Tools for assembling the cell: Towards the era of cell structural bioinformatics

Mengzhou Hu[1†], Xikun Zhang[2†], Andrew Latham[3†],
Andrej Šali[3*], Trey Ideker[1*], and Emma Lundberg[2*]

[1]*Department of Medicine, University of California San Diego, San Diego, CA, USA*
[2]*Department of Bioengineering, Stanford University, Stanford, CA, USA*
[3]*Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA, USA.*

† These authors contributed equally

* To whom correspondence should be addressed:
sali@salilab.org, tideker@health.ucsd.edu, emmalu@stanford.edu

Cells consist of large components, such as organelles, that recursively factor into smaller systems, such as condensates and protein complexes, forming a dynamic multi-scale structure of the cell. Recent technological innovations have paved the way for systematic interrogation of subcellular structures, yielding unprecedented insights into their roles and interactions. In this workshop, we discuss progress, challenges, and collaboration to marshal various computational approaches toward assembling an integrated structural map of the human cell.

Keywords: cell mapping, subcellular structures, computational modeling of cell

1. Overview

A fundamental objective of cell biology is to decode the intricate multi-scale structures within cells, ranging from macroscopic organelles to microscopic condensates and protein complexes. This goal necessitates a comprehensive understanding of the spatial and functional organizations of subcellular components, particularly within the context of cell function and diseases.

In recent years, a plethora of advanced technologies have emerged, enabling systematic interrogation of subcellular structures and providing unprecedented insights into their functional significance. For example, immunofluorescence imaging[1–3] facilitates the real-time visualization of static and dynamic subcellular interactions at high resolution. Similarly, cryo-electron tomography[4,5] and microscopy[6–9] capture intricate structural details of subcellular components in their native, hydrated state, thus preserving their functional context. On the biochemical front, affinity purification,[10,11] co-elution,[12] and crosslinking mass spectrometry[13,14] techniques have provided avenues for elucidating the complex networks of protein interactions within cells. The emerging machine learning pipelines[15–18] associated with these technologies have further augmented the systematic interpretation of cell architecture and association with diseases.

The integration of these complementary technologies represents a promising avenue for mapping the architecture of cells across a broad range of scales. Using two of these techniques, protein imaging and affinity purification, the session organizers have recently published a novel framework, called MuSIC (Multi-Scale Integrated Cell),[19] for assembling hierarchical maps of human subcellular components spanning the multiple scales of cell biology. The timely and distinct opportunity that emerges from this work is to assemble a key group of thought leaders in a suitable location to discuss progress, open challenges, and, most importantly, how collaborative teams can be established to marshal the various technologies toward an integrated structural map of the human cell.

Hence, this workshop, "Tools for assembling the cell: Towards the era of cell structural bioinformatics," aims to be a catalyst for scientific discourse and collaboration, providing a platform for eminent professionals from varying domains to explore and strategize the future of subcellular structure mapping. We invited seven distinguished speakers (Drs. David Baker, Markus Covert, Jan Ellenberg, Rachel Karchin, Tychele Turner, Aubrey Weigel, and Marinka Zitnik) to share insights on data acquisition and computational approaches for cellular modeling. This workshop is designed to provide attendees with a deep dive into the present technological innovations and highlight avenues for potential collaboration and exploration.

2. Navigating the Workshop

Developing a spatiotemporal map of the cell necessitates integrating various sources of data into a single model. To enable communication and synergy between experimental scientists and computational modelers, this workshop features incisive talks from seven experts in procuring spatiotemporal biological data and advancing computational modeling of cellular architecture across multiple scales.

Dr. David Baker is a Henrietta and Aubrey Davis Endowed Professor of Biochemistry at University of Washington, Director of Institute for Protein Design and an Investigator at Howard Hughes Medical Institute. His research focuses on developing protein design software and using it to create molecules that solve challenges in medicine, technology and sustainability. His group developed the Rosetta algorithm for ab initio protein structure prediction.[16,20] Most recently, his group has developed RoseTTAFold, a three-track network to process sequence, distance, and coordinate information simultaneously, and achieved more accurate protein structure prediction.[21]

Dr. Markus Covert, a Professor of Bioengineering and, by courtesy, of Chemical and Systems Biology at Stanford University, focuses on building computational models of complex biological processes and using these models to guide an experimental program. His lab pioneered the "whole-cell" model encoding all known information about each gene and molecule to predict cell behaviors.[22] His lab has also made significant contributions to live-cell imaging of immune signaling, including a game-changing method to analyze microscopy images using deep learning[15] and a technique that traces cellular behavior from the initial stimulus, through the signaling pathways, down to genome-wide changes in gene expression, within the single cell.[23]

Dr. Jan Ellenberg is Head of Cell Biology and Biophysics, and Head of the European

Molecular Biology Laboratory (EMBL) Imaging Center, at EMBL Heidelberg. He has developed state-of-the-art quantitative fluorescence-based imaging techniques,[2] and combined these technologies with subsequent automation and analysis platforms.[3] His lab leveraged these four-dimensional imaging approaches to enable characterization of processes within human cells, such as protein localization during cell division[24] and nuclear pore complex assembly.[25]

Dr. Rachel Karchin is a professor at Johns Hopkins University and the Institute for Computational Medicine. She has made significant contributions to the field of cancer genomics by leveraging 3D protein structure for variant interpretation developing tools to detect somatic mutation hotspot regions in 3D protein structures.[17,26] Similarly, **Dr. Tychele Turner**, an Assistant Professor at Washington University in St. Louis, worked on precision genomics in neurodevelopmental disorders, determining all possible relevant variations within an individual to the precise nucleotide.[27] Together, both of them focused on mapping mutations in 3D and aimed to compare the 3D mutation clusters between neurodevelopmental diseases and cancers, bringing new insight into genomics research.

Dr. Aubrey Weigel is a Project Scientist of the Cellular Organelle Segmentation in Electron Microscopy (COSEM) Project Team at Howard Hughes Medical Institute (HHMI) - Janelia Research Campus. She has pioneered a pipeline that combines focused ion beam scanning electron microscopy (FIB-SEM) with deep learning annotation methods to reconstruct maps of entire cells at 4-8 nm resolution.[8,9] Such data and models are available to the scientific community through an open-sourced platform, called OpenOrganelle. These data acquisition and analysis techniques can provide insight into complicated cellular processes, and similar analyses revealed the dynamics of endoplasmic reticulum (ER)-to-Golgi protein delivery.[28]

Dr. Marinka Zitnik is an Assistant Professor at Harvard Medical School, and affiliated with several Harvard-based institutes. She investigates the foundations of AI to enhance scientific discovery and to realize individualized diagnosis and treatment. She proposed Decagon, a graph-convolution-network-based model to model polypharmacy side effects.[18] She also founded Therapeutics Data Commons (TDC), an initiative to access and evaluate AI capability across therapeutic modalities and stages of discovery. Their aim is to establish which AI methods are most suitable for advancing therapeutic science and why these techniques are advantageous.[29,30]

3. Discussion and Implications

In this workshop, we delve into cutting-edge technologies designed to illuminate the spatial and functional organizations of subcellular components. Drs. David Baker, Markus Covert, Jan Ellenberg, Rachel Karchin, Tychele Turner, Aubrey Weigel, and Marinka Zitnik are the distinguished speakers contributing their extensive knowledge to this workshop. They elucidate the advancements in data acquisition, sophisticated analysis techniques, and computational tools essential for the assembly of human subcellular components at various scales. This workshop provides a platform not only as a repository of knowledge but also as a forum for academic exchange. Scientists are welcome to discuss the promises, pitfalls, and challenges of modeling the subcellular structures. In addition, the insights of the distinguished speakers can foster the promise of interdisciplinary projects using cell mapping techniques, encouraging potential

collaborations to drive cell structural biology further.

References

1. P. J. Thul, *et al.*, A subcellular map of the human proteome, *Science* **356** (2017).
2. J. K. Hériché, *et al.*, Integrating Imaging and Omics: Computational Methods and Challenges, *Annual Review of Biomedical Data Science* **2**, 175 (2019).
3. D. D. Nogare, *et al.*, Using AI in bioimage analysis to elevate the rate of scientific discovery as a community, *Nature methods* **20**, 973 (2023).
4. M. Turk, *et al.*, The promise and the challenges of cryo-electron tomography, *FEBS Letters* **594**, 3243 (2020).
5. S. Zheng, *et al.*, AreTomo: An integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction, *Journal of Structural Biology: X* **6** (2022).
6. X. Li, *et al.*, Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM, *Nature Methods* **10**, 584 (2013).
7. S. Q. Zheng, *et al.*, MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy, *Nature Methods* **14**, 331 (2017).
8. L. Heinrich, *et al.*, Whole-cell organelle segmentation in volume electron microscopy, *Nature* **599**, 141 (2021).
9. C. S. Xu, *et al.*, An open-access volume electron microscopy atlas of whole cells and tissues, *Nature* **599**, 147 (2021).
10. J. H. Morris, *et al.*, Affinity purification-mass spectrometry and network analysis to understand protein-protein interactions, *Nat. Protoc.* **9**, 2539 (2014).
11. E. L. Huttlin, *et al.*, Architecture of the human interactome defines protein communities and disease networks, *Nature* **545**, 505 (2017).
12. D. Salas, *et al.*, Next-generation interactomics: Considerations for the use of co-elution to measure protein interaction networks, *Mol. Cell. Proteomics* **19**, 1 (2020).
13. F. Liu, *et al.*, Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry, *Nat. Methods* **12**, 1179 (2015).
14. L. Piersimoni, *et al.*, Cross-linking mass spectrometry for investigating protein conformations and protein–protein interactions-a method for all seasons, *Chemical Reviews* **122**, 7500 (2022), PMID: 34797068.
15. D. A. Van Valen, *et al.*, Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments, *PLoS computational biology* **12**, p. e1005177 (2016).
16. C. A. Rohl, *et al.*, Protein structure prediction using rosetta, in *Methods in enzymology*, eds. L. Brand, *et al.* (Academic Press, 2004) pp. 66–93.
17. C. Tokheim, *et al.*, Exome-Scale Discovery of Hotspot Mutation Regions in Human Cancer Using 3D Protein Structure, *Cancer Research* **76**, 3719 (2016).
18. M. Zitnik, *et al.*, Modeling polypharmacy side effects with graph convolutional networks, *Bioinformatics* **34**, i457 (2018).
19. Y. Qin, *et al.*, A multi-scale map of cell structure fusing protein images and interactions, *Nature* **600**, 536 (2021).
20. D. E. Kim, *et al.*, Protein structure prediction and analysis using the robetta server, *Nucleic acids research* **32**, W526 (2004).
21. M. Baek, *et al.*, Accurate prediction of protein structures and interactions using a three-track neural network, *Science* **373**, 871 (2021).
22. J. R. Karr, *et al.*, A whole-cell computational model predicts phenotype from genotype, *Cell* **150**, 389 (2012).

23. K. Lane, *et al.*, Measuring signaling and rna-seq in the same cell links gene expression to dynamic patterns of nf-κb activation, *Cell systems* **4**, 458 (2017).
24. Y. Cai, *et al.*, Experimental and computational framework for a dynamic protein atlas of human cell division, *Nature* **561**, 411 (2018).
25. S. Otsuka, *et al.*, A quantitative map of nuclear pore assembly reveals two distinct mechanisms, *Nature* **613**, 575 (2023).
26. N. Niknafs, *et al.*, Mupit interactive: webserver for mapping variant positions to annotated, interactive 3d structures, *Human genetics* **132**, 1235 (2013).
27. T. N. Turner, *et al.*, Genomic patterns of de novo mutation in simplex autism, *Cell* **171**, 710 (2017).
28. A. V. Weigel, *et al.*, ER-to-Golgi protein delivery through an interwoven, tubular network extending from ER, *Cell* **184**, 2412 (2021).
29. K. Huang, *et al.*, Therapeutics data commons: Machine learning datasets and tasks for drug discovery and development, *Proceedings of Neural Information Processing Systems, NeurIPS Datasets and Benchmarks* (2021).
30. K. Huang, *et al.*, Artificial intelligence foundation for therapeutic science, *Nature Chemical Biology* **18**, 1033 (2022).