The Biodiversity Cell Atlas: mapping the tree of life at cellular resolution

https://doi.org/10.1038/s41586-025-09312-4

Received: 13 November 2024

Accepted: 20 June 2025

Arnau Sebé-Pedrós[1,2,3,4] ✉, Amos Tanay[5,6], Mara K. N. Lawniczak[2], Detlev Arendt[7], Stein Aerts[8,9,10], John Archibald[11], Maria Ina Arnone[12], Mark Blaxter[2], Phillip Cleves[13,14], Susana M. Coelho[15], Mafalda Dias[1,3], Casey Dunn[16], Anamaria Elek[1], Jonathan Frazer[1,3], Toni Gabaldón[4,17,18], Jesse Gillis[19,20], Xavier Grau-Bové[1], Roderic Guigó[1,3], Oliver Hobert[21], Jaime Huerta-Cepas[22], Manuel Irimia[1,3,4], Allon Klein[23], Harris Lewin[24,25], Christopher J. Lowe[26], Heather Marlow[27], Jacob M. Musser[28,29], László G. Nagy[30], Sebastián R. Najle[1], Lior Pachter[31,32], Sadye Paez[33], Irene Papatheodorou[34,35], Michael J. Passalacqua[36], Nikolaus Rajewsky[37,38,39,40], Seung Y. Rhee[41,42,43,44], Thomas A. Richards[45], Tatjana Sauka-Spengler[46,47], Lauren M. Saunders[48], Eve Seuntjens[49], Jordi Solana[50,51], Yuyao Song[52], Ulrich Technau[53], Bo Wang[54] & Biodiversity Cell Atlas meeting participants*

Cell types are fundamental functional units that can be traced across the tree of life. Rapid advances in single-cell technologies, coupled with the phylogenetic expansion in genome sequencing, present opportunities for the molecular characterization of cells across a broad range of organisms. Despite these developments, our understanding of eukaryotic cell diversity remains limited and we are far from decoding this diversity from genome sequences. Here we introduce the Biodiversity Cell Atlas initiative, which aims to create comprehensive single-cell molecular atlases across the eukaryotic tree of life. This community effort will be phylogenetically informed, rely on high-quality genomes and use shared standards to facilitate comparisons across species. The Biodiversity Cell Atlas aspires to deepen our understanding of the evolution and diversity of life at the cellular level, encompassing gene regulatory programs, differentiation trajectories, cell-type-specific molecular profiles and inter-organismal interactions.

Genome sequencing has transformed biology in the past three decades. The sequencing of the first eukaryotic genome, *Saccharomyces cerevisiae*[1], in 1996 was soon followed by the first animal and plant genomes: the nematode *Caenorhabditis elegans*[2] in 1998 and *Arabidopsis thaliana*[3] in 2000, and with the Human Genome Project releasing a first draft in 2001 (ref. 4). In recent years, we have witnessed a rapid phylogenetic expansion of high-quality, chromosome-scale genome sequences, driven by the Earth BioGenome Project[5,6]. With more than 10,000 eukaryotic species sequenced so far, the achievements of biodiversity genome sequencing have been extraordinary, transforming fields such as molecular and population genetics, evolutionary biology, biotechnology and medical research[7]. Moreover, biodiversity genomics is key to decipher evolutionary relationships among organisms[8], providing an indispensable phylogenetic framework for comparative biology[9].

One of the most important applications of genome sequencing is to define and compare gene repertoires, transposable elements, conserved non-coding sequences and chromosomal organization among species. However, genomes can also inspire a deeper understanding of biological function, acting as encoders of complex phenotypes driven by the collective activity of thousands of genes. Unlike the genetic code that defines proteins from DNA sequences, the codes underlying higher-order genomic function remain exceedingly difficult to decipher[10]. This has led to an increasing divide between our knowledge of genomes across the tree of life and our limited understanding of how these genomes encode diverse biological functions[11].

This gap in functional genomics also underlies a fundamental challenge in evolutionary biology[12]: connecting molecular evolution with phenotypic changes observed in subcellular structures, cell types, life cycles, morphologies and behaviours. The comprehensive scope of modern genomics, combined with centuries of detailed phenotypic observations, demands consolidation and synthesis. However, the complexity of the code linking genome sequence to functional traits at present impedes this synthesis. The Biodiversity Cell Atlas (BCA) initiative will attempt to bridge this gap by characterizing cellular diversity across the tree of life through the lens of molecular cell atlases. To achieve this, the BCA aims to coordinate sampling priorities and fieldwork, standardize experimental and data analysis procedures, scale up single-cell data production across species and, when needed, drive technological development to support these efforts.

Decoding genomes through single cells

Cells are the native 'decoders' of genomes, robustly integrating developmental and environmental signals to functionally interpret genomic information. This dynamic and flexible process allows a single genome to give rise to many cellular functions by modulating the activity of all genes. Although genomes can theoretically attain an almost unlimited

---

A list of affiliations appears at the end of the paper. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: arnau.sebe@crg.eu

Box 1

Cell types, cell states and gene modules

The morphological and molecular regularities observed in cells within and across organisms suggest the existence of cell types[52,129–131]: distinct entities with structural and functional properties that are reproducible across generations through embryonic development or throughout the life cycle. Cell types can be viewed as evolutionary units, traceable through phylogeny by the specific deployment of conserved genetic programs. Within a species, cells of the same type may show considerable variability—referred to as cell states—arising from developmental processes, environmental stimuli or disease conditions. Clustering of molecular profiles obtained by single-cell methods enable us to map the rich repertoire of cell states that comprise this variation. By broadly sampling across species, the BCA will contribute to refining the conceptual framework that describes how cell types originate from cell states and diversify through evolutionary processes. At an even finer level of granularity, we may define gene modules: groups of genes that work together to drive specific functions (for example, neurotransmission or digestive metabolism) or to form distinct cellular structures (for example, flagella, hyphae or extracellular matrix components). In contrast to the hierarchical definition of cell types and states, some of these modules operate combinatorially across different cell types, as seen with processes such as the cell cycle or innate immune responses.

number of activity states—given that the number of possible states is at least exponential to the number of genes—natural selection and other evolutionary forces notably restrict this functional repertoire. The result is a rich, yet limited, set of genome-encoded cellular states in every species, which can be classified into cell types (Box 1). Functionally, these cell types underlie organized tissues and life cycles. Experimentally, the bounded complexity of cellular states implies that it is possible to map them comprehensively and derive detailed cell atlases, defined here as systematic catalogues of molecular profiles from individual cells within a biological system. Indeed, the advent of single-cell omics technologies has enabled the classification of cell types based on genome activity states within an organism (Fig. 1). These technologies can quantify thousands of transcripts in individual cells, either after dissociation (single-cell RNA sequencing (scRNA-seq))[13] or in their tissue context (spatial transcriptomics[14]); or profile chromatin accessibility at thousands of *cis*-regulatory elements (single-cell assay for transposase-accessible chromatin with sequencing)[15]. When scaled to millions of cells and combined with computational methods for grouping single-cell profiles into cell states and types, single-cell genomics approaches facilitate the charting of cell atlases, as demonstrated by the Human Cell Atlas consortium[16,17] and similar efforts in model organisms[18–24].

By integrating biodiversity genome sequencing and single-cell omics, the BCA initiative aims to acquire molecular snapshots for millions of cells to infer thousands of genomic activity states across hundreds of species. Even before considering the effect these data will have on our understanding of eukaryotic functional diversity and evolution, the availability of such an unprecedented corpus of ‘decoded’ genomes will propel functional genomics into the artificial intelligence era. To understand why, one can revisit the factors contributing to solving the protein folding problem[25]. The adaptation of artificial intelligence technologies to biology relies on massive datasets that sample across phylogenetic diversity. In the case of protein structure, the curation of vast numbers of protein sequences from across the tree of life enabled the inference of covariation signals, which in turn could be used to predict the three-dimensional structure of the protein from its sequence[26]. These predictions were then refined by leveraging thousands of solved protein structures sampling a large fraction of known families of proteins across the tree of life. The BCA aims to deliver a similar level of quality in measuring genomic molecular activities and cellular states across species representing eukaryotic phylodiversity. By doing this, the BCA data will empower new tools to link genomes with molecular states through computational models that could not be foreseen thus far. Ultimately, this has the potential to enable the in silico decoding of phenotypes from genomes and to reveal how genomic changes affect the evolution of cellular phenotypes.

Why map cell types across eukaryotes?

The systematic interrogation of cell type molecular phenotypes across the tree of life offers unique opportunities for impactful discoveries in many fields, including biotechnology and biomedicine applications, basic research and evolutionary theory, and environmental sciences (Fig. 2).

An important area of BCA applications will involve the development of new technologies based on the unprecedented characterization of new biomolecules and biological systems in understudied organisms. Cell atlases put genes in context: when and where a gene is expressed, and which genes are co-expressed together. This information will underpin new hypotheses about the functions of uncharacterized genes across biodiversity, potentially leading to industrial, agricultural or biomedical applications. These could include the discovery of proteolytic enzymes, various antimicrobial peptides, viral defence systems, genes involved in efficient metabolism and energy use (for example, C3 and C4 photosynthesis in plants[27,28]) and much more. In addition to this reservoir of evolutionary ‘biotechnological’ innovation that the BCA will help to uncover, the comprehensive functional annotation of genomes will provide a rich source of high-quality data for developing artificial intelligence models in biology[29–33], with applications that are only recently start to emerge such as the design of synthetic regulatory sequences and circuits[34,35]. Comparative cell atlases will also empower the inference of gene-to-gene functional dependencies by revealing conserved co-expression patterns across species. Using evolutionary diversification as a perturbation experiment[9], phylogenetic gene expression data can eventually be used to develop models that predict expression variation pathogenicity in humans, analogous to how protein sequence conservation is used to predict disease-associated protein variants[36].

The BCA will have diverse effects on basic research. For example, cell atlases will uncover new biology in under-investigated species, becoming a powerful tool for generating hypotheses about the function of both genes and cells. Atlases will help to discover previously unknown cell types and gene modules, while providing information about the degree of cell type specificity and the pleiotropic usage of genes, enabling the efficient selection of cell type markers. In addition, a cell atlas facilitates unbiased phenotyping of experimental perturbations and environmental variation by means of projection of the re-sampled conditions to the reference atlas. Therefore, a reference atlas, much like a reference genome, becomes an indispensable resource for emerging model organisms. Single-cell atlases are also an essential first step towards a comparative molecular biology of cell types that will transform the study of cell type evolution. A deeper understanding of the evolutionary processes that define cell types can underpin more robust cell type classification schemes, on the basis of maximally informative molecular traits[37]. This phylogenetic taxonomy of cell types will be immediately useful for developing unified cell type nomenclatures, guiding cross-species comparative analyses, and refining the definition of what constitutes a cell type.

Finally, we anticipate that advances within the BCA initiative will enable the application of single-cell technologies to study and monitor

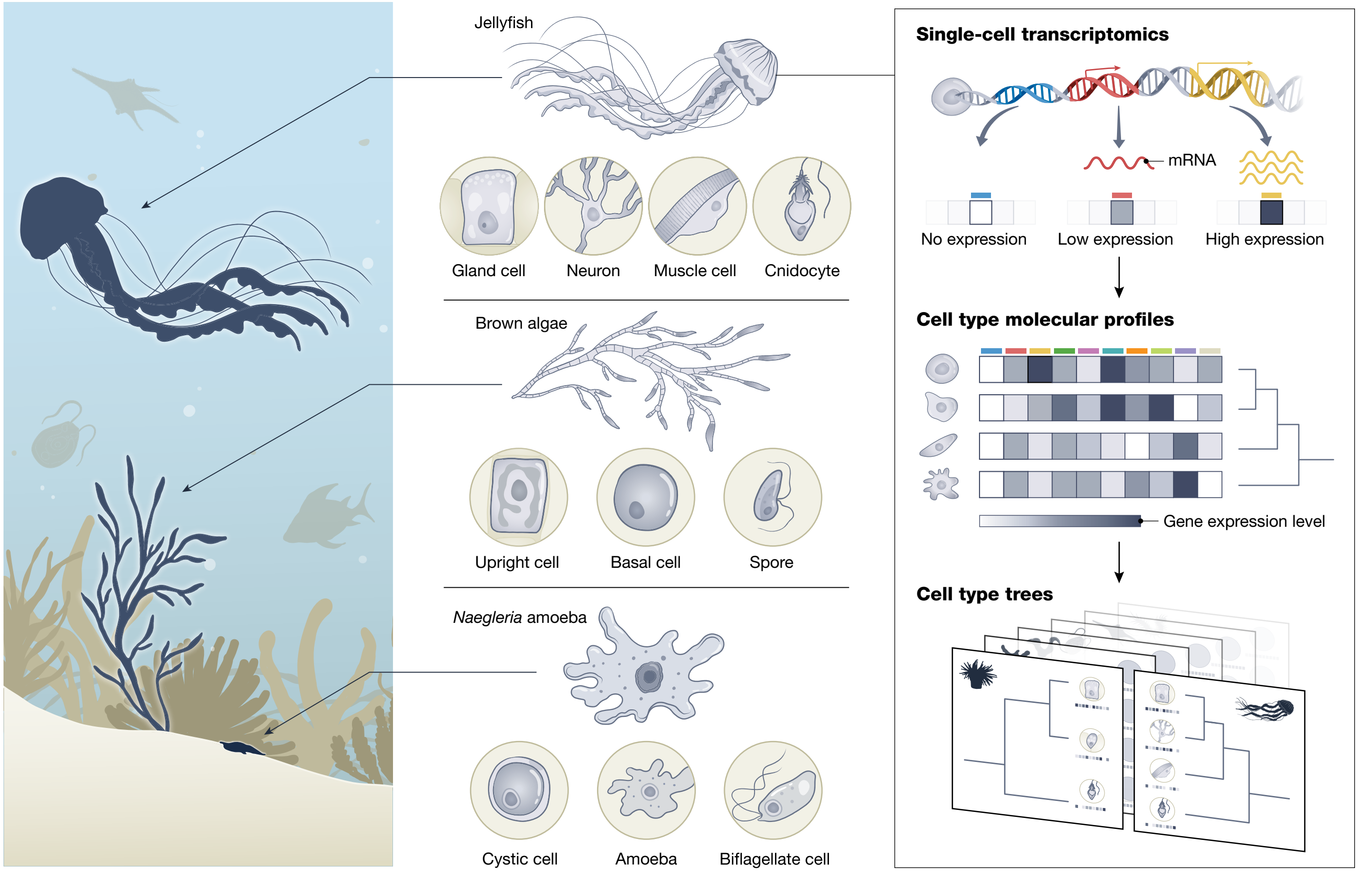

Fig. 1 | The BCA aims at molecularly characterizing cell types across the eukaryotic tree of life. Molecular profiles derived from single-cell transcriptomics (or other data modalities such as chromatin accessibility) capture the gene expression patterns of individual cells. By clustering similar cells, cell types and states can be identified and organized into hierarchies that reflect their molecular similarities. mRNA, messenger RNA.

species in their natural environment. For example, single-cell methods can reveal the often-cryptic life cycles of microbial eukaryotes by dissecting the diversity and developmental dynamics of cell states in the wild. This is crucial for two related reasons: only a small fraction of microbial eukaryotes can be cultured in the laboratory[38], and cell type transitions in protists are often triggered by unknown environmental cues absent in laboratory culture conditions. Cell atlases can also uncover symbiotic interactions between organisms in a high-throughput manner, identifying the constituents of holobiont assemblies and their molecular interactions with cellular resolution[39,40]. This has been demonstrated by single-cell studies of coral–dinoflagellate symbiosis[41,42], of arbuscular mycorrhizal symbioses[43] and of marine giant virus infection dynamics[44,45]. Beyond basic research, we predict further development of single-cell methods as powerful biomonitoring tools. Compared with environmental metagenome[46–48] and metatranscriptome[49] sequencing methods, single-cell approaches should allow researchers to determine which species are present in a sample while also interrogating organismal interactions, transcriptional dynamics and life-cycle stages (for example, metabolic contributions, cell cycle, dormant cystic states). Overall, the capacity to characterize complex ecosystems at single-cell resolution and monitor how they change over time and under different environmental stressors will help to inform conservation strategies and policymaking.

Both basic and translational research applications of the BCA represent vast opportunities. However, as the BCA is focused on charting unknown and new biology, it is also poised to bring about completely unexpected discoveries, driving new theories, ideas and interdisciplinary science.

Eukaryotic cellular diversity

Specialized cell types underlie cooperative functions in multicellular organisms and complex life cycles in microbial eukaryotes[12,50–52]. In animals, plants, fungi and multicellular algae, cells co-exist spatially and typically arise from successive cell divisions coupled with differentiation from an initial, pluripotent single cell. In addition, aggregative multicellularity occurs throughout the eukaryotic tree of life[51], as seen, for example, in the dictyostelid amoebozoans. Unicellular eukaryotes, or protists, show cells with intricate morphologies and unique physiological adaptations[53] that are far from static; most have life cycles involving temporally differentiated cell types. For example, the discoban *Naegleria gruberi* transitions between amoeba, cystic and bi-flagellated cells[54]; choanoflagellates show transitions involving rosette colonies, solitary swimmers, thecate and amoeboid forms[55–57]; and many parasitic protists have intricated temporal adaptations to different hosts, tissues and symbiotic partners[58].

Scientists have morphologically characterized cells across organisms since the invention of early microscopes. This led to the identification and classification of protist species[59,60] and the description of diverse cell types, including neurons[61,62] and other specialized cells[63]. Cell types were first recognized morphologically by size, shape, organelle content, and structure and tissue context (Box 1). These cellular phenotypes can be conserved across vast evolutionary distances. For instance, neurons, sensory cells, muscle fibres and epidermal cells are morphologically recognizable in distant animal phyla[64] and these observations inspired the idea of 'cell families' and the possibility of studying cell type evolution through comparative cytology[65].

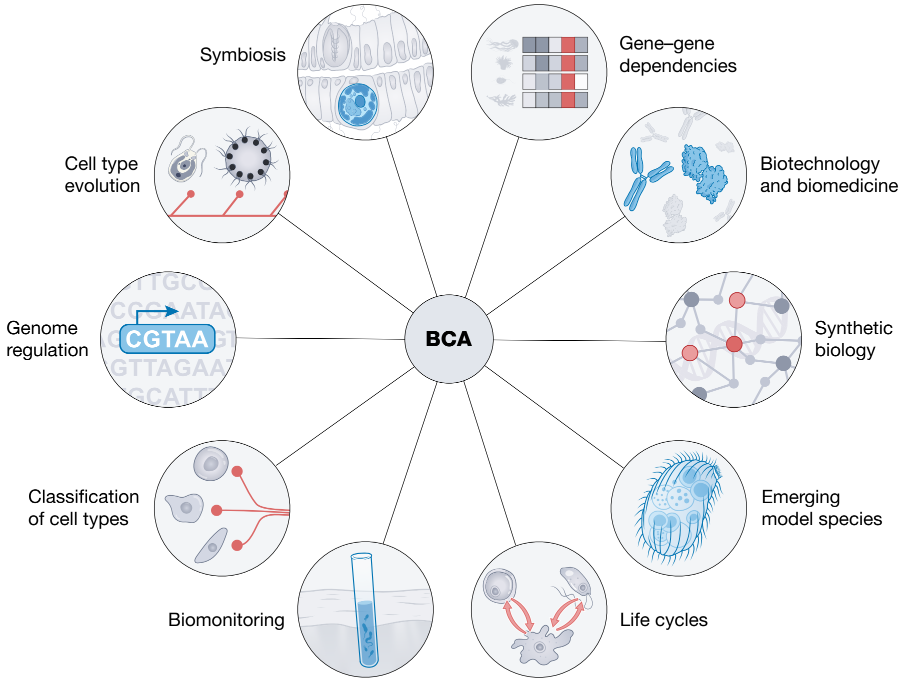

Fig. 2 | BCA anticipated impacts. Systematic single-cell atlasing across species have the potential of transformative discoveries in biotechnology, biomedicine, evolutionary biology and environmental science. By contextualizing gene expression across cell types and organisms, cell atlases can uncover new biology, support functional genome annotation, inform artificial intelligence models and facilitate biomonitoring of ecosystems, while opening doors to unexpected, interdisciplinary breakthroughs.

The advent of histochemical and molecular profiling tools further extended the taxonomic profiling of cell types across species[66–68]. These tools include detecting specific proteins using antibody-based immunostaining[69] and specific transcripts using RNA in situ hybridization for marker genes. Molecular characterization then extended to bulk transcriptomics and epigenomics profiling for isolated or enriched cell types[70–72]. However, it was only with the advent of single-cell omics methods that the systematic molecular characterization of cell types within and across organisms became feasible. Whole-organism single-cell expression profiles can be organized into cell type hierarchies and, in some cases, differentiation trajectories, thus creating a cell atlas for the organism.

The full potential of single-cell approaches for discovering unexpected biology and understanding cellular diversity will only be realized with a substantial increase in the taxonomic sampling of cell atlases[73] (Fig. 3). The first organism-wide cell atlases across the tree of life demonstrate the power of single-cell analysis to uncover new biology in understudied species[42,74–86], in a way comparable to the discoveries enabled by genome sequencing. Cell atlases provide a data-driven, operational definition for cell types and states (Box 1) and create comprehensive molecular catalogues expected and new cell types. Cell atlases facilitate further validation and visualization of cell types using atlas-derived markers (for example, cell-type-specific transcripts and/or proteins) and enable the formulation of hypotheses about functional states and roles on the basis of expressed gene repertoires. Furthermore, atlas-defined cell states and types, especially when analysed jointly with the genome sequence, can reveal gene–gene regulatory relationships and putative functional gene programs[87], supporting comparative analysis between types and species at multiple levels[37].

Cell atlases in evolutionary studies

Cell atlases are poised to be central to our understanding of organismal biology and evolution[52,73]. Historically, the phenotypic similarities observed between cells within and across organisms suggested the existence of conserved cell types and inspired the first hierarchical cell type taxonomies[64,65]. Within organisms, the hierarchical nature of cell types derives from shared ontogenetic trajectories[88] as well as from functional similarities and shared effector gene expression in differentiated cells[87]. However, quantifying cell type similarities (and dissimilarities) across organisms has been traditionally challenging due to the lack of adequate cell-type-defining traits that can be systematically measured and compared. For example, comparing ultrastructural features or biochemical compositions of cell types across distantly related species is not straightforward and is limited to a low number of traits. Cell atlases provide quantitative traits to compare cell types and study cell type evolution, revealing the genetic basis of cellular identity through thousands of molecular measurements. For example, similarities between cell types in different organisms can be defined by comparing the expression of orthologous genes[73,89] or the *cis*-regulatory sequences that control cell-type-specific gene expression[90,91]. These comparisons based on molecular traits can be used to build cross-species cell type hierarchies[42,74,77,92,93] and formulate hypotheses about cell type evolution[94].

However, whereas cell type hierarchies can be derived from molecular data, such inference is not based on explicit evolutionary models[37,73,95]. This is because we lack reliable estimates for the divergence rates and the evolutionary regimes (drift, selection) of both lower-level traits (gene expression, *cis*-regulatory codes, coregulated gene programs) and the broader molecular phenotype (specialized cell types). Consequently, it remains difficult to assess cell type homology, differentiate convergent similarities in non-homologous cell types and identify rapidly diverging homologous cell types. To address these challenges, we need to systematically measure comparable cellular traits in densely sampled phylogenetic trees, along a gradient of divergence times. The BCA will collect and curate such comprehensive datasets, offering a unique opportunity for studying cell type evolutionary relationships, developing new theoretical models and linking genome-level evolutionary processes to phenotypic changes. These standardized datasets will empower the study of cell type origins,

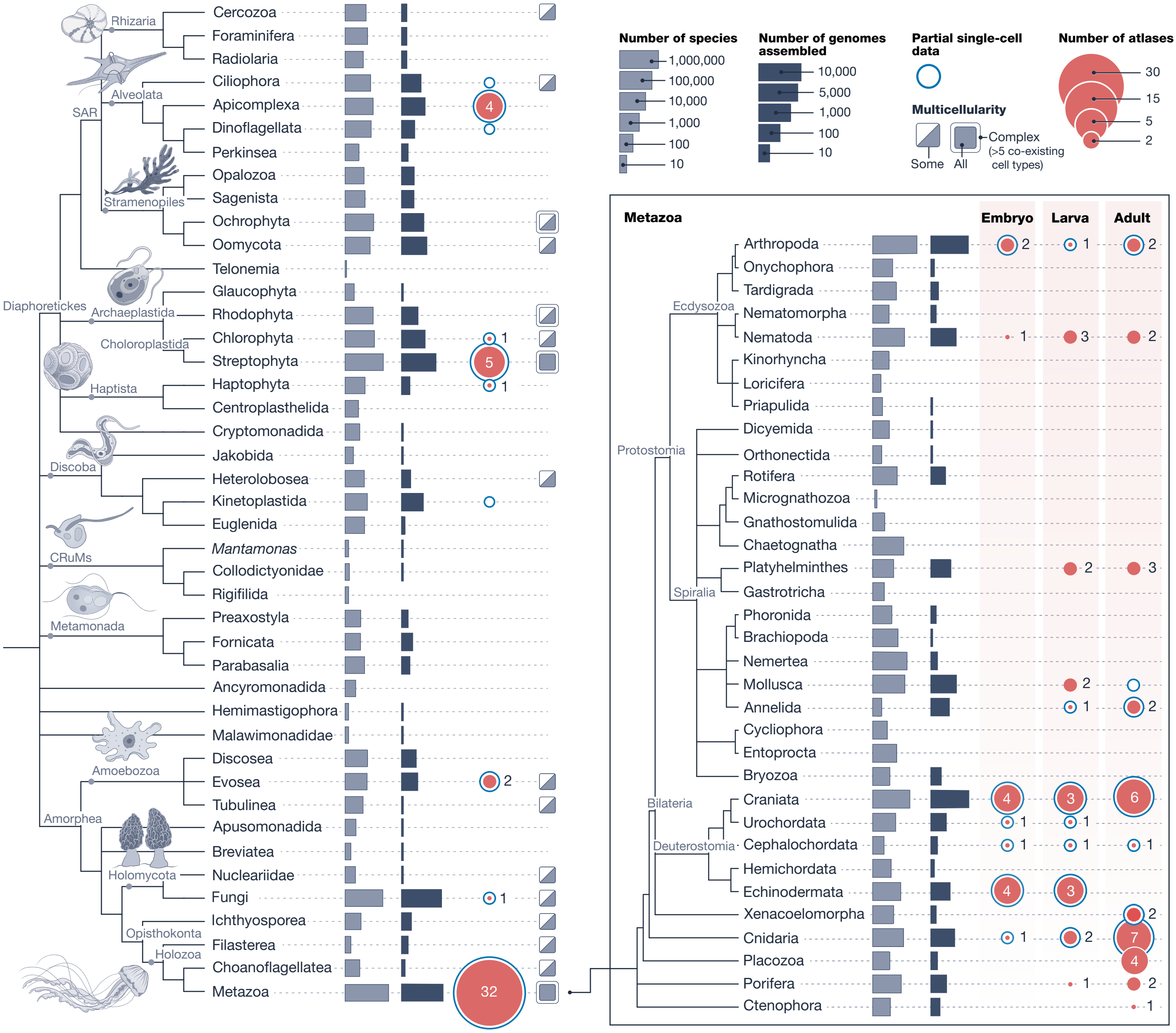

Fig. 3 | Phylogenetic state-of-the-art of single-cell atlases across eukaryotes. For major eukaryotic (left) and animal (right) lineages, bar plots indicate the number of estimated species and sequenced genomes (based on GoaT[128] (https://goat.genomehubs.org/)). The presence of single-cell data is marked by blue circumferences and red circles denote the number of available whole-organism cell atlases. Partial single-cell data (blue circumference only) refers to datasets limited to specific tissues (for example, animal blood or brain, or plant roots) or particular life stages (for example, in unicellular eukaryotes). For animals, we distinguish the availability of whole-body atlases for embryonic stages, postembryonic larval stages (where applicable) and adults. CRuMs, Collodictyonidae, Rigifilida and Mantamonadida supergroup; SAR, Stramenopila, Alveolata, Rhizaria supergroup.

novelties and the evolution of functional gene programs and cellular differentiation trajectories. The relevance to numerous long-standing evolutionary questions will be immediate: for example, it may become possible to address the origin of neurons[96,97] from several perspectives, including functional gene modules, neurogenesis programs and cell identity *cis*-regulatory codes. Ultimately, the BCA will allow us to connect emerging cell types and states to genome sequence evolution, linking the continuity of cellular phenotypes to the historical continuity of genetic information[98].

BCA goals and strategies

The BCA initiative is a coordinated international effort aimed at molecularly characterizing cell types across the eukaryotic tree of life (Fig. 3). The BCA initiative was launched at a meeting in 2023 that brought together leading experts in biodiversity genomics, existing cell atlas initiatives and comparative single-cell genomics. During this and subsequent meetings, the challenges and opportunities were discussed, and this Perspective sets out to share these. Working groups have been established to focus on three key areas: taxonomic prioritization, single-cell technologies, and data analysis and integration (www.biodiversitycellatlas.org has more information on how to participate). The initial BCA phase 0 (2024–2026) focuses on developing experimental workflows and computational resources essential to scale up data production in phase 1 (2026 and beyond).

A first goal of the BCA is to establish species sampling criteria and to coordinate taxonomic coverage among initiative members. A primary criterion is to maximize phylogenetic diversity by prioritizing unsampled lineages across the tree of life. A practical approach to this is focusing on species with high-quality genomes either available or being sequenced under the Earth BioGenome Project[5] and associated projects. A second criterion is to study organisms at key phylogenetic positions to gain insights into major transitions in cellular complexity, such as independent origins of multicellularity (for example, brown algae, plants, animals, fungi), bursts in cell type diversity (for example, neurons in bilaterian nervous systems, new germ layers) and complex unicellular life cycles and symbiotic interactions. The BCA will also prioritize the creation of high-quality cell atlases as foundational resources to support emerging model species. In addition, we will undertake deep taxonomic samplings through pilot projects aimed at extensive phylogenetic analysis to develop quantitative models for understanding cell type molecular evolution. Finally, the BCA will target organisms that pose considerable experimental challenges (for example, species with unknown cell wall compositions, exoskeletons, low cell numbers or hard tissues) to foster technological developments that can eventually enhance single-cell sampling in other difficult-to-study species.

A second goal of the BCA is to consolidate technologies and standardize procedures to scale up production of cell atlases. As part of BCA phase 0, benchmarking studies are underway to develop a BCA methods decision tree. This resource will offer guidance and detailed protocols for sampling, preserving, dissociating and sequencing new species. These studies will also identify experimental bottlenecks, highlighting the need for future technology developments. Initially, BCA efforts will focus on producing scRNA-seq atlases, with the goal of progressively incorporating other technologies, such as single-cell assay for transposase-accessible chromatin with sequencing, as they become scalable to a broader range of species. In addition, we contemplate the extension of single-cell atlases (which do not provide spatial context) into spatial maps of cell types within tissues, organs and entire organisms using spatial transcriptomics and high-throughput imaging techniques[99,100].

Finally, the BCA will develop new data infrastructure to analyse, model and compare cell atlases, as well as to efficiently disseminate the generated data and knowledge. Best practices in data analysis, quality control metrics and data standards will be established by the BCA working groups, along with standardized cell type annotations and classifications for any eukaryotic species. To achieve these goals, phase 0 of the BCA involves building unified single-cell data processing pipelines using Nextflow[101], designed to work across species and platforms. Furthermore, a dedicated database and portal are under development to enable flexible access and multi-level exploration of BCA datasets across species.

The challenges faced by the BCA are complex and will require interdisciplinary collaboration, new experimental and computational methodologies, and coordinated fieldwork. Thus, a primary objective of the BCA initiative is to establish a platform that bridges single-cell genomics with biodiversity expertise, creating a forum to exchange and discuss practical information and first-hand experiences on these critical aspects. Recognizing that invaluable genetic resources—species for which the atlases will be developed—are often found in low and middle income countries, the BCA is committed to adhering to the benefit-sharing principles outlined in the Nagoya Protocol and will involve local communities throughout the cell atlas creation process, from sample collection to data analysis. This global community will facilitate the efficient and rapid expansion of single-cell atlases across the eukaryotic tree of life.

BCA technical challenges

Several technical challenges contribute to explain the at-present limited phylogenetic representation of single-cell atlases across eukaryotes (Fig. 3). The BCA initiative must tackle these constraints to scale time-consuming species-specific optimizations and drive rapid expansion of cell atlases.

BCA experimental challenges

One of the primary challenges in single-cell methods is the extraction of single cells or nuclei from intact organisms. Methods and conditions can vary between species and can severely affect the quality of the single-cell data at several levels, such as inducing transcriptional stress responses[102], biasing against fragile cell types and causing the release of RNA and RNases into the sample. The most common cell dissociation strategy involves enzymatic digestion of fresh tissue. After digestion, cell death is evaluated using cell vitality staining, and samples with more than 5–10% of dead cells are generally discarded. In addition, large clumps of undissociated cells are filtered out before capture, although this step may not eliminate small clusters of non-dissociated cells. An extra consideration for marine organisms is osmotic stress. High salt concentrations can inhibit scRNA-seq reactions, such as reverse transcription, and alternative osmotic agents such as mannitol[103] or sorbitol[77] have been successfully used as substitutes.

There are two main alternatives to dissociating fresh tissues: fixation strategies and nuclei extraction. Whole-organism or tissue fixation prevents stress responses induced by digestion, osmotic stress and other dissociation-related artefacts. Three main types of fixation have been successfully used in scRNA-seq: methanol-based methods (for example, pure methanol[78,104,105] or methanol mixed with acetic acid and glycerol[106]); crosslinker-based methods using agents such as formaldehyde[107], glyoxal[108] and dithiobis-succinimidyl propionate (DSP)[109]; and methods using deep eutectic solvents[110]. After fixation, cells are dissociated mechanically using sonication[110], digestive enzymes[109] or a combination of the two. An extra advantage of these methods is that fixed cells can be sorting using fluorescence-activated cell sorting[106] to remove multiplets and ambient RNA. Alternatively, nuclei can be extracted from fresh or flash-frozen tissues[111]. Single-nucleus RNA-seq is often used for challenging tissues such as the brain[112] and muscle fibres[113], and it is the most feasible option for multicellular organisms with cell walls, such as plants[18], brown and red algae, and fungi. However, a potential drawback of single-nucleus RNA-seq is its reduced sensitivity, as the nucleus contains only a fraction of the cellular messenger RNAs and a possibility is to combine single-nucleus RNA-seq for breadth and scRNA-seq for depth[114].

Another important consideration is the efficiency of cell lysis. Whereas mild detergents and hypoosmotic conditions are sufficient to lyse animal cells, this is not the case for other eukaryotes with cell walls. Enzymes such as chitinases and glucanases in yeasts[115] and cellulases in plants[116] can be used to digest cell walls. However, for many eukaryotes, the composition of their cell walls remains unknown and no enzymes are available to generate protoplasts. Two solutions to work with difficult-to-lyse cells are to isolate nuclei or to use plate-based methods[42,45], in which sorted cells can be physically lysed using high temperature, freeze–thaw cycles or sonication. In addition, many single-cell methods require relatively large numbers of cells and have low cell recovery rates, which complicates the sampling of small specimens. A potential solution is to mix genetically different specimens and deconvolve single-cell data by genotype[117].

Whereas a universal protocol for single-cell experiments across all species is unrealistic, a key goal of the BCA initiative is to define a methods decision tree to guide the main steps when sampling a new species. This will include recommendations on the advantages and disadvantages of each approach, specific experimental protocols for each strategy and suggested quality controls to evaluate outcomes (Supplementary Table 1). Applying a unified sampling framework to an increasing number of species will also help identify further constraints and guide technology-development efforts to overcome these challenges.

BCA computational challenges

The BCA initiative aims to standardize and streamline all aspects of single-cell data analysis, including unified pipelines, data formats and data quality descriptors, to enable construction of comparable whole-organism atlases (Supplementary Table 1). A common challenge in single-cell data analysis and interpretation is the inaccuracy of gene annotations, such as missing or partial genes[118,119]. To address this issue, we will modify and extend existing gene annotations and complement short-read scRNA-seq data with long-read sequencing, working together with biodiversity genomics projects to improve existing genome annotations when needed. Beyond gene annotation, accurate gene orthology inference[120,121] will be another important aspect to enable cross-species cell atlas comparisons.

Another key problem is the use of standardized data quality and coverage metrics for single-cell atlases. The BCA will work to establish and promote the use of metrics for atlas coverage and precision, analogous to those used in genome assembly and annotation. For instance, we will develop metrics to assess the saturation of atlases, helping to determine the target numbers of cells needed to ensure robust coverage of an organism's diversity of cell types and states. We will also promote the development of algorithms for imputing missing cell types based on a species' genomic sequence and complete atlases from related species.

The data representations developed by the BCA, as well as standards and software implementing them, will enable access to and flexible manipulation of single-cell atlases across species. To this end, the BCA database will connect existing nomenclatures, such as species and gene names, with new higher-level objects such as cell types and states, co-expressed gene modules or co-accessible regulatory elements. We expect that coherent data generation and processing will strategically foster the development of new ideas and methods for comparative cell atlas analyses, including tools to compare cell type transcriptomes[94,122,123], cross-species cell embeddings[122,123], gene expression evolution models[124–127] and new phylogenetic methods[95]. Even more ambitiously, models for gene regulatory networks, developmental trajectories or modular cellular programs, initially defined for individual species, could be generalized in a comparative context and inform new evolutionary models. Standardized ontologies, quality and coverage metrics are essential to make substantial progress towards these goals.

Overall, adopting shared data processing and metadata standards will significantly enhance the reusability and meta-analysis of cell atlases, whether the aim is to infer the biology of a species of interest, characterize components and interactions within an ecosystem, or tackle the challenge of building a large language model that predicts function from DNA.

Towards a cell type tree of life

The BCA initiative represents the coalescence of biodiversity genomics and single-cell biology, with the ambition to generate cellular atlases at a large phylogenetic scale. We predict initial steps to be focused on optimizing experimental strategies, while simultaneously building the computational infrastructure to efficiently analyse and interpret variation in cell atlases within a phylogenetically informed framework. These initial experiences and design considerations, collectively shared and discussed within the BCA community, will pave the way for the efficient expansion of cell atlases to hundreds of species. Similarly, whereas we initially plan to focus on single-cell gene expression profiling, the methodological expertise developed in this process will lay the groundwork for incorporating extra data modalities, such as single-cell resolved chromatin accessibility and spatial omics.

The BCA comparative perspective, rooted in phylogenetic methods and evolutionary concepts, will help us to systematically describe and interpret cell types, cell states and gene modules across organisms and timescales. This conceptual and practical foundation will have far-reaching effects beyond the creation of molecular catalogues of eukaryotic cell types. The BCA outcomes will provide transformative insights into genome function and evolution, gene regulation, organismal complexity, multicellular ontogeny, eukaryotic life cycles, symbiosis and other biological interactions both in natural and altered environmental conditions. The BCA will venture into unknown biology with tools that have never before been applied systematically to explore it at this scale, holding the potential for unexpected discoveries in one of the major frontiers of modern biology.

1. Goffeau, A. et al. Life with 6000 genes. *Science* **274**, 546–567 (1996).
2. C. elegans Sequencing Consortium.Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
3. Initiative, T. A. G. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
4. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
5. Lewin, H. A. et al. Earth BioGenome Project: sequencing life for the future of life. *Proc. Natl Acad. Sci. USA* **115**, 4325–4333 (2018).
6. Lewin, H. A. et al. The Earth BioGenome Project 2020: starting the clock. *Proc. Natl Acad. Sci. USA* **119**, e2115635118 (2022).
7. Sullivan, P. F. et al. Leveraging base-pair mammalian constraint to understand genetic variation and human disease. *Science* **380**, eabn2937 (2024).
8. Kapli, P., Yang, Z. & Telford, M. J. Phylogenetic tree building in the genomic age. *Nat. Rev. Genet.* **21**, 428–444 (2020).
9. Smith, S. D., Pennell, M. W., Dunn, C. W. & Edwards, S. V. Phylogenetics is the new genetics (for most of biodiversity). *Trends Ecol. Evol.* **35**, 415–425 (2020).
10. Kim, S. & Wysocka, J. Deciphering the multi-scale, quantitative *cis*-regulatory code. *Mol. Cell* **83**, 373–392 (2023).
11. Oliver, S. G. From DNA sequence to biological function. *Nature* **379**, 597–600 (1996).
12. Arendt, D. The evolution of cell types in animals: emerging principles from molecular studies. *Nat. Rev. Genet.* **9**, 868–882 (2008).
13. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620 (2015).
14. Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).
15. Minnoye, L. et al. Chromatin accessibility profiling methods. *Nat. Rev. Meth. Primers* **1**, 10 (2021).
16. Regev, A. et al. The Human Cell Atlas. *eLife* **6**, e27041 (2017).
17. Haniffa, M. et al. A roadmap for the Human Developmental Cell Atlas. *Nature* **597**, 196–205 (2021).
18. Lee, T. A. et al. A single-nucleus atlas of seed-to-seed development in *Arabidopsis*. Preprint at *bioRxiv* https://doi.org/10.1101/2023.03.23.533992 (2023).
19. Schaum, N. et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
20. Li, H. et al. Fly Cell Atlas: a single-nucleus transcriptomic atlas of the adult fruit fly. *Science* **375**, eabk2432 (2022).
21. Zhang, X. et al. A spatially resolved multi-omic single-cell atlas of soybean development. *Cell* **188**, 550–567 (2025).
22. Marand, A. P., Chen, Z., Gallavotti, A. & Schmitz, R. J. A cis-regulatory atlas in maize at single-cell resolution. *Cell* **184**, 3041–3055 (2021).
23. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
References 18–23 represent comprehensive whole-organism cell atlases for animal and plant model species.
24. Toker, I. A. et al. Divergence in neuronal signaling pathways despite conserved neuronal identity among *Caenorhabditis* species. *Curr. Biol.* **35**, 2927–2945 (2025).
25. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
26. Marks, D. S. et al. Protein 3D structure computed from evolutionary sequence variation. *PLoS ONE* **6**, e28766 (2011).
27. Mendieta, J. P. et al. Investigating the *cis*-regulatory basis of C_3 and C_4 photosynthesis in grasses at single-cell resolution. *Proc. Natl Acad. Sci. USA* **121**, e2402781121 (2024).
28. Triesch, S. et al. Single-nuclei sequencing of *Moricandia arvensis* reveals bundle sheath cell function in the photorespiratory shuttle of C_3-C_4 intermediate Brassicaceae. *J. Exp. Bot.* https://doi.org/10.1093/jxb/eraf245 (2025).
29. Li, J. et al. Deep learning of cross-species single-cell landscapes identifies conserved regulatory programs underlying cell types. *Nat. Genet.* **54**, 1711–1720 (2022).
30. Agarwal, V. & Shendure, J. Predicting mRNA abundance directly from genomic sequence using deep convolutional neural networks. *Cell Rep.* **31**, 107663 (2020).
31. Avsec, Ž. et al. Effective gene expression prediction from sequence by integrating long-range interactions. *Nat. Methods* **18**, 1196–1203 (2021).
32. Janssens, J. et al. Decoding gene regulation in the fly brain. *Nature* **601**, 630–636 (2022).
33. Pearce, J. D. et al. A cross-species generative cell atlas across 1.5 billion years of evolution: the TranscriptFormer single-cell model. Preprint at *bioRxiv* https://doi.org/10.1101/2025.04.25.650731 (2025).
34. Taskiran, I. I. et al. Cell-type-directed design of synthetic enhancers. *Nature* **626**, 212–220 (2024).
35. Mich, J. K. et al. Functional enhancer elements drive subclass-selective expression from mouse to primate neocortex. *Cell Rep.* **34**, 108754 (2021).
36. Frazer, J. et al. Disease variant prediction with deep generative models of evolutionary data. *Nature* **599**, 91–95 (2021).
37. Church, S. H., Mah, J. L. & Dunn, C. W. Integrating phylogenies into single-cell RNA sequencing analysis allows comparisons across species, genes, and cells. *PLoS Biol.* **22**, e3002633 (2024).
38. Del Campo, J. et al. The others: our biased perspective of eukaryotic genomes. *Trends Ecol. Evol.* **29**, 252–259 (2014).
39. Ku, C. & Sebé-Pedrós, A. Using single-cell transcriptomics to understand functional states and interactions in microbial eukaryotes. *Philos. Trans. R. Soc. B* **374**, 20190098 (2019).

40. Alacid, E. & Richards, T. A. A cell–cell atlas approach for understanding symbiotic interactions between microbes. *Curr. Opin. Microbiol.* **64**, 47–59 (2021).
41. Hu, M., Zheng, X., Fan, C.-M. & Zheng, Y. Lineage dynamics of the endosymbiotic cell type in the soft coral Xenia. *Nature* **582**, 534–538 (2020).
42. Levy, S. et al. A stony coral cell atlas illuminates the molecular and cellular basis of coral symbiosis, calcification, and immunity. *Cell* **184**, 2973–2987 (2021).
43. Serrano, K. et al. Spatial co-transcriptomics reveals discrete stages of the arbuscular mycorrhizal symbiosis. *Nat. Plants* **10**, 673–688 (2024).
44. Fromm, A. et al. Single-cell RNA-seq of the rare virosphere reveals the native hosts of giant viruses in the marine environment. *Nat. Microbiol.* **9**, 1619–1629 (2024).
45. Ku, C. et al. A single-cell view on alga-virus interactions reveals sequential transcriptional programs and infection states. *Sci. Adv.* **6**, eaba4137 (2020).
References 42–45 represent the first examples of using single-cell methods to study symbiotic interactions, simultaneously mapping host and symbiont gene expression programs within the same cells.
46. Burki, F., Sandin, M. M. & Jamy, M. Diversity and ecology of protists revealed by metabarcoding. *Curr. Biol.* **31**, 1267–1280 (2021).
47. Delmont, T. O. et al. Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. *Cell Genom.* **2**, 100123 (2022).
48. Cordier, T., Lanzén, A., Apothéloz-Perret-Gentil, L., Stoeck, T. & Pawlowski, J. Embracing environmental genomics and machine learning for routine biomonitoring. *Trends Microbiol.* **27**, 387–397 (2019).
49. Marchetti, A. et al. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc. Natl Acad. Sci. USA* **109**, E317–E325 (2012).
50. Brunet, T. & King, N. The origin of animal multicellularity and cell differentiation. *Dev. Cell* **43**, 124–140 (2017).
51. Sebé-Pedrós, A., Degnan, B. M. & Ruiz-Trillo, I. The origin of Metazoa: a unicellular perspective. *Nat. Rev. Genet.* **18**, 498–512 (2017).
52. Arendt, D. et al. The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016).
Establishes a foundational conceptual framework for the study of cell type evolution and outlines key open questions and future research directions in the field.
53. Archibald, J. M., Simpson, A. G. B. & Slamovits, C. H. *Handbook of the Protists* (Springer, 2017).
54. Fritz-Laylin, L. K. et al. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* **140**, 631–642 (2010).
55. Brunet, T. et al. Light-regulated collective contractility in a multicellular choanoflagellate. *Science* **366**, 326–334 (2019).
56. Brunet, T. et al. A flagellate-to-amoeboid switch in the closest living relatives of animals. *eLife* **10**, 61037 (2021).
57. Dayel, M. J. et al. Cell differentiation and morphogenesis in the colony-forming choanoflagellate *Salpingoeca rosetta*. *Dev. Biol.* **357**, 73–82 (2011).
58. Phillips, M. A. et al. Malaria. *Nat. Rev. Dis. Primers* **3**, 17050 (2017).
59. Häckel, E. Monograph of Monera. *J. Cell Sci.* **s2-9**, 327–341 (1869).
60. Saville-Kent, W. *A Manual of the Infusoria: Including a Description of All Known Flagellate, Ciliate, and Tentaculiferous Protozoa, British and Foreign, and an Account of the Organization and the Affinities of the Sponges* Vol. 1 (D. Bogue, 1880).
61. Ramón y Cajal, S. *Histologie Du Système Nerveux de l'homme & Des Vertébrés: Cervelet, Cerveau Moyen, Rétine, Couche Optique, Corps Strié, Écorce Cérébrale Générale & Régionale, Grand Sympathique* Vol. 2 (A. Maloine, 1911).
62. Ramón y Cajal, S. Estructura de los centros nerviosos de las Aves. *Revista Trimestral de Histología Normal y Patológica* **1**, 1–10 (1888).
63. Virchow, R. *Cellular Pathology as Based Upon Physiological and Pathological Histology* (John Churchill, 1860).
64. Hyman, L. H. *The Invertebrates: Protozoa through Ctenophora* (McGraw-Hill, 1940).
65. Willmer, E. N. *Cytology and Evolution* (Academic, 1970).
66. Vergara, H. M. et al. Whole-body integration of gene expression and single-cell morphology. *Cell* **184**, 4819–4837 (2021).
67. Steinmetz, P. R. H. et al. Independent evolution of striated muscles in cnidarians and bilaterians. *Nature* **487**, 231–234 (2012).
68. Ogino, K., Tsuneki, K. & Furuya, H. Distinction of cell types in *Dicyema japonicum* (phylum Dicyemida) by expression patterns of 16 genes. *J. Parasitol.* **97**, 596–601 (2011).
69. Novershtern, N. et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296–309 (2011).
70. Minnoye, L. et al. Cross-species analysis of enhancer logic using deep learning. *Genome Res.* **30**, 1815–1834 (2020).
71. Alié, A. et al. The ancestral gene repertoire of animal stem cells. *Proc. Natl Acad. Sci. USA* **112**, E7093–E7100 (2015).
72. Cherbas, L. et al. The transcriptional diversity of 25 *Drosophila* cell lines. *Genome Res.* **21**, 301–314 (2011).
73. Tanay, A. & Sebé-Pedrós, A. Evolutionary cell type mapping with single-cell genomics. *Trends Genet.* **37**, 919–932 (2021).
74. Musser, J. M. et al. Profiling cellular diversity in sponges informs animal cell type and nervous system evolution. *Science* **374**, 717–723 (2021).
75. Sebé-Pedrós, A. et al. Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-seq. *Cell* **173**, 1520–1534 (2018).
76. Sebé-Pedrós, A. et al. Early metazoan cell type diversity and the evolution of multicellular gene regulation. *Nat. Ecol. Evol.* **2**, 1176–1188 (2018).
77. Najle, S. R. et al. Stepwise emergence of the neuronal gene expression program in early animal evolution. *Cell* **186**, 4676–4693 (2023).
78. Plass, M. et al. Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science* **1723**, eaaq1723 (2018).
79. Fincher, C. T., Wurtzel, O., de Hoog, T., Kravarik, K. M. & Reddien, P. W. Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. *Science* **360**, eaaq1736 (2018).
References 75, 76, 78 and 79 represent the first whole-adult cell atlases for non-model animal species.
80. Robertson, H. E. et al. Single cell atlas of *Xenoturbella bocki* highlights limited cell-type complexity. *Nat. Commun.* **15**, 2469 (2024).
81. Álvarez-Campos, P. et al. Annelid adult cell type diversity and their pluripotent cellular origins. *Nat. Commun.* **15**, 3194 (2024).
82. Ghaddar, A. et al. Whole-body gene expression atlas of an adult metazoan. *Sci. Adv.* **9**, 358 (2023).
83. Dogga, S. K. et al. A single cell atlas of sexual development in *Plasmodium falciparum*. *Science* **384**, eadj4088 (2024).
Exemplifies the power of single-cell analysis to molecularly characterize cell states across the life cycle of unicellular eukaryotes.
84. Wang, S. Y. et al. Role of epigenetics in unicellular to multicellular transition in *Dictyostelium*. *Genome Biol.* **22**, 134 (2021).
85. Howick, V. M. et al. The Malaria Cell Atlas: single parasite transcriptomes across the complete *Plasmodium* life cycle. *Science* **365**, eaaw2619 (2019).
86. Wang, L. et al. The maturation and aging trajectory of *Marchantia polymorpha* at single-cell resolution. *Dev. Cell* **58**, 1429–1444 (2023).
87. Tanay, A. & Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338 (2017).
88. Domcke, S. & Shendure, J. A reference cell tree will serve science better than a reference cell atlas. *Cell* **186**, 1103–1114 (2023).
89. Shafer, M. E. R. Cross-species analysis of single-cell transcriptomic data. *Front. Cell Dev. Biol.* **7**, 175 (2019).
90. Yan, H. et al. Evolution of plant cell-type-specific *cis*-regulatory elements. Preprint at *bioRxiv* https://doi.org/10.1101/2024.01.08.574753 (2024).
91. Hecker, N. et al. Enhancer-driven cell type comparison reveals similarities between the mammalian and bird pallium. *Science* **387**, eadp3957 (2025).
One of the first examples of cross-species cell type comparisons based on regulatory sequence information rather than gene expression data.
92. Bakken, T. E. et al. Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).
93. Song, Y., Miao, Z., Brazma, A. & Papatheodorou, I. Benchmarking strategies for cross-species integration of single-cell RNA sequencing data. *Nat. Commun.* **14**, 6495 (2023).
94. Tarashansky, A. J. et al. Mapping single-cell atlases throughout Metazoa unravels cell type evolution. *eLife* **10**, e66747 (2021).
95. Mah, J. L. & Dunn, C. W. Cell type evolution reconstruction across species through cell phylogenies of single-cell RNA sequencing data. *Nat. Ecol. Evol.* **8**, 325–338 (2024).
Lays out important considerations for cell type phylogenetic reconstruction and evolutionary models.
96. Burkhardt, P. & Jékely, G. Evolution of synapses and neurotransmitter systems: the divide-and-conquer model for early neural cell-type evolution. *Curr. Opin. Neurobiol.* **71**, 127–138 (2021).
97. Arendt, D., Bertucci, P. Y., Achim, K. & Musser, J. M. Evolution of neuronal types and families. *Curr. Opin. Neurobiol.* **56**, 144–152 (2019).
98. Wagner, G. P. The developmental genetics of homology. *Nat. Rev. Genet.* **8**, 473–479 (2007).
99. Pacureanu, A., Silva, J. C. da, Yang, Y., Bohic, S. & Cloetens, P. Nanoscale three-dimensional imaging of biological tissue with X-ray holographic tomography. In *Proc. SPIE Vol. 10711 Biomedical Imaging and Sensing Conf.* (eds Yatagai, T. et al.) 107112B (SPIE, 2018).
100. Zinchenko, V., Hugger, J., Uhlmann, V., Arendt, D. & Kreshuk, A. MorphoFeatures for unsupervised exploration of cell types, tissues, and organs in volume electron microscopy. *eLife* **12**, e80918 (2023).
101. Di Tommaso, P. et al. Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **35**, 316–319 (2017).
102. Denisenko, E. et al. Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol.* **21**, 130 (2020).
103. Scully, T. & Klein, A. A mannitol-based buffer improves single-cell RNA sequencing of high-salt marine cells. Preprint at *bioRxiv* https://doi.org/10.1101/2023.04.26.538465 (2023).
104. Chari, T. et al. Whole-animal multiplexed single-cell RNA-seq reveals transcriptional shifts across *Clytia medusa* cell types. *Sci. Adv.* **7**, eabh1683 (2021).
105. Alles, J. et al. Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol.* **15**, 44 (2017).
106. García-Castro, H. et al. ACME dissociation: a versatile cell fixation-dissociation method for single-cell transcriptomics. *Genome Biol.* **22**, 89 (2021).
107. Rosenberg, A. B. et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).
108. Bageritz, J. et al. Glyoxal as an alternative fixative for single-cell RNA sequencing. *G3* **13**, jkad160 (2023).
109. Jiménez-Gracia, L. et al. FixNCut: single-cell genomics through reversible tissue fixation and dissociation. *Genome Biol.* **25**, 81 (2024).
110. Fortmann, S. D. et al. Fixation before dissociation using a deep eutectic solvent preserves in vivo states and phospho-signaling in single-cell sequencing. Preprint at *bioRxiv* https://doi.org/10.1101/2023.02.13.528370 (2023).
111. Martin, B. K. et al. Optimized single-nucleus transcriptional profiling by combinatorial indexing. *Nat. Protoc.* **18**, 188–207 (2023).
112. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
113. Petrany, M. J. et al. Single-nucleus RNA-seq identifies transcriptional heterogeneity in multinucleated skeletal myofibers. *Nat. Commun.* **11**, 6374 (2020).
114. Guillotin, B. et al. A pan-grass transcriptome reveals patterns of cellular divergence in crops. *Nature* **617**, 785–791 (2023).
115. Jariani, A. et al. A new protocol for single-cell RNA-seq reveals stochastic gene expression during lag phase in budding yeast. *eLife* **9**, e55320 (2020).
116. Grones, C. et al. Best practices for the execution, analysis, and data storage of plant single-cell/nucleus transcriptomics. *Plant Cell* **36**, 812–828 (2024).
117. Heaton, H. et al. Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat. Methods* **17**, 615–620 (2020).

118. Guigó, R. Genome annotation: from human genetics to biodiversity genomics. *Cell Genom.* **3**, 100375 (2023).
119. Weisman, C. M., Murray, A. W. & Eddy, S. R. Mixing genome annotation methods in a comparative analysis inflates the apparent number of lineage-specific genes. *Curr. Biol.* **32**, 2632–2639 (2022).
120. Altenhoff, A. M. et al. Standardized benchmarking in the quest for orthologs. *Nat. Methods* **13**, 425–430 (2016).
121. Glover, N. et al. Advances and applications in the quest for orthologs. *Mol. Biol. Evol.* **36**, 2157–2164 (2019).
122. Rosen, Y. et al. Toward universal cell embeddings: integrating single-cell RNA-seq datasets across species with SATURN. *Nat. Methods* **21**, 1492–1500 (2024).
123. Rosen, Y. et al. Universal cell embeddings: a foundation model for cell biology. Preprint at *bioRxiv* https://doi.org/10.1101/2023.11.28.568918 (2023).
124. Price, P. D. et al. Detecting signatures of selection on gene expression. *Nat. Ecol. Evol.* **6**, 1035–1045 (2022).
125. Bertram, J. et al. CAGEE: computational analysis of gene expression evolution. *Mol. Biol. Evol.* **40**, msad106 (2023).
126. Rohlfs, R. V., Harrigan, P. & Nielsen, R. Modeling gene expression evolution with an extended Ornstein–Uhlenbeck process accounting for within-species variation. *Mol. Biol. Evol.* **31**, 201–211 (2014).
127. Rohlfs, R. V. & Nielsen, R. Phylogenetic ANOVA: the expression variance and evolution model for quantitative trait evolution. *Syst. Biol.* **64**, 695–708 (2015).
128. Challis, R., Kumar, S., Sotero-Caio, C., Brown, M. & Blaxter, M. Genomes on a Tree (GoaT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life. *Wellcome Open Res.* **8**, 24 (2023).
129. Zeng, H. What is a cell type and how to define it? *Cell* **185**, 2739–2755 (2022).
130. Osumi-Sutherland, D. et al. Cell type ontologies of the Human Cell Atlas. *Nat. Cell Biol.* **23**, 1129–1135 (2021).
131. Morris, S. A. The evolving concept of cell identity in the single cell era. *Development* **146**, dev169748 (2019).

Acknowledgements We thank J. C. Carvalho for her expert work on the scientific illustrations and figure design. The BCA phase 0 is supported by the Gordon and Betty Moore Foundation (Grant GBMF12189). We also acknowledge financial support from the Wellcome Trust for the inaugural BCA meeting.

Author contributions A.S.-P., A.T., M.K.N.L. and D.A. wrote the paper based on discussions, comments and edits from S.A., J.A., M.I.A., M.B., P.C., S.M.C., M.D., C.D., A.E., J.F., T.G., J.G., X.G.-B., R.G., O.H., J.H.-C., M.I., A.K., H.L., C.J.L., H.M., J.M.M., L.G.N. S.R.N., L.P., S.P., I.P., M.J.P., N.R., S.Y.R., T.A.R., T.S.-S., L.M.S., E.S., J.S., Y.S., U.T. and B.W.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-09312-4.
Correspondence and requests for materials should be addressed to Arnau Sebé-Pedrós.
Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.

---

[1]Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. [2]Tree of Life, Wellcome Sanger Institute, Cambridge, UK. [3]Universitat Pompeu Fabra, Barcelona, Spain. [4]ICREA, Barcelona, Spain. [5]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. [6]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel. [7]Developmental Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. [8]Laboratory of Computational Biology, VIB Center for AI and Computational Biology, Leuven, Belgium. [9]VIB-KU Leuven Center for Brain and Disease Research, Leuven, Belgium. [10]Department of Human Genetics, KU Leuven, Leuven, Belgium. [11]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada. [12]Stazione Zoologica Anton Dohrn, Napoli, Italy. [13]Department of Embryology, Carnegie Science, Baltimore, MD, USA. [14]Department of Biology, Johns Hopkins University, Baltimore, MD, USA. [15]Department of Algal Development and Evolution, Max Planck Institute for Biology Tübingen, Tübingen, Germany. [16]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA. [17]Barcelona Supercomputing Centre (BSC-CNS), Plaça Eusebi Güell, Barcelona, Spain. [18]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain. [19]Donnelly Centre, University of Toronto, Toronto, Ontario, Canada. [20]Department of Physiology, University of Toronto, Toronto, Ontario, Canada. [21]Department of Biological Sciences, Howard Hughes Medical Institute, Columbia University, New York, NY, USA. [22]Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid (UPM)—Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA-CSIC), Campus de Montegancedo-UPM, Madrid, Spain. [23]Department of Systems Biology, Harvard Medical School, Boston, MA, USA. [24]Global Futures Laboratory, Arizona State University, Tempe, AZ, USA. [25]Department of Evolution and Ecology, University of California, Davis, CA, USA. [26]Department of Biology, Stanford University, Stanford, CA, USA. [27]Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL, USA. [28]Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT, USA. [29]Systems Biology Institute, Yale University, New Haven, CT, USA. [30]HUN-REN Biological Research Centre Szeged, Szeged, Hungary. [31]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. [32]Department of Computer Science, California Institute of Technology, Pasadena, CA, USA. [33]The Rockefeller University, New York, NY, USA. [34]Earlham Institute, Norwich Research Park, Norwich, UK. [35]Medical School, University of East Anglia, Norwich Research Park, Norwich, UK. [36]Genomics Department, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. [37]Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Berlin, Germany. [38]Charité—Universitätsmedizin, Berlin, Germany. [39]German Center for Cardiovascular Research (DZHK), Berlin, Germany. [40]NeuroCluster of Excellence, Berlin, Germany. [41]Plant Resilience Institute, Michigan State University, East Lansing, MI, USA. [42]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, USA. [43]Department of Plant Biology, Michigan State University, East Lansing, MI, USA. [44]Department of Plant, Soil, and Microbial Sciences, Michigan State University, East Lansing, MI, USA. [45]Department of Biology, University of Oxford, Oxford, UK. [46]Stowers Institute for Medical Research, Kansas City, MO, USA. [47]Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [48]Centre for Organismal Studies, Heidelberg University, Heidelberg, Germany. [49]Department of Biology, KU Leuven, Leuven, Belgium. [50]Living Systems Institute, University of Exeter, Exeter, UK. [51]Department of Biosciences, University of Exeter, Exeter, UK. [52]European Molecular Biology Laboratory—European Bioinformatics Institute, Hinxton, UK. [53]Department of Neurosciences and Developmental Biology, Faculty of Life Sciences, University of Vienna, Vienna, Austria. [54]Department of Bioengineering, Stanford University, Stanford, CA, USA.

Biodiversity Cell Atlas meeting participants

Stein Aerts[8,9,10], **John Archibald**[11], **Detlev Arendt**[7], **Maria Ina Arnone**[12], **Mark Blaxter**[2], **Phillip Cleves**[13,14], **Susana M. Coelho**[15], **Mafalda Dias**[1,3], **Casey Dunn**[16], **Anamaria Elek**[1], **Jonathan Frazer**[1,3], **Toni Gabaldón**[4,17,18], **Jesse Gillis**[19,20], **Xavier Grau-Bové**[1], **Roderic Guigó**[1,3], **Oliver Hobert**[21], **Jaime Huerta-Cepas**[22], **Manuel Irimia**[1,3,4], **Allon Klein**[23], **Mara K. N. Lawniczak**[2], **Harris Lewin**[24,25], **Christopher J. Lowe**[26], **Heather Marlow**[27], **Jacob M. Musser**[28,29], **László G. Nagy**[30], **Sebastián R. Najle**[1], **Lior Pachter**[31,32], **Sadye Paez**[33], **Irene Papatheodorou**[34,35], **Michael J. Passalacqua**[36], **Nikolaus Rajewsky**[37,38,39,40], **Seung Y. Rhee**[41,42,43,44], **Thomas A. Richards**[45], **Tatjana Sauka-Spengler**[46,47], **Lauren M. Saunders**[48], **Arnau Sebé-Pedrós**[1,2,3,4], **Eve Seuntjens**[49], **Jordi Solana**[50,51], **Yuyao Song**[52], **Amos Tanay**[5,6], **Ulrich Technau**[53] **& Bo Wang**[54]

A full list of members and their affiliations appears in the Supplementary Information.