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

Mary Mirvis[1], Brooke Weingard[1], Steven N. 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 is spatially arranged 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, resulting in a corpus consisting of 89 studies and 118 image datasets. We describe the trajectory and current state of the field between 2004 and 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 (<10 cells) and small cell types. We describe common quantitative analyses across studies, including volumetric ratios of organelles and inter-organelle contact analyses. This work establishes the initial iteration of 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 in the context of an emerging sub-field of cell biology. Our curated dataset now provides the basis for a plethora of future aggregate and comparative analyses to reveal larger patterns and generalized hypotheses about the systems behavior and regulation of whole-cell organelle networks. More broadly, we showcase the potential of new rigorous secondary research methods to strengthen

cell biology's literature review and reproducibility toolkit, create 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 defining 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 at the intersection of cell, systems, and physical biology. 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 directly governed by genetically programmed or energetically active mechanisms versus emerging or self-organizing 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?

Organelles, the largest mesoscale structural components, together represent the highest level of organization within the mesoscale. The fine ultrastructure of organelles has been extensively characterized through electron and light microscopy since the 1950s, and a rich understanding of the complex molecular composition, cellular functions, and dynamics of each individual organelle has been amassed since the 1980s. In the last decade, a new picture of organelles as interconnected components in a complex and dynamic network mediated by many contact site proteins and communication mechanisms has emerged, summarized in many reviews (Cohen et al., 2018; Scorrano et al., 2019; Voeltz et al., 2024). A systems view of organelles spawns new questions around the structure, dynamics, and regulation of the organelle network; its responsiveness to perturbations and mechanism of disruption in disease; and the role of each organelle component and pairwise interaction in network structure and function. Questions of this nature are being increasingly addressed through systems imaging (Bray et al., 2016; Rohban et al., 2017; Wang & Mukherji, 2022), spatial proteomics (Cho et al., 2022; Hein et al., 2023; Lundberg & Borner, 2019; Qin et al., 2021), and computational modeling (Donovan-Maiye et al., 2022; Earnest et al., 2017; Majarian et al., 2019; Murphy, 2012; Viana et al., 2023; Zhao & Murphy, 2007).

A new construct: Cell Anatomy

Studying the structure of organelle networks is a challenging task. Highly dynamic and morphologically heterogeneous organelles interact physically through a combination of protein-based membrane contact sites (MCS), mechanical forces, spatial constraints, membrane dynamics, and potentially other unidentified mechanisms. In aggregate, the coalescence of organelle morphologies, spatial distributions, and inter-relationships produces whole-cell patterns, which we term “cell anatomy”. Cell anatomy is a fundamental, but still incompletely defined cellular property bridging the mesoscale and whole cell, which is presumably both unique to specific cell types, as well as variable and dynamic within and across cell states. By

defining and observing cell anatomy across many cells and contexts, we can formulate novel questions. Which types, features, or variations of cell anatomy are common, and which are context-specific? Which are genetically regulated versus determined by simple geometry or intrinsic physics? Answering such questions requires the generation of microscopy images of multiple organelles in tandem in the context of the entire cellular volume, revealing the morphology and positioning of each organelle in its native, three-dimensional context.

Whole-cell imaging has been accomplished across a range of cell types using a variety of imaging modalities, including electron, X-ray, and light microscopy methods. Among image data types, whole-cell imaging is particularly technically challenging. Images encompassing large biological volumes[1] 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 automatically, which requires sufficient amounts of high-quality manually-segmented 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 requires data from large numbers of cells and is therefore generally beyond the scope of single whole-cell imaging studies. Aggregating the existing set of relatively small whole-cell imaging studies may reveal trends in cell anatomy across cell types and biological contexts, providing an avenue for novel analyses and insights into cell anatomy and organelle interaction networks.

This goal requires finding and curating relevant whole-cell imaging studies across the literature as comprehensively as possible, extracting useful information, and analyzing image data and reported measurements. Under the broad umbrella of secondary research, there exists a wide array of methods to synthesize and review published work, yielding summaries of existing knowledge, meta-scientific insights into the state of a field or subset of literature, and

identification of trends and gaps in the research landscape to motivate future work. A large family of systematic, rigorous, transparent, reproducible approaches to literature review collectively known as *evidence synthesis* or *research synthesis*, offers an array of rigorously developed methods for comprehensively and reproducibly curating and analyzing trends and knowledge across studies (Munn et al., 2018, 2023; Tricco et al., 2016). Here, we present a first step towards leveraging these tools for synthesizing mesoscale data from whole-cell imaging studies.

Our evidence synthesis approach ultimately seeks to uncover general and context-specific trends in cell anatomical patterns and organelle interdependencies. Combining information from many publications and datasets is a promising avenue to answer larger-scale questions than could be addressed by the results of individual published works. In order to formulate specific testable hypotheses, it is necessary to first have a clear picture of the amount and content of data currently available, and the ways it can be merged and compared across studies. Therefore, a crucial first step is to produce a thorough characterization of the volume, quality, and technical and biological features of the whole-cell image data landscape. In this work, 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 publication trends and the range of methodologies, reporting practices, and biological conditions included in the rapidly growing corpus of whole-cell imaging data. We highlight key points of overlap in the data which lay the foundation for future systematic review work. 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.

WORKFLOW

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, and full details into each step of the workflow, including screening criteria, categorization, and analysis, are described in Methods. In brief, we developed a search strategy consisting of a set of microscopy terms and a set of cell- and organelle- scale terms, based on an initial set of 11 target papers. Our search strategy returned over 11,000 results from PubMed. We defined two main inclusion criteria: 1) generation of original three-dimensional *whole-cell* image data in eukaryotic cells, including complete cell boundary delineation, and 2) volumetric information for at least two *organelles* concurrently must be included in the whole-cell volume. After three screening stages – title and PubMed summary, abstract, and full text – conducted by a team of three reviewers, 89 studies were fully included. We extracted information from studies to summarize publication trends and study design, including imaging modality and resolution, cell types, sample sizes, organelles, comparative conditions, and quantitative analyses,

RESULTS

Metadata and bibliometrics.

The potential for effective study curation and data extraction, and ultimately, novel insights into whole-cell structure, depends on the overall volume, accessibility, and type of studies screened. 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 complete 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 (70/89, 78.7%) (Figure 2B).

We reasoned that, while many of the studies we include may be similar in nature to our initial set, in that they perform extensive data collection and analysis in pursuit of a particular biological question, whole-cell image data may be found in other publication types. Technical studies which describe some advance in microscopy, image processing, or analysis, often provide small proof-of-concept datasets. While such studies may or may not undertake extensive biologically-driven analysis, the images themselves may include valuable structural information and be available for further analysis in combination with other datasets. 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 authors share the original imaging data 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). Of these, seven datasets are available in their entirety as reported in the study (Cortese et al., 2020; Ezzedine et al., 2023; Laundon et al., 2019; Mocaer K et al., 2023; Müller et al., 2021; Rudlaff et al., 2020; Uwizeye et al., 2021a), while five are available in partial form (Antao NV et al., 2023; Hair et al., 2024; McQuate A et al., 2023; Sakaguchi M et al., 2016; Uwizeye C et al., 2021). A summary of publicly available datasets is provided in Additional File 3. Overall, 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.

Can whole-cell imaging represent an emerging field? 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 (Figure S1A), and one large cluster of collaborators among prolific whole-cell microscopy researchers (authors on 2+ studies) (Figure 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 cited by many subsequent papers that used other microscopy methods, indicating that there has been 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

Structural biology through volumetric imaging has made great strides in recent years. The rapid improvement of image resolution, volumetric and statistical throughput, structure identification, and integration of cross-modal data, including correlative and dynamics approaches, have enabled exciting advancements in the overall accessibility and applicability of many imaging modalities to a wider and more complex array of problems than ever before. Whole-cell imaging approaches the upper limit of capabilities of most high resolution, 3D microscopy and reconstruction techniques. Each technique offers a slightly different set of strengths in terms of resolution, throughput, accessibility, cell type amenability and ultrastructural preservation,

207 organelle identity and number, potential for quantitative analysis, and so on. Therefore, in
208 characterizing the whole-cell imaging corpus, it is crucial to understand which imaging methods
209 are being used, which are more common, and how they data they produce compare and
210 contrast. This may motivate recommendations for modalities best suited for future whole-cell
211 imaging studies.
212
213 In total, we identified 13 distinct volumetric imaging types of modalities used to produce whole-
214 cell image data. Focused ion beam SEM (FIB-SEM) (also called ion-abrasion SEM, IA-SEM)
215 and soft X-ray tomography (SXT) and are the most common modalities, represented in 22 and
216 21 studies, respectively (Figure 3A). These were followed by serial block-face SEM (SBF-SEM),
217 serial sectioning TEM (ssTEM), electron tomography (ET), fluorescence (including 3D SIM, and
218 multispectral lattice light sheet), optical tomography (including optical diffraction and optical
219 computed tomography/live cell CT), hard X-ray tomography (HXT, X-ray ptychography),
220 scanning TEM (STEM), and X-ray diffraction (XRD) techniques. Field emission SEM (FE-SEM),
221 scanning transmission X-ray microscopy (STXM), and SEM array tomography (SEM-AT) were
222 each found in one study. In total, 55 studies used electron microscopy (EM) techniques (FIB-
223 SEM, SBF-SEM, ssTEM, ET, STEM, FE-SEM, SEM-AT), 27 (30%) used X-ray techniques
224 (SXT, HXT, XRD, STXM), and 10 used optical techniques (fluorescence, optical tomography).
225
226 Each modality excels within a limited range of spatial resolution, determining the size scale,
227 volumetric completeness, throughput, and potential for artifacts of anisotropic resolution in the
228 resulting images. Generally, optical imaging methods are limited by the wavelength of light
229 (~200nm), typically capable of achieving higher lateral (xy) than axial (z) resolution, but can be
230 used to produce sample sizes in the hundreds or thousands. Electron microscopy modalities are
231 capable of achieving resolution in the sub-10nm range, but are typically limited in volumetric
232 scope and statistical throughput due to the high information density and cumbersome sample

233 preparation, imaging, and image processing. X-ray based methodologies typically fall
234 somewhere in between light and electron microscopy in throughput and resolution (Guo &
235 Larabell, 2019; McDermott et al., 2012). The range of cell types potentially accessible for whole-
236 cell reconstruction, information content contained in the resulting images, comparability across
237 datasets and modalities, and capacity for secondary analysis is strongly dependent on image
238 resolution. Whole-cell imaging requires the acquisition of large volumes, so even the best
239 resolution possible with a particular modality or microscopy may need to be compromised to
240 increase the reconstruction volume and/or throughput.
241
242 We extracted the technical parameter information for each distinct dataset across the corpus,
243 where a dataset is considered a distinct set of cells of the same type imaged with the same
244 modality, such that multiple datasets may appear within one study (e.g. four datasets
245 comprising two cell types (*C. pepo, N. tabacum*) each imaged by two modalities each (ssTEM,
246 FIB-SEM) in (Zechmann et al., 2022). We found 118 total datasets across the 89 studies
247 (Additional File 4). The reported resolutions are summarized as lateral (L) and axial (A)
248 resolutions for electron microscopy (EM), X-ray microscopy (X-ray), and light microscopy (LM)
249 methods in Figure 3B. Each dataset is plotted as a separate datapoint in Figure 3C. Overall, 72
250 datasets (61%) were generated using EM techniques, 29 (24.6%) with X-ray techniques, and 17
251 (14.4%) with optical techniques. As expected, the highest resolution was achieved by electron
252 microscopy modalities (ET, FIB-SEM, SBF-SEM, ssTEM, STEM, FE-SEM, SEM-AT), followed
253 by X-ray based methodologies (SXT, HXT, X-ray diffraction, STXM), and the lowest and most
254 variable resolution in optical data (fluorescence microscopy, optical tomography). Isotropic
255 resolution, which is highly advantageous for structural accuracy and morphological analyses
256 (Chen et al., 2024; Le Gros et al., 2005), was far more common in X-ray based techniques
257 (26/27 datasets, 92.9%, counting only those reporting both axial and lateral resolution), than in
258 EM-based (25/56, 53.6%) or light-based (0/16) 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 for 18 datasets from 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).

In summary, our results show that many modalities are capable of producing whole-cell image data, but X-ray based techniques, particularly soft X-ray tomography, have so far been most successfully leveraged to produce high resolution, isotropic, high-throughput whole cell datasets.

Cell type and sample size

The potential to pool and compare findings across studies depends largely on the biological similarity of the data, in particular, the cell type or model organism imaged. The same cell type imaged in multiple studies (with varying sample sizes) offers the potential to cross-reference and pool data across modalities and experimental conditions. The range of cell types represented across the corpus will point to appropriate groupings and comparisons for cross-cell type syntheses.

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 80 distinct cell types across 118

285 datasets. The most strongly-represented cell types (in terms number of separate datasets and
286 number of cells imaged in total across datasets, which we call “data density”), are highlighted in
287 Table 1.

288 **Table 1. Cell types with highest data density. A summary of data available for cell**
289 **types/species recurring in multiple studies across the corpus.**

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

290

291 Most studies had small sample sizes, defined as the number of cells fully reconstructed in 3D
292 with multiple organelle volumetric information compatible with our inclusion criteria, with 81
293 (68.6%) of datasets reporting n<=10 whole cell images, and 36 (30.5%) studies consisting of a
294 single cell (Figure 4B). Single-cell datasets were more common in earlier papers, and frequently
295 occurred in technical papers – 16 of the 89 studies were technical papers reporting a single
296 whole-cell reconstruction, representing half (51.6%) of the 31 technical studies and 66.7% of the
297 single-cell studies. This suggests that the first ten or so years of our dataset (~2005-2015) could
298 be described as technical early days, but the advances during this time paved the way for more
299 biologically-focused research studies with larger datasets in the following decade (~2015-
300 today). Sample sizes were largest in SXT (up to n=200) and SBF-SEM (up to n=141) studies. In
301 14 datasets, the study sample size, was either reported inconsistently or incompletely, in which

302 case we reported a minimum known sample size, or else was not reported at all, resulting in
303 missing data in our analysis (Figure 4A, white bubbles).
304
305 *Organelles and anatomical completeness*
306 Our primary inclusion criterion was the presence of at least two organelles within the same
307 whole-cell reconstruction. We argue that for visualizing cell anatomy, at least two different
308 structures must be present in order to provide spatial context for one another. Furthermore, as
309 pairwise spatial and morphological inter-organelle interactions are building blocks of the whole-
310 cell organelle network, the presence of two organelles reveals at least one pair whose
311 interaction can be described. We therefore extracted the list of organelles reconstructed in each
312 dataset to assess the potential for analysis at the single-organelle, inter-organelle, and cell
313 anatomical levels.
314
315 Figure 5A shows the cumulative organelle data per cell type across the corpus. The nucleus
316 and mitochondria are the most commonly segmented organelles across all datasets (106 and
317 91 instances respectively, accounting for multiple distinct cell types per paper in some cases),
318 followed by vacuoles/lysosomes (53 instances) endoplasmic reticulum (28 instances), lipid
319 droplets (27 instances), Golgi (26 instances), chloroplast/plastid (27 instances), vesicles (18
320 instances), flagella/cilia (17 instances), peroxisome (7 instances), centrosome/basal body (5
321 instances), endosome (5 instances), multivesicular bodies (1 instance), and autophagosome (1
322 instance). Some organisms may appear to have only one or no organelles included despite our
323 inclusion criterion requiring imaging of two or more organelles . This is due to the fact that the
324 heatmap only includes “common organelles” present in all or most eukaryotes. The blank lines
325 represent studies which image and reconstruct only specialized or cell type-specific organelles,
326 such as the specialized secretory organelles (microneme, rhoptry, dense granules) or conoid
327 structures in *Toxoplasma gondii* (Koutsogiannis Z *et al*., 2023; Paredes-Santos TC *et al*., 2012).

328 In some cases, we find one common organelle in combination with specialized organelles, such
329 as the study in C57BL/6 resting mouse platelets (mitochondria, dense granules, alpha-granules,
330 canalicular system) (Pokrovskaya I *et al.*, 2021). The recurrence of cell types in multiple studies
331 enables potential examination of organelle morphologies across datasets, experimental
332 designs, and biological contexts. While not all studies imaging the same cell type segment the
333 same set of organelles, the most frequently-imaged specific organelles are found among the
334 best-represented cell types: *S. cerevisiae* vacuoles (12 instances), nuclei (11), mitochondria
335 (10), ER (7),lipid droplets (7), and vesicles (5); *P. falciparum* nuclei (5) and vacuoles (4), and *T.*
336 *cruzi* nuclei (5).
337
338 The rationale for requiring at least two organelles per segmented cell in the dataset is 1) the
339 supposition that a single internal landmark is insufficient to constitute an anatomical pattern, and
340 2) to reveal spatial interactions between organelles in the 3D context of the entire cell. Overall,
341 we found 59 pairwise organelle combinations occurring across the corpus, of which 35 had at
342 least five independent instances, suggesting rich potential for comparison and synthesis of each
343 organelle-organelle relationship across many cell types and contexts (Figure 5B). The most
344 common organelle co-occurrences are mitochondria-nucleus (82 instances across 65 cell
345 types), nucleus-vacuole/lysosome (48 instances across 36 cell types), and mitochondria-
346 vacuole/lysosome (40 instances across 33 cell types). These cases will enable within-cell type
347 synthesis of similar data (e.g. analysis of a specific organelle pair, as well as cross-study
348 comparison, which can potentially isolate biological and technical differences.
349
350 The ultimate requirement for an understanding of cell anatomy and modeling whole-cell patterns
351 and inter-organelle interaction networks is to be able to visualize as many organelles
352 simultaneously as possible. The number of “common organelles” per reconstruction, termed
353 “anatomical completeness” is distributed with a peak at 3-4 organelles (Figure 5C). The most

354 anatomically complete reconstructions are in choanoflagellate *S. rosetta* (Laundon et al., 2019)
355 with 8 organelles, and 7 organelles in *L. mexicana* (Hair et al., 2024), sponge *O. carmella*
356 choanocyte (Laundon et al., 2019), *A. thaliana* cortical root cell (Cui Y et al., 2019), and *S.*
357 *cerevisiae* (Wei et al., 2012).
358
359 *Morphological metrics and Quantitative analyses*
360 The unique advantage of high-resolution 3D structural data of whole cells is their capacity to
361 provide rich and precise quantitative morphological and spatial information. Future meta-
362 analysis of organelle features relies on the presence of comparable quantitative measurements
363 across studies. An assessment of quantitative analyses performed across studies revealed a
364 limited set of organelle geometric features measured (Figure 6A). Volume was the most
365 frequently measured property (65 studies, 73%), followed by volume fraction (organelle-to-cell
366 volume ratio, 50 studies, 56%). Also common were surface area (34 studies, 38%), count/copy
367 number (29 studies, 32.6%) and linear dimensions such as length and diameter (27 studies,
368 30%). Less frequent, and more variable, were measures of density such as LAC or electron
369 density and positioning/distance (n=18 each, 20.2%), organelle contacts/associations and
370 shape (n=14 each, 15.7%). Organelle volume scaling trends were calculated in 7 studies (7.9%)
371 and three studies measured refractive index. One study devised a method to quantify
372 subcellular crowding of organelles ("congestion index", Loconte V *et al*., 2022). These
373 frequencies roughly reflect the general ease of measurement and conceptual familiarity of these
374 metrics - volume and surface area are generally trivial to measure, albeit with variable accuracy,
375 with many standard software tools automatically generating these calculations. Count and
376 dimensions can be straightforwardly manually or semi-automatically calculated. Distance and
377 contact frequency, shape, and crowding require more complex calculations and computational
378 operations, and can be interpreted in variable ways. A sizable fraction of studies (n=14, 15.7%)
379 performed no quantitative analysis.

380

381 *Comparators*

382 While whole-cell structural data is largely static, cellular structure is in fact highly dynamic and
383 variable across cells. We recorded the experimentally compared conditions between samples
384 ("comparators"), reasoning that variety in conditions for a single cell type could suggest
385 functional organelle and anatomical variability, while overlap in conditions across cell types
386 could reveal generalizable context-specific organelle and anatomical trends. Most studies (71,
387 79.8%) included comparator data across at least one cellular perturbation or condition, which
388 we grouped into broad classes (Figure 6B). The most common of these are cell cycle stages
389 and metabolic conditions, such as glucose stimulation of pancreatic cells (14 studies each).
390 Other comparators, in decreasing order of frequency, were cross-cell type comparisons (n=12),
391 stressors such as drug treatments (n=9), methodology, such as fixation methods or imaging
392 modalities (n=6), stages of life cycle, development, or differentiation (n=6), environmental
393 conditions, such as light exposure (n=3), morphology, such as shape or size (n=3) , disease or
394 infection states (n=2) , functional or behavioral states, such as cell-cell communication (n=1),
395 and genetic mutation (n=1). The predominance of cell cycle stages and cell type as the primary
396 comparators highlights the potential of whole-cell imaging to serve as a basis for defining cell
397 type and functional cell state-related cellular reorganizations.

398

399 In summary, the whole-cell multi-organelle imaging corpus includes 89 published studies
400 between 2004 and 2024 from over 11,000 initial PubMed results. In total, the dataset represents
401 the whole-cell imaging and reconstruction of at least 1690 individual cells, of 80 eukaryotic cell
402 types from protists to humans, across 118 independent datasets, 13 imaging modalities, at least
403 13 common organelles in many different combinations, 71 perturbative or comparative
404 conditions, and at least 12 types of morphological metrics quantitatively analyzed.

405

406 **LIMITATIONS**

407 Our scoping study is, to our knowledge, one of the first applications of scoping review
408 methodology to the field of whole-cell microscopy, and to fundamental cell biology more broadly.
409 Our method comes with limitations arising from the iterative nature of proof-of-concept methods
410 development, the breadth and heterogeneity of the literature corpus, and the size of our team.
411 One of the goals of a formal systematic literature synthesis is to be as comprehensive as
412 possible, spanning peer-reviewed and grey literature sources as thoroughly as possible. Our
413 approach focused specifically on PubMed, but ideally, other databases such as Web of Science
414 and EMBASE, would be included as well. Although our PubMed search was extremely broad
415 relative to our final inclusion set, which comprised only less than 1% of the original search
416 results, we cannot be sure how many relevant studies might have been missed by the search
417 entirely. Furthermore, grey literature sources such as dissertations and online image data
418 hosting sites without corresponding publications would be of relevance to a comprehensive set
419 of cell anatomy data, but would be missed by our strategy. For example, the Allen Institute for
420 Cell Science hosts a collection of 3D confocal images in human iPSCs, a subset of which would
421 fit our inclusion criteria, but these datasets have not been directly thoroughly described in a
422 publication, despite their inclusion within a large dataset underlying many impactful publications
423 (*Cell Line Catalog*, 2024). Furthermore, as of January 2025, our search string retrieves 300 new
424 results since our last round of curation in May 2024. The dataset and analysis presented here
425 should be seen only as an initial characterization which will be continually updated as the
426 literature base expands. We intend to establish our growing curated corpus and analysis as a
427 publicly available, living database of cell anatomy data and field trends, inviting community input
428 to identify gaps and curate the corpus as thoroughly as possible.
429

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. 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 and include them under an adjusted set of criteria.

DISCUSSION

455 We have created a comprehensive corpus of whole-cell imaging literature spanning a wide
456 breadth of modalities, cell types, organelles, and contexts through a novel application of formal
457 scoping review methodology. There is substantial comparability across studies, highlighting
458 opportunities for further meta-scientific and synthetic studies building on this corpus. Future
459 directions include systematic review and meta-analysis approaches to understand how
460 organelles typically self-organize and interact within the whole cell space across biological
461 contexts, including qualitative categorizations of cell anatomy, and meta-analyses of key
462 morphometric measurements.

463 Although our search is not exhaustive, this corpus comprises a foundational evidence base
464 revealing an emerging field of significant, and rapidly increasing, volume and scholarly
465 interconnectedness, unified by a specific data type – whole cell, multi-organelle 3D
466 reconstruction (Figures 1 and S1). Our systematic curation and meta-scientific characterization
467 of these remarkably information-dense structural data, from raw data to processed images to
468 derived measurements, enables a range future secondary analysis and synthesis efforts. The
469 substantial methodological and biological overlap opens a wide range of possibilities for within-
470 and cross-group synthetic studies to uncover novel insights that would otherwise be difficult to
471 access through traditional primary research methods. Data in a specific cell type pooled across
472 datasets, modalities, and conditions provide opportunities to query the consistency of similar
473 metrics across studies, integrate disparate findings into an organism-centric model, and
474 compare distinct imaging modalities and parameters applied to the same experimental system.
475 Summarizing quantitative analyses and extending them where possible allows for a more
476 comprehensive quantitative understanding of generalizable principles of organization, an
477 assessment of single organelle, between-organelle, and cell-wide feature variability across
478 contexts, and provide a strong empirical foundation for integrative modeling efforts (Loconte et
479 al., 2023). Publicly shared original data enable novel analyses and the development of new data

integration and secondary analysis workflows. Ultimately, work building on our corpus promises to reveal emergent patterns of cell anatomy across cell types and contexts, and insights into the constraints and variations embedded in these patterns as a result of cellular geometry and organelle interactions.

We intend for our dataset and analysis to become a growing tool for the scientific community. In addition to generalized, high level synthesis, the corpus can be subdivided in many ways to enable inquiry into specific cell types, organelles and organelle interactions, biological perturbations, imaging and image analysis approaches. In the near future, we intend to make the corpus easily accessible, interactive, and kept up-to-date so that researchers interested in any component of the corpus can easily contribute to and draw upon it for their own investigations. We hope this work will galvanize whole-cell imaging and analysis researchers, synthesis researchers, and anyone else interested in organelle-to-whole cell mesoscale structural biology to create a community, collaborating to refine the ongoing curation and re-use of whole-cell data, refine and unify the generation and reporting of novel work, and engaging in discourse around terminology, best practices, optimal applications for existing imaging technologies and needs for technical advancement, and a broad vision of the future of the whole-cell imaging and cell anatomy field.

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,

505 reconstruction and multi-organelle 3D segmentation, particularly at sub-100nm resolution. The
506 13 imaging modalities represented each offer particular advantages and disadvantages, with
507 unique tradeoffs between resolution, throughput, native structure preservation, and overall
508 tractability depending on the application. For example, the two dominant modalities, SXT and
509 FIB-SEM, are both capable of achieving high isotropic resolution with some degree of
510 quantitative throughput. SXT generally achieves isotropic 30-60nm resolution with minimal
511 sample damage or tomographic artifacts, and is capable of producing throughput in the
512 hundreds per study. However, cell size is strictly capped by the ~10um diameter of the capillary
513 used to hold the samples, limiting the cell types one can study by this method, and access is
514 limited by the requirement for beamline time. FIB-SEM excels for very high-resolution
515 reconstructions (~5 nm), but is limited by throughput and, as a sample-destructive method, is
516 more likely to introduce ultrastructural artifacts. Extensive reviews of strengths and weaknesses
517 across volumetric imaging modalities have been discussed in depth elsewhere (Hanssen et al.,
518 2012; Loconte et al., 2021; Loconte V et al., 2022), and technical capabilities for all modalities
519 are constantly being improved.
520
521 However, the limiting factor for maximizing the utility of whole-cell image data is perhaps not
522 data quality or abundance itself, but rather the limited use and reuse of these incredibly rich
523 datasets, potentially leaving much of the embedded information untapped. We found that the
524 publication of whole-cell imaging studies is accelerating, with growing trends toward open-
525 access publication and data sharing, suggesting good current and future potential for meta-
526 scientific studies and secondary re-analysis of datasets (Fig. 2). However, only a small subset of
527 our corpus made original data publicly available, and we found inconsistent reporting of key
528 methodological details, such as imaging resolution and sample size. Such metadata items are
529 crucial for meta-scientific analysis, including secondary analysis and synthesis research as well
530 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. 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 costly-to-produce data.

While interest in open science and data-reuse is quite mainstream and rapidly growing in the sphere of cell and molecular biology, the range of approaches currently used are largely focused on big -omics data informatics. 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). There is still a clear need for robust knowledge and data synthesis in biology, driven by widely recognized challenges associated with "data deluges"

557 (“The Data Deluge,” 2012; Zhang, 2024) and the exponential growth of scientific output (Hanson
558 et al., 2024; Pautasso, 2012). The monumental goal of constructing meaningful holistic
559 understandings of highly complex biological systems by merging a rich tradition of narrowly
560 focused mechanistic studies and molecular parts lists, with interdisciplinary systems,
561 biophysical, evolutionary, synthetic, and engineering perspectives (Rafelski & Theriot, 2024),
562 requires robust frameworks to comprehensively and rigorously synthesize data, knowledge, and
563 interpretations across disciplinary boundaries, aided by a high-level understanding of field
564 trends revealed by meta-scientific analyses.
565
566 There is at present limited familiarity in the fundamental biology field of evidence synthesis
567 methods widely used in many disciplines including the clinical and social sciences. These are
568 extensive, carefully developed toolkits supported by synthesis researchers and large
569 international collaborations and initiatives such as Cochrane and the Joanna Briggs Institute.
570 Systematic research syntheses in fundamental cell biology, particularly intermediate-scale
571 syntheses to address specific biological questions or hypotheses, are beginning to come on to
572 the scene in recent years, still rare and methodologically disparate (Gillooly et al., 2015; Opalek
573 et al., 2023; Rutkowska et al., 2020; Sercel et al., 2023; Trumpff et al., 2021).
574
575 The present work illustrates the unique considerations, challenges, and scientific and meta-
576 scientific potential of developing an evidence synthesis toolkit tailored to fundamental cell
577 biology. In contrast to translational and applied disciplines, particularly those based on animal or
578 human subjects, exploratory fields such as fundamental cell biology marked by high
579 heterogeneity in their technical repertoire, study design, experimental and reporting practices,
580 philosophical and sociological standards for epistemology and argument construction, and
581 conceptual and linguistic frameworks. This heterogeneity presents unique and significant
582 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 iteratively refined, and robust stepwise methods for screening and data extraction, ideally carried out by teams of 2-4 expert investigators working in parallel to minimize bias, are needed. In general, a typical systematic review process can take 1-1.5 years of painstaking manual work conducted and validated by a dedicated team.

The prospect of employing AI tools to aid in these intensive steps is attractive, and being actively explored by many scientific groups and software 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

608 ecosystem for novel discovery and meta-research to derive the maximal benefit from the
609 immense volume and richness of biological literature and data.
610

611 **METHODS**

612 ***Search strategy:***

613 We developed our search strategy starting with an initial set of 11 target papers (Decelle J et al.,
614 2021; Heinrich L et al., 2021; Henderson et al., 2007; Laundon et al., 2019; Müller et al., 2021;
615 Uchida et al., 2009, 2011; Uwizeye et al., 2021a; Valm et al., 2017; Wei et al., 2012; White et
616 al., 2020)], thoroughly combed for cell attribute- (e.g. “cell organization”, “subcellular
617 architecture”), mesoscale-, and imaging-related terms. These terms were used to devise an
618 iterative series of search strings for PubMed aimed at retrieving a defined corpus that includes
619 the full initial set. While an initial attempt focused on cell anatomy-related terms was
620 unsuccessful due to the heterogeneity and inconsistent usage of such conceptual language,
621 iterations on search strategies consisting of 3D imaging and mesoscale (cell/subcellular and
622 organelle-scale) terms successfully retrieved the initial papers.
623

624 (("Imaging, Three-Dimensional"[MH] OR "3D imaging"[tiab] OR "three-dimensional
625 imaging"[tiab] OR "3-D imaging"[tiab] OR "Soft X-ray tomography"[tiab] OR (3-D[tiab] OR
626 3D[tiab] OR "three dimensional"[tiab] OR "three-dimensional"[tiab]) AND (tomography[tiab]
627 OR microscopy[tiab] OR Microscopy[MH] OR Tomography[MH] OR imaging[tiab]))
628 AND
629 ((ultrastructure[MeSH Subheading]) OR (Organelles[mh] OR organell*[tiab]) OR
630 (intracell*[tiab] OR intra-cell*[tiab]) OR (subcell*[tiab] OR sub-cell*[tiab]) OR (((whole[tiab]
631 OR entire[tiab]) AND cell[tiab]) OR "whole-cell"[tiab] or "whole cell" OR "entire cell"[tiab])))
632 NOT Review[pt]

633

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

638

639 ***Screening***

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

646

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

649 1. Cell anatomy is necessarily a whole-cell property. The study must have collected original
650 three-dimensional *whole-cell* images in eukaryotic systems, described in the primary study
651 in the form of direct visualizations (still images and/or supplementary movies displaying
652 volumetric rotations), and/or quantitative analyses necessarily taken from whole-cell data
653 (e.g. organelle-to-cell volume ratios). “Whole-cell” requires that the entire cellular boundary
654 (cell membrane or cell wall, or other marker showing the continuous cell periphery) is
655 captured within the image, such that the cellular volume can be calculated.
656 2. Volumetric information for at least two *organelles* concurrently must be included in the
657 whole-cell volume. We reasoned that at least two subcellular structures must be visible in
658 relation to one another for the concept of “anatomy” to be applicable. This criterion requires

659 a full volumetric reconstruction of the multi-organelle cell, with organelle identities clearly
660 annotated (typically in the form of a segmentation). The reconstructed data should either be
661 directly shown as an image in figures, movie files, or associated data, or necessarily
662 generated to produce published analysis, such as volume calculations. Although the usage
663 of the term "organelle" has expanded in recent years, we used a traditional definition,
664 referring to a structurally and functionally distinct, membrane-bound, intracellular
665 compartment.
666
667 Curating a set of target papers from this large set of search results required a series of
668 increasingly thorough screening steps – title and PubMed summary, abstract, and full text –
669 conducted by a team of three reviewers. Search results and screening decisions for each stage,
670 including exclusion reasons, along with information extracted from initial target papers are
671 provided in Supporting File 1.
672
673 *Title/summary.* The first screening step was a single-reviewer rapid screen through search
674 result pages in the web browser (Mozilla Firefox) of titles and short three-line PubMed-
675 generated summary statements. Studies that were obviously irrelevant, focused on the wrong
676 biological scale, or purely technical with no biological data were excluded. Studies that clearly
677 described whole-cell imaging and/or structures of entire cells were included. If unclear, a skim of
678 the abstract (and PMC figures, if available) was done, and studies that were still ambiguous
679 were included for further screening. This step was performed by one reviewer (MM), and
680 excluded 95% of results.
681
682 *Abstract*. A custom Python (v3) script using BioEntrez module of Biopython (Cock et al., 2009)
683 was used to pull abstracts from PubMed e-utilities (*Entrez Programming Utilities Help*, 2010)
684 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 general 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 Additional File 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 inaccessible. This includes several studies that partially reconstructed prohibitively large cell types (Trebichalská et al., 2021), as well as

711 reconstructed tissue volumes from which while relevant metrics are quantified, such as
712 organelle volume/volumetric ratios, calculated across the tissue block rather than in individual
713 cells, as distinct cell boundaries could not be delineated. In some cases, the whole-cell images
714 had slightly clipped cell edges due to nuances such as cells in cultures spreading their edges
715 and protrusions far from the main cell body on a culture substrate, or an elongated cell type
716 such as myocytes. We chose to exclude such examples but include others such as images of
717 protists with flagella partially cut-off, due to the fact a missing edge of the cell body leaves key
718 morphometrics, such as cell volume and shape, impossible to calculate, and creates
719 uncertainties about holistic cell anatomy due missing information, whereas missing part of a
720 cilium or flagellum would only affect analyses of ciliary length, but would preserve the ability to
721 calculate organelle to cell size ratios. Other borderline categories include: organelle identities
722 uncertain; ineligible structures defined as "organelles"; a very minor volume of the cell edge is
723 cut off in the image, technically violating the “whole cell” criterion; cell membrane not clearly
724 delineated; the paper itself does not include the relevant data, but accompanies the
725 establishment of a public repository or resource containing relevant data.
726
727 Disagreements in inclusion/exclusion decisions were discussed between reviewers, prompting
728 slight adjustments to our protocol, such as establishing new exclusion reasons or borderline
729 categories. One earlier study was added *post hoc* (Larabell & Le Gros 2004). Ultimately, 89
730 studies were fully included.
731
732 ***Data Extraction and Analysis***
733 We extracted information from studies to summarize study design, including imaging modality
734 and resolution, cell types, sample sizes, organelles, comparative conditions, and quantitative
735 analyses, only as pertains to information within the study directly describing or derived from
736 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. The full set of final data extraction items, including 89 included studies and 39 borderline studies, is provided in Additional File 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.

Acknowledgements

This project was funded by the UCSF Program for Breakthrough Biomedical Research, funded in part by the Sandler Foundation, and NIH grant R35 GM130327. 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.

FIGURE LEGENDS

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. Data are represented between 2004 and May 2024.

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. A) The number of studies using each of the 13 imaging modalities. B) Box-and-whisker plots 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 values or

789 a range of values were reported, the value plotted is the median of the reported values. C)
790 Scatterplot of axial vs lateral resolution in nanometers for each dataset, where reported, colored
791 by modality. Error bars represent the reported range, when applicable, with the datapoint
792 representing the median. The dotted line represents isotropic resolution (axial = lateral). Right
793 panel represents datasets for which only axial resolution was reported.
794
795 **Figure 4.** A) Datasets summarized according to cell type and modality. Each bubble represents
796 one dataset, and bubble size represents reported sample size (refer to legend, top left), colored
797 according to modality for clearly-reported sample size, and white for minimum known sample
798 size when not clearly reported. B) Distribution of sample sizes, according to minimum known
799 sample size. Inset show distribution for studies with sample sizes <=10.

800 **Figure 5.** Representation of common organelles across datasets. A) Organelle data collected
801 for each cell type represented across 118 datasets, excluding specialized or cell type-specific
802 organelles. Heatmap color represents the number of independent datasets in which the
803 organelle appears. Summary bar graphs show the total number of organelle instances per cell
804 type (right) and total number of instances per organelle across cell types (bottom) (Nuc –
805 nucleus, Mito – mitochondria, Vac/Lyso – vacuole/lysosome, ER – endoplasmic reticulum, LD –
806 lipid droplet, Golgi – Golgi body, Plast – plastid/chloroplast, Ves – vesicle, Cilia – cilia/flagella,
807 Perox – peroxisome, Cent – centrosome/centriole/basal body, Endo – endosome, MVB –
808 multivesicular body, Autophag – autophagosome.) B) Organelle combinations. Heatmap
809 represents the number of instances of each organelle pair co-imaged in the same dataset. C)
810 Distribution of anatomical completeness, defined by the number of distinct common organelles
811 coinciding in a single dataset.

812 **Figure 6.** A) Quantitative analyses performed on whole-cell reconstruction data. Dimensions,
813 density, positioning/distance, contact/association, and shape are all groupings including

814 different measurements conveying similar properties. (No quant – studies with no quantitative
815 analysis; RI – refractive index) B) Frequency of comparator classes by study. Within-study
816 comparable conditions or perturbations were grouped into general classes. ("No comp" –
817 studies with no comparator; Life Cycle/Dev/Dif – stages of life cycle, development, or
818 differentiation)

819

820 **ADDITIONAL FILES**

821 **Additional File 1.** Step by step details of corpus construction, with each step of searching and
822 screening listed in a separate tab. 0_Initial Target Papers: list of initial papers and extracted
823 features exemplifying target article types, used to construct PubMed Search Strategy.
824 1_Pubmed Search (2005-5.28.24): Complete list of 11,104 PubMed search results. 2_Title
825 Screening: PubMed metadata for 570 titles remaining after first screening step based on titles
826 and PubMed summary statements. 3_Abstract Screening: Results of second screening step
827 based on abstracts with PMID, inclusion decision (Include (I), Exclude (E), and Check Full Text
828 (C) and brief justification for each study. 4a_Full Text Screening Includes: Metadata for 127
829 studies included after full text screening. 4b_Full Text Screening Excludes: Metadata for 256
830 studies excluded after full text screening. 5_Post hoc additions: Metadata for 1 study included
831 after completion of bulk screening.

832 **Additional File 2.** Complete data extraction information for included studies after full text
833 screening, with full includes (Data Extraction_Includes) and borderlines (Data
834 Extraction_Borderlines) separated.

835 **Additional File 3.** Publicly available datasets. Parameters for 12 studies reporting public data
836 availability, including dataset completeness, study design, links, amount and format of data.

Additional File 4. Complete information for 118 datasets. Each individual dataset from 89 studies is listed including specifics of imaging modality, resolution, cell type and taxonomy, sample size, and organelles.

Additional File 5. Data underlying analysis plotted in figures. Specific subsets of data used to plot figures is separated into tabs.

REFERENCES

Antao NV, Lam C, Davydov A, Riggi M, Sall J, Petzold C, Liang FX, Iwasa JH, Ekiert DC, & Bhabha G. (2023). 3D reconstructions of parasite development and the intracellular niche of the microsporidian pathogen Encephalitozoon intestinalis. *Nat Commun*, *14*(1), 7662. https://doi.org/10.1038/s41467-023-43215-0

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

Bray, M.-A., Singh, S., Han, H., Davis, C. T., Borgeson, B., Hartland, C., Kost-Alimova, M., Gustafsdottir, S. M., Gibson, C. C., & Carpenter, A. E. (2016). Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nature Protocols*, *11*(9), 1757–1774. https://doi.org/10.1038/nprot.2016.105

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

Cho, N. H., Cheveralls, K. C., Brunner, A.-D., Kim, K., Michaelis, A. C., Raghavan, P., Kobayashi, H., Savy, L., Li, J. Y., Canaj, H., Kim, J. Y. S., Stewart, E. M., Gnann, C., McCarthy, F., Cabrera, J. P., Brunetti, R. M., Chhun, B. B., Dingle, G., Hein, M. Y., … Leonetti, M. D. (2022). OpenCell: Endogenous tagging for the cartography of human cellular organization. *Science*, *375*(6585), eabi6983. https://doi.org/10.1126/science.abi6983

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

Cohen, S., Valm, A. M., & Lippincott-Schwartz, J. (2018). Interacting organelles. *Current Opinion in Cell Biology*, *53*, 84–91. https://doi.org/10.1016/j.ceb.2018.06.003

Cortese, M., Lee, J.-Y., Cerikan, B., Neufeldt, C. J., Oorschot, V. M. J., Köhrer, S., Hennies, J., Schieber, N. L., Ronchi, P., Mizzon, G., Romero-Brey, I., Santarella-Mellwig, R., Schorb, M., Boermel, M., Mocaer, K., Beckwith, M. S., Templin, R. M., Gross, V., Pape, C., … Bartenschlager, R. (2020). Integrative Imaging Reveals SARS-CoV-2-Induced Reshaping of Subcellular Morphologies. *Cell Host & Microbe*, *28*(6), 853-866.e5. https://doi.org/10.1016/j.chom.2020.11.003

Cui Y, Cao W, He Y, Zhao Q, Wakazaki M, Zhuang X, Gao J, Zeng Y, Gao C, Ding Y, Wong HY, Wong WS, Lam HK, Wang P, Ueda T, Rojas-Pierce M, Toyooka K, Kang BH, & Jiang L. (2019). A whole-cell electron tomography model of vacuole biogenesis in Arabidopsis root cells. *Nat Plants*, *5*(1), 95–105. https://doi.org/10.1038/s41477-018-0328-1

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

Donovan-Maiye, R. M., Brown, J. M., Chan, C. K., Ding, L., Yan, C., Gaudreault, N., Theriot, J. A., Maleckar, M. M., Knijnenburg, T. A., & Johnson, G. R. (2022). A deep generative model of 3D single-cell organization. *PLOS Computational Biology*, *18*(1), e1009155. https://doi.org/10.1371/journal.pcbi.1009155

Earnest, T. M., Watanabe, R., Stone, J. E., Mahamid, J., Baumeister, W., Villa, E., & Luthey-Schulten, Z. (2017). Challenges of Integrating Stochastic Dynamics and Cryo-Electron Tomograms in Whole-Cell Simulations. *The Journal of Physical Chemistry B*, *121*(15), 3871–3881. https://doi.org/10.1021/acs.jpcb.7b00672

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

Ezzedine, J. A., Uwizeye, C., Si Larbi, G., Villain, G., Louwagie, M., Schilling, M., Hagenmuller, P., Gallet, B., Stewart, A., Petroutsos, D., Devime, F., Salze, P., Liger, L., Jouhet, J., Dumont, M., Ravanel, S., Amato, A., Valay, J.-G., Jouneau, P.-H., … Maréchal, E. (2023). Adaptive traits of cysts of the snow alga Sanguina nivaloides unveiled by 3D subcellular imaging. *Nature Communications*, *14*(1), 7500. https://doi.org/10.1038/s41467-023-43030-7

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

Hein, M. Y., Peng, D., Todorova, V., McCarthy, F., Kim, K., Liu, C., Savy, L., Januel, C., Baltazar-Nunez, R., Bax, S., Vaid, S., Vangipuram, M., Ivanov, I. E., Byrum, J. R., Pradeep, S., Gonzalez, C. G., Aniseia, Y., Wang, E., Creery, J. S., … Leonetti, M. D. (2023). *Global organelle profiling reveals subcellular localization and remodeling at proteome scale*. https://doi.org/10.1101/2023.12.18.572249

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

989 Loconte, V., Chen, J.-H., Vanslembrouck, B., Ekman, A. A., McDermott, G., Le Gros, M. A., &
990 Larabell, C. A. (2023). Soft X-ray tomograms provide a structural basis for whole-cell
991 modeling. *The FASEB Journal*, *37*(1), e22681. https://doi.org/10.1096/fj.202200253R
992 Loconte V, Singla J, Li A, Chen JH, Ekman A, McDermott G, Sali A, Le Gros M, White KL, &
993 Larabell CA. (2022). Soft X-ray tomography to map and quantify organelle interactions at
994 the mesoscale. *Structure*, *30*(4), 510-521.e3. https://doi.org/10.1016/j.str.2022.01.006
995 Lundberg, E., & Borner, G. H. H. (2019). Spatial proteomics: A powerful discovery tool for cell
996 biology. *Nature Reviews Molecular Cell Biology*, *20*(5), 285–302.
997 https://doi.org/10.1038/s41580-018-0094-y
998 Majarian, T. D., Cao-Berg, I., Ruan, X., & Murphy, R. F. (2019). CellOrganizer: Learning and
999 Using Cell Geometries for Spatial Cell Simulations. In W. S. Hlavacek (Ed.), *Modeling*
1000 *Biomolecular Site Dynamics* (Vol. 1945, pp. 251–264). Springer New York.
1001 https://doi.org/10.1007/978-1-4939-9102-0_11
1002 McDermott, G., Le Gros, M. A., & Larabell, C. A. (2012). Visualizing Cell Architecture and
1003 Molecular Location Using Soft X-Ray Tomography and Correlated Cryo-Light
1004 Microscopy. *Annual Review of Physical Chemistry*, *63*(1), 225–239.
1005 https://doi.org/10.1146/annurev-physchem-032511-143818
1006 McQuate A, Knecht S, & Raible DW. (2023). Activity regulates a cell type-specific mitochondrial
1007 phenotype in zebrafish lateral line hair cells. *Elife*, *12*.
1008 https://doi.org/10.7554/eLife.80468
1009 Meta-analysis in basic biology. (2016). *Nature Methods*, *13*(12), 959–959.
1010 https://doi.org/10.1038/nmeth.4102
1011 Mocaer K, Mizzon G, Gunkel M, Halavatyi A, Steyer A, Oorschot V, Schorb M, Le Kieffre C, Yee
1012 DP, Chevalier F, Gallet B, Decelle J, Schwab Y, & Ronchi P. (2023). Targeted volume
1013 correlative light and electron microscopy of an environmental marine microorganism. *J*
1014 *Cell Sci*, *136*(15). https://doi.org/10.1242/jcs.261355

1015 Moore, J., Allan, C., Besson, S., Burel, J.-M., Diel, E., Gault, D., Kozlowski, K., Lindner, D.,
1016 Linkert, M., Manz, T., Moore, W., Pape, C., Tischer, C., & Swedlow, J. R. (2021). OME-
1017 NGFF: A next-generation file format for expanding bioimaging data-access strategies.
1018 *Nature Methods*, *18*(12), Article 12. https://doi.org/10.1038/s41592-021-01326-w
1019 Moore, J., Basurto-Lozada, D., Besson, S., Bogovic, J., Bragantini, J., Brown, E. M., Burel, J.-
1020 M., Casas Moreno, X., de Medeiros, G., Diel, E. E., Gault, D., Ghosh, S. S., Gold, I.,
1021 Halchenko, Y. O., Hartley, M., Horsfall, D., Keller, M. S., Kittisopikul, M., Kovacs, G., …
1022 Swedlow, J. R. (2023). OME-Zarr: A cloud-optimized bioimaging file format with
1023 international community support. *Histochemistry and Cell Biology*, *160*(3), 223–251.
1024 https://doi.org/10.1007/s00418-023-02209-1
1025 Müller, A., Schmidt, D., Xu, C. S., Pang, S., D'Costa, J. V., Kretschmar, S., Münster, C., Kurth,
1026 T., Jug, F., Weigert, M., Hess, H. F., & Solimena, M. (2021). 3D FIB-SEM reconstruction
1027 of microtubule–organelle interaction in whole primary mouse β cells. *Journal of Cell*
1028 *Biology*, *220*(2), e202010039. https://doi.org/10.1083/jcb.202010039
1029 Munn, Z., Pollock, D., Barker, T. H., Stone, J., Stern, C., Aromataris, E., Schünemann, H. J.,
1030 Clyne, B., Khalil, H., Mustafa, R. A., Godfrey, C., Booth, A., Tricco, A. C., & Pearson, A.
1031 (2023). The Pandora's Box of Evidence Synthesis and the case for a living Evidence
1032 Synthesis Taxonomy. *BMJ Evidence-Based Medicine*, *28*(3), 148–150.
1033 https://doi.org/10.1136/bmjebm-2022-112065
1034 Munn, Z., Stern, C., Aromataris, E., Lockwood, C., & Jordan, Z. (2018). What kind of systematic
1035 review should I conduct? A proposed typology and guidance for systematic reviewers in
1036 the medical and health sciences. *BMC Medical Research Methodology*, *18*(1), Article 1.
1037 https://doi.org/10.1186/s12874-017-0468-4
1038 Murphy, R. F. (2012). CellOrganizer: Image-Derived Models of Subcellular Organization and
1039 Protein Distribution. In *Methods in Cell Biology* (Vol. 110, pp. 179–193). Elsevier.
1040 https://doi.org/10.1016/B978-0-12-388403-9.00007-2

1041 Opalek, M., Tutaj, H., Pirog, A., Smug, B. J., Rutkowska, J., & Wloch-Salamon, D. (2023). A
1042 Systematic Review on Quiescent State Research Approaches in S. cerevisiae. *Cells*,
1043 *12*(12), Article 12. https://doi.org/10.3390/cells12121608
1044 Paredes-Santos TC, de Souza W, & Attias M. (2012). Dynamics and 3D organization of
1045 secretory organelles of Toxoplasma gondii. *J Struct Biol*, *177*(2), 420–430.
1046 https://doi.org/10.1016/j.jsb.2011.11.028
1047 Pautasso, M. (2012). Publication Growth in Biological Sub-Fields: Patterns, Predictability and
1048 Sustainability. *Sustainability*, *4*(12), Article 12. https://doi.org/10.3390/su4123234
1049 Pokrovskaya I, Tobin M, Desai R, Aronova MA, Kamykowski JA, Zhang G, Joshi S, Whiteheart
1050 SW, Leapman RD, & Storrie B. (2021). Structural analysis of resting mouse platelets by
1051 3D-EM reveals an unexpected variation in α-granule shape. *Platelets*, *32*(5), 608–617.
1052 https://doi.org/10.1080/09537104.2020.1799970
1053 Qin, Y., Huttlin, E. L., Winsnes, C. F., Gosztyla, M. L., Wacheul, L., Kelly, M. R., Blue, S. M.,
1054 Zheng, F., Chen, M., Schaffer, L. V., Licon, K., Bäckström, A., Vaites, L. P., Lee, J. J.,
1055 Ouyang, W., Liu, S. N., Zhang, T., Silva, E., Park, J., … Ideker, T. (2021). A multi-scale
1056 map of cell structure fusing protein images and interactions. *Nature*, *600*(7889), 536–
1057 542. https://doi.org/10.1038/s41586-021-04115-9
1058 Rafelski, S. M., & Theriot, J. A. (2024). Establishing a conceptual framework for holistic cell
1059 states and state transitions. *Cell*, *187*(11), 2633–2651.
1060 https://doi.org/10.1016/j.cell.2024.04.035
1061 Rohban, M. H., Singh, S., Wu, X., Berthet, J. B., Bray, M.-A., Shrestha, Y., Varelas, X., Boehm,
1062 J. S., & Carpenter, A. E. (2017). Systematic morphological profiling of human gene and
1063 allele function via Cell Painting. *eLife*, *6*, e24060. https://doi.org/10.7554/eLife.24060
1064 Rudlaff, R. M., Kraemer, S., Marshman, J., & Dvorin, J. D. (2020). Three-dimensional
1065 ultrastructure of Plasmodium falciparum throughout cytokinesis. *PLOS Pathogens*,
1066 *16*(6), e1008587. https://doi.org/10.1371/journal.ppat.1008587

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

Sakaguchi M, Miyazaki N, Fujioka H, Kaneko O, & Murata K. (2016). Three-dimensional analysis of morphological changes in the malaria parasite infected red blood cell by serial block-face scanning electron microscopy. *J Struct Biol*, *193*(3), 162–171. https://doi.org/10.1016/j.jsb.2016.01.003

Scorrano, L., De Matteis, M. A., Emr, S., Giordano, F., Hajnóczky, G., Kornmann, B., Lackner, L. L., Levine, T. P., Pellegrini, L., Reinisch, K., Rizzuto, R., Simmen, T., Stenmark, H., Ungermann, C., & Schuldiner, M. (2019). Coming together to define membrane contact sites. *Nature Communications*, *10*(1), 1287. https://doi.org/10.1038/s41467-019-09253-3

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

1092 Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters,
1093 M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J.,
1094 Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018).
1095 PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation.
1096 *Annals of Internal Medicine*, *169*(7), 467–473. https://doi.org/10.7326/M18-0850
1097 Tricco, A. C., Soobiah, C., Antony, J., Cogo, E., MacDonald, H., Lillie, E., Tran, J., D'Souza, J.,
1098 Hui, W., Perrier, L., Welch, V., Horsley, T., Straus, S. E., & Kastner, M. (2016). A
1099 scoping review identifies multiple emerging knowledge synthesis methods, but few
1100 studies operationalize the method. *Journal of Clinical Epidemiology*, *73*, 19–28.
1101 https://doi.org/10.1016/j.jclinepi.2015.08.030
1102 Trumpff, C., Michelson, J., Lagranha, C. J., Taleon, V., Karan, K. R., Sturm, G., Lindqvist, D.,
1103 Fernström, J., Moser, D., Kaufman, B. A., & Picard, M. (2021). Stress and circulating
1104 cell-free mitochondrial DNA: A systematic review of human studies, physiological
1105 considerations, and technical recommendations. *Mitochondrion*, *59*, 225–245.
1106 https://doi.org/10.1016/j.mito.2021.04.002
1107 Uchida, M., McDermott, G., Wetzler, M., Le Gros, M. A., Myllys, M., Knoechel, C., Barron, A. E.,
1108 & Larabell, C. A. (2009). Soft X-ray tomography of phenotypic switching and the cellular
1109 response to antifungal peptoids in *Candida albicans*. *Proceedings of the National*
1110 *Academy of Sciences*, *106*(46), 19375–19380. https://doi.org/10.1073/pnas.0906145106
1111 Uchida, M., Sun, Y., McDermott, G., Knoechel, C., Le Gros, M. A., Parkinson, D., Drubin, D. G.,
1112 & Larabell, C. A. (2011). Quantitative analysis of yeast internal architecture using soft X-
1113 ray tomography. *Yeast*, *28*(3), 227–236. https://doi.org/10.1002/yea.1834
1114 Uwizeye, C., Decelle, J., Jouneau, P.-H., Flori, S., Gallet, B., Keck, J.-B., Bo, D. D., Moriscot,
1115 C., Seydoux, C., Chevalier, F., Schieber, N. L., Templin, R., Allorent, G., Courtois, F.,
1116 Curien, G., Schwab, Y., Schoehn, G., Zeeman, S. C., Falconet, D., & Finazzi, G.
1117 (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

Uwizeye C, Mars Brisbin M, Gallet B, Chevalier F, LeKieffre C, Schieber NL, Falconet D, Wangpraseurt D, Schertel L, Stryhanyuk H, Musat N, Mitarai S, Schwab Y, Finazzi G, & Decelle J. (2021). Cytoklepty in the plankton: A host strategy to optimize the bioenergetic machinery of endosymbiotic algae. *Proc Natl Acad Sci U S A*, *118*(27). https://doi.org/10.1073/pnas.2025252118

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

Viana, M. P., Chen, J., Knijnenburg, T. A., Vasan, R., Yan, C., Arakaki, J. E., Bailey, M., Berry, B., Borensztejn, A., Brown, E. M., Carlson, S., Cass, J. A., Chaudhuri, B., Cordes Metzler, K. R., Coston, M. E., Crabtree, Z. J., Davidson, S., DeLizo, C. M., Dhaka, S., …

Rafelski, S. M. (2023). Integrated intracellular organization and its variations in human iPS cells. *Nature*, *613*(7943), 345–354. https://doi.org/10.1038/s41586-022-05563-7

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

Voeltz, G. K., Sawyer, E. M., Hajnóczky, G., & Prinz, W. A. (2024). Making the connection: How membrane contact sites have changed our view of organelle biology. *Cell*, *187*(2), 257–270. https://doi.org/10.1016/j.cell.2023.11.040

Wang, S., & Mukherji, S. (2022). *Uncovering the principles coordinating systems-level organelle biogenesis with cellular growth* [Preprint]. Systems Biology. https://doi.org/10.1101/2022.11.01.514705

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

Zhao, T., & Murphy, R. F. (2007). Automated learning of generative models for subcellular location: Building blocks for systems biology. *Cytometry Part A*, *71A*(12), 978–990. https://doi.org/10.1002/cyto.a.20487

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

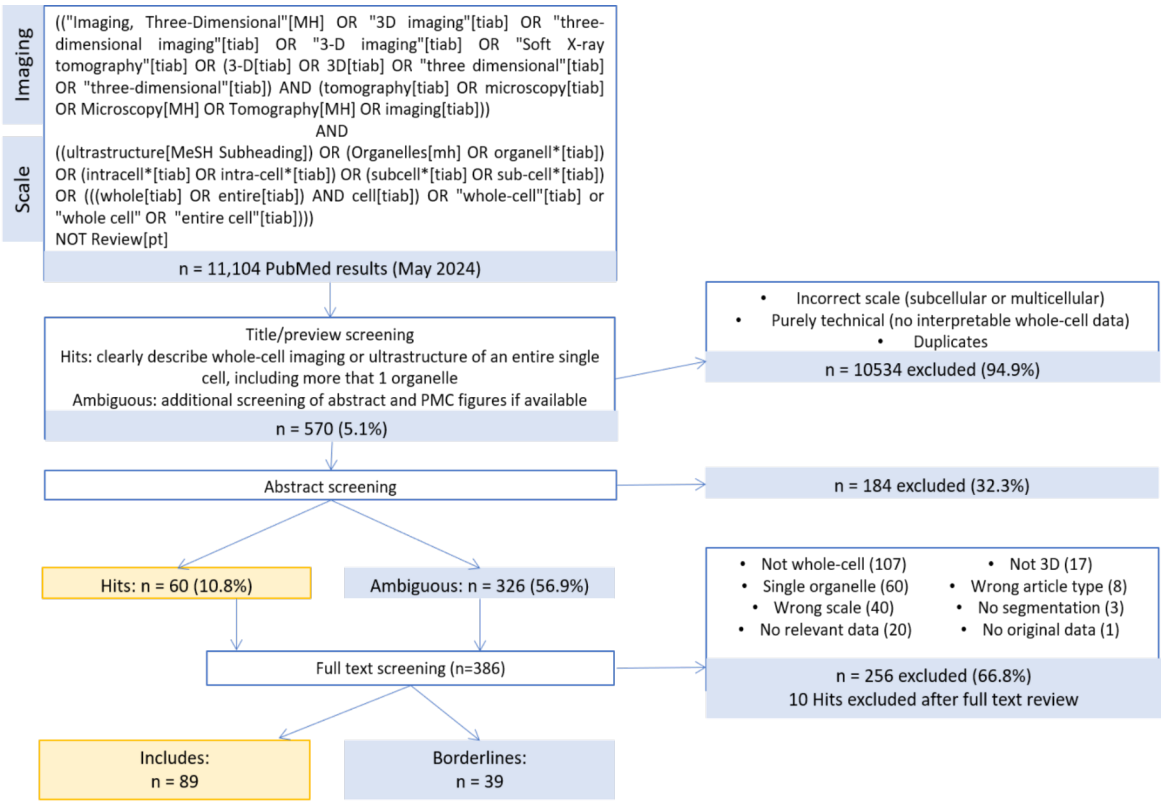
Imaging
Scale
(("Imaging, Three-Dimensional"[MH] OR "3D imaging"[tiab] OR "three-dimensional imaging"[tiab] OR "3-D imaging"[tiab] OR "Soft X-ray tomography"[tiab] OR (3-D[tiab] OR 3D[tiab] OR "three dimensional"[tiab] OR "three-dimensional"[tiab]) AND (tomography[tiab] OR microscopy[tiab] OR Microscopy[MH] 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]
n = 11,104 PubMed results (May 2024)
Title/preview screening
Hits: clearly describe whole-cell imaging or ultrastructure of an entire single cell, including more that 1 organelle
Ambiguous: additional screening of abstract and PMC figures if available
n = 570 (5.1%)
• Incorrect scale (subcellular or multicellular)
• Purely technical (no interpretable whole-cell data)
• Duplicates
n = 10534 excluded (94.9%)
Abstract screening
n = 184 excluded (32.3%)
Hits: n = 60 (10.8%)
Ambiguous: n = 326 (56.9%)
Full text screening (n=386)
• Not whole-cell (107)
• Single organelle (60)
• Wrong scale (40)
• No relevant data (20)
• Not 3D (17)
• Wrong article type (8)
• No segmentation (3)
• No original data (1)
n = 256 excluded (66.8%)
10 Hits excluded after full text review
Includes:
n = 89
Borderlines:
n = 39

Figure 1.

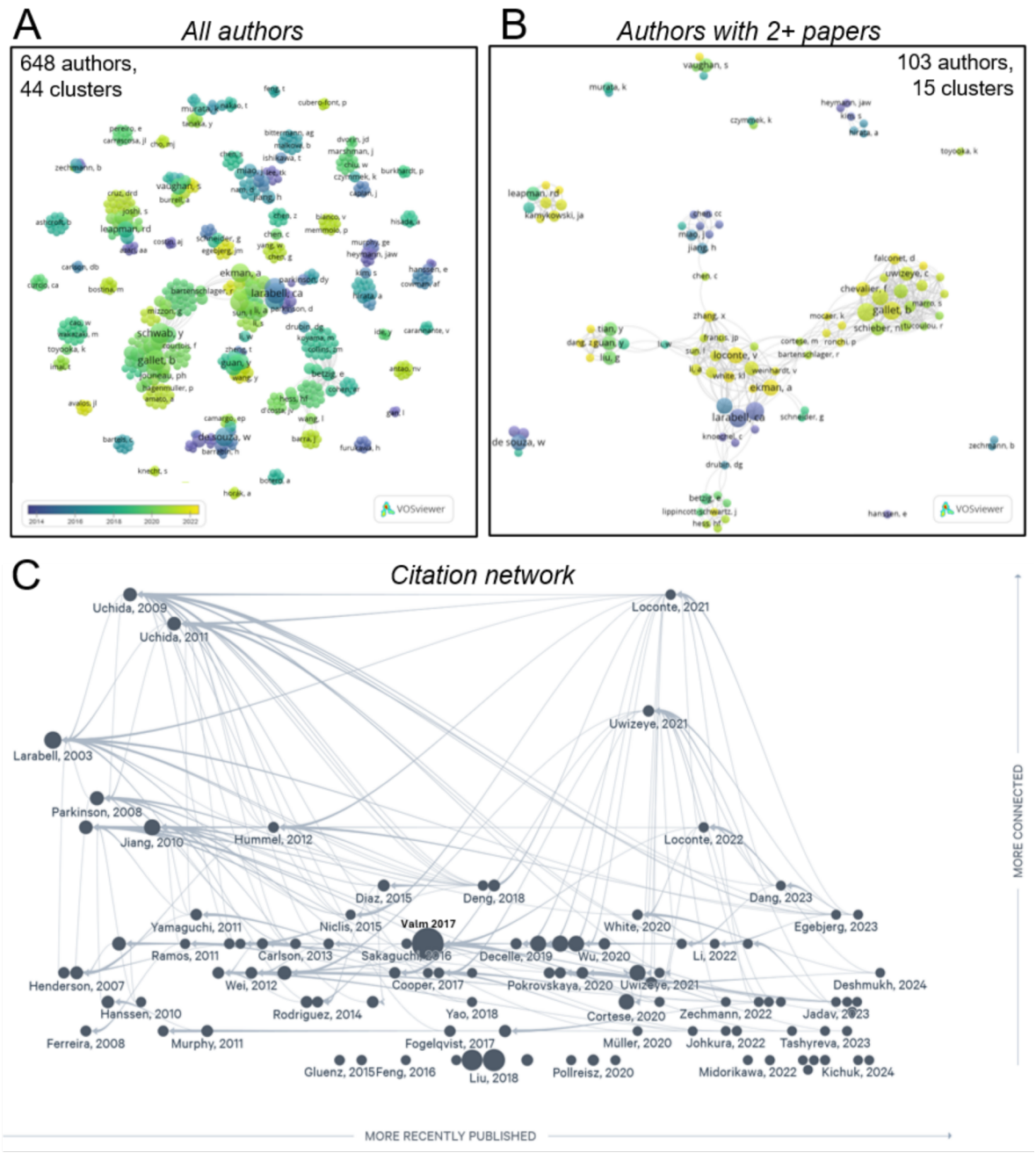
A
All authors
648 authors,
44 clusters
VOSviewer
B
Authors with 2+ papers
103 authors,
15 clusters
VOSviewer
C
Citation network
Uchida, 2009
Uchida, 2011
Loconte, 2021
Uwizeye, 2021
Larabell, 2003
Parkinson, 2008
Jiang, 2010
Hummel, 2012
Loconte, 2022
Diaz, 2015
Deng, 2018
Dang, 2023
Yamaguchi, 2011
Niclis, 2015
Valm 2017
White, 2020
Egebjerg, 2023
Ramos, 2011
Carlson, 2013
Sakaguchi, 2016
Decelle, 2019
Wu, 2020
Li, 2022
Henderson, 2007
Wei, 2012
Cooper, 2017
Pokrovskaya, 2020
Uwizeye, 2021
Deshmukh, 2024
Hanssen, 2010
Rodriguez, 2014
Yao, 2018
Cortese, 2020
Zechmann, 2022
Jadav, 2023
Ferreira, 2008
Murphy, 2011
Fogelqvist, 2017
Müller, 2020
Johkura, 2022
Tashyreva, 2023
Gluenz, 2015
Feng, 2016
Liu, 2018
Pollreisz, 2020
Midorikawa, 2022
Kichuk, 2024
MORE CONNECTED
MORE RECENTLY PUBLISHED

Figure S1.

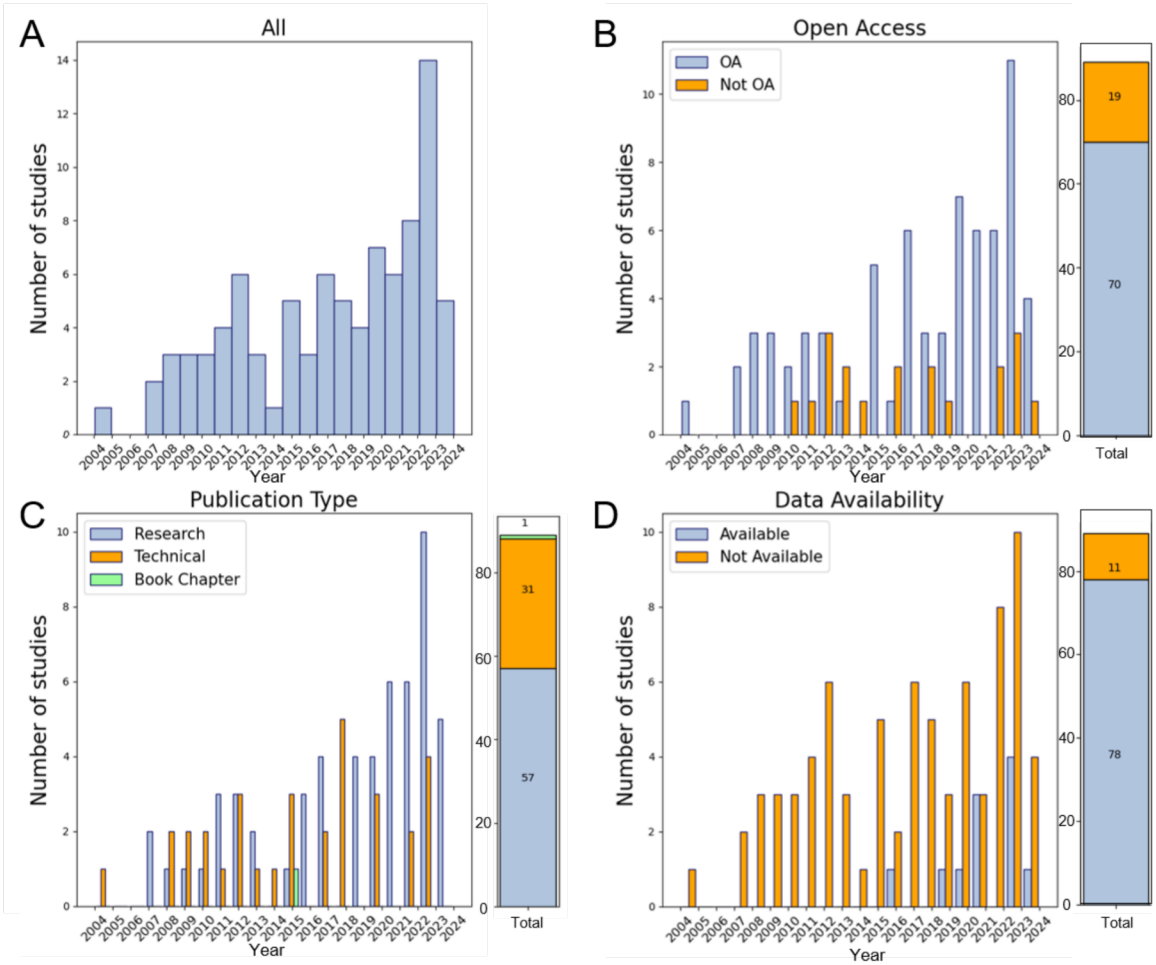
A
All
Number of studies
Year
B
Open Access
OA
Not OA
Number of studies
Year
Total
19
70
C
Publication Type
Research
Technical
Book Chapter
Number of studies
Year
Total
1
31
57
D
Data Availability
Available
Not Available
Number of studies
Year
Total
11
78

Figure 2.

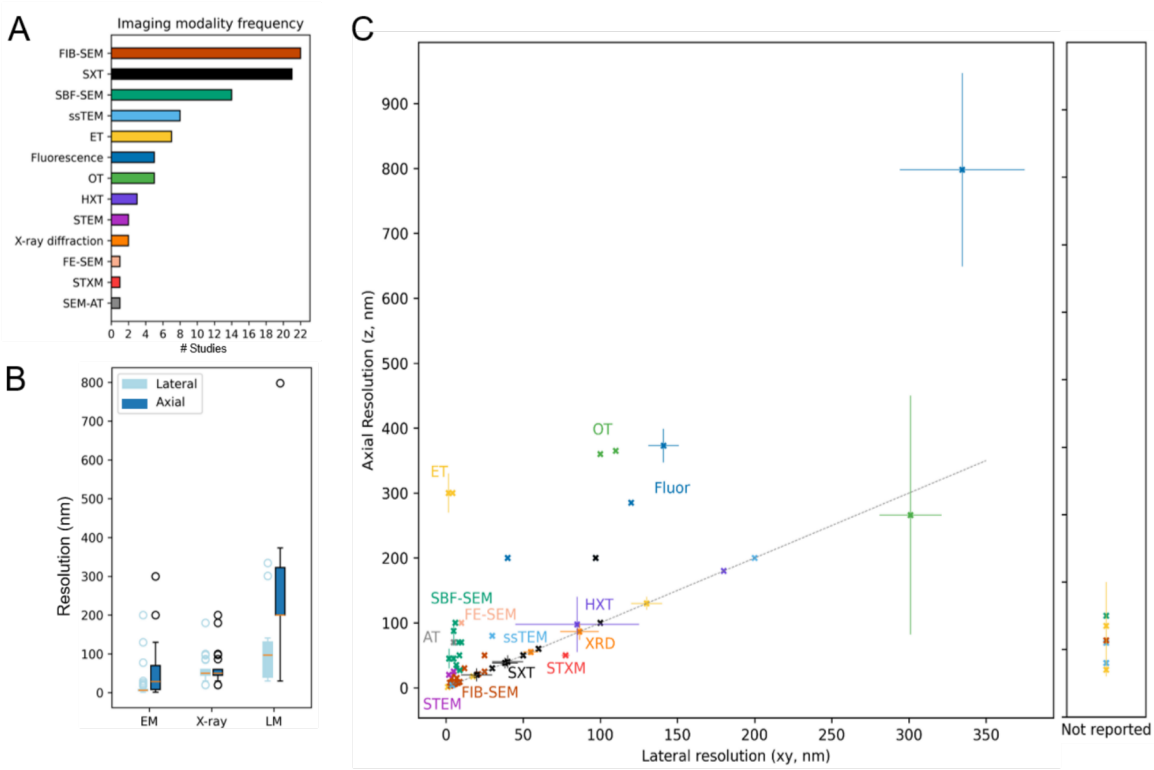
A
Imaging modality frequency
FIB-SEM
SXT
SBF-SEM
ssTEM
ET
Fluorescence
OT
HXT
STEM
X-ray diffraction
FE-SEM
STXM
SEM-AT
0 2 4 6 8 10 12 14 16 18 20 22
Studies
B
Lateral
Axial
Resolution (nm)
EM
X-ray
LM
C
Axial Resolution (z, nm)
Lateral resolution (xy, nm)
OT
ET
Fluor
SBF-SEM
FE-SEM
HXT
AT
ssTEM
XRD
SXT
STXM
FIB-SEM
STEM
Not reported

Figure 3.

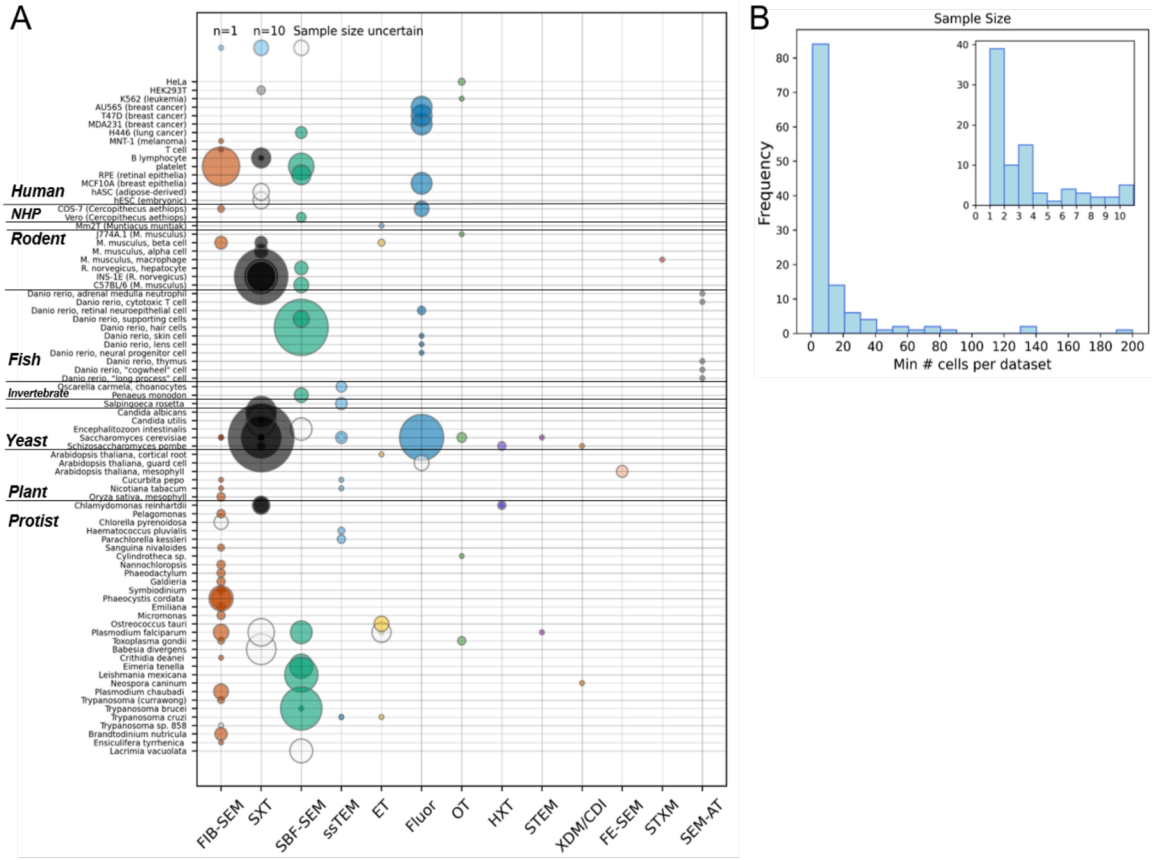
A
n=1
n=10
Sample size uncertain
Human
NHP
Rodent
Fish
Invertebrate
Yeast
Plant
Protist
FIB-SEM
SXT
SBF-SEM
ssTEM
ET
Fluor
OT
HXT
STEM
XDM/CDI
FE-SEM
STXM
SEM-AT
B
Sample Size
Frequency
Min # cells per dataset

Figure 4.

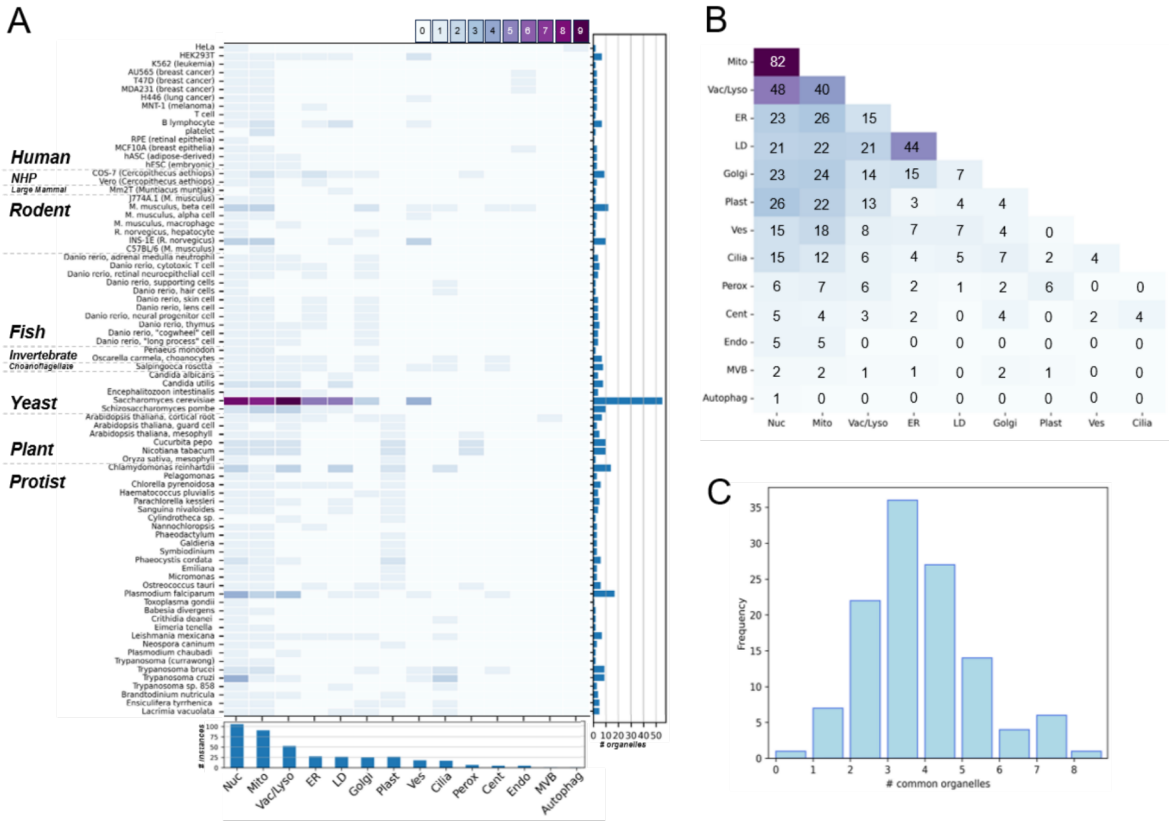
A
0 1 2 3 4 5 6 7 8 9
Human
NHP
Large Mammal
Rodent
Fish
Invertebrate
Choanoflagellate
Yeast
Plant
Protist
instances
organelles
Nuc
Mito
Vac/Lyso
ER
LD
Golgi
Plast
Ves
Cilia
Perox
Cent
Endo
MVB
Autophag
B
Mito 82
Vac/Lyso 48 40
ER 23 26 15
LD 21 22 21 44
Golgi 23 24 14 15 7
Plast 26 22 13 3 4 4
Ves 15 18 8 7 7 4 0
Cilia 15 12 6 4 5 7 2 4
Perox 6 7 6 2 1 2 6 0 0
Cent 5 4 3 2 0 4 0 2 4
Endo 5 5 0 0 0 0 0 0 0
MVB 2 2 1 1 0 2 1 0 0
Autophag 1 0 0 0 0 0 0 0 0
Nuc Mito Vac/Lyso ER LD Golgi Plast Ves Cilia
C
Frequency
common organelles

Figure 5.

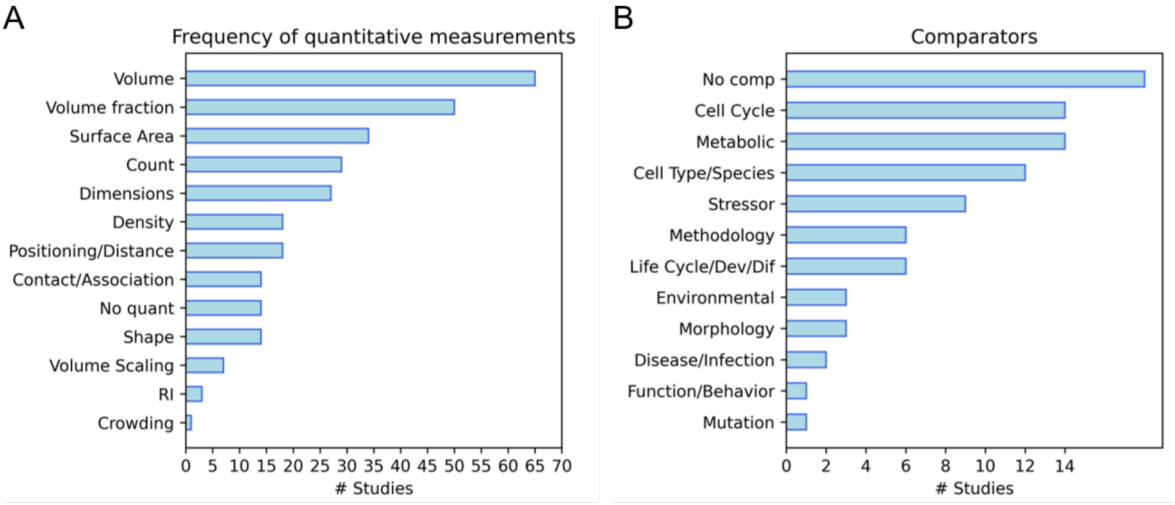
A
Frequency of quantitative measurements
Volume
Volume fraction
Surface Area
Count
Dimensions
Density
Positioning/Distance
Contact/Association
No quant
Shape
Volume Scaling
RI
Crowding
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70
Studies
B
Comparators
No comp
Cell Cycle
Metabolic
Cell Type/Species
Stressor
Methodology
Life Cycle/Dev/Dif
Environmental
Morphology
Disease/Infection
Function/Behavior
Mutation
0 2 4 6 8 10 12 14
Studies

Figure 6.