately, and these embeddings were integrated into a coembedding with supervised machine learning trained with the Gene Ontology (GO) knowledgebase to determine predicted proximities between each pair of proteins. These resulting proximities were clustered at multiple resolutions (64, 65) to reveal a data-driven hierarchy of cellular subsystems, calibrated to physical distance using GO components of known sizes in the literature.

3. GENETICS IN THE ERA OF SYSTEMS BIOLOGY

Multiscale maps enable the interpretation of gene-level data to reveal important biological information from genetic information, i.e., to connect genotype to phenotype. In studies that focus on gene-level information, much is often lost due to solely looking for enrichment of certain phenotypes on a per-gene level. One way systems biology has augmented genetic analyses is by finding modules from network-level data (66–71). One group of methods to perform this network-level enrichment is called network propagation; in this approach a signal is allowed to propagate along a network of prior associations of genes for a phenotype of interest (72). Propagating signals for gene-level information across these types of networks can result in gene scores that are more consistent across datasets, enable robust module discovery, and better rank genes to prioritize them for follow-up analyses.

Performing multiscale community detection on networks organizes genes into hierarchical clusters in cell maps, which can then be used for interpreting genetic mutations across the scales of cellular organization. We recently used this type of approach to reveal cellular subsystems that are recurrently mutated across cancer patients (37). Notably, these subsystems often contain genes that are sparsely mutated at the gene level and are only revealed as important when the associations of proteins into higher-order assemblies are considered. This is an important motivating example of the power of using systems-level analyses through multiscale maps for the analysis of genetic results.

Systems-level analyses do not need to be limited to interpretation of genetic mutations; any sort of inputs, such as differential transcriptomic or proteomic analyses upon perturbation, can be analyzed in the context of the multiscale proteomic organization to better understand the underlying biology. Many times in these analyses, a list of differential abundant proteins or transcripts upon a perturbation is obtained. How to interpret this list of proteins to understand the underlying subcellular phenotype remains a challenge; enrichment against the GO knowledgebase has typically been used for attempting to interpret sets of genes or proteins, but the associated GO terms may be biased by heavily annotated genes (73). We believe data-derived cell maps will have key utility in these analyses for understanding where groups of proteins interact in the cell relative to others, especially in different cell types or contexts.

4. EXPANDING THE SCALES AND MODALITIES OF MULTISCALE MAPS OF PROTEIN ASSEMBLIES

Understanding proteomic organization will involve an expansion of the scales and data modalities that can be included in cell models (**Figure 1**). Smaller-scale changes include protein sequence variations and PTMs that define proteoforms and minute structural changes, while tissue architecture is an example of a larger-scale change that can be captured in proteomic data and represented in multiscale models. Proteins also interact with other biomolecules such as nucleic acids, lipids, and metabolites, and including these can capture new layers of protein assemblies and regulation.

4.1. Understanding Proteoform Diversity and Organization

New biological problems will present themselves as systems are analyzed across scales, in which case previously generated data will likely yield new discoveries. One important avenue of future work is understanding the relationship between proteoforms and subcellular structure. Since immunofluorescent images are created with broadly targeting antibodies for the gene of interest, they depict the localization of the superset of proteoforms (74, 75) for each gene. Thus, there is an opportunity to better understand proteoform systems biology by deconvoluting these images into the localization of different proteoforms from the same gene. These proteoform differences can change subcellular localization; for example, phosphorylation of the tau protein regulates its binding to microtubules (76), phosphorylation of 4E-BP2 has been shown to determine whether it exists in a disordered or structured state (77), cells and tissues show notable spatial regulation of glycosylation (78), and alternatively spliced isoforms of the same gene can produce proteoforms with rewired protein interactions that are as different as those of entirely different genes (79). In cases where a modification affects a protein's structure or localization, this conditional information may need to be reflected in the multiscale model as an attribute on a node or by representing different proteoforms or isoforms separately in interaction networks or cell maps (79–83). Efforts to expand the maps of protein assemblies to the scales of proteoforms and PTMs are likely to yield new understanding of the spatial proteomic regulation.

4.2. Zeroing In On Smaller Scales by Incorporating Structural Shifts

Protein structural shifts can be induced by PTMs, sequence changes, and PPIs, and they can induce functional changes in proteins. Notably distinct bond angles and conformational patterns can be revealed by methods such as cryo-EM (14), MS techniques like XL-MS (84), trapped ion mobility spectrometry (TIMS) (85), tandem-TIMS (86), and MS-based accessibility measurements such as hydrogen deuterium exchange (87) and covalent protein painting (CPP) (88). One study using CPP found that conformational changes can affect protein accessibility over tumorigenesis and are distinct between cell lines, and alterations in proteins such as heat shock proteins are predictors of drug efficacy (89). Just as protein accessibility can affect interactions with drugs, it can affect interactions with other proteins. Thus, we expect that integrating proteoform sequence and structural information will clarify the variation underlying some or many pleiotropy and moonlighting events. These proteoforms or structural conformers could be represented in future cell maps in the appropriate subcellular systems. Other useful information may come from predictive modeling approaches. For example, protein language models are already being used to generate three-dimensional protein structures (90), and these structures are being analyzed for their propensity for interactions (91, 92). These predictive modeling approaches could provide new avenues to investigate how the cell utilizes these subtle structural variations to regulate complex biological phenomena.

4.3. Broadening the Scales: Cell Types, Neighborhoods, Tissues, and Beyond

Toward the larger end of the scale, we expect these hierarchies to include information at the tissue, organismal, and possibly even ecological levels. There is a growing field of multiscale modeling at the tissue level using spatial single-cell RNA sequencing (93–97) and multiplexed proteomic imaging (98) to determine the spatial locations and relationships of different cell states. Many of the current interactomic and imaging datasets listed in **Table 1** were collected in cell line models, which may miss key biological factors, such as secreted proteins or interactions between different cell states. Determining the protein assemblies in different cell states remains challenging, although recent developments in spatial single-cell proteomics analyses provide the opportunity to reveal the proteins expressed in different cell states (99, 100). A recently reported strategy for dissecting complex cellular phenotypes using the deep visual proteomics technique (99, 101) classifies cell types with label-free imaging and uses laser capture microdissection to cut out regions of the tissue with specific cell types to prepare biologically relevant subsamples for high-sensitivity MS-based proteomics. Multiplexed protein imaging across tissues using technologies like CODEX (102), MIBI-TOF (103), and others (17, 104) can also reveal these cellular phenotypes and relationships. Constructing multiscale models with these types of data will reveal how the proteome is organized across tissue architectures.

The relationships between the cell types present in different tissues, such as immune cells in the blood and intestinal (98) or fetal tissues (105), can reveal connections between the phenotypes of these tissues. There may even be an opportunity to combine modalities like whole-organism imaging (106, 107) with other data types to represent cellular phenotypes over the development of whole organisms, such as has been performed for zebrafish embryos (108) and is proposed for human embryos (35). Higher order still, we may be able to represent our interactions with other organisms, including microbes in our gut and host–pathogen interactions (109). Overall, expanding multiscale models to account for these larger-scale interactions will improve our ability to decode organismal biology, with important applications for understanding human health and disease in the clinic (110).

4.4. Integrating More Types of Biomolecules Will Improve Our Understanding of Proteomic Organization

Proteins are not the only types of molecules that contribute to proteomic organization and regulation. Other types of biomolecules can interact with and even scaffold protein assemblies and should eventually be included in multiscale maps. For example, data showing protein–nucleic acid and nucleic acid–nucleic acid interactions could be integrated to represent interfaces that regulate much of gene expression, transcription, and translation [e.g., protein–DNA with ChIP (chromatin immunoprecipitation) or EMSA (electrophoretic mobility shift assay); protein–RNA with eCLIP (enhanced cross-linking and immunoprecipitation), loRNA (localization of RNA), or FISH (fluorescence in situ hybridization)/ISH; DNA–DNA with Hi-C or 3C (chromosome conformation capture); RNA–RNA with RIC-seq (RNA in situ conformation sequencing) or proximity ligation; and RNA–DNA or proximity ligation]. The subcellular locations of these molecules can also be visualized using spatial transcriptomics, which is approaching single-cell resolution (111, 112), including in whole organs (113), and with subcellular and superresolution transcript sequencing techniques including seqFISH+ (114), MERFISH (multiplexed error-robust FISH) (115), and APEX-seq (116). Similarly, there is progress in studying protein–metabolite (117) and protein–lipid interactions (118), as well as MS-based imaging techniques that can resolve spatial distributions of these biomolecules (119, 120). Revealing the proteomic organization in the context of these diverse biomolecules will inform our understanding of cellular substructures, metabolism, and gene regulation.

4.5. Expanding in the Temporal Dimension

Another critical component is the expansion of multiscale models in the temporal direction. A spatiotemporal map of cell structure may be able to show how cell structure organizes over time during the cell cycle (21), circadian rhythms, metabolic oscillations, organismal development (35, 108), or hitherto undiscovered autonomous clocks (121). Temporal analyses could also include mapping the subcellular reorganization upon stimuli such as drug response or viral infection (122). Live-cell imaging enables tracking the subcellular localization of proteins over time. Fluorescent markers with intensities that are indicative of cellular processes have allowed a temporal dimension to be calculated and correlated to protein expression, such as using FUCCI (fluorescent ubiquitination-based cell cycle indicator) markers to interrogate protein expression over the cell cycle (21). To determine changes in the interactome over time, light-activated proximity labeling strategies like LOV-Turbo (123) have been developed to capture the proteome of organelles and show promise for revealing PPIs.

Coembedding: transformation of multimodal data into a shared embedding (a lower-dimensional representation of high-dimensional vectors) to reveal relationships and commonalities

Incorporating these data into spatial multiscale maps will facilitate understanding how the cell is structured temporally in a dynamic fashion. Maps could include recent work on elucidating the structure of proteins over time, such as modeling the structures of dynamic complexes like the nuclear pore complex (124). At a smaller timescale, nanosecond structural fluctuations of proteins have been mapped using X-ray free-electron lasers and infrared techniques (125). At a larger timescale, whole-organism-level analyses in model species such as the zebrafish (108) have modeled spatiotemporal development. A recent study has found that different organs within humans can age at different rates (126), inviting further study of the drivers of these cell and tissue changes over larger timescales.

5. ADVANCES IN MULTISCALE MODELING AND REPRESENTATION

5.1. Advances in Multimodal Data Integration for Mapping Proteomic Organization

Recent advances in artificial intelligence research have led to the development of deep learning approaches for systematically integrating data modalities that could readily be applied to the integration of protein modalities for protein assembly construction. In the realm of multimodal data integration, a prominent category of deep learning techniques centers on coembedding (or joint embedding), which aims to translate diverse, high-dimensional data types into a shared, low-dimensional space. Importantly, many of these approaches can integrate data in self-supervised or unsupervised learning schemas that do not require labeled training data such as known protein functions or complexes, which may be unavailable for many proteins or biased toward well-studied proteins. Coembedding approaches also offer the possibility of integrating unpaired data, such as datasets where proteins are present in one modality but missing in another. Unpaired data integration provides an avenue to include modalities that reveal other biomolecules, such as RNAs or glycans, which will be inherently missing from one data type or the other.

One technique that has been applied to align multiple data modalities is contrastive learning, which brings corresponding pairs (e.g., the same protein in two different modalities) together in a joint latent space while pushing dissimilar pairs (e.g., different proteins) apart (127). This approach has been applied to tasks like predicting drug–target interactions from protein sequence embeddings (128) and coembedding drug structures and transcriptional responses (129). Methods such as ImageBind have been used to align data from many data modalities, without samples that are fully paired across all domains (e.g., audio, visual, thermal); this method finds a common representation by linking all data to their paired sample in the image domain (130).

Another approach for finding unified embeddings from various input data modalities is through multimodal autoencoder models, which generate a latent embedding space from the input data that can then be used to reconstruct the original data (131). These models have naturally lent themselves to translation tasks, such as transforming transcriptomic data into single-cell nuclear images using DAPI (4′,6-diamidino-2-phenylindole) staining in a common embedding space (131). This technique could potentially be useful in a situation where data are missing in one modality but closely related ground truth exists in another, providing broader coverage over all relevant proteins (132).

Coembedding approaches have been readily applied to integrate imaging and single-cell RNA sequencing data for the characterization of cells (131, 133–135) and other multiomics data (136–143). We anticipate that such methods will be instrumental to integrating biological data for mapping protein assemblies across scales (**Figure 2*a***).

5.2. Advances in Hierarchical Representation

A critical step in constructing multiscale models involves clustering at different resolutions to capture communities across scales, for which many algorithms have been developed (64, 65, 144, 145). A hierarchy of protein assemblies in the cell should allow parent nodes to have multiple children representing different hierarchical containments within a protein subsystem. These models must also represent pleiotropic proteins that perform multiple functions and localize to more than one subcellular compartment or assembly. These characteristics are widespread, with over half of all proteins localizing to multiple locations (20) and moonlighting proteins performing multiple functions (146) in the cell. For example, a large number of Golgi and nuclear proteins are dual-localizing, including the regulatory homologous recombination repair protein RAD51C (147). This overlap in biological systems also occurs at larger scales, such as immune cell types that appear in more than one tissue type in the human body. Representing pleiotropy and other types of multifunctionality in multiscale maps may entail either displaying multiple parent–child relations in a hierarchy or representing entities in maps as distributions of conformations or other properties as in statistical mechanics (148). Ideally, multiscale maps should accurately reflect these diverse functions of proteins, protein assemblies, and cell types.

5.3. Adding Dynamics to Multiscale Maps

Many observations provide a snapshot of cellular processes. Incorporating dynamics into multiscale maps will help us represent and better understand dynamic processes like signaling and environmental response (**Figure 2*b***). They may also allow us to predict protein interaction remodeling over these processes, which remains poorly understood (9). Many protein assemblies have conditional states, such as those dependent on rhythms like the cell cycle, PTMs and their effects on protein structure or localization, and chemical environments that may drive protein folding. Multiscale models may be able to represent the cascade of changes in cells and even tissues, capturing the dynamic changes downstream of a cell surface event or perhaps a response to nutrient gradients or other stimuli. How much of the interactome is rewired between cells and over cell state shifts like differentiation or the cell cycle remains an open question.

Integrating dynamics into multiscale models of cell structure presents a host of new challenges in data integration and coembedding. Challenges include representing data over time, combining bulk and single-cell information, and including cell conditions that are not represented in all datasets. Solutions might include projecting results for these types of experiments onto a static map for that system, using factor analysis (149), or using transfer learning to leverage pretrained models (150). Alternatively, integration approaches like neural ordinary differential equations (151) and

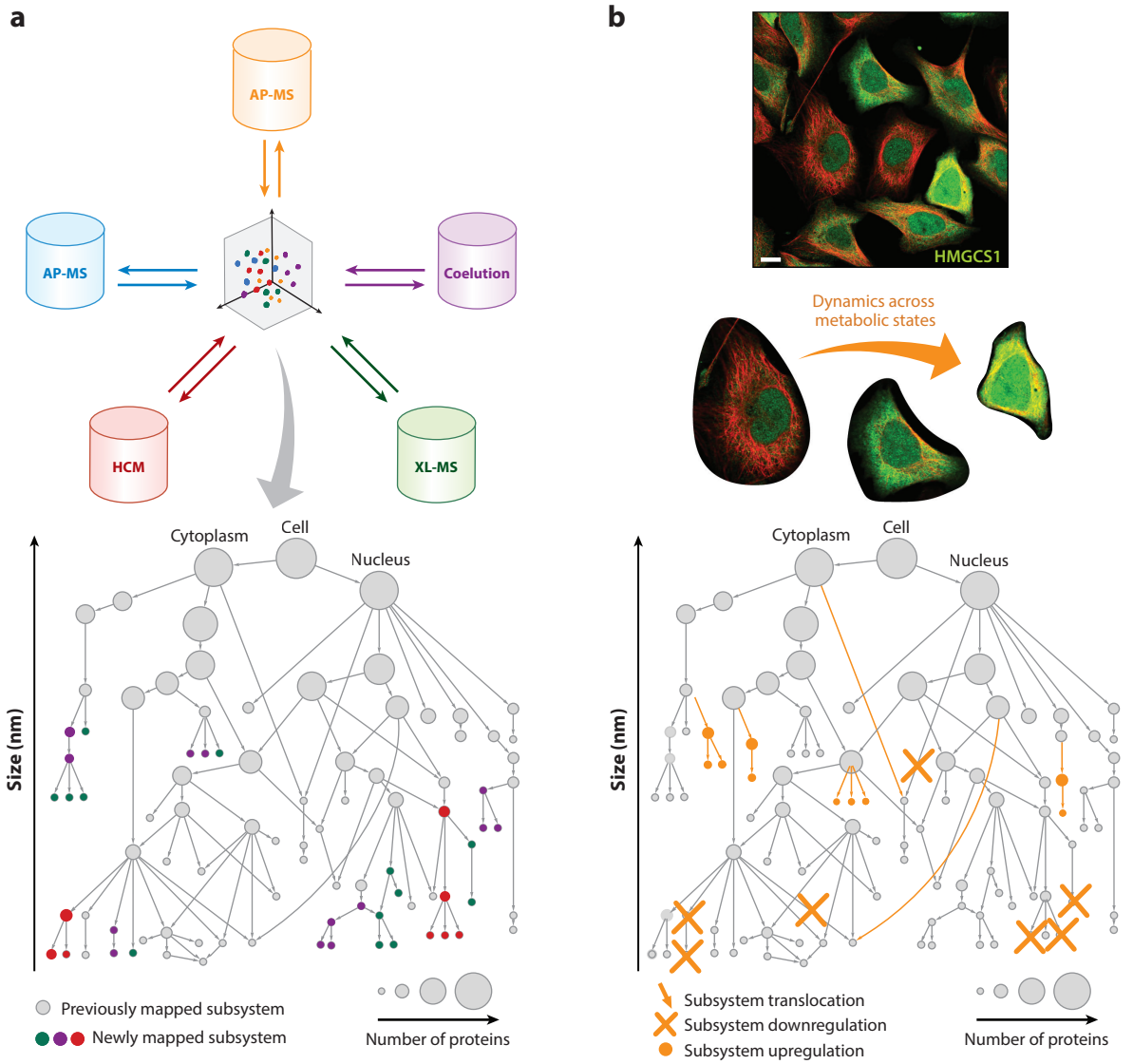

Figure 2

Advancing cell maps with new data modalities for understanding proteomic organization across space and time. (*a*) New multimodal coembedding approaches will be key to the development of multiscale modeling of protein assemblies. Imaging and various interactomic data types are represented in cylinders. Coembedding approaches provide an opportunity to tolerate missing proteins in different input modalities. Expanding to integrate additional data modalities will reveal newly mapped assemblies. (*b*) Key challenges in cell mapping include the representation and embedding of dynamic cellular processes in multiscale models. Targeting dynamic maps, with unique cellular imaging proteomics and protein–protein interaction maps, possibly over cell processes, may reveal deep insights into cellular phenotypes and disease. Shown are images of HMGCS1 in U-2 OS, which is involved in cholesterol biosynthesis. Scale bar is 10 μm. Abbreviations: AP-MS, affinity purification mass spectrometry; HCM, Human Cell Map; XL-MS, crosslinking mass spectrometry. Cell images in panel *b* from The Human Protein Atlas (CC BY-SA 3.0). Cell map hierarchy image adapted from Reference 10.

geometric deep learning (152, 153) could be used to generate embeddings with explicit physical constraints on the system that might include elements such as the density of the cell, diffusion rates, and metabolite mass balance equations. An initial study reported that only approximately 50% of interactions are similar between two cell lines (9), which indicates there is much to learn about context-specific and dynamic interactions (154). We expect to see a deepening understanding of this rewiring concurrently with the development of dynamic multiscale maps.

SUMMARY POINTS

1. Cellular proteomes are organized in a hierarchical manner across scales from protein interaction complexes to subcellular locations, but much of this organization remains unknown.
2. Technologies exist that capture the organization of proteins at different scales, including protein–protein interactions and their subcellular localization.
3. Integrating data from different technologies provides the opportunity to better map the organization of the proteome across scales and can increase the power of multiscale maps of protein assemblies.
4. Multiscale maps empower genetics and other biological analyses by revealing the subsystems upon which a set of individual proteins or genes converge, such as mutated proteins or differentially abundant proteins across conditions.
5. We believe multiscale maps will expand to smaller scales to account for post-translational modifications and protein structural shifts, and larger scales to account for different cell states and interactions between cell types.
6. Recent machine learning advances have enabled approaches for integrating multimodal datasets that will be highly important when integrating different proteomic data modalities to create unified multiscale maps.
7. Dynamic multiscale maps will provide an exciting opportunity to understand how the proteome reorganizes across spatiotemporal phenomena, such as cell division, drug response, and metabolism.

DISCLOSURE STATEMENT

T.I. is a co-founder and member of the advisory board and has an equity interest in Data4Cure and Serinus Biosciences. T.I. is a consultant for and has an equity interest in Ideaya Biosciences. The terms of these arrangements have been reviewed and approved by the University of California San Diego in accordance with its conflict-of-interest policies. E.L. is an advisor for Cartography Biosciences, Nautilus Biotechnology, Element Biosciences, Santa Ana Bio, Pixelgen Technologies, and Moleculent AB. The terms of these arrangements have been reviewed and approved by Stanford University in accordance with its conflict-of-interest policies.

ACKNOWLEDGMENTS

We thank the reviewer for their insightful comments and Ernst Pulido for the engaging discussions of interactomic techniques. This work was supported by the Erling-Persson Foundation, the Knut and Alice Wallenberg Foundation (2021.0346), Schmidt Futures, and the National Institutes of Health's Bridge to Artificial Intelligence program under grant 1OT2OD032742.

LITERATURE CITED

1. Shapiro L. 2022. A half century defining the logic of cellular life. *Annu. Rev. Genet.* 56:1–15
2. Curtis PD, Brun YV. 2010. Getting in the loop: regulation of development in *Caulobacter crescentus*. *Microbiol. Mol. Biol. Rev.* 74(1):13–41
3. Subramanian A, Miller DM. 2000. Structural analysis of α-enolase. *J. Biol. Chem.* 275(8):5958–65
4. Pancholi V, Fischetti VA. 1998. α-Enolase, a novel strong plasmin(ogen) binding protein on the surface of pathogenic streptococci. *J. Biol. Chem.* 273(23):14503–15
5. Nott TJ, Craggs TD, Baldwin AJ. 2016. Membraneless organelles can melt nucleic acid duplexes and act as biomolecular filters. *Nat. Chem.* 8(6):569–75
6. Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, et al. 2015. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* 57(5):936–47
7. Lazebnik Y. 2002. Can a biologist fix a radio?—Or, what I learned while studying apoptosis. *Cancer Cell* 2(3):179–82
8. He J, Mo D, Chen J, Luo L. 2020. Combined whole-mount fluorescence in situ hybridization and antibody staining in zebrafish embryos and larvae. *Nat. Protoc.* 15(10):3361–79
9. Huttlin EL, Bruckner RJ, Navarrete-Perea J, Cannon JR, Baltier K, et al. 2021. Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell* 184(11):3022–40.e28
10. Qin Y, Huttlin EL, Winsnes CF, Gosztyla ML, Wacheul L, et al. 2021. A multi-scale map of cell structure fusing protein images and interactions. *Nature* 600:536–42
11. Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, et al. 2015. The BioPlex network: a systematic exploration of the human interactome. *Cell* 162(2):425–40
12. Lundberg E, Borner GHH. 2019. Spatial proteomics: a powerful discovery tool for cell biology. *Nat. Rev. Mol. Cell Biol.* 20(5):285–302
13. Stadler C, Rexhepaj E, Singan VR, Murphy RF, Pepperkok R, et al. 2013. Immunofluorescence and fluorescent-protein tagging show high correlation for protein localization in mammalian cells. *Nat. Methods* 10(4):315–23
14. Cheng Y. 2018. Single-particle cryo-EM—How did it get here and where will it go. *Science* 361(6405):876–80
15. Thalassinos K, Pandurangan AP, Xu M, Alber F, Topf M. 2013. Conformational states of macromolecular assemblies explored by integrative structure calculation. *Structure* 21(9):1500–8
16. Doll S, Burlingame AL. 2015. Mass spectrometry-based detection and assignment of protein posttranslational modifications. *ACS Chem. Biol.* 10(1):63–71
17. Hickey JW, Neumann EK, Radtke AJ, Camarillo JM, Beuschel RT, et al. 2022. Spatial mapping of protein composition and tissue organization: a primer for multiplexed antibody-based imaging. *Nat. Methods* 19(3):284–95
18. Hum. Microbiome Proj. Consort. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486(7402):207–14
19. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, et al. 2010. Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* 28(12):1248–50
20. Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, et al. 2017. A subcellular map of the human proteome. *Science* 356(6340):eaal3321
21. Mahdessian D, Cesnik AJ, Gnann C, Danielsson F, Stenström L, et al. 2021. Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. *Nature* 590(7847):649–54
22. Cho NH, Cheveralls KC, Brunner A-D, Kim K, Michaelis AC, et al. 2022. OpenCell: endogenous tagging for the cartography of human cellular organization. *Science* 375(6585):eabi6983
23. Geladaki A, Kočevar Britovšek N, Breckels LM, Smith TS, Vennard OL, et al. 2019. Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics. *Nat. Commun.* 10(1):331
24. Orre LM, Vesterlund M, Pan Y, Arslan T, Zhu Y, et al. 2019. SubCellBarCode: proteome-wide mapping of protein localization and relocalization. *Mol. Cell* 73(1):166–82.e7
25. Bray M-A, Singh S, Han H, Davis CT, Borgeson B, et al. 2016. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* 11(9):1757–74

26. Chandrasekaran SN, Cimini BA, Goodale A, Miller L, Kost-Alimova M, et al. 2022. Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. bioRxiv 2022.01.05.475090. **https://doi.org/10.1101/2022.01.05.475090**
27. Go CD, Knight JDR, Rajasekharan A, Rathod B, Hesketh GG, et al. 2021. A proximity-dependent biotinylation map of a human cell. *Nature* 595(7865):120–24
28. Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, et al. 2012. A census of human soluble protein complexes. *Cell* 150(5):1068–81
29. Heusel M, Bludau I, Rosenberger G, Hafen R, Frank M, et al. 2019. Complex-centric proteome profiling by SEC-SWATH-MS. *Mol. Syst. Biol.* 15(1):e8438
30. Wheat A, Yu C, Wang X, Burke AM, Chemmama IE, et al. 2021. Protein interaction landscapes revealed by advanced in vivo cross-linking-mass spectrometry. *PNAS* 118(32):e2023360118
31. Shilts J, Severin Y, Galaway F, Müller-Sienerth N, Chong Z-S, et al. 2022. A physical wiring diagram for the human immune system. *Nature* 608(7922):397–404
32. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, et al. 2015. Tissue-based map of the human proteome. *Science* 347(6220):1260419
33. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, et al. 2017. The Human Cell Atlas. *eLife* 6:e27041
34. Karlsson M, Zhang C, Méar L, Zhong W, Digre A, et al. 2021. A single-cell type transcriptomics map of human tissues. *Sci. Adv.* 7(31):eabh2169
35. Haniffa M, Taylor D, Linnarsson S, Aronow BJ, Bader GD, et al. 2021. A roadmap for the Human Developmental Cell Atlas. *Nature* 597(7875):196–205
36. Qin Y, Huttlin EL, Winsnes CF, Gosztyla ML, Wacheul L, et al. 2021. A multi-scale map of cell structure fusing protein images and interactions. *Nature* 600(7889):536–42
37. Zheng F, Kelly MR, Ramms DJ, Heintschel ML, Tao K, et al. 2021. Interpretation of cancer mutations using a multiscale map of protein systems. *Science* 374(6563):eabf3067
38. Kratz A, Kim M, Kelly MR, Zheng F, Koczor CA, et al. 2023. A multi-scale map of protein assemblies in the DNA damage response. *Cell Syst.* 14(6):447–63.e8
39. Weber TP. 2000. Biological objects, units of selection and character decomposition. *Trends Ecol. Evol.* 15(8):304–5
40. ENCODE Proj. Consort. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74
41. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, et al. 2022. The complete sequence of a human genome. *Science* 376(6588):44–53
42. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, et al. 2014. Mass-spectrometry-based draft of the human proteome. *Nature* 509(7502):582–87
43. Kelleher NL. 2012. A cell-based approach to the Human Proteome Project. *J. Am. Soc. Mass Spectrom.* 23(10):1617–24
44. Tabula Muris Consort. 2020. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* 583:590–95
45. Tabula Sapiens Consort., Jones RC, Karkanias J, Krasnow MA, Pisco AO, et al. 2022. The Tabula Sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. *Science* 376(6594):eabl4896
46. Smith LM, Agar JN, Chamot-Rooke J, Danis PO, Ge Y, et al. 2021. The Human Proteoform Project: defining the human proteome. *Sci. Adv.* 7(46):eabk0734
47. Melani RD, Gerbasi VR, Anderson LC, Sikora JW, Toby TK, et al. 2022. The Blood Proteoform Atlas: a reference map of proteoforms in human hematopoietic cells. *Science* 375(6579):411–18
48. Sharifi Tabar M, Parsania C, Chen H, Su X-D, Bailey CG, Rasko JEJ. 2022. Illuminating the dark protein-protein interactome. *Cell Rep. Methods* 2(8):100275
49. Richards AL, Eckhardt M, Krogan NJ. 2021. Mass spectrometry-based protein–protein interaction networks for the study of human diseases. *Mol. Syst. Biol.* 17(1):e8792
50. Greco TM, Kennedy MA, Cristea IM. 2020. Proteomic technologies for deciphering local and global protein interactions. *Trends Biochem. Sci.* 45(5):454–55
51. Gnann C, Cesnik AJ, Lundberg E. 2021. Illuminating non-genetic cellular heterogeneity with imaging-based spatial proteomics. *Trends Cancer* 7(4):278–82

52. Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, et al. 2015. A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* 163(3):712–23
53. Fossati A, Li C, Uliana F, Wendt F, Frommelt F, et al. 2021. PCprophet: a framework for protein complex prediction and differential analysis using proteomic data. *Nat. Methods* 18(5):520–27
54. Jiao F, Yu C, Wheat A, Wang X, Rychnovsky SD, Huang L. 2022. Two-dimensional fractionation method for proteome-wide cross-linking mass spectrometry analysis. *Anal. Chem.* 94(10):4236–42
55. Bogdanow B, Gruska I, Mühlberg L, Protze J, Hohensee S, et al. 2023. Spatially resolved protein map of intact human cytomegalovirus virions. *Nat. Microbiol.* 8(9):1732–47
56. Piersimoni L, Kastritis PL, Arlt C, Sinz A. 2022. Cross-linking mass spectrometry for investigating protein conformations and protein–protein interactions—a method for all seasons. *Chem. Rev.* 122(8):7500–31
57. Simpson JC, Neubrand VE, Wiemann S, Pepperkok R. 2001. Illuminating the human genome. *Histochem. Cell Biol.* 115(1):23–29
58. Simpson JC, Wellenreuther R, Poustka A, Pepperkok R, Wiemann S. 2000. Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep.* 1(3):287–92
59. Liebel U, Starkuviene V, Erfle H, Simpson JC, Poustka A, et al. 2003. A microscope-based screening platform for large-scale functional protein analysis in intact cells. *FEBS Lett.* 554(3):394–98
60. Chen B, Zou W, Xu H, Liang Y, Huang B. 2018. Efficient labeling and imaging of protein-coding genes in living cells using CRISPR-Tag. *Nat. Commun.* 9(1):5065
61. Dunkley TPJ, Watson R, Griffin JL, Dupree P, Lilley KS. 2004. Localization of organelle proteins by isotope tagging (LOPIT). *Mol. Cell Proteom.* 3(11):1128–34
62. Sali A. 2021. From integrative structural biology to cell biology. *J. Biol. Chem.* 296:100743
63. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, et al. 2007. The molecular architecture of the nuclear pore complex. *Nature* 450(7170):695–701
64. Zheng F, Zhang S, Churas C, Pratt D, Bahar I, Ideker T. 2021. HiDeF: identifying persistent structures in multiscale 'omics data. *Genome Biol.* 22(1):21
65. Kramer M, Dutkowski J, Yu M, Bafna V, Ideker T. 2014. Inferring gene ontologies from pairwise similarity data. *Bioinformatics* 30(12):i34–42
66. Kim Y-A, Cho D-Y, Przytycka TM. 2016. Understanding genotype-phenotype effects in cancer via network approaches. *PLOS Comput. Biol.* 12(3):e1004747
67. Creixell P, Reimand J, Haider S, Wu G, Shibata T, et al. 2015. Pathway and network analysis of cancer genomes. *Nat. Methods* 12(7):615–21
68. Horn H, Lawrence MS, Chouinard CR, Shrestha Y, Hu JX, et al. 2018. NetSig: network-based discovery from cancer genomes. *Nat. Methods* 15(1):61–66
69. Dimitrakopoulos C, Hindupur SK, Häfliger L, Behr J, Montazeri H, et al. 2018. Network-based integration of multi-omics data for prioritizing cancer genes. *Bioinformatics* 34(14):2441–48
70. Reyna MA, Leiserson MDM, Raphael BJ. 2018. Hierarchical HotNet: identifying hierarchies of altered subnetworks. *Bioinformatics* 34(17):i972–80
71. Reyna MA, Haan D, Paczkowska M, Verbeke LPC, Vazquez M, et al. 2020. Pathway and network analysis of more than 2500 whole cancer genomes. *Nat. Commun.* 11(1):729
72. Cowen L, Ideker T, Raphael BJ, Sharan R. 2017. Network propagation: a universal amplifier of genetic associations. *Nat. Rev. Genet.* 18(9):551–62
73. Haynes WA, Tomczak A, Khatri P. 2018. Gene annotation bias impedes biomedical research. *Sci. Rep.* 8(1):1362
74. Smith LM, Kelleher NL, Consort. for Top Down Proteom. 2013. Proteoform: a single term describing protein complexity. *Nat. Methods* 10(3):186–87
75. Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, et al. 2018. How many human proteoforms are there? *Nat. Chem. Biol.* 14(3):206–14
76. Morris M, Maeda S, Vossel K, Mucke L. 2011. The many faces of tau. *Neuron* 70(3):410–26
77. Dawson JE, Bah A, Zhang Z, Vernon RM, Lin H, et al. 2020. Non-cooperative 4E-BP2 folding with exchange between eIF4E-binding and binding-incompatible states tunes cap-dependent translation inhibition. *Nat. Commun.* 11(1):3146

78. Williams SE, Noel M, Lehoux S, Cetinbas M, Xavier RJ, et al. 2022. Mammalian brain glycoproteins exhibit diminished glycan complexity compared to other tissues. *Nat. Commun.* 13(1):275
79. Li Y, Sahni N, Pancsa R, McGrail DJ, Xu J, et al. 2017. Revealing the determinants of widespread alternative splicing perturbation in cancer. *Cell Rep*. 21(3):798–812
80. Buljan M, Chalancon G, Dunker AK, Bateman A, Balaji S, et al. 2013. Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. *Curr. Opin. Struct. Biol.* 23(3):443–50
81. Buljan M, Chalancon G, Eustermann S, Wagner GP, Fuxreiter M, et al. 2012. Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol. Cell* 46(6):871–83
82. Ellis JD, Barrios-Rodiles M, Colak R, Irimia M, Kim T, et al. 2012. Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol. Cell* 46(6):884–92
83. Biswas SK. 1989. Cultivation of *Mycobacterium leprae* in artificial culture medium. *Indian J. Med. Sci.* 43(1):5–10
84. Herzog F, Kahraman A, Boehringer D, Mak R, Bracher A, et al. 2012. Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science* 337(6100):1348–52
85. May JC, McLean JA. 2015. Ion mobility-mass spectrometry: time-dispersive instrumentation. *Anal. Chem.* 87(3):1422–36
86. Liu FC, Ridgeway ME, Wootton CA, Theisen A, Panczyk EM, et al. 2023. Top-down protein analysis by tandem-trapped ion mobility spectrometry/mass spectrometry (tandem-TIMS/MS) coupled with ultraviolet photodissociation (UVPD) and parallel accumulation/serial fragmentation (PASEF) MS/MS analysis. *J. Am. Soc. Mass Spectrom.* 34(10):2232–46
87. Englander SW. 2006. Hydrogen exchange and mass spectrometry: a historical perspective. *J. Am. Soc. Mass Spectrom.* 17(11):1481–89
88. Bamberger C, Pankow S, Martínez-Bartolomé S, Ma M, Diedrich J, et al. 2021. Protein footprinting via covalent protein painting reveals structural changes of the proteome in Alzheimer's disease. *J. Proteome Res.* 20(5):2762–71
89. Bamberger C, Diedrich J, Martìnez-Bartholomé S, Yates JR. 2022. Cancer conformational landscape shapes tumorigenesis. *J. Proteome Res.* 21(4):1017–28
90. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873):583–89
91. Bennett NR, Coventry B, Goreshnik I, Huang B, Allen A, et al. 2023. Improving de novo protein binder design with deep learning. *Nat. Commun.* 14(1):2625
92. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, et al. 2021. Protein complex prediction with AlphaFold-Multimer. bioRxiv 2021.10.04.463034. **https://doi.org/10.1101/2021.10.04.463034**
93. Uzquiano A, Kedaigle AJ, Pigoni M, Paulsen B, Adiconis X, et al. 2022. Proper acquisition of cell class identity in organoids allows definition of fate specification programs of the human cerebral cortex. *Cell* 185(20):3770–88.e27
94. Fischer DS, Schaar AC, Theis FJ. 2023. Modeling intercellular communication in tissues using spatial graphs of cells. *Nat. Biotechnol.* 41(3):332–36
95. Chidester B, Zhou T, Alam S, Ma J. 2023. SpiceMix enables integrative single-cell spatial modeling of cell identity. *Nat. Genet.* 55(1):78–88
96. Kramer BA, Sarabia Del Castillo J, Pelkmans L. 2022. Multimodal perception links cellular state to decision-making in single cells. *Science* 377(6606):642–48
97. Van Ineveld RL, Kleinnijenhuis M, Alieva M, De Blank S, Barrera Roman M, et al. 2021. Revealing the spatio-phenotypic patterning of cells in healthy and tumor tissues with mLSR-3D and STAPL-3D. *Nat. Biotechnol.* 39(10):1239–45
98. Hickey JW, Becker WR, Nevins SA, Horning A, Perez AE, et al. 2023. Organization of the human intestine at single-cell resolution. *Nature* 619(7970):572–84
99. Rosenberger FA, Thielert M, Strauss MT, Schweizer L, Ammar C, et al. 2023. Spatial single-cell mass spectrometry defines zonation of the hepatocyte proteome. *Nat. Methods* 20(10):1530–36
100. Zhang H, Delafield DG, Li L. 2023. Mass spectrometry imaging: the rise of spatially resolved single-cell omics. *Nat. Methods* 20(3):327–30

101. Mund A, Coscia F, Kriston A, Hollandi R, Kovács F, et al. 2022. Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nat. Biotechnol.* 40(8):1231–40
102. Goltsev Y, Samusik N, Kennedy-Darling J, Bhate S, Hale M, et al. 2018. Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell* 174(4):968–81.e15
103. Keren L, Bosse M, Thompson S, Risom T, Vijayaragavan K, et al. 2019. MIBI-TOF: a multiplexed imaging platform relates cellular phenotypes and tissue structure. *Sci. Adv.* 5(10):eaax5851
104. Mund A, Brunner A-D, Mann M. 2022. Unbiased spatial proteomics with single-cell resolution in tissues. *Mol. Cell* 82(12):2335–49
105. Greenbaum S, Averbukh I, Soon E, Rizzuto G, Baranski A, et al. 2023. A spatially resolved timeline of the human maternal–fetal interface. *Nature* 619(7970):595–605
106. Perens J, Salinas CG, Skytte JL, Roostalu U, Dahl AB, et al. 2021. An optimized mouse brain atlas for automated mapping and quantification of neuronal activity using iDISCO+ and light sheet fluorescence microscopy. *Neuroinformatics* 19(3):433–46
107. Renier N, Wu Z, Simon DJ, Yang J, Ariel P, Tessier-Lavigne M. 2014. iDISCO: a simple, rapid method to immunolabel large tissue samples for volume imaging. *Cell* 159(4):896–910
108. Lange M, Granados A, VijayKumar S, Bragantini J, Ancheta S, et al. 2023. Zebrahub – multimodal zebrafish developmental atlas reveals the state-transition dynamics of late-vertebrate pluripotent axial progenitors. bioRxiv 2023.03.06.531398. **https://doi.org/10.1101/2023.03.06.531398**
109. Galeano Niño JL, Wu H, LaCourse KD, Kempchinsky AG, Baryiames A, et al. 2022. Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. *Nature* 611(7937):810–17
110. Sorin M, Rezanejad M, Karimi E, Fiset B, Desharnais L, et al. 2023. Single-cell spatial landscapes of the lung tumour immune microenvironment. *Nature* 614(7948):548–54
111. Fan Y, Andrusivová Ž, Wu Y, Chai C, Larsson L, et al. 2022. Expansion spatial transcriptomics. bioRxiv 2022.10.25.513696. **https://doi.org/10.1101/2022.10.25.513696**
112. Kishi JY, Liu N, West ER, Sheng K, Jordanides JJ, et al. 2022. Light-Seq: light-directed in situ barcoding of biomolecules in fixed cells and tissues for spatially indexed sequencing. *Nat. Methods* 19(11):1393–402
113. Shi H, He Y, Zhou Y, Huang J, Wang B, et al. 2022. Spatial atlas of the mouse central nervous system at molecular resolution. bioRxiv 2022.06.20.496914. **https://doi.org/10.1101/2022.06.20.496914**
114. Eng C-HL, Lawson M, Zhu Q, Dries R, Koulena N, et al. 2019. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* 568(7751):235–39
115. Xia C, Fan J, Emanuel G, Hao J, Zhuang X. 2019. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *PNAS* 116(39):19490–99
116. Fazal FM, Han S, Parker KR, Kaewsapsak P, Xu J, et al. 2019. Atlas of subcellular RNA localization revealed by APEX-seq. *Cell* 178(2):473–90.e26
117. Hicks KG, Cluntun AA, Schubert HL, Hackett SR, Berg JA, et al. 2023. Protein-metabolite interactomics of carbohydrate metabolism reveal regulation of lactate dehydrogenase. *Science* 379(6636):996–1003
118. Saliba A-E, Vonkova I, Gavin A-C. 2015. The systematic analysis of protein-lipid interactions comes of age. *Nat. Rev. Mol. Cell Biol.* 16(12):753–61
119. Saunders KDG, Lewis H-M, Beste DJ, Cexus O, Bailey MJ. 2023. Spatial single cell metabolomics: current challenges and future developments. *Curr. Opin. Chem. Biol.* 75:102327
120. Vicari M, Mirzazadeh R, Nilsson A, Shariatgorji R, Bjärterot P, et al. 2023. Spatial multimodal analysis of transcriptomes and metabolomes in tissues. bioRxiv 2023.01.26.525195. **https://doi.org/10.1101/2023.01.26.525195**
121. Mofatteh M, Echegaray-Iturra F, Alamban A, Dalla Ricca F, Bakshi A, Aydogan MG. 2021. Autonomous clocks that regulate organelle biogenesis, cytoskeletal organization, and intracellular dynamics. *eLife* 10:e72104
122. Hein MY, Peng D, Todorova V, McCarthy F, Kim K, et al. 2023. Global organelle profiling reveals subcellular localization and remodeling at proteome scale. bioRxiv 2023.12.18.572249. **https://doi.org/10.1101/2023.12.18.572249**

123. Lee S-Y, Cheah JS, Zhao B, Xu C, Roh H, et al. 2023. Engineered allostery in light-regulated LOV-Turbo enables precise spatiotemporal control of proximity labeling in living cells. *Nat. Methods* 20(6):908–17
124. Kim SJ, Fernandez-Martinez J, Nudelman I, Shi Y, Zhang W, et al. 2018. Integrative structure and functional anatomy of a nuclear pore complex. *Nature* 555(7697):475–82
125. Shimada A, Kubo M, Baba S, Yamashita K, Hirata K, et al. 2017. A nanosecond time-resolved XFEL analysis of structural changes associated with CO release from cytochrome c oxidase. *Sci. Adv.* 3(7):e1603042
126. Oh HS-H, Rutledge J, Nachun D, Pálovics R, Abiose O, et al. 2023. Organ aging signatures in the plasma proteome track health and disease. *Nature* 624(7990):164–72
127. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, et al. 2021. Learning transferable visual models from natural language supervision. arXiv:2103.00020 [cs.CV]
128. Singh R, Sledzieski S, Bryson B, Cowen L, Berger B. 2023. Contrastive learning in protein language space predicts interactions between drugs and protein targets. *PNAS* 120(24):e2220778120
129. Jang G, Park S, Lee S, Kim S, Park S, Kang J. 2021. Predicting mechanism of action of novel compounds using compound structure and transcriptomic signature coembedding. *Bioinformatics* 37(Suppl. 1):i376–82
130. Girdhar R, El-Nouby A, Liu Z, Singh M, Alwala KV, et al. 2023. ImageBind: one embedding space to bind them all. arXiv:2305.05665 [cs.CV]
131. Yang KD, Belyaeva A, Venkatachalapathy S, Damodaran K, Katcoff A, et al. 2021. Multi-domain translation between single-cell imaging and sequencing data using autoencoders. *Nat. Commun.* 12(1):31
132. Tu X, Cao Z-J, Chenrui X, Mostafavi S, Gao G. 2022. Cross-linked unified embedding for cross-modality representation learning. *Adv. Neural Inf. Process. Syst.* 35:15942–55
133. Bao F, Deng Y, Wan S, Shen SQ, Wang B, et al. 2022. Integrative spatial analysis of cell morphologies and transcriptional states with MUSE. *Nat. Biotechnol.* 40(8):1200–9
134. Ash JT, Darnell G, Munro D, Engelhardt BE. 2021. Joint analysis of expression levels and histological images identifies genes associated with tissue morphology. *Nat. Commun.* 12(1):1609
135. Zhang X, Wang X, Shivashankar GV, Uhler C. 2022. Graph-based autoencoder integrates spatial transcriptomics with chromatin images and identifies joint biomarkers for Alzheimer's disease. *Nat. Commun.* 13(1):7480
136. Cao K, Gong Q, Hong Y, Wan L. 2022. A unified computational framework for single-cell data integration with optimal transport. *Nat. Commun.* 13(1):7419
137. Forster DT, Li SC, Yashiroda Y, Yoshimura M, Li Z, et al. 2022. BIONIC: biological network integration using convolutions. *Nat. Methods* 19(10):1250–61
138. Benkirane H, Pradat Y, Michiels S, Cournède P-H. 2023. CustOmics: a versatile deep-learning based strategy for multi-omics integration. *PLOS Comput. Biol.* 19(3):e1010921
139. Qi L, Wang W, Wu T, Zhu L, He L, Wang X. 2021. Multi-omics data fusion for cancer molecular subtyping using sparse canonical correlation analysis. *Front. Genet.* 12:607817
140. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, et al. 2014. Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods* 11(3):333–37
141. Yu T. 2022. AIME: autoencoder-based integrative multi-omics data embedding that allows for confounder adjustments. *PLOS Comput. Biol.* 18(1):e1009826
142. Ma T, Zhang A. 2019. Integrate multi-omics data with biological interaction networks using Multi-view Factorization AutoEncoder (MAE). *BMC Genom.* 20(Suppl. 11):944
143. Cao Z-J, Gao G. 2022. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat. Biotechnol.* 40(10):1458–66
144. Lancichinetti A, Radicchi F, Ramasco JJ, Fortunato S. 2011. Finding statistically significant communities in networks. *PLOS ONE* 6(4):e18961
145. Edler D, Bohlin L, Rosvall M. 2017. Mapping higher-order network flows in memory and multilayer networks with Infomap. *Algorithms* 10(4):112
146. Snaebjornsson MT, Schulze A. 2018. Non-canonical functions of enzymes facilitate cross-talk between cell metabolic and regulatory pathways. *Exp. Mol. Med.* 50(4):1–16

147. Galea G, Kuodyte K, Khan MM, Thul P, Neumann B, et al. 2022. The Golgi complex is a regulatory hub for homologous recombination-mediated DNA repair. bioRxiv 2022.10.17.512236. **https://doi.org/10.1101/2022.10.17.512236**
148. Zhao Y, Cortes-Huerto R, Kremer K, Rudzinski JF. 2020. Investigating the conformational ensembles of intrinsically disordered proteins with a simple physics-based model. *J. Phys. Chem. B* 124(20):4097–113
149. Velten B, Braunger JM, Argelaguet R, Arnol D, Wirbel J, et al. 2022. Identifying temporal and spatial patterns of variation from multimodal data using MEFISTO. *Nat. Methods* 19(2):179–86
150. Theodoris CV, Xiao L, Chopra A, Chaffin MD, Al Sayed ZR, et al. 2023. Transfer learning enables predictions in network biology. *Nature* 618(7965):616–24
151. Chen RTQ, Rubanova Y, Bettencourt J, Duvenaud D. 2018. Neural ordinary differential equations. arXiv:1806.07366 [cs.LG]
152. Bronstein MM, Bruna J, LeCun Y, Szlam A, Vandergheynst P. 2017. Geometric deep learning: going beyond Euclidean data. *IEEE Signal Process. Mag.* 34(4):18–42
153. Bronstein MM, Bruna J, Cohen T, Veličković P. 2021. Geometric deep learning: grids, groups, graphs, geodesics, and gauges. arXiv:2104.13478 [cs.LG]
154. Bludau I, Aebersold R. 2020. Proteomic and interactomic insights into the molecular basis of cell functional diversity. *Nat. Rev. Mol. Cell Biol.* 21(6):327–40