A scoping study of the whole-cell imaging literature: a foundational corpus, potential for data-mining and research synthesis, and a call for standardization of an emerging field

Mary Mirvis[1], Brooke Weingard[1], Steven Goodman[2], Wallace F. Marshall[1]

[1]Department of Biochemistry & Biophysics, University of California, San Francisco, San Francisco, CA

[2]Department of Epidemiology and Population Health, Stanford University, Stanford, CA

ABSTRACT

The level of cellular organization bridging the mesoscale and whole-cell scale is coming into focus as a new frontier in cell biology. Great progress has been made in unraveling the complex physical and functional interconnectivity of organelles, but how the entire organelle network spatially arranges within the cytoplasm is only beginning to be explored. Drawing on cross-disciplinary research synthesis methods, we systematically curated the whole-cell volumetric imaging literature through 3 rounds of screening involving 3 independent reviewers, resulting in 89 top hits and 38 “borderline” studies. We describe the trajectory and current state of the field (2004-2024). A broad characterization, or “scoping review”, of bibliometrics, study design, and reporting practices shows accelerating technological development and research output. We find high variability in study design and reporting practices, including imaging modality, model organism, cellular contexts, organelles imaged, and analyses. Due to the laborious, low-throughput nature of most volumetric imaging methods, we find trends toward small sample sizes (<30 cells) and small cell types. We find common quantitative analyses across studies, including volumetric ratios of organelles and inter-organelle contact analyses. Our curated dataset now enables future aggregate and comparative analyses to potentially reveal larger patterns and generate more generalized hypotheses. This work establishes a growing dataset of whole-cell imaging literature and data, and motivates a call for standardized whole-cell imaging study design, reporting, and data sharing practices. More broadly, we showcase the potential of new rigorous secondary research methods to strengthen cell biology’s literature review and reproducibility toolkit, create a new avenues for discovery, and promote open research practices that support secondary data-reuse and integration.

INTRODUCTION

Biological organization spans many physical scales, from the atomic and molecular, to global ecosystems. In cell biology, the dominant relevant scale is that of the cell as the indivisible unit of life, where questions of cellular function, behavior, motion, communication, and shape dominate. At the subcellular scale of molecular biology and biochemistry, the structure, function, information flow of biological molecules, from ions to large macromolecular complexes such as nucleic acids and ribosomes reigns. Between these two scales lies the less clearly defined mesoscale, encompassing all subcellular structures larger than ~100 nm including condensates, cytoskeletal

assemblies, organelles, and the cytoplasm (Sear et al., 2015). Great progress has been made in identifying the relevant structural components of the mesoscale, but the driving forces and principles governing the complex spatial patterning, dynamics, interactions and systems properties of mesoscale entities is now a pressing challenge in cell biology. At this frontier of biological organization, new types of questions arise. How do well-understood molecular mechanisms scale *en masse* across the entirety of the cell? How do disparate mesoscale structures, all well known to be highly dynamic and structurally complex, interact with each other? To what extent are mesoscale structure and function governed by genetically programmed or energetically active mechanisms versus emergent from intrinsic mechanical interactions, self-organization(Laughlin et al., 2000)? Can the entire set of mesoscale components be understood as a whole interconnected system, and what conceptual frameworks and experimental and theoretical approaches are necessary for this understanding?

Answering such questions requires the generation of microscopy images of multiple mesoscale structures in tandem in the context of the entire cellular volume. Whole-cell imaging has been accomplished across a range of cell types using a variety of imaging modalities. Among image data types, whole-cell imaging is particularly technically challenging. Large volumes must be acquired and reconstructed, at resolutions high enough to accurately capture and segment multiple subcellular structures. Segmentation must be done either manually, which requires specialist expertise and many hours of labor per image, or automated, which requires sufficient amounts of high-quality training data and rigorous validation (Chen et al., 2024; Ekman et al., 2020; Erozan et al., 2024). Addressing questions about the variability and regulation of 3D mesoscale patterns is therefore generally beyond the scope of single whole-cell imaging studies. Here, we aimed to develop an approach to aggregate small whole-cell imaging studies to assemble a data- and knowledge-base capable of robustly addressing such large-scale questions.

Under the broad umbrella of secondary research, there exists a diverse and expanding ecosystem of methods to synthesize and review published work, yielding summaries of existing knowledge, generalized insights, and identification of trends and gaps in the research landscape. In the discipline of fundamental cell biology, the dominant forms of secondary research are narrative reviews, perspective articles, and books. These works act as authoritative landmarks for students and researchers at all stages, serving as starting points for newcomers to a field, presenting general state-of-the-field views, and pointing readers toward key open questions and critical further reading. One limitation of such publication types, however, is the lack of standard and clear methodology – they are typically produced from the extensive, but admittedly subjective, expertise of field leaders and groups. We rarely consider the *transparency* or *reproducibility* of such works – how were reviewed references chosen, and what was excluded? Would another author or team of authors come to similar conclusions?

A large family of systematic, rigorous, transparent, reproducible approaches to literature review collectively known as *evidence synthesis* or *research synthesis*, offers an alternative approach (Munn et al., 2018, 2023; Tricco et al., 2016). Although less well

known in the sphere of fundamental biology, evidence synthesis methods are widely used across many disciplines including the clinical and social sciences, carefully developed and supported by synthesis researchers and large international collaborations and initiatives such as Cochrane and the Joanna Briggs Institute. Nevertheless, there is a clear need for robust knowledge and data synthesis in biology, driven by widely recognized challenges associated with “data deluges” (“The Data Deluge,” 2012; Zhang, 2024) and the exponential growth of scientific output (Hanson et al., 2024; Pautasso, 2012). Further, the monumental goal of constructing meaningful holistic understandings of highly complex biological systems by merging a rich tradition of narrowly focused mechanistic studies and molecular parts lists, with interdisciplinary systems, biophysical, evolutionary, synthetic, and engineering perspectives (Rafelski & Theriot, 2024), requires robust frameworks to synthesize data, knowledge, and interpretations across disciplinary boundaries. The past two decades have yielded the independent development of synthesis and integration methods in some specific contexts, such as GWAS meta-analyses in genomics, computationally-assembled gene- and protein interaction networks in bioinformatics, and massive consortium collaborations to map cell types and beyond (“Meta-Analysis in Basic Biology,” 2016). Systematic research syntheses in fundamental cell biology, particularly intermediate-scale syntheses to address specific biological questions or hypotheses, are beginning to come on to the scene in recent years, still rare and methodologically disparate ((Gillooly et al., 2015; Opalek et al., 2023; Rutkowska et al., 2020; Sercel et al., 2023; Trumpff et al., 2021).

Developing a formal, standardized evidence synthesis toolkit for any new field requires special consideration of unique features of the field and efforts to adapt existing tools and approaches to those needs. Fundamental cell biology deals with highly complex living systems, yielding a landscape of feature-rich data, qualitative observations, and carefully controlled experimentation to identify meaningful relationships from noisy data. Conceptual frameworks are flexible and constantly shifting with the frequent introduction of new terminology, evolving definitions, and continuous debate over models, hypotheses, and attempts to draw meaningful distinctions among necessarily imperfect and unstable categorizations of nature. More fundamentally, data types, analysis methods, study design, practices for reporting and sharing of methods, data, and tools, and publication practices, are highly heterogeneous. This heterogeneity poses unique challenges to the findability, systematic aggregation, and meaningful synthesis of studies, data, and knowledge in cell biology.

Here, we undertake a synthesis of the whole-cell imaging literature. As a crucial first step to extracting novel cross-study insights from this field, we aim here to produce a thorough characterization of the volume, quality, and technical and biological features of the whole-cell imaging data landscape, in order to design more specific hypotheses and analyses based on the knowledge actually available for synthesis. Therefore, we describe a “scoping study”, based on “scoping review” methods previously described and standardized (Arksey & O’Malley, 2005; Tricco et al., 2018). We present the development of our scoping methodology, including iterative search strategy development and screening, and delineate the contours of the whole-cell imaging field.

We summarize scientometric trends and the range of methodologies, reporting practices, and biological conditions included in the existing, but rapidly growing, set of whole-cell imaging data. Our work produces a foundational corpus of whole-cell imaging data with rich potential for comprehensive knowledge synthesis, novel biological insights, and a bird's eye view of the changing trends and needs of this promising and rapidly-crystallizing sub-field of mesoscale cell biology.

METHODS

We adapted our approach from the well-defined scoping review methodology (Arksey & O'Malley, 2005; Peters et al., 2022). Our workflow for this study is summarized in Figure 1.

Search strategy:

We developed our search strategy starting with an initial set of 11 target papers (Decelle J et al., 2021; Heinrich L et al., 2021; Henderson et al., 2007; Laundon et al., 2019; Müller et al., 2021; Uchida et al., 2009, 2011; Uwizeye et al., 2021a; Valm et al., 2017; Wei et al., 2012; White et al., 2020)], thoroughly combed for cell attribute- (e.g. "cell organization", "subcellular architecture"), mesoscale-, and imaging-related terms. These terms were used to devise an iterative series of search strings for PubMed aimed at retrieving a defined corpus that includes the full initial set. While an initial attempt focused on cell anatomy-related terms was unsuccessful due to the heterogeneity and inconsistent usage of such conceptual language, iterations on search strategies consisting of 3D imaging and mesoscale (cell/subcellular and organelle-scale) terms successfully retrieved the initial papers.

> (("Imaging, Three-Dimensional"[MH] OR "3D imaging"[tiab] OR "three-dimensional imaging"[tiab] OR "3-D imaging"[tiab] OR "Soft X-ray tomography"[tiab] OR (3-D[tiab] OR 3D[tiab] OR "three dimensional"[tiab] OR "three-dimensional"[tiab]) AND (tomography[tiab] OR microscopy[tiab] OR Microscopy[MH] OR Tomography[MH] OR imaging[tiab]))
>
> AND
>
> ((ultrastructure[MeSH Subheading]) OR (Organelles[mh] OR organell*[tiab]) OR (intracell*[tiab] OR intra-cell*[tiab]) OR (subcell*[tiab] OR sub-cell*[tiab]) OR (((whole[tiab] OR entire[tiab]) AND cell[tiab]) OR "whole-cell"[tiab] or "whole cell" OR "entire cell"[tiab])))
> NOT Review[pt]

We set our search to begin at 2005 due to the emergence of technological capability to perform whole-cell microscopy and reconstructions around this time (Le Gros et al., 2005). However, this missed one key earlier study, which was added *post hoc* (Larabell & Le Gros, 2004). The most recent PubMed search update was performed on May 23, 2024.

Screening

Our search strategy returned over 11000 results from PubMed. Because our search was aimed at a particular data type (whole-cell volumetric microscopy) rather than a specific research topic, our target corpus is highly heterogeneous, spanning a wide range of publication venues, key terminology, and other bibliometric features. Therefore, our primary objective was to retrieve our initial set as well as any studies with even a marginal potential of containing whole-cell microscopy data, yielding a highly sensitive but not specific search.

To curate our dataset, we performed three rounds of screening according to the following strictly-defined inclusion criteria:

1. Cell anatomy is necessarily a whole-cell property. The study must have collected original three-dimensional *whole-cell* images in eukaryotic systems, described in the primary study in the form of direct visualizations (still images and/or supplementary movies displaying volumetric rotations), and/or quantitative analyses necessarily taken from whole-cell data (e.g. organelle-to-cell volume ratios). “Whole-cell” requires that the entire cellular boundary (cell membrane or cell wall, or other marker showing the continuous cell periphery) is captured within the image, such that the cellular volume can be calculated.
2. Volumetric information for at least two *organelles* concurrently must be included in the whole-cell volume. We reasoned that at least two subcellular structures must be visible in relation to one another for the cellular property of anatomy to be applicable. This criterion requires a full volumetric reconstruction of the multi-organelle cell, with organelle identities clearly annotated (typically in the form of a segmentation). The reconstructed data should either be directly shown as an image in figures, movie files, or associated data, or necessarily generated to produce published analysis, such as volume calculations. Although the usage of the term “organelle” has expanded in recent years, we used a traditional definition, referring to a structurally and functionally distinct, membrane-bound, intracellular compartment.

Curating a set of target papers from this large set of search results required a series of increasingly thorough screening steps – title and PubMed summary, abstract, and full text – conducted by a team of three reviewers. Search results and screening decisions for each stage, including exclusion reasons, along with information extracted from initial target papers are provided in Supporting File 1.

Title/summary. The first screening step was a single-reviewer rapid screen through search result pages in the web browser (Mozilla Firefox) of titles and short three-line PubMed-generated summary statements. Studies that were obviously irrelevant, focused on the wrong biological scale, or purely technical with no biological data were excluded. Studies that clearly described whole-cell imaging and/or structures of entire cells were included. If unclear, a skim of the abstract (and PMC figures, if available) was done, and studies that were still ambiguous were included for further screening. This step was performed by one reviewer (MM), and excluded 95% of results.

Abstract. A custom Python (v3) script using Bio.Entrez module of Biopython (Cock et al., 2009) was used to pull abstracts from PubMed e-utilities (*Entrez Programming Utilities Help*, 2010) and record screening decisions from one reviewer. Studies were either included, excluded with exclusion reasoning recorded, or marked for full text screening. A 10% subset was screened in parallel by a second reviewer (WM) to check inter-rater consistency. A low level of consistency was initially found due to differing definitions of "organelle". This led to a refinement of our working definition of "organelle" for the purpose of screening criteria (see above), and the full set of abstract screening was performed in parallel by two independent reviewers (MM & BW). This step yielded 60 clear hits (10.8%), excluded 184 (32.3%), and sequestered 326 studies for full text review.

Full text screening. Full texts, including supplementary data, were screened in parallel by two reviewers (MM & BW). Screening decisions were recorded in Covidence software (Veritas Health Innovation, n.d.), along with reasons for exclusion. Exclusion reasons were categorized as: missing 3D data, including cases where 3D data is textually described but not included as data in the paper (“Not 3D”); imaging only at subcellular or multicellular/tissue scales (“Wrong scale”); cells only partially imaged and/or reconstructed, or with cut-off edges such that cell volume cannot be calculated (“Not whole-cell”); raw data missing organelle segmentation (“Not segmentation”); fewer than 2 organelles segmented (“Not multi-organelle”); data reused from other studies (“No original data”); review or perspective articles (“Wrong article type”); and a complete lack of whole-cell, multi-organelle reconstruction data (“No relevant data”).

We also designated 39 studies as "Borderlines" if they represented edge cases with regards to our inclusion criteria, but nevertheless contained relevant data, analysis, or insights, therefore not warranting complete exclusion from our dataset, summarized in Source Data/Supplementary Table 1. One borderline category is “whole-cell data not shown”, including studies in which the target data type is implied or explicitly described in the text but not included in figures or associated data files, or provided in supplement where files are corrupted/inaccessible. This includes several studies that partially reconstructed prohibitively large cell types (Trebichalská et al., 2021), as well as reconstructed tissue volumes from which while relevant metrics are quantified, such as organelle volume/volumetric ratios, calculated across the tissue block rather than in individual cells, as distinct cell boundaries could not be delineated. In some cases, the whole-cell images had slightly clipped cell edges due to nuances such as cells in cultures spreading their edges and protrusions far from the main cell body on a culture substrate, or an elongated cell type such as myocytes. We chose to exclude such examples, but include others such as images of protists with flagella partially cut-off, due to the fact a missing edge of the cell body leaves key morphometrics, such as cell volume and shape, impossible to calculate, and uncertainties about holistic cell anatomy due missing information. Other borderline categories include: organelle identities uncertain; ineligible structures defined as "organelles"; a very minor volume of the cell edge is cut off in the image, technically violating the “whole cell” criterion; cell membrane not clearly delineated; the paper itself does not include the relevant data, but

accompanies the establishment of a public repository or resource containing relevant data.

Disagreements in inclusion/exclusion decisions were discussed between reviewers, prompting slight adjustments to our protocol, such as establishing new exclusion reasons or borderline categories. One earlier study was added *post hoc* (Larabell & Le Gros 2004). Ultimately, 89 studies were fully included.

Data Extraction and Analysis

We extracted information from studies to summarize study design, including imaging modality and resolution, cell types, sample sizes, organelles, comparative conditions, and quantitative analyses, only as pertains to information within the study directly describing or derived from whole-cell imaging data, as opposed to other types of data within the same paper. Data extraction was performed manually by two researchers independently (MM & BW), and subsequently discussed observations to reach consensus.

Data items were annotated on full text PDFs in Zotero (*Zotero | Your Personal Research Assistant*, n.d.) and entered into the data extraction template in Covidence (Veritas Health Innovation, n.d.), followed by inter-rater consensus discussions to resolve disagreements and inconsistencies. Extracted data were exported from Covidence as .csv, and manually processed and formatted in Excel. Full set of final data extraction items, including 89 included studies and 39 borderline studies, is provided in Supplementary/Source Data Table 2. Co-authorship networks were generated in VOSViewer (van Eck & Waltman, 2010) using PubMed files as input and the following settings: bibliographic data, all co-authorship, include documents with a large number of authors, allow unconnected items, and overlay visualization. Citation network was created in Litmaps (*Litmaps (Version 2025-01-16) [Search Tool].*, 2024) using the following settings: X axis: Publication Date, compact; Y axis: Map Connectivity; Size: Cite Count, linear interpolation; disable Position Editing, Show citation arrows, Avoid collisions. All other data analysis and plotting was performed in Python using Jupyter notebooks and the following packages: Matplotlib (Hunter, 2007), SciPy (Virtanen et al., 2020), Pandas (The pandas development team, 2020), Numpy (Harris et al., 2020), Seaborn (Waskom, 2021). Code and source data files are available at https://github.com/bdwucsf/Cell-Anatomy-Scoping-Review.

RESULTS

Metadata and bibliometrics.

We first aimed to broadly characterize the publishing trends within our corpus. The publication rate of studies containing whole-cell imaging data has risen steadily since 2005, with a peak at 14 papers in 2023 (the most recent full year, as our search stopped in May 2024), and 49.4% of the included studies (44/89) published in the last 5 years (2019-2024) (Figure 2A). The majority are open-access publications, which is important for study amenability to systematic search and screening approaches like ours (70/89, 78.7%) (Figure 2B). The majority (57, 64%) were "research" articles, focused on

answering a biological question, while 31 articles were considered "technical", focused on method development, although some biological observations may be included (Figure 2C). One publication is a technical book chapter, which would have been otherwise excluded from our dataset as a secondary source, but included some original image data (Gluenz E et al., 2015). We asked whether papers made the original imaging data available via data repositories or other open-access platforms, which would enable secondary analysis of this information-rich data. Across the core included datasets and borderlines, only 11 included studies (12.4%) and 1 borderline study reported that underlying whole-cell image datasets were publicly available, although those that did were more recently published, from 2015 onwards (Figure 2D.). A summary of publicly available datasets is provided in Supplementary Table 3. In conclusion, we see a progressive expansion of study and data volume, as well as open science practices, with a pronounced increase in all of these trends in the past 5 years, suggesting good current and future potential for meta-scientific studies and secondary re-analysis of whole-cell imaging datasets.

We asked whether the corpus is scientifically interconnected, showing a substantial degree of scientific collaboration and scholarly acknowledgement. Analysis of co-authorship networks shows many tightly-linked co-authorship clusters (Fig. S1A), and one large cluster of collaborators among prolific whole-cell microscopy researchers (authors on 2+ studies) (Fig S1B). A citation network shows that most (71/89), but not all studies in the corpus cite each other. A few studies (Larabell & Le Gros, 2004; Loconte et al., 2021; Uchida et al., 2009, 2011; Uwizeye et al., 2021b) emerged as the central, most highly connected studies. The most cited study (Valm et al., 2017), with over 1000 citations, does not cite previous studies in this corpus, but is then cited along with other studies in the corpus by subsequent publications, representing a convergence of whole-cell imaging efforts from disparate research streams. This bibliometric overview illustrates the growing pace and cohesion of the whole-cell imaging literature.

Imaging modality and resolution

In total, we identified 13 distinct volumetric imaging types of modalities used to produce whole-cell image data. Focused ion beam SEM (FIB-SEM) (also called ion-abrasion SEM, IA-SEM) and soft X-ray tomography (SXT) and are the most common modalities, represented in 22 and 19 studies, respectively (Figure 3A). These were followed by serial block-face SEM (SBF-SEM), serial sectioning TEM (ssTEM), electron tomography (ET), fluorescence (including 3D SIM, and multispectral lattice light sheet), optical tomography (including optical diffraction and optical computed tomography/live cell CT), hard X-ray tomography (HXT, including X-ray ptychography), scanning TEM (STEM), and X-ray diffraction (XRD) techniques. Field emission SEM (FE-SEM), scanning transmission X-ray microscopy (STXM), and SEM array tomography (SEM-AT) were each found in one study.

Each modality excels within a limited range of spatial resolution, determining the size scale, volumetric completeness, throughput, and potential for anisotropic artifacts in the resulting images. Generally, optical imaging methods are limited by the wavelength of

light (~200nm), typically capable of achieving higher lateral (xy) than axial (z) resolution, but can be used to produce sample sizes in the hundreds or thousands from manual experiments. Electron microscopy modalities are capable of achieving resolution in the sub-10nm range, but are typically limited in volumetric scope and statistical throughput due to the high information density and cumbersome sample preparation, imaging, and image processing. X-ray based methodologies typically fall somewhere in between light and electron microscopy in throughput and resolution (Guo & Larabell, 2019; McDermott et al., 2012). The range of cell types potentially accessible for whole-cell reconstruction, information content contained in the resulting images, comparability across datasets and modalities, and capacity for secondary analysis is strongly dependent on image resolution. Whole-cell imaging requires the acquisition of large volumes, so even the best resolution possible with a particular modality or microscopy may need to be compromised to increase the reconstruction volume and/or throughput.

The reported resolution for each study is plotted in Fig. 3C, including separate datapoint for multiple datasets with different resolutions from the same study, and summarized as lateral (L) and axial (A) resolutions for electron microscopy (EM), X-ray microscopy (X-ray), and light microscopy (LM) methods in Figure 3B. As expected, the highest resolution was achieved by electron microscopy modalities (ET, FIB-SEM, SBF-SEM, ssTEM, STEM, FE-SEM, AT, followed by X-ray based methodologies (SXT, HXT, X-ray diffraction, STXM), and the lowest resolution in optical data (fluorescence microscopy, optical tomography). Isotropic resolution, which is highly advantageous for structural accuracy and morphological analyses (Chen et al., 2024; Le Gros et al., 2005), was far more common in X-ray based techniques (20/22 studies, 90.1%), than in EM-based (17/43, 39.5%) or light-based (1/9, 11.1%) modalities.

Imaging resolution is therefore unique for each study, chosen within the available range to optimize feasibility and information density for the particular question or goal, and is a critical methodological parameter in methods reporting for critical interpretation of results, reproducibility, and evidence synthesis. In extracting the reported resolution from each study, we found that imaging resolution was not clearly reported in 14 studies, in some cases inconsistently stated (conflicting and/or ambiguous mentions throughout the text), while in others, resolution was incompletely reported (e.g. axial (z), but not lateral (xy), resolution clearly stated, Figure 3C, right panel).

Image datasets: Cell type and sample size

A wide range of cell types are covered in the corpus, including yeast, protists, plants, invertebrate animals, and both cultured and *in situ* mammalian cell types from rodents, non-human primates, and humans (Fig 4A). In total, we found 83 distinct cell types across 121 datasets, where a dataset is considered a distinct set of cells of the same type imaged with the same modality, and multiple datasets may appear within one study (e.g. four datasets comprising two cell types (*C. pepo, N. tabacum*) each imaged by two modalities each (ssTEM, FIB-SEM) in Zechmann *et al.*, 2022). The most strongly-represented cell types (in terms number of separate datasets and number of cells imaged in total across datasets, which we call “data density”), are highlighted in Table 1.

Table 1. Cell types with highest data density.

Cell type	# of studies	Total # cells
S. cerevisiae	11 (13 datasets)	380
P. falciparum	5 (6 datasets)	84
C. reinhardtii	4	30
T. Cruzi	4	4
Rat insulinoma pancreatic β cell (INS-1E)	3	228
Mouse pancreatic β cell	3	16
S. pombe	3	8
human platelets	2	94
Trypanosoma brucei	2	80
Candida utilis	2	42
human B-lymphocytes	2	18
COS-7	2	12
Toxoplasma gondii	2	5
HEK293T	2	5

Most studies had small sample sizes, with 88 (72.7%) of datasets reporting n<=10 whole cell images, and 38 (31.4%) studies consisting of a single cell (Figure 4B). Single-cell datasets were more common in earlier papers, and frequently occurred in technical papers – 15 of the 89 studies were technical papers reporting a single whole-cell reconstruction, representing nearly half (48.4%) of the technical studies and 75% of the single-cell studies. This suggests that the first ten or so years of our dataset (~2005-2015) could be described as technical early days, but the advances during this time paved the way for more biologically-focused research studies with larger datasets in the following decade (~2015-today). Sample sizes were largest in SXT (up to n=200) and SBF-SEM (up to n=141) studies (range). In 14 of the 121 datasets (11.6%), the study sample size, specifically the number of cells fully reconstructed in 3D with multiple organelle volumetric information compatible with our inclusion criteria, was either reported inconsistently or incompletely, in which case we reported a minimum known sample size, or else was not reported at all, resulting in missing data in our analysis (Fig 4A, gray bubbles).

Organelles and anatomical completeness

Our primary inclusion criterion was the presence of at least two organelles within the same whole-cell reconstruction. Figure 5A shows the cumulative organelle data per cell type across the corpus. The nucleus and mitochondria are the most commonly segmented organelles across all studies (78 and 76 instances respectively, accounting for multiple distinct cell types per paper in some cases), followed by vacuoles/lysosomes (41 instances) endoplasmic reticulum (27 instances), Golgi (23 instances), chloroplast/plastid (21 instances), lipid droplets (20 instances), vesicles (12 instances), flagella/cilia (9 instances), peroxisome (5 instances), centrosome/basal body (5 instances), multivesicular bodies (2 instances), and autophagosome (1 instance). Some organisms may appear to have only one or no organelles included. This is due to

the fact that the heatmap only includes “common organelles” present in all or most eukaryotes. The blank lines represent studies which image and reconstruct only specialized or cell type-specific organelles, such as the specialized secretory organelles (microneme, rhoptry, dense granules) or conoid structures in *Toxoplasma gondii* (Koutsogiannis Z *et al*., 2023; Paredes-Santos TC *et al*., 2012). In some cases, we find one common organelle in combination with specialized organelles, such as the study in C57BL/6 resting mouse platelets (mitochondria, dense granules, alpha-granules, canalicular system) (Pokrovskaya I *et al.*, 2021). The recurrence of cell types in multiple studies enables potential examination of organelle properties across datasets, experimental designs, and biological contexts. While not all studies imaging the same cell type segment the same set of organelles, the most frequently-imaged specific organelles are found among the best-represented cell types: *S. cerevisiae* vacuoles (9 instances), nuclei (8), mitochondria (9), and lipid droplets (5); *P. falciparum* nuclei (5) and mitochondria (4), *T. cruzi* nuclei (4), *S. pombe* vacuoles (4).

The rationale for requiring at least two organelles per segmented cell in the dataset is 1) the supposition that a single internal landmark is insufficient to constitute an anatomical pattern, and 2) to reveal spatial interactions between organelles in the 3D context of the entire cell. Overall, we found 58 pairwise organelle combinations occurring across the corpus, of which 32 had at least five independent instances (of the 121 datasets), suggesting rich potential for comparison and synthesis of each organelle-organelle relationship across many cell types and contexts (Figure 5B). The most common organelle co-occurrences are mitochondria-nucleus (76 instances across 61 cell types), nucleus-vacuole/lysosome (42 instances across 30 cell types), and mitochondria-vacuole/lysosome (35 instances across 26 cell types). Of these combinations, 20 were found recurring in different studies in the same cell type. These cases will enable within-cell type synthesis of similar data, as well as cross-study comparison, which can potentially isolate biological context- and technical factor-dependent differences.

The ultimate requirement for an understanding of cell anatomy and modeling whole-cell patterns and inter-organelle interaction networks is to be able to visualize as many organelles simultaneously as possible. The number of “common organelles” per reconstruction, termed “anatomical completeness” is distributed with a peak at 3-4 organelles (Figure 5C). The most anatomically complete reconstructions are parasitic protist *Leishmania Mexicana* with 8 organelles (Hair et al., 2024), choanoflagellate *S. rosetta* and sponge *O. carmella* choanocyte with 7 organelles (Laundon et al., 2019).

Comparators

While whole-cell structural data is largely static, cellular structure is in fact highly dynamic and variable between cell types, cellular states and functions, and genetic and internal perturbations. Most studies (74, 83%) included comparable data across at least one cellular perturbation or condition, which we grouped into broad classes (Figure 6A). The most common of these are cell cycle stages (15 studies) and metabolic conditions, such as glucose stimulation of pancreatic cells (14 studies). Other comparators, in decreasing order of frequency, were cross-cell type comparisons (n=11), stressors such as drug treatments (n=9), methodology, such as fixation methods or imaging modalities

(n=7), stages of life cycle or differentiation (n=6), environmental conditions, such as light exposure (n=5), morphology, such as shape or size (n=3) , disease or infection states (n=2) , functional or behavioral states, such as cell-cell communication (n=1), and mutations (n=1).

Morphological metrics and Quantitative analyses
The unique advantage of high-resolution 3D structural data of whole cells is their capacity to provide rich and precise quantitative morphological information. An assessment of quantitative analyses performed across studies revealed a limited set of organelle geometric features measured (Figure 6B). Volume was the most frequently measured property (66 studies, 74%), followed by volume fraction (organelle-to-cell volume ratio, 51 studies, 57.3%). Also common were surface area (30 studies, 33.7%), count/copy number and linear dimensions such as length and diameter (29 studies each, 32.6%). Less frequent, and more variable, were measures of density such as LAC or electron density (n=18, 20.2%), positioning/distance and organelle contacts/associations (n=17 each, 19.1%), shape (n=14, 15.7%), and organelle volume scaling trends (n=7, 7.9%). Three studies measured refractive index, and one study devised a method to quantify subcellular crowding of organelles ("congestion index", Loconte V *et al*., 2022)). These frequencies roughly reflect the general ease of measurement and conceptual familiarity of these metrics - volume and surface area are generally trivial to measure, albeit with variable accuracy, with many standard software tools automatically generating these calculations. Count and dimensions can be straightforwardly manually or semi-automatically calculated. Distance and contact frequency, shape, and crowding require more complex calculations and computational operations, and can be interpreted in variable ways. A sizable fraction of studies (n=14, 15.7%) performed no quantitative analysis.

In summary, the whole-cell multi-organelle imaging corpus includes 89 published studies between 2004 and 2024 from over 11,000 initial PubMed results. In total, the dataset represents the whole-cell imaging and reconstruction of at least 1689 individual cells, of 83 eukaryotic cell types from protists to humans, across 121 independent datasets, 13 imaging modalities, at least 13 common organelles in many different combinations, 71 perturbative or comparative conditions, and at least 12 types of morphological metrics quantitatively analyzed.

LIMITATIONS

Our scoping study is, to our knowledge, one of the first applications of scoping review methodology to the field of cell microscopy and fundamental cell biology more broadly. Our method comes with limitations arising from the iterative nature of proof-of-concept methods development, the breadth and heterogeneity of the literature corpus, and the size of our team. One of the goals of a formal systematic literature synthesis is to be as comprehensive as possible, spanning peer-reviewed and grey literature sources as thoroughly as possible. Our approach focused specifically on PubMed, but ideally, other databases such as Web of Science and EMBASE, would be included as well. Although our PubMed search was extremely broad relative to our final inclusion set, which

comprised only less than 1% of the original search results, we cannot be sure how many relevant studies might have been missed by the search entirely. Furthermore, grey literature sources such as dissertations and online image data hosting sites without corresponding publications would be of relevance to a comprehensive set of cell anatomy data, but would be missed by our strategy. For example, the Allen Institute for Cell Science hosts a collection of 3D confocal images in human iPSCs, a subset of which would fit our inclusion criteria, but these datasets have not been directly thoroughly described in a publication, despite their inclusion within a large dataset underlying many impactful publications (*Cell Line Catalog*, 2024). Furthermore, as of January 2025, our search string retrieves 300 new results since our last round of curation in May 2024. The dataset and analysis presented here should be seen only as an initial characterization, which will be continually updated as the literature base continues to expand at an accelerating rate. We intend to establish our growing curated corpus and analysis as a publicly available, living database of cell anatomy data and field trends, inviting community input to identify gaps and curate the corpus as thoroughly as possible.

Another limitation of our approach, as of any systematic research synthesis, is that the choice of strict inclusion criteria necessarily excludes potentially valuable studies. These boundary cases help to define the field of whole-cell reconstructions and reveal peripheral sources of knowledge that can be incorporated into follow-up studies depending on the specific biological question. One core criterion was the need to define "*whole cell*" strictly, requiring that the full cell periphery is clearly delineated. The rationale for this is that, first, the full anatomy of the cell requires that no part of the cell is missing from the image, and second, the intact cell boundary is needed for accurate measurement of cell volume, which we consider to be the critical common measurement required for any cross-study quantitative analysis as the denominator for organelle volume ratio. Therefore, many studies reporting mostly complete cell volumes, in some cases even describing the data as "whole cell" despite missing small cell edges, had to be excluded. This was often the case for adherent cultured cells, which spread out flat on culture dishes with thin, faint, irregular protrusions when sparsely plated, and for very large, irregularly shaped cell types such as muscle cells and neurons, which are extremely difficult to perfectly capture in their entirety *in situ* or in culture. In other cases, we found studies of tissue sections containing partial reconstructions of multiple cells clipped at the section boundary, including organelle volumes and volume ratios pooled and normalized across the entire section volume, but no single cell fully reconstructed. Therefore, our strict "whole cell" criterion introduces a cell type bias into our corpus. Moving forward, a case could be justified for softening this criterion to estimate rather than directly calculate cell volume, and the systematic nature of our screening process means that steps can be back-tracked to retrieve specific studies previously excluded for either of these reasons.

DISCUSSION

We have described a novel application of formal scoping review methodology to comprehensively curate the whole-cell microscopy literature, yielding a corpus spanning

a wide breadth of imaging modalities, model systems and cell types, organelles visualized, and conditions compared. There is substantial comparability across studies, highlighting opportunities for further meta-scientific and synthetic studies building on this corpus. Future directions include systematic review and meta-analysis approaches to understand how mesoscale structures such as organelles typically self-organize and interact within the whole cell space across biological contexts, including qualitative categorizations of cell anatomy (the aggregate appearance of a cell type cumulating from the holistically morphological properties and interactions of all organelles), and meta-analyses of key morphometric measurements.

Although our search is not exhaustive, this corpus comprises a foundational evidence base revealing an emerging field of significant, and rapidly increasing, volume and scholarly interconnectedness, unified by a specific data type – whole cell, multi-organelle 3D reconstruction (Figures. 1 and S1). Our systematic curation and meta-scientific characterization of these remarkably information-dense structural data, from raw data to processed images to derived measurements, enables future secondary analysis and synthesis efforts to unlock their full potential. The diversity of methods, model systems, and biological themes illustrate an opportunity to establish community discourse around best practices and standards to grow this potential further as output accelerates.

In aggregate, our corpus contains a unique set of qualitative and quantitative information into common and context-specific patterns of cell anatomy. The substantial methodological and biological overlap and comparability across studies opens a wide range of possibilities for within- and cross-group synthetic studies to uncover novel technical and biological insights that would otherwise be difficult to access through traditional primary research methods. Measurements, such as organelle volume ratios and inter-organelle interaction frequencies, can be assessed across many cell types, biological contexts, and technical parameters. Data in a specific cell type pooled across disparate datasets, modalities, and conditions provide opportunities to query the consistency of similar metrics across studies, integrate disparate findings into an organism-centric model, and compare distinct imaging modalities and parameters applied to the same experimental system. Summarizing quantitative analyses and extending them where possible allows for a more comprehensive quantitative understanding of generalizable principles of organization, an assessment of single organelle, between-organelle, and cell-wide feature variability across contexts, and provide a strong empirical foundation for integrative modeling efforts (Loconte et al., 2023). Publicly shared original data enable novel analyses and the development of new data integration and secondary analysis workflows. We intend for our dataset and analysis to become a growing tool for the scientific community interested in particular cell types, organelles and organelle interactions, biological perturbations, imaging and image analysis approaches, and beyond.

The range of imaging modalities and technical parameters represented across our corpus demonstrates the versatility of approaches available to achieve whole-cell anatomy data. It also provides opportunities to directly compare these approaches and

trends in how they are applied, identify common challenges, and motivate recommendations for future work. We find that the current whole-cell imaging landscape, on the whole, is biased toward small cell types and small sample sizes, with a single cell reconstruction reported in nearly one-third of studies. This is largely due to the time-, cost-, and labor-intensive nature of whole-cell image acquisition, reconstruction and multi-organelle 3D segmentation, particularly at sub-100nm resolution. The 13 imaging modalities represented each offer particular advantages and disadvantages, with unique tradeoffs between resolution, throughput, native structure preservation, and overall tractability depending on the application. For example, the two dominant modalities, SXT and FIB-SEM, are both capable of achieving high isotropic resolution with some degree of quantitative throughput. SXT generally nearly always achieves isotropic 30-60nm resolution with minimal impact on ultrastructure or tomographic artifacts, and is capable of producing throughput in the hundreds within a single study. However, cell size is strictly capped by ~10um capillary diameter, limiting the cell types one can study by this method, and access is limited by the requirement for beamline time. FIB-SEM excels for very high resolution reconstructions (~5 nm), but is limited by throughput and, as a sample-destructive method, is more likely to introduce ultrastructural artifacts. Extensive discussions of strengths and weaknesses across volumetric imaging modalities have been discussed in depth elsewhere (Hanssen et al., 2012; Loconte et al., 2021; Loconte V et al., 2022), and technical capabilities for all modalities are constantly being improved.

Our scientometric analysis demonstrates that the publication of whole-cell imaging studies is accelerating, with growing trends toward open-access publication and data sharing], suggesting good current and future potential for meta-scientific studies and secondary re-analysis of datasets. However, only a small subset of our corpus made original data publicly available, and we found inconsistent reporting of key methodological details, such as imaging resolution and sample size. Such metadata items are crucial for meta-scientific analysis, including secondary analysis and synthesis research as well as within-study reproducibility and cross-study replicability assessment. Currently, there is increasing momentum in the development of open science and data sharing guidelines and implementation led by international collaborations of scientists (FAIR, DORA), funding bodies (NIH, NSF), and publishers. Many data type-specific metadata standards have been developed for bioimage data by community consensus (Global BioImaging, REMBI, foundingGIDE, Open Microscopy Environment) and the growth of image dataset repositories (EMBD, EMPIAR, IDR, SSBD, NIH Imaging Data Commons). Image data, particularly high-content volumetric imaging such as whole-cell reconstruction datasets, present particular challenges for data sharing, such as the large size and complexity of files and datasets, irregular file formats. Ongoing efforts to develop novel bioimage file formats (OME-Zarr (Moore et al., 2023); OME-NGFF (Moore et al., 2021)) and accessibility tools (BioFileFinder) promise to help tackle some of these barriers, and lead to new solutions as the public image data ecosystem scales up. There may be factors of particular interest to the emerging scientific community generating and using whole-cell imaging data with regards to specific imaging modalities and data types, such as segmentation and 3D reconstruction parameters. Particularly as the publication of whole-cell imaging studies

accelerates, we envision that community discourse and establishment of specific standards for methodological design, reporting, and data sharing could be of broad benefit. We hope to see a coordinated effort to develop open science practices to maximize the utility of these laboriously-generated data.

The present work illustrates the unique considerations, challenges, and scientific and meta-scientific potential of developing an evidence synthesis toolkit tailored to fundamental cell biology. In contrast to translational and applied disciplines, particularly those based on animal or human subjects, exploratory fields such as fundamental cell biology marked by high heterogeneity in their technical repertoire, study design, experimental and reporting practices, philosophical and sociological standards for epistemology and argument construction, and conceptual and linguistic frameworks. This heterogeneity presents unique and significant challenges for the application of highly standardized, guideline-driven evidence synthesis methods. For example, development of a narrow and specific bibliographic database search strategy is particularly challenging. Firstly, deep and broad technical expertise is required to understand the scope of research likely to contain a particular data type, specific experiment, or conceptual theme. Second, highly varied terminology used across the literature, and particularly in the lack of concrete norms for titles and abstracts results in a tendency toward very large search result corpuses, preferring comprehensiveness over specificity. The search scope and resulting corpus may therefore be larger than is the accepted norm in other disciplines. Therefore, inclusion criteria must be carefully justified, and perhaps iteratively developed, and robust stepwise methods for screening and data extraction, ideally carried out by teams of 2-4 expert investigators, are needed.

The prospect of employing AI tools to aid in these intensive steps is attractive, and being actively explored by many scientific groups and software tool companies (Bolaños et al., 2024; Yip et al., 2024). However, due to the goal of maximum comprehensiveness and need for a careful balance between rigor and pragmatic flexibility to handle the highly heterogeneous and nuanced nature of the fundamental biology literature, the requirement for human judgement and validation is unlikely to be replaced by AI in the immediate future. We could envision a manually curated scoping study, such as this one, forming the foundation for a novel synthesis effort, and in tandem, creating a robust training dataset for an AI synthesis assistant, while further refinement, establishment of a living corpus, and updating with novel publications as they are released, can be accomplished with the aid of the trained AI model and community curation and validation. We look forward to a bright future in which specialized synthesis researchers, in collaboration with the expert community and automated tools, help to create a thriving ecosystem for novel discovery and meta-research to derive the maximal benefit from the immense volume and richness of biological literature and data.

Acknowledgements
This project was funded by the UCSF Program for Breakthrough Biomedical Research, funded in part by the Sandler Foundation, by NIH grant R35 GM130327, and by the Center for Cellular Construction, funded by NSF grant DBI-1548297. We owe a debt of

gratitude to the UCSF librarians, Peggy Tahir, Josephine Tan, and Evans Whitaker, for evidence synthesis methods consultations and assistance with developing our search strategy. We thank Dr. Jane Maienschein, Dr. Susanne Rafelski, Dr. Sabina Leonelli, Dr. Matthew Akamatsu, Dr. Shane Jinson, Dr. Sam Church, and members of the Marshall lab for constructive feedback and support.

FIGURES

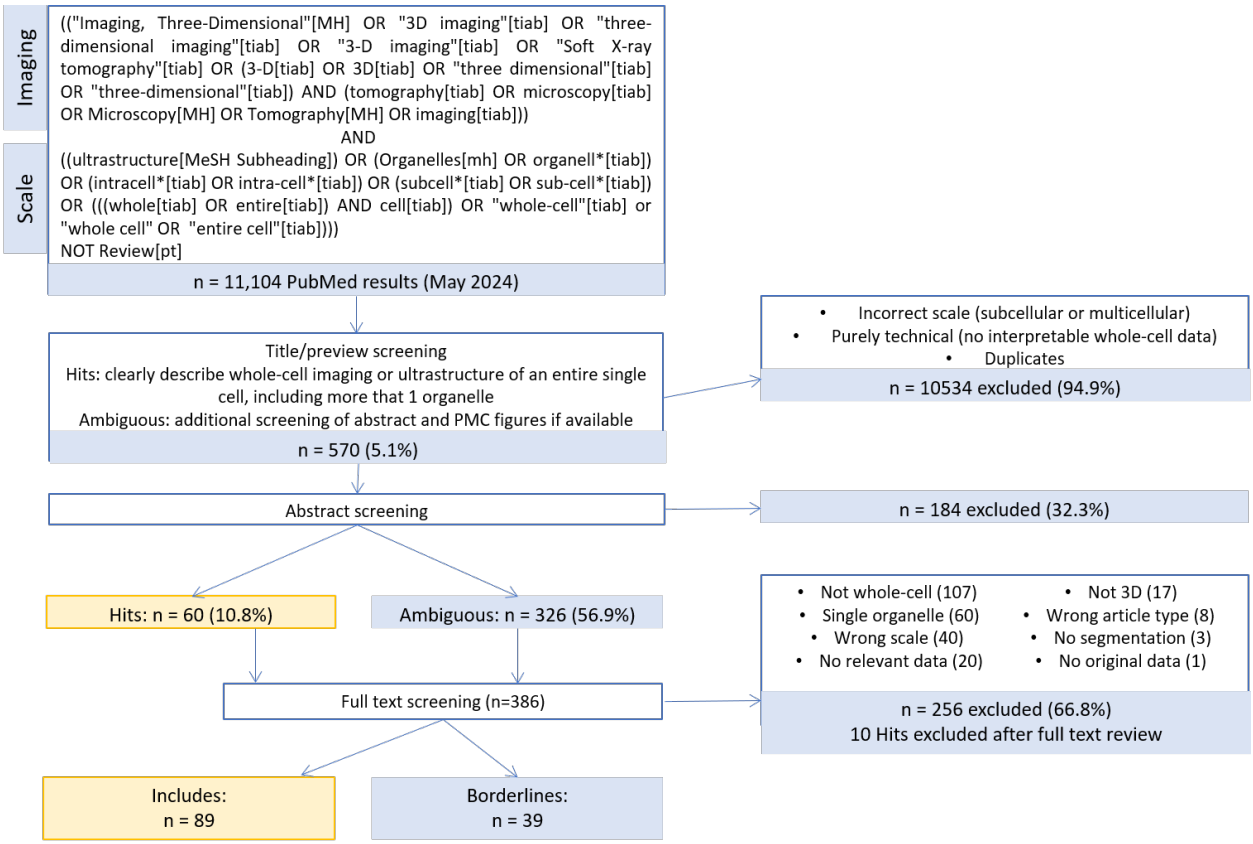

Figure 1. Study workflow to curate whole-cell imaging literature. PRISMA diagram detailing the PubMed search strategy and screening steps used, with included and excluded study counts at each stage. In search strategy, asterisk (*) indicates wildcard character, [MH] indicates Mesh header, [tiab] indicates title/abstract field, [pt] indicates publication type.

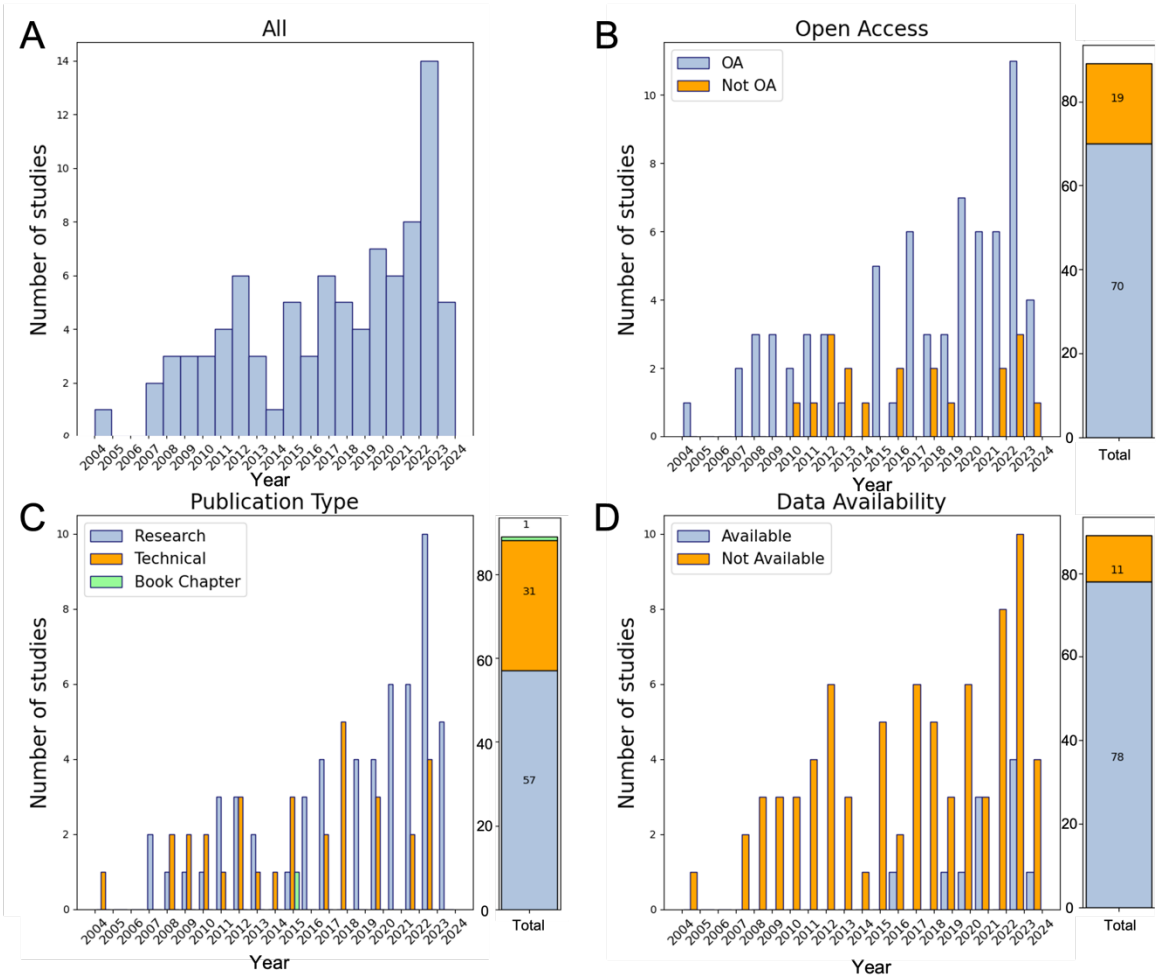

Figure 2. Scientometric summary of the whole-cell imaging literature. A) Histogram of total publication frequency by year. B) Open access (OA) publishing frequency by year. C) Publication type by year. D) Data availability, as reported, by year.

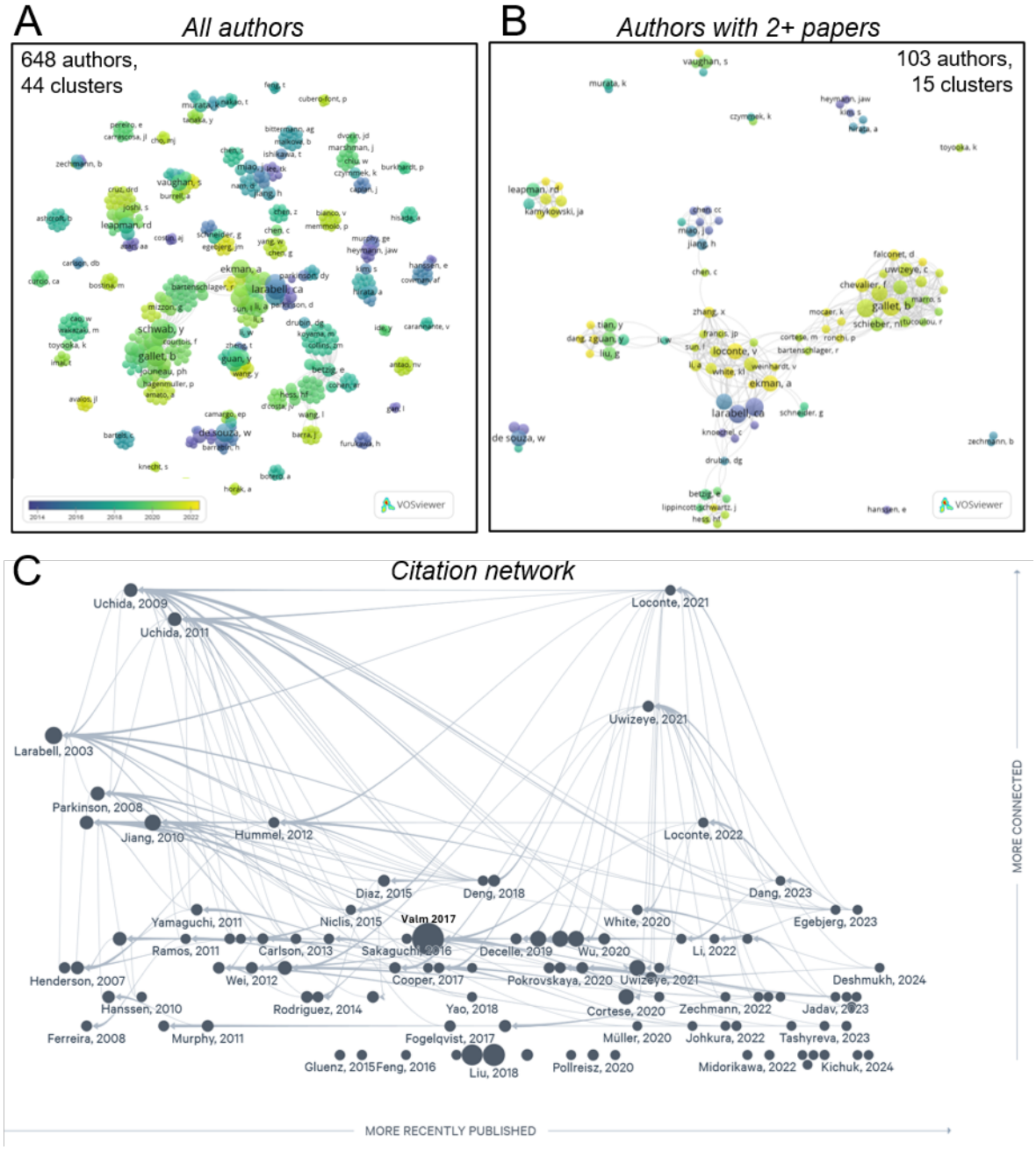

Figure S1. Bibliometric network analysis. A-B) Co-authorship network consisting of authors with A) any number of publications and B) those with at least two publications represented in the corpus. Color gradient represents average publication year per author and bubble size indicates relative number of articles per author. Visualizations created in VOSViewer. C) Citation network, created in Litmaps, representing most connected papers at the top, and most recent papers on the right. Bubble size reflects relative citation number, and arrows point to articles cited.

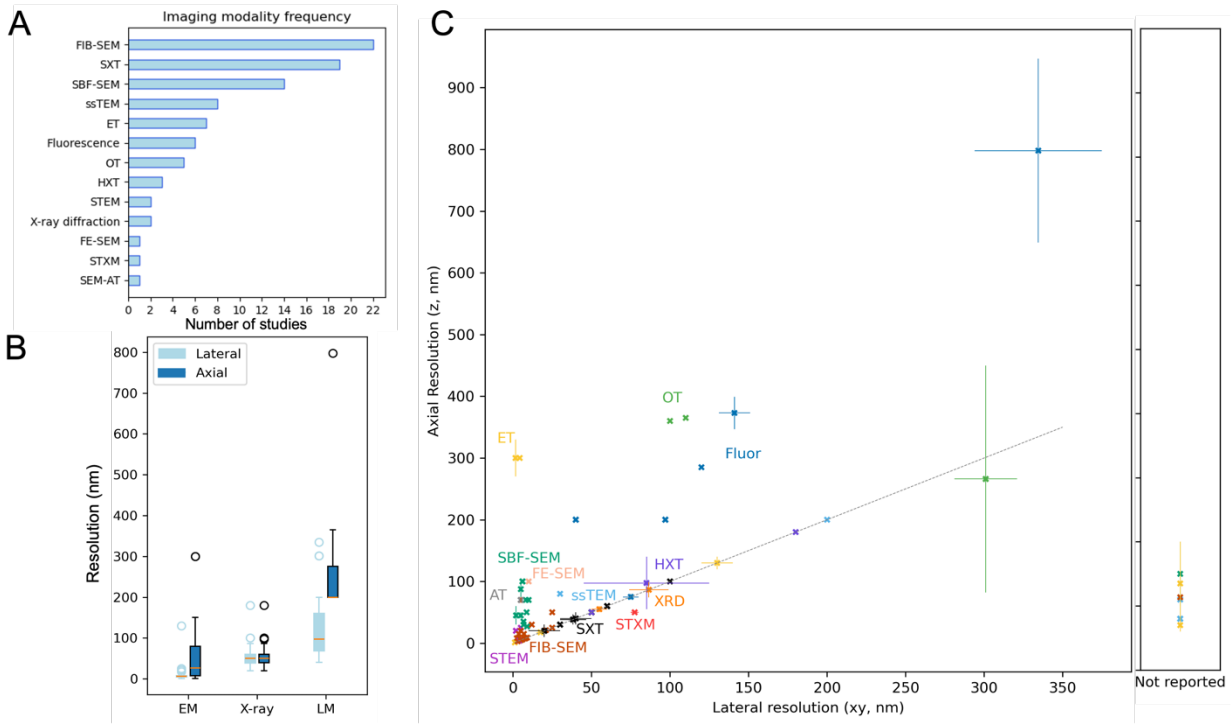

Figure 3. Summary of imaging methodology and resolution by dataset (n=121). A) The number of studies using each of the 13 imaging modalities. B) Box-and-whisker plot of reported lateral and axial imaging resolutions, in nanometers (nm), grouped by modality type: electron microscopy (EM), X-ray, or light microscopy (LM). A single value was plotted for each dataset. If multiple resolutions, or inconsistent reportion of resolution values were reported, the value plotted is the median of the reported values. C) Scatterplot of axial vs lateral resolution in nanometers for each dataset, where reported, colored by modality. The line represents isotropic resolution (axial = lateral). Right panel represents datasets for which only axial resolution was reported.

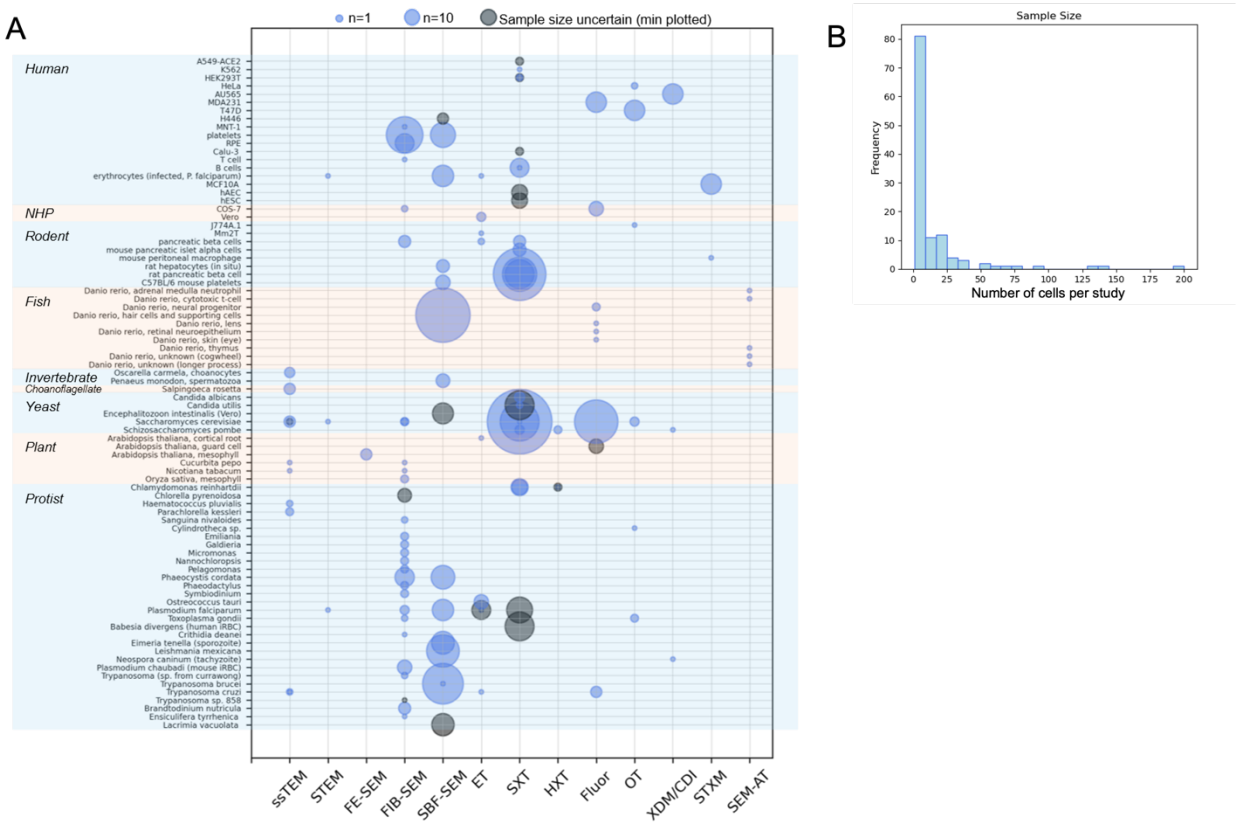

Figure 4. A) Datasets summarized according to cell type and modality. Each bubble represents one dataset, and bubble size represents reported sample size, colored blue for clearly-reported sample size, and gray for minimum known sample size when not clearly reported. B) Distribution of sample sizes.

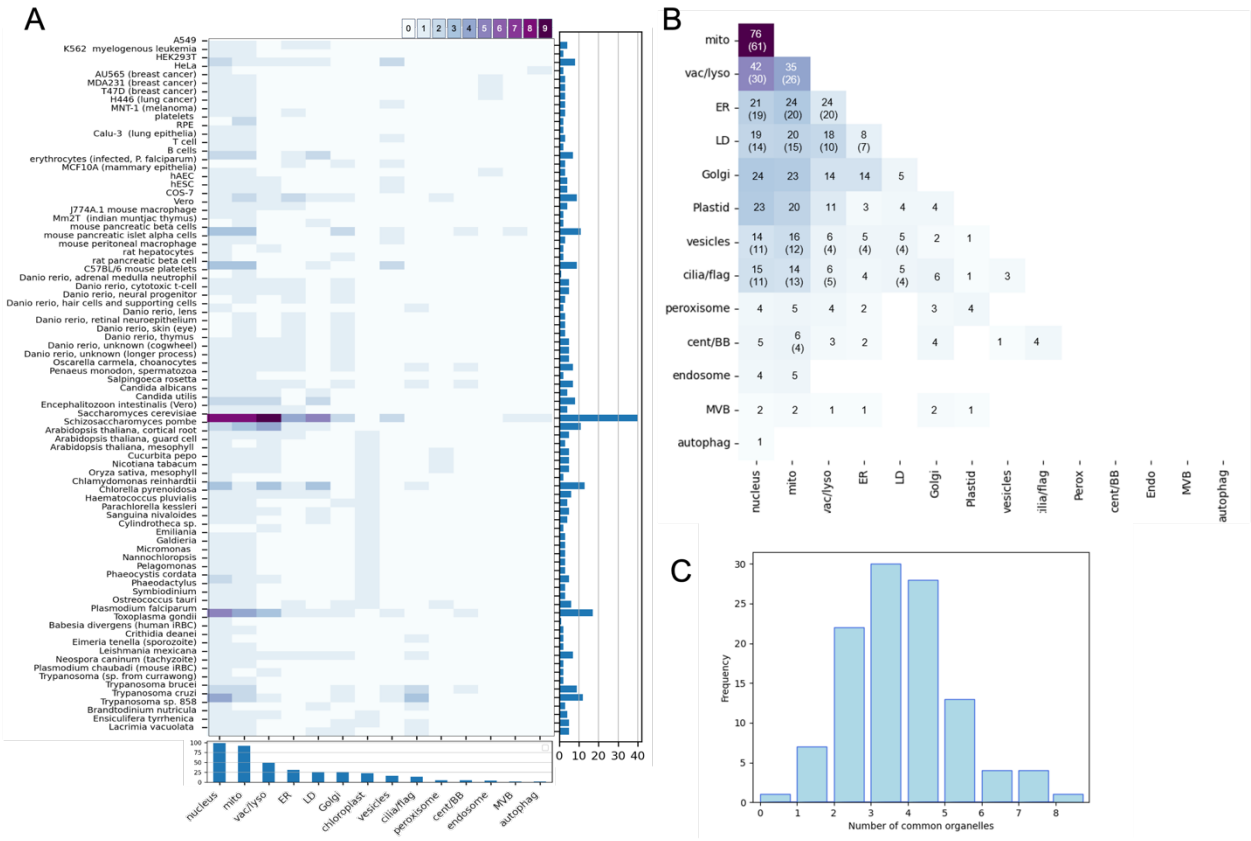

Figure 5. Representation of common organelles across datasets. A) Organelles included in datasets by cell type. Heatmap color represents the number of instances, or independent datasets in which the organelle appears. Summary bar graphs show the total number of organelles per cell type (right) and total number of instances per organelle (bottom). B) Organelle combination instances. Where two numbers are reported per cell, the top number is separate instances across all datasets, and the bottom number is the number of distinct cell types in which the combination appears. The differential between these numbers indicates the number of instances of the same organelle combination in the same cell type. C) Distribution of anatomical completeness factor, defined by the number of different common organelles reconstructed in tandem in a single image.

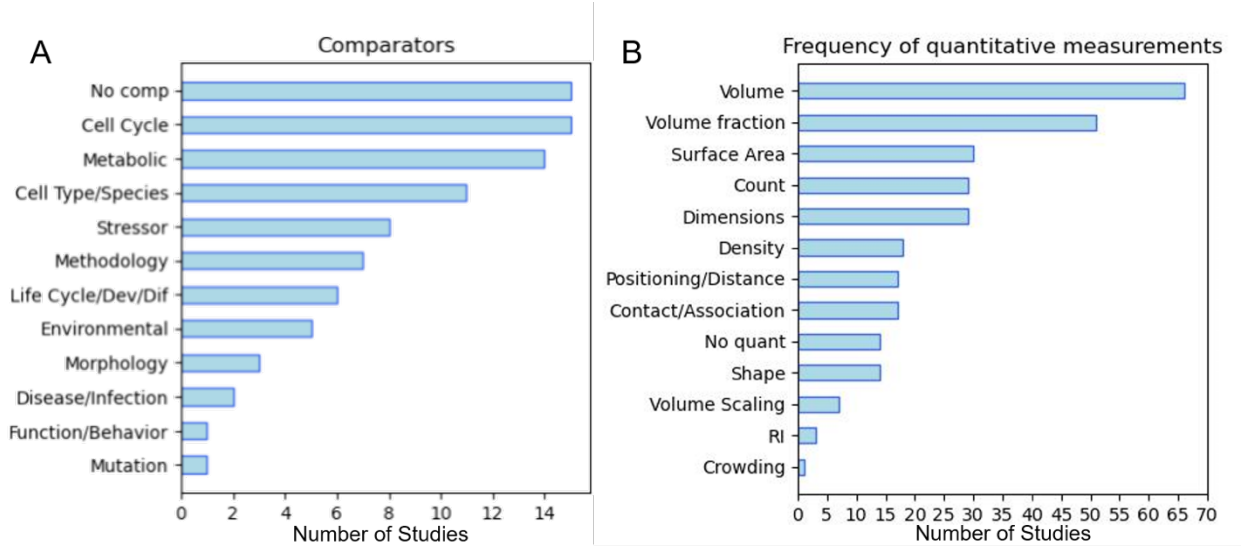

Figure 6. A) Frequency of comparator classes by study. Within-study comparable conditions or perturbations were grouped into general classes. “No comp” – studies with no comparator. B) Quantitative analyses performed on whole-cell reconstruction data. Dimensions, density, positioning/distance, contact/association, and shape are all groupings including different measurements conveying similar properties.

REFERENCES

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Bolaños, F., Salatino, A., Osborne, F., & Motta, E. (2024). Artificial intelligence for literature reviews: Opportunities and challenges. *Artificial Intelligence Review*, *57*(10), 259. https://doi.org/10.1007/s10462-024-10902-3

Cell Line Catalog. (2024). ALLEN CELL EXPLORER. https://www.allencell.org/cell-catalog.html

Chen, J.-H. F., Mirvis, M., Ekman, A., Vanslembrouck, B., Gros, M. L., Larabell, C., & Marshall, W. F. (2024). *Automated segmentation of soft X-ray tomography: Native cellular structure with sub-micron resolution at high throughput for whole-cell quantitative imaging in yeast* (p. 2024.10.31.621371). bioRxiv. https://doi.org/10.1101/2024.10.31.621371

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & de Hoon, M. J. L. (2009). Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, *25*(11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163

Decelle J, Veronesi G, LeKieffre C, Gallet B, Chevalier F, Stryhanyuk H, Marro S, Ravanel S, Tucoulou R, Schieber N, Finazzi G, Schwab Y, & Musat N. (2021). Subcellular architecture and metabolic connection in the planktonic photosymbiosis between Collodaria (radiolarians) and their microalgae. *Environ Microbiol*, *23*(11), 6569–6586. https://doi.org/10.1111/1462-2920.15766

Ekman, A., Chen, J.-H., Dermott, G. M., Le Gros, M. A., & Larabell, C. (2020). Task Based Semantic Segmentation of Soft X-ray CT Images Using 3D Convolutional Neural Networks. *Microscopy and Microanalysis*, *26*(S2), 3152–3154. https://doi.org/10.1017/S1431927620023983

Entrez Programming Utilities Help. (2010). National Center for Biotechnology Information (US). https://www.ncbi.nlm.nih.gov/books/NBK25501/

Erozan, A., Lösel, P. D., Heuveline, V., & Weinhardt, V. (2024). Automated 3D cytoplasm segmentation in soft X-ray tomography. *iScience*, *27*(6), 109856. https://doi.org/10.1016/j.isci.2024.109856

Gillooly, J. F., Hein, A., & Damiani, R. (2015). Nuclear DNA Content Varies with Cell Size across Human Cell Types. *Cold Spring Harbor Perspectives in Biology*, *7*(7), a019091. https://doi.org/10.1101/cshperspect.a019091

Gluenz E, Wheeler RJ, Hughes L, & Vaughan S. (2015). Scanning and three-dimensional electron microscopy methods for the study of Trypanosoma brucei and Leishmania mexicana flagella. *Methods Cell Biol*, *127*, 509–542. https://doi.org/10.1016/bs.mcb.2014.12.011

Guo, J., & Larabell, C. A. (2019). Soft X-ray tomography: Virtual sculptures from cell cultures. *Current Opinion in Structural Biology*, *58*, 324–332. https://doi.org/10.1016/j.sbi.2019.06.012

Hair, M., Yanase, R., Moreira-Leite, F., Wheeler, R. J., Sádlová, J., Volf, P., Vaughan, S., & Sunter, J. D. (2024). Whole cell reconstructions of Leishmania mexicana through the cell cycle. *PLOS Pathogens*, *20*(2), e1012054. https://doi.org/10.1371/journal.ppat.1012054

Hanson, M. A., Barreiro, P. G., Crosetto, P., & Brockington, D. (2024). The strain on scientific publishing. *Quantitative Science Studies*, 1–21. https://doi.org/10.1162/qss_a_00327

Hanssen, E., Knoechel, C., Dearnley, M., Dixon, M. W. A., Le Gros, M., Larabell, C., & Tilley, L. (2012). Soft X-ray microscopy analysis of cell volume and hemoglobin content in erythrocytes infected with asexual and sexual stages of Plasmodium falciparum. *Journal of Structural Biology*, *177*(2), 224–232. https://doi.org/10.1016/j.jsb.2011.09.003

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Heinrich L, Bennett D, Ackerman D, Park W, Bogovic J, Eckstein N, Petruncio A, Clements J, Pang S, Xu CS, Funke J, Korff W, Hess HF, Lippincott-Schwartz J, Saalfeld S, & Weigel AV. (2021). Whole-cell organelle segmentation in volume electron microscopy. *Nature*, *599*(7883), 141–146. https://doi.org/10.1038/s41586-021-03977-3

Henderson, G. P., Gan, L., & Jensen, G. J. (2007). 3-D Ultrastructure of O. tauri: Electron Cryotomography of an Entire Eukaryotic Cell. *PLoS ONE*, *2*(8), e749. https://doi.org/10.1371/journal.pone.0000749

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. Computing in Science & Engineering. https://doi.org/10.1109/MCSE.2007.55

Koutsogiannis Z, Mina JGM, Suman R, & Denny PW. (2023). Assessment of Toxoplasma gondii lytic cycle and the impact of a gene deletion using 3D label-free optical diffraction holotomography. *Front Cell Infect Microbiol*, *13*, 1237594. https://doi.org/10.3389/fcimb.2023.1237594

Larabell, C. A., & Le Gros, M. A. (2004). X-ray Tomography Generates 3-D Reconstructions of the Yeast, Saccharomyces cerevisiae, at 60-nm Resolution. *Molecular Biology of the Cell*, *15*(3), 957–962. https://doi.org/10.1091/mbc.e03-07-0522

Laughlin, R. B., Pines, D., Schmalian, J., Stojković, B. P., & Wolynes, P. (2000). The middle way. *Proceedings of the National Academy of Sciences*, *97*(1), 32–37. https://doi.org/10.1073/pnas.97.1.32

Laundon, D., Larson, B. T., McDonald, K., King, N., & Burkhardt, P. (2019). The architecture of cell differentiation in choanoflagellates and sponge choanocytes. *PLOS Biology*, *17*(4), e3000226. https://doi.org/10.1371/journal.pbio.3000226

Le Gros, M. A., McDermott, G., & Larabell, C. A. (2005). X-ray tomography of whole cells. *Current Opinion in Structural Biology*, *15*(5), 593–600. https://doi.org/10.1016/j.sbi.2005.08.008

Litmaps. (2024). Litmaps. https://app.litmaps.com

Loconte, V., Chen, J.-H., Cortese, M., Ekman, A., Le Gros, M. A., Larabell, C., Bartenschlager, R., & Weinhardt, V. (2021). Using soft X-ray tomography for

rapid whole-cell quantitative imaging of SARS-CoV-2-infected cells. *Cell Reports Methods*, *1*(7), 100117. https://doi.org/10.1016/j.crmeth.2021.100117

Loconte, V., Chen, J.-H., Vanslembrouck, B., Ekman, A. A., McDermott, G., Le Gros, M. A., & Larabell, C. A. (2023). Soft X-ray tomograms provide a structural basis for whole-cell modeling. *The FASEB Journal*, *37*(1), e22681. https://doi.org/10.1096/fj.202200253R

Loconte V, Singla J, Li A, Chen JH, Ekman A, McDermott G, Sali A, Le Gros M, White KL, & Larabell CA. (2022). Soft X-ray tomography to map and quantify organelle interactions at the mesoscale. *Structure*, *30*(4), 510-521.e3. https://doi.org/10.1016/j.str.2022.01.006

McDermott, G., Le Gros, M. A., & Larabell, C. A. (2012). Visualizing Cell Architecture and Molecular Location Using Soft X-Ray Tomography and Correlated Cryo-Light Microscopy. *Annual Review of Physical Chemistry*, *63*(1), 225–239. https://doi.org/10.1146/annurev-physchem-032511-143818

Meta-analysis in basic biology. (2016). *Nature Methods*, *13*(12), 959–959. https://doi.org/10.1038/nmeth.4102

Moore, J., Allan, C., Besson, S., Burel, J.-M., Diel, E., Gault, D., Kozlowski, K., Lindner, D., Linkert, M., Manz, T., Moore, W., Pape, C., Tischer, C., & Swedlow, J. R. (2021). OME-NGFF: A next-generation file format for expanding bioimaging data-access strategies. *Nature Methods*, *18*(12), Article 12. https://doi.org/10.1038/s41592-021-01326-w

Moore, J., Basurto-Lozada, D., Besson, S., Bogovic, J., Bragantini, J., Brown, E. M., Burel, J.-M., Casas Moreno, X., de Medeiros, G., Diel, E. E., Gault, D., Ghosh, S.

S., Gold, I., Halchenko, Y. O., Hartley, M., Horsfall, D., Keller, M. S., Kittisopikul, M., Kovacs, G., … Swedlow, J. R. (2023). OME-Zarr: A cloud-optimized bioimaging file format with international community support. *Histochemistry and Cell Biology*, *160*(3), 223–251. https://doi.org/10.1007/s00418-023-02209-1

Müller, A., Schmidt, D., Xu, C. S., Pang, S., D'Costa, J. V., Kretschmar, S., Münster, C., Kurth, T., Jug, F., Weigert, M., Hess, H. F., & Solimena, M. (2021). 3D FIB-SEM reconstruction of microtubule–organelle interaction in whole primary mouse β cells. *Journal of Cell Biology*, *220*(2), e202010039. https://doi.org/10.1083/jcb.202010039

Munn, Z., Pollock, D., Barker, T. H., Stone, J., Stern, C., Aromataris, E., Schünemann, H. J., Clyne, B., Khalil, H., Mustafa, R. A., Godfrey, C., Booth, A., Tricco, A. C., & Pearson, A. (2023). The Pandora's Box of Evidence Synthesis and the case for a living Evidence Synthesis Taxonomy. *BMJ Evidence-Based Medicine*, *28*(3), 148–150. https://doi.org/10.1136/bmjebm-2022-112065

Munn, Z., Stern, C., Aromataris, E., Lockwood, C., & Jordan, Z. (2018). What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. *BMC Medical Research Methodology*, *18*(1), Article 1. https://doi.org/10.1186/s12874-017-0468-4

Opalek, M., Tutaj, H., Pirog, A., Smug, B. J., Rutkowska, J., & Wloch-Salamon, D. (2023). A Systematic Review on Quiescent State Research Approaches in S. cerevisiae. *Cells*, *12*(12), Article 12. https://doi.org/10.3390/cells12121608

Paredes-Santos TC, de Souza W, & Attias M. (2012). Dynamics and 3D organization of secretory organelles of Toxoplasma gondii. *J Struct Biol*, *177*(2), 420–430. https://doi.org/10.1016/j.jsb.2011.11.028

Pautasso, M. (2012). Publication Growth in Biological Sub-Fields: Patterns, Predictability and Sustainability. *Sustainability*, *4*(12), Article 12. https://doi.org/10.3390/su4123234

Pokrovskaya I, Tobin M, Desai R, Aronova MA, Kamykowski JA, Zhang G, Joshi S, Whiteheart SW, Leapman RD, & Storrie B. (2021). Structural analysis of resting mouse platelets by 3D-EM reveals an unexpected variation in α-granule shape. *Platelets*, *32*(5), 608–617. https://doi.org/10.1080/09537104.2020.1799970

Rafelski, S. M., & Theriot, J. A. (2024). Establishing a conceptual framework for holistic cell states and state transitions. *Cell*, *187*(11), 2633–2651. https://doi.org/10.1016/j.cell.2024.04.035

Rutkowska, J., Lagisz, M., Bonduriansky, R., & Nakagawa, S. (2020). Mapping the past, present and future research landscape of paternal effects. *BMC Biology*, *18*(1), 183. https://doi.org/10.1186/s12915-020-00892-3

Sear, R. P., Pagonabarraga, I., & Flaus, A. (2015). Life at the mesoscale: The self-organised cytoplasm and nucleoplasm. *BMC Biophysics*, *8*(1), 4. https://doi.org/10.1186/s13628-015-0018-6

Sercel, A. J., Sturm, G., Shaulson, E. D., Gallagher, D., St-Onge, M.-P., Kempes, C. P., Pontzer, H., Hirano, M., & Picard, M. (2023). *Accelerated physiology and increased energy expenditure in animals and humans with mitochondrial defects:*

A meta-analysis (p. 2023.09.09.556754). bioRxiv. https://doi.org/10.1101/2023.09.09.556754

The data deluge. (2012). *Nature Cell Biology*, *14*(8), 775–775. https://doi.org/10.1038/ncb2558

The pandas development team. (2020). *pandas-dev/pandas: Pandas* (Version 2.2.3) [Computer software]. https://doi.org/10.5281/zenodo.3509134

Trebichalská, Z., Javůrek, J., Tatíčková, M., Kyjovská, D., Kloudová, S., Otevřel, P., Hampl, A., & Holubcová, Z. (2021). High-Resolution 3D Reconstruction of Human Oocytes Using Focused Ion Beam Scanning Electron Microscopy. *Frontiers in Cell and Developmental Biology*, *9*. https://www.frontiersin.org/articles/10.3389/fcell.2021.755740

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, *169*(7), 467–473. https://doi.org/10.7326/M18-0850

Tricco, A. C., Soobiah, C., Antony, J., Cogo, E., MacDonald, H., Lillie, E., Tran, J., D'Souza, J., Hui, W., Perrier, L., Welch, V., Horsley, T., Straus, S. E., & Kastner, M. (2016). A scoping review identifies multiple emerging knowledge synthesis methods, but few studies operationalize the method. *Journal of Clinical Epidemiology*, *73*, 19–28. https://doi.org/10.1016/j.jclinepi.2015.08.030

Trumpff, C., Michelson, J., Lagranha, C. J., Taleon, V., Karan, K. R., Sturm, G., Lindqvist, D., Fernström, J., Moser, D., Kaufman, B. A., & Picard, M. (2021). Stress and circulating cell-free mitochondrial DNA: A systematic review of human studies, physiological considerations, and technical recommendations. *Mitochondrion*, *59*, 225–245. https://doi.org/10.1016/j.mito.2021.04.002

Uchida, M., McDermott, G., Wetzler, M., Le Gros, M. A., Myllys, M., Knoechel, C., Barron, A. E., & Larabell, C. A. (2009). Soft X-ray tomography of phenotypic switching and the cellular response to antifungal peptoids in *Candida albicans*. *Proceedings of the National Academy of Sciences*, *106*(46), 19375–19380. https://doi.org/10.1073/pnas.0906145106

Uchida, M., Sun, Y., McDermott, G., Knoechel, C., Le Gros, M. A., Parkinson, D., Drubin, D. G., & Larabell, C. A. (2011). Quantitative analysis of yeast internal architecture using soft X-ray tomography. *Yeast*, *28*(3), 227–236. https://doi.org/10.1002/yea.1834

Uwizeye, C., Decelle, J., Jouneau, P.-H., Flori, S., Gallet, B., Keck, J.-B., Bo, D. D., Moriscot, C., Seydoux, C., Chevalier, F., Schieber, N. L., Templin, R., Allorent, G., Courtois, F., Curien, G., Schwab, Y., Schoehn, G., Zeeman, S. C., Falconet, D., & Finazzi, G. (2021a). Morphological bases of phytoplankton energy management and physiological responses unveiled by 3D subcellular imaging. *Nature Communications*, *12*(1), 1049. https://doi.org/10.1038/s41467-021-21314-0

Uwizeye, C., Decelle, J., Jouneau, P.-H., Flori, S., Gallet, B., Keck, J.-B., Bo, D. D., Moriscot, C., Seydoux, C., Chevalier, F., Schieber, N. L., Templin, R., Allorent,

G., Courtois, F., Curien, G., Schwab, Y., Schoehn, G., Zeeman, S. C., Falconet, D., & Finazzi, G. (2021b). Morphological bases of phytoplankton energy management and physiological responses unveiled by 3D subcellular imaging. *Nature Communications*, *12*(1), 1049. https://doi.org/10.1038/s41467-021-21314-0

Valm, A. M., Cohen, S., Legant, W. R., Melunis, J., Hershberg, U., Wait, E., Cohen, A. R., Davidson, M. W., Betzig, E., & Lippincott-Schwartz, J. (2017). Applying systems-level spectral imaging and analysis to reveal the organelle interactome. *Nature*, *546*(7656), 162–167. https://doi.org/10.1038/nature22369

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Veritas Health Innovation. (n.d.). *Covidence systematic review software*. Covidence. Retrieved February 3, 2025, from www.covidence.org

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … van Mulbregt, P. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Waskom, M. L. (2021). seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

Wei, D., Jacobs, S., Modla, S., Zhang, S., Young, C. L., Cirino, R., Caplan, J., & Czymmek, K. (2012). High-resolution three-dimensional reconstruction of a whole yeast cell using focused-ion beam scanning electron microscopy. *BioTechniques*, *53*(1), 41–48. https://doi.org/10.2144/000113850

White, K. L., Singla, J., Loconte, V., Chen, J.-H., Ekman, A., Sun, L., Zhang, X., Francis, J. P., Li, A., Lin, W., Tseng, K., McDermott, G., Alber, F., Sali, A., Larabell, C., & Stevens, R. C. (2020). Visualizing subcellular rearrangements in intact β cells using soft x-ray tomography. *Science Advances*, *6*(50), eabc8262. https://doi.org/10.1126/sciadv.abc8262

Yip, R., Sun, Y. J., Bassuk, A. G., & Mahajan, V. B. (2024). *Artificial Intelligence's Contribution to Biomedical Literature Search: Revolutionizing or Complicating?* https://doi.org/10.1101/2024.10.07.617112

Zechmann, B., Möstl, S., & Zellnig, G. (2022). Volumetric 3D reconstruction of plant leaf cells using SEM, ion milling, TEM, and serial sectioning. *Planta*, *255*(6), 118. https://doi.org/10.1007/s00425-022-03905-3

Zhang, Z. (2024). Expanding bioinformatics: Toward a paradigm shift from data to theory. *Fundamental Research*. https://doi.org/10.1016/j.fmre.2024.11.019

Zotero | Your personal research assistant. (n.d.). Retrieved February 3, 2025, from https://www.zotero.org/