Bioinformatics, 2025, **41(6)**, btaf205
https://doi.org/10.1093/bioinformatics/btaf205
Advance Access Publication Date: 9 June 2025
Applications Note

Systems biology

Cell Mapping Toolkit: an end-to-end pipeline for mapping subcellular organization

Joanna Lenkiewicz[1,†], Christopher Churas[1,†], Mengzhou Hu[1], Gege Qian[1], Mayank Jain[1], Maxwell Adam Levinson[2], Sadnan Al Manir[2], Yue Qin[1,3], Dylan Fong[1], Keiichiro Ono[1], Jing Chen[1], Chengzhan Gao[1], Dexter Pratt[1], Jillian A. Parker[1], Timothy Clark[2,4,5], Trey Ideker[1,6,7,*], Leah V. Schaffer[1,*]

[1]Department of Medicine, University of California San Diego, La Jolla, CA, 92037, United States
[2]Department of Public Health Sciences (Biomedical Informatics), University of Virginia School of Medicine, Charlottesville, VA, 22903, United States
[3]Eric and Wendy Schmidt Center, Broad Institute of MIT and Harvard, Boston, MA, 02142, United States
[4]Center for Advanced Medical Analytics, University of Virginia School of Medicine, Charlottesville, VA, 22903, United States
[5]University of Virginia School of Data Science, Charlottesville, VA, 22903, United States
[6]Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, 92037, United States
[7]Department of Bioengineering, University of California San Diego, La Jolla, CA, 92037, United States

*Corresponding authors. Leah V. Schaffer. Department of Medicine, University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92037, United States. E-mail: leahvschaffer@gmail.com; Trey Ideker. Department of Medicine, University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92037, United States. E-mail: tideker@health.ucsd.edu.

†= equal contribution.

Associate Editor: Pier Luigi Martelli

Abstract

Summary: Cells are organized as a hierarchy of macromolecular assemblies, ranging from small protein complexes to entire organelles. Various technologies have been developed to elucidate subcellular architecture at different scales, such as mass spectrometry approaches for mapping protein biophysical interactions and immunofluorescence imaging for mapping protein localization. We present the Cell Mapping Toolkit, which is designed to systematically integrate data from different modalities into unified hierarchical maps of subcellular organization. The toolkit facilitates an end-to-end pipeline including processing datasets, integrating modalities, and visualizing the final cell map with rich metadata including provenance documentation at each step. The Cell Mapping Toolkit provides researchers with tools for analyzing, integrating, and visualizing diverse protein datasets in a robust and reproducible framework.

Availability and implementation: The code is freely available and is hosted on GitHub at https://github.com/idekerlab/cellmaps_pipeline. Comprehensive documentation and practical examples are provided at https://cellmaps-pipeline.readthedocs.io/.

1 Introduction

A fundamental goal in biology is mapping protein assemblies and their spatial distribution within cells, with downstream applications including understanding disease phenotypes, revealing drug targets, and interpreting genetics (Karr *et al.* 2012, Johnson *et al.* 2023, Cesnik *et al.* 2024). Various technologies currently exist for mapping biological systems, each measuring different biological scales ranging from nanometers to microns (Wilhelm *et al.* 2014, Mulvey *et al.* 2017, Thul and Lindskog 2018, Luck *et al.* 2020, Richards *et al.* 2021, Skinnider *et al.* 2021, Reed *et al.* 2024). For example, approaches including affinity purification coupled with mass spectrometry (AP-MS) (Choi *et al.* 2012, Huttlin *et al.* 2015, 2021, Gordon *et al.* 2020) or size exclusion chromatography mass spectrometry (SEC-MS) (Havugimana *et al.* 2012, Bludau *et al.* 2020, Fossati *et al.* 2023) enable the identification of protein-protein interactions (PPIs) and protein complexes. At larger biological scales, approaches including subcellular fractionation (Dunkley *et al.* 2004, Mulvey *et al.* 2017) and immunofluorescence (IF) (Thul *et al.* 2017) or endogenous fluorescent-tagged imaging (Cho *et al.* 2022) determine the specific localization of proteins within larger cell compartments. There are also approaches for mapping protein functional associations, such as genome-wide CRISPR perturbations (Dixit *et al.* 2016, Replogle *et al.* 2022) that determine pairs of proteins with similar transcriptional effects upon knockdown.

These technologies have typically been applied separately, each revealing different information about protein organization and with unique advantages and challenges (Christopher *et al.* 2021, Richards *et al.* 2021). Integrating data from multiple protein mapping technologies presents an opportunity to generate a more comprehensive understanding of subcellular structure. Toward this goal, we recently developed an approach for integrating diverse data modalities into a hierarchical map of protein assemblies (Qin *et al.* 2021, Schaffer

Received: 26 November 2024; **Revised:** 25 March 2025; **Editorial Decision:** 10 April 2025

et al. 2025), robustly revealing more protein assemblies in the cell than any individual dataset alone. We developed the Cell Mapping Toolkit to streamline and productionize this process of integrating datasets into hierarchical cell maps and to make the tools accessible to a broad research community. The toolkit is a scalable and user-friendly software tool consisting of a set of Python packages. In what follows, we describe the Cell Mapping Toolkit and present a tutorial demonstrating its application to currently available datasets.

2 Software implementation

The Cell Mapping Toolkit comprises a set of Python packages that are pip installable and facilitate data downloading, processing, co-embedding, and cell map hierarchy generation and evaluation (Fig. 1A). The toolkit provides auto-generated documentation hosted on ReadTheDocs, includes unit testing that runs automatically on code commits, and adheres to a strict version control policy to minimize integration issues.

The main architecture follows a pattern where each step creates a directory on the filesystem that stores one or more data files (Fig. 1B). Subsequent steps use these data files, and each directory is registered as an RO-crate (Soiland-Reyes *et al.* 2022) via FAIRSCAPE framework (Levinson *et al.* 2022) for provenance. As part of the RO-crate, each tool registers the code used to generate the data, as well as required provenance information for any imported data. This provenance and metadata ensure that every step implemented with the toolkit is documented and reproducible, which is important for downstream analysis and interpretability of cell maps (Wilson *et al.* 2021, Clark *et al.* 2024). Schemas defining the format for each file are available at Zenodo (https://doi.org/10.5281/zenodo.14200177). The organization of the toolkit enables users to substitute any step with external code and new methods, as long as the output matches the required format specified by each step. Each tool in the pipeline has a command line interface, as well as a programmatic interface that can be called individually or as a whole. Here, we describe each step in the cell mapping process and the associated tool in the Cell Mapping Toolkit.

2.1 Step 1: Image and Protein-Protein interaction data downloaders

The download process is managed by scripts that ensure the data is fetched, followed by validation against predefined schemas and packaging with rich metadata including provenance into standard RO-Crate packages by the FAIRSCAPE client. We developed an Image Data Downloader, which currently supports downloads from the Human Protein Atlas (HPA) (Thul and Lindskog 2018) using a .tsv file that specifies the required images or a text file with a list of proteins. We also created a PPI Data Downloader, which formats gene names and attributes for an input edge list.

2.2 Step 2: Embed each data modality

We generated tools to create embeddings (a low-dimensional representation extracted from complex high-dimensional input) for each data modality, implementing algorithms to support image and network-based data. For images, the default embedding is the penultimate layer of an HPA image classification model [densenet (Ouyang *et al.* 2019)] which captures information about protein subcellular localization. For network-based data modalities, we developed a PPI Embedding tool that runs the node2vec (Grover and Leskovec 2016) algorithm on the network, which generates an embedding for each node (i.e. protein) that captures relative relationships about the interaction neighborhoods.

2.3 Step 3: Co-embed the data modalities

The embeddings for each data modality—generated either by our toolkit for image and network embeddings or externally for other data types—are integrated using the co-embedding tool (Schaffer *et al.* 2025). The integration uses a self-supervised learning approach (Bao *et al.* 2022) to learn a unified embedding for each protein. The toolkit provides utilities for evaluating the co-embeddings, including assessing the similarities of protein pairs present in known complexes and visualizing the embedding space using the UMAP method (McInnes *et al.* 2018) (Fig. 1C).

2.4 Step 4: Generate hierarchy of protein assemblies

Hierarchy generation within the toolkit begins by calculating cosine similarities of the co-embedding between each protein pair. A set of protein-protein similarity networks is generated at various thresholds, and pan-resolution community detection is performed using Hierarchical community Decoding Framework (Zheng *et al.* 2021) to generate a multi-scale hierarchy.

2.5 Step 5: Evaluate the hierarchy

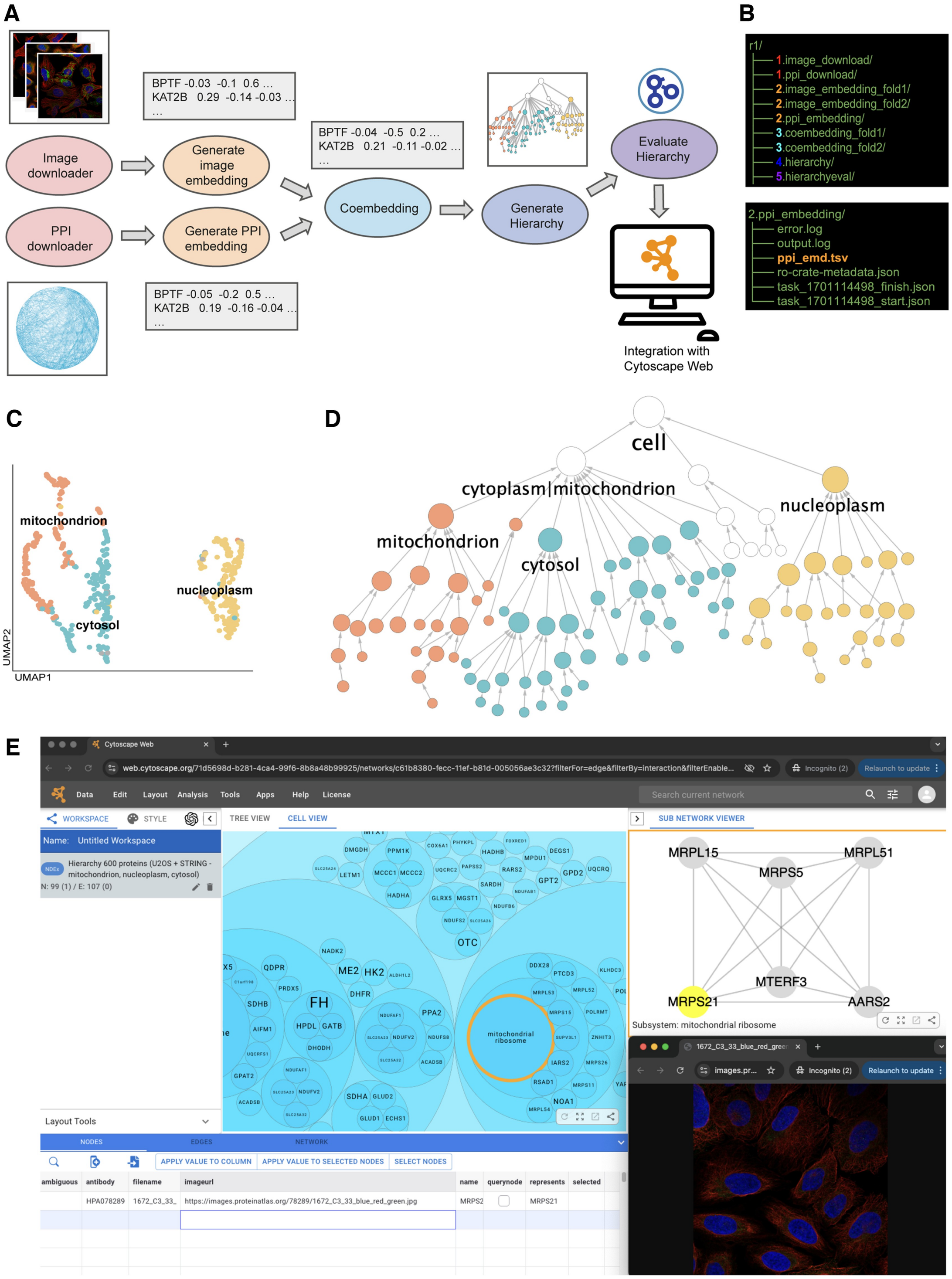
The hierarchy is evaluated for overlap with documented protein assemblies using multiple resources including Gene Ontology (Ashburner *et al.* 2000, Aleksander *et al.* 2023) cellular component terms, the comprehensive resource of mammalian protein complexes [CORUM (Tsitsiridis *et al.* 2023)], and HPA cellular compartments. Additionally, the toolkit provides the option to annotate assemblies in the cell maps using a large language model (LLM) approach to name sets of proteins and assign a name confidence score using a designed prompt (Hu *et al.* 2023). The final annotated hierarchy is saved in a format allowing visualization in Cytoscape, and can be uploaded to the Network Data Exchange (NDEx, https://www.ndexbio.org/) for storage, sharing, manipulation, and publication (Pratt *et al.* 2015, Pillich *et al.* 2021). Finally, the toolkit provides documentation and utilities to assess the robustness of protein assemblies across multiple jackknife resamplings.

3 Results

3.1 Environmental setup

The Cell Maps Pipeline Python package can be installed using the following command: `pip install cellmaps_pipeline`.

This package is compatible with Python versions 3.8 through 3.11. For optimal performance and isolation of dependencies, it is strongly recommended to utilize an Anaconda environment (docs.anaconda.com).



```
BPTF -0.03  -0.1  0.6 ...
KAT2B   0.29  -0.14 -0.03 ...
...
```

```
BPTF -0.05  -0.2  0.5 ...
KAT2B   0.19  -0.16 -0.04 ...
...
```

```
BPTF -0.04  -0.5  0.2 ...
KAT2B   0.21  -0.11 -0.02 ...
...
```

```
r1/
├── 1.image_download/
├── 1.ppi_download/
├── 2.image_embedding_fold1/
├── 2.image_embedding_fold2/
├── 2.ppi_embedding/
├── 3.coembedding_fold1/
├── 3.coembedding_fold2/
├── 4.hierarchy/
├── 5.hierarchyeval/
```

```
2.ppi_embedding/
├── error.log
├── output.log
├── ppi_emd.tsv
├── ro-crate-metadata.json
├── task_1701114498_finish.json
└── task_1701114498_start.json
```

web.cytoscape.org/71d5698d-b281-4ca4-99f6-8b8a48b99925/networks/c61b8380-fecc-11ef-b81d-005056ae3c32?filterFor=edge&filterBy=interaction&filterEnable...

Name: Untitled Workspace

Hierarchy 600 proteins (U2OS + STRING - mitochondrion, nucleoplasm, cytosol)
N: 99 (1) / E: 107 (0)

Subsystem: mitochondrial ribosome

ambiguous	antibody	filename	imageurl	name	querynode	represents	selected
	HPA078289	1672_C3_33_	https://images.proteinatlas.org/78289/1672_C3_33_blue_red_green.jpg	MRPS2	□	MRPS21	

Figure 1. Demonstration of the Cell Mapping Toolkit. (A) Overview of processing steps in Cell Mapping Toolkit. (B) Directory structure and outputs of the Cell Mapping Toolkit. The upper panel displays all output directories created after running the full pipeline, with each folder corresponding to a specific step in the process. The lower panel shows the resulting files in the directory generated by the PPI embedding step. These include log files, an embedding file, and a RO-Crate metadata file capturing provenance information. (C) Multimodal embedding of proteins based on integration of AP-MS and imaging data, reduced to two dimensions using the UMAP method. (D) Proof-of-concept cell map generated with 600 proteins using associations in STRING and images from the HPA. The hierarchy is represented in a tree view. The size of nodes is proportional to the number of proteins. Nodes are colored based on subcellular location, as defined by HPA. (E) Cell map in cell view (circle packing) on Cytoscape Web. Selecting a protein assembly cluster in the cell map shows underlying interaction data and links to the images.

3.2 Data acquisition

To create a proof-of-concept cell map, we randomly selected a set of 600 proteins, including 200 localized to each of three different cell compartments (nucleoplasm, mitochondria, and cytosol), as defined by HPA. A list of these proteins was provided as an input in the Image Downloader to obtain images (see below). Protein-protein Interactions (PPIs) were obtained from the high-confidence (score ≥ 0.7) STRING (Snel *et al.*

2000, Szklarczyk *et al.* 2019) interactome. We selected a subnetwork for the same 600 proteins from the network on NDEx (https://ndexbio.org/, uuid: 24823fd3-6ebb-11ef-a7fd-005056ae23aa), saved as an edgelist in a .tsv file.

3.3 Data provenance

A provenance file detailing the information about input data must be provided to adhere to FAIR principles. Users have the option to generate a sample provenance file using the following command:

```
cellmaps_pipelinecmd.py . --example_provenance
>provenance.json
```

Once the sample provenance file is generated, the user should edit it to include the necessary information, including name, organization name, project name, cell line, treatment, gene set, and information about individual input files.

3.4 Running the cell mapping toolkit

The Cell Mapping Toolkit can be executed by running cellmaps_pipelinecmd.py with required arguments including the output directory, provenance file, and input data.

```
cellmaps_pipelinecmd.py ./cellmaps_pipeline_outdir [FLAGS WITH PARAMETERS]
```

Alternatively, individual toolkit steps can be run separately through their respective Python packages.

1) Downloading images from HPA
   ```
   cellmaps_imagedownloadercmd.py ./1.image_downloader --protein_list proteins.txt
   --cell_line U2OS --provenance provenance_images.json
   ```
2) Generating embeddings in image and PPI data
   ```
   cellmaps_image_embeddingcmd.py ./2.image_embedding --inputdir ./1.image_downloader
   cellmaps_ppi_embeddingcmd.py ./2.ppi_embedding --inputdir ./string_ppi_dir
   --provenance provenance_ppi.json
   ```
3) Integrating the embeddings (co-embedding)
   ```
   cellmaps_coembeddingcmd.py ./3.coembedding --embeddings ./2.ppi_embedding ./2.image_embedding
   ```
4) Generating the hierarchical cell map
   ```
   cellmaps_generate_hierarchycmd.py ./4.hierarchy --coembedding_dirs ./3.coembedding
   ```
5) Evaluating cell map for known components (Fig. 1D)
   ```
   cellmaps_hierarchyevalcmd.py ./5.hierarchyeval --hierarchy_dir ./4.hierarchy
   ```

3.5 Visualization and sharing

The Cell Mapping Toolkit can be used to upload the final hierarchy to NDEx, a platform for sharing biological network data (Pratt *et al.* 2015, Pillich *et al.* 2021). NDEx provides other users easy access to the hierarchy, making it accessible to a broader community and facilitating collaboration. Users can upload their hierarchy using the cellmaps_generate_hierarchy tool included in the toolkit, using their credentials to the NDEx account with the following command:

```
cellmaps_generate_hierarchycmd.py ./5.hierarchyeval --mode ndexsave --ndexuser < USER >
```

Once the hierarchy is uploaded, a link is generated that allows the user to access and interact with the hierarchy through Cytoscape Web (web.cytoscape.org), a new web application based on the desktop application (Shannon *et al.* 2003, Smoot *et al.* 2011). This platform provides a visual interface where users can explore and interact with the cell map in two views, the tree hierarchy (Fig. 1D) and a cell view (Fig. 1E). Users can also browse the underlying subnetworks and links to view images for each protein assembly.

3.6 Cell mapping toolkit test users

As part of the National Institutes of Health (NIH) Bridge2AI program (Clark *et al.* 2024), we have hosted a series of in-person and virtual codefests where users implement and test the Cell Mapping Toolkit. These codefests resulted in a total of approximately 50 participants who ran the toolkit and created cell maps from different sample datasets. We used feedback from the users to fix unexpected issues and improve the documentation and guides. This number of test users highlights the stability of the toolkit on a variety of computational systems and by personnel of varying computational experiences.

4 Conclusions

We have developed the Cell Mapping Toolkit to build and analyze hierarchical maps of cell architecture via integration of diverse data modalities. The toolkit's modularity and flexibility enable users to adapt the pipeline to their specific datasets and applications. The tool is user-friendly, extensible, and ensures the creation of trackable and reproducible results.

Acknowledgements

We acknowledge collaborations with the laboratories of Emma Lundberg, Nevan Krogan, and Prashant Mali and thank them for helpful discussions regarding input data. We thank Wade Schulz for leading and planning codefest events, as well as the participants of the codefests for their valuable feedback.

Author contributions

Joanna Lenkiewicz (Conceptualization, Software, Writing Original Draft), Christopher Churas (Conceptualization, Software, Writing Review Editing), Mengzhou Hu (Conceptualization, Software, Writing Review Editing), Gege Qian (Conceptualization, Software, Writing Review Editing), Mayank Jain (Software, Writing Review Editing), Maxwell Adam Levinson (Software, Writing Review Editing), Sadnan Al Manir (Software, Writing Review Editing), Yue Qin (Conceptualization, Software, Writing Review Editing), Dylan Fong (Software, Writing Review Editing), Keiichiro Ono (Software, Writing Review Editing), Jing Chen (Software, Writing Review Editing), Chengzhan Gao (Software, Writing Review Editing), Dexter Pratt (Conceptualization, Writing Review Editing), Jillian A. Parker (Conceptualization, Supervision, Writing Review

Editing), Timothy Clark (Conceptualization, Supervision, Writing Review Editing), Trey Ideker (Conceptualization, Supervision, Writing Review Editing), and Leah V. Schaffer (Conceptualization, Software, Writing Original Draft, Writing Review Editing)

Conflict of interest: T.I. is a co-founder, 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.

Funding

This work was supported by the Bridge2AI Program [NIH Common Fund; OT2 OD032742, T.I., T.C.], NIH NCI U24 CA269436 [to D.P., T.I.], NIH NHGRI U24 HG012107 [to T.I.], and Schmidt Futures [to T.I.]. Y.Q. was supported by NIH F99 CA264422.

Data availability

The protein interaction data used for the proof-of-concept cell map are available at https://ndexbio.org under uuid 24823fd3-6ebb-11ef-a7fd-005056ae23aa. Images are available at the Human Protein Atlas (https://www.proteinatlas.org/).

References

Aleksander SA, Balhoff J, Carbon S *et al.*; Gene Ontology Consortium. The gene ontology knowledgebase in 2023. *Genetics* 2023;**224**. https://doi.org/10.1093/genetics/iyad031.

Ashburner M, Ball CA, Blake JA *et al.* Gene ontology: tool for the unification of biology. *Nat Genet* 2000;**25**:25–9.

Bao F, Deng Y, Wan S *et al.* Integrative spatial analysis of cell morphologies and transcriptional states with MUSE. *Nat Biotechnol* 2022; **40**:1200–9.

Bludau I, Heusel M, Frank M *et al.* Complex-centric proteome profiling by SEC-SWATH-MS for the parallel detection of hundreds of protein complexes. *Nat Protoc* 2020;**15**:2341–86.

Cesnik A, Schaffer LV, Gaur I *et al.* Mapping the multiscale proteomic organization of cellular and disease phenotypes. *Annu Rev Biomed Data Sci* 2024;**7**:369–89.

Cho NH, Cheveralls KC, Brunner A-D *et al.* OpenCell: endogenous tagging for the cartography of human cellular organization. *Science* 2022;**375**:eabi6983.

Choi H, Liu G, Mellacheruvu D *et al.* Analyzing protein–protein interactions from affinity purification-mass spectrometry data with SAINT. *Curr Protoc Bioinf* 2012;Chapter 8:8.15.1–23.

Christopher JA, Stadler C, Martin CE *et al.* Subcellular proteomics. *Nat Rev Methods Primers* 2021;**1**. https://doi.org/10.1038/s43586-021-00029-y

Clark T, Mohan J, Schaffer L *et al.* Cell maps for artificial intelligence: AI-ready maps of human cell architecture from disease-relevant cell lines. BioRxiv, https://doi.org/10.1101/2024.05.21.589311, 2024, preprint: not peer reviewed.

Dixit A, Parnas O, Li B *et al.* Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 2016;**167**:1853–66.e17.

Dunkley TPJ, Watson R, Griffin JL *et al.* Localization of organelle proteins by isotope tagging (LOPIT). *Mol Cell Proteomics* 2004; **3**:1128–34.

Fossati A, Mozumdar D, Kokontis C *et al.* Next-Generation proteomics for quantitative Jumbophage-Bacteria interaction mapping. *Nat Commun* 2023;**14**:5156.

Gordon DE, Jang GM, Bouhaddou M *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 2020; **583**:459–68.

Grover A, Leskovec J. Node2vec: scalable feature learning for networks. In: *KDD: Proceedings/International Conference on Knowledge Discovery & Data Mining. International Conference on Knowledge Discovery & Data Mining 2016 (August)*. 2016, 855–64.

Havugimana PC, Hart GT, Nepusz T *et al.* A census of human soluble protein complexes. *Cell* 2012;**150**:1068–81.

Hu M, Alkhairy S, Lee I *et al.* Evaluation of large language models for discovery of gene set function. *Nat Methods* 2024;**22**:82–91.

Huttlin EL, Bruckner RJ, Navarrete-Perea J *et al.* Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell* 2021;**184**:3022–40.e28.

Huttlin EL, Ting L, Bruckner RJ *et al.* The BioPlex network: a systematic exploration of the human interactome. *Cell* 2015;**162**:425–40.

Johnson GT, Agmon E, Akamatsu M *et al.* Building the next generation of virtual cells to understand cellular biology. *Biophys J* 2023; **122**:3560–9.

Karr JR, Sanghvi JC, Macklin DN *et al.* A whole-cell computational model predicts phenotype from genotype. *Cell* 2012;**150**:389–401.

Levinson MA, Niestroy J, Al Manir S *et al.* FAIRSCAPE: a framework for FAIR and reproducible biomedical analytics. *Neuroinformatics* 2022;**20**:187–202.

Luck K, Kim D-K, Lambourne L *et al.* A reference map of the human binary protein interactome. *Nature* 2020;**580**:402–8.

McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. arXiv, arXiv:1802.03426, 2018, preprint: not peer reviewed.

Mulvey CM, Breckels LM, Geladaki A *et al.* Using HyperLOPIT to perform high-resolution mapping of the spatial proteome. *Nat Protoc* 2017;**12**:1110–35.

Ouyang W, Winsnes CF, Hjelmare M *et al.* Analysis of the human protein atlas image classification competition. *Nat Methods* 2019; **16**:1254–61.

Pillich RT, Chen J, Churas C *et al.* NDEx: accessing network models and streamlining network biology workflows. *Curr Protoc* 2021; **1**:e258.

Pratt D, Chen J, Welker D *et al.* NDEx, the network data exchange. *Cell Syst* 2015;**1**:302–5.

Qin Y, Huttlin EL, Winsnes CF *et al.* A multi-scale map of cell structure fusing protein images and interactions. *Nature* 2021;**600**:536–42. https://doi.org/10.1038/s41586-021-04115-9

Reed TJ, Tyl MD, Tadych A *et al.* Tapioca: a platform for predicting de novo protein–protein interactions in dynamic contexts. *Nat Methods* 2024;**21**:488–500.

Replogle JM, Saunders RA, Pogson AN *et al.* Mapping information-rich genotype-phenotype landscapes with genome-scale perturb-seq. *Cell* 2022;**185**:2559–75.e28.

Richards AL, Eckhardt M, Krogan NJ. Mass spectrometry-based protein–protein interaction networks for the study of human diseases. *Mol Syst Biol* 2021;**17**:e8792.

Schaffer LV, Hu M, Qian G *et al.* Multimodal cell maps as a foundation for structural and functional genomics. *Nature* 2025. https://doi.org/10.1038/s41586-025-08878-3

Shannon P, Markiel A, Ozier O *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;**13**:2498–504.

Skinnider MA, Scott NE, Prudova A *et al.* An atlas of protein–protein interactions across mouse tissues. *Cell* 2021;**184**:4073–89.e17.

Smoot ME, Ono K, Ruscheinski J *et al.* Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 2011; **27**:431–2.

Snel B, Lehmann G, Bork P *et al.* STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res* 2000;**28**:3442–4.

Soiland-Reyes S, Sefton P, Crosas M *et al.* Packaging research artefacts with RO-Crate. *Data Science* 2022;**5**:97–138.

Szklarczyk D, Gable AL, Lyon D *et al.* STRING V11: protein–protein association networks with increased coverage, supporting functional discovery in Genome-Wide experimental datasets. *Nucleic Acids Res* 2019;**47**:D607–13.

Thul PJ, Åkesson L, Wiking M *et al.* A subcellular map of the human proteome. *Science* 2017;**356**. https://doi.org/10.1126/science.aal3321

Thul PJ, Lindskog C. The human protein atlas: a spatial map of the human proteome. *Protein Sci* 2018;**27**:233–44.

Tsitsiridis G, Steinkamp R, Giurgiu M *et al.* CORUM: the comprehensive resource of mammalian protein complexes–2022. *Nucleic Acids Res* 2023;**51**:D539–45.

Wilhelm M, Schlegl J, Hahne H *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* 2014;**509**:582–7.

Wilson SL, Way GP, Bittremieux W *et al.* Sharing biological data: why, when, and how. *FEBS Lett* 2021;**595**:847–63.

Zheng F, Zhang S, Churas C *et al.* HiDeF: identifying persistent structures in multiscale 'omics data. *Genome Biol* 2021;**22**:21.

Bioinformatics, 2025, 41, 1–6
https://doi.org/10.1093/bioinformatics/btaf205
Applications Note