THEME ARTICLE: BREAKTHROUGHS, CHALLENGES, AND OPPORTUNITIES IN BIOLOGICAL DATA VISUALIZATION

The State of Single-Cell Atlas Data Visualization in the Biological Literature

Mark S. Keller, Eric Mörth, and Thomas C. Smits, *Harvard Medical School, Boston, MA, 02115, USA*
Simon Warchol and Grace Guo, *Harvard University, Allston, MA, 02134, USA*
Qianwen Wang, *University of Minnesota, Twin Cities, Minneapolis, MN, 55455, USA*
Robert Krueger, *New York University, Brooklyn, NY, 11201, USA*
Hanspeter Pfister, *Harvard University, Allston, MA, 02134, USA*
Nils Gehlenborg, *Harvard Medical School, Boston, MA, 02115, USA*

Recent advancements have enabled tissue samples to be profiled at the unprecedented level of detail of a single cell. Analysis of these data has enabled discoveries that are relevant to understanding disease and developing therapeutics. Large-scale profiling efforts are underway, which aim to generate "atlas" resources that catalog cellular archetypes, including biomarkers and spatial locations. While the problem of cellular data visualization is not new, the size, resolution, and heterogeneity of single-cell atlas datasets present challenges and opportunities. We survey the usage of visualization to interpret single-cell atlas datasets by assessing over 1800 figure panels from 45 biological publications. We intend for this report to serve as a foundational resource for the visualization community as atlas-scale single-cell datasets are emerging rapidly with aims of advancing our understanding of biological function in health and disease.

The cell is the fundamental unit of life, responsible for structure, function, and division in organisms.[1] The disruption of cellular processes can lead to dysfunction and disease at the tissue, organ, or organism level. In recognition, researchers have set out to measure biological processes across millions of cells from many different contexts (e.g., organisms and diseases) and publish them in the form of *atlas* datasets.[2] Life scientists envision single-cell atlases as reference resources, which include characteristic features (biomarkers, such as gene expression), locations, and abundances.[1] More comprehensive than any individual experiment (which may profile cells from only a single tissue sample), atlases combine data from multiple experiments to provide more power to generate and test hypotheses.

Atlas datasets are emerging rapidly but remain daunting to analyze and interpret.[3] One obstacle is the high dimensionality of the data, as each cell can have thousands of measured features. Moreover, single-cell data bring challenges that do not necessarily emerge in other high-dimensional settings: the presence of associated imaging or genomic data, batch effects,[4] and concerns about dimensionality reduction (DR).[5] Other challenges include contextualization of atlas information with prior biological literature, connection of atlases with differing organization,[6] and alignment to differing coordinate frameworks.[7] The domain also requires visualizations that integrate multiple data types (e.g., gene expression, genomics, and microscopy images) and facilitate tasks, such as relation and comparison.

We survey the state of visualization for single-cell atlas data for a primary audience of visualization

This article has supplementary downloadable material available at https://doi.org/10.1109/MCG.2025.3583979, provided by the authors.
Digital Object Identifier 10.1109/MCG.2025.3583979
Date of publication 27 June 2025; date of current version 29 September 2025.

researchers. In this article, we contribute the following.

- An introduction to the single-cell domain.
- A systematic review of over 1800 visualizations contained in figures within single-cell atlas publications from the biological literature.
- Discussion of opportunities at the intersection of the single-cell and visualization domains.

Related Work

All prior reviews related to single-cell atlas data analysis have been written for an audience of biological researchers. Two survey papers for this audience discuss the landscape of visualization tools for single-cell data. Cakir et al.[8] analyzed 13 interactive tools for single-cell transcriptomics data visualization. Zappia and Theis[9] analyzed the scRNA-tools database, revealing that 437 of the 1059 cataloged tools (41%) were annotated with the "visualization" category. Neither survey considers aspects of visualization design and both are limited to the transcriptomics data type.

Within visualization, there are state-of-the-art reports (STARs) on related topics under the umbrella of biology and medicine. The most closely related STARs are on live cell imaging[10] and connectomics,[11] as these two domains deal with cellular data, but these areas differ in that they lack the high dimensionality and heterogeneity that single-cell atlases present. A recent STAR on visualization for physiology contains sections on cellular-scale and tissue-scale data, but focuses on spatially resolved (SR) and temporally resolved experimental modalities, such as imaging (as opposed to sequencing-based data), with Garrison et al.[12] acknowledging, "We found comparatively few cell- and tissue-scale visualization works." STARs for genomics[13] and molecular structures[14] are also closely related, but the primary entities analyzed in these domains are sequences and atoms, respectively, rather than cells. There are no reviews focusing on the single-cell domain from a visualization research perspective, even though visualization is a key part of single-cell data analysis. The importance of this topic and its challenges motivates a survey of current practices and the need to chart a path forward for visualization research applied to problems in the single-cell domain.

BIOLOGICAL BACKGROUND

Computational analysis, including visualization, of single-cell atlas data requires familiarity with the goals and logistics of atlasing efforts. It is also beneficial to understand the basics of experimental methods.

Experiment Versus Atlas

A single-cell *experiment* refers to quantitatively profiling multiple features of individual cells, typically via sequencing, imaging, or mass spectrometry-based techniques, in which different cells can be distinguished from one another. An *atlas* undoubtedly refers to something larger and more complex, but what exactly?

Despite numerous perspective articles on the topic of single-cell atlases, this term has not been well defined. In some contexts, a singular "atlas" is described as a "cell census" and explained using analogies, such as Wikipedia and the periodic table, while in other contexts, the plural "atlases" are used to refer to separate publications or data resources (with each atlas presumably serving a different purpose). The initial outline for the Human Cell Atlas project called for publication of intermediate draft atlases, which build toward a "full" human atlas, with "sister" atlases for model organisms.[1] More recently, the idea of a cell atlas has been framed as a foundation model.[15] We note that in geography, the term "atlas" is defined as a "collection of maps," in which geographic information and political boundaries may also be accompanied by statistics.

We define a "human single-cell atlas" as a collection of maps consisting of data from *multiple single-cell experiments from multiple tissue donors that are intended to summarize and communicate this data through hierarchical organization* (e.g., cell → region → sample → tissue donor, as well as via ontology- or metadata-based hierarchy) *and spatial arrangement* (in physical or latent space), *which may also be accompanied by statistics at each hierarchy level or at different spatial partitions*. There are multiple dimensions along which we might further classify an atlas, including whether it comprises one versus multiple organs, one versus multiple experimental modalities, physical versus only latent space organization, temporal versus not, and the count of cells profiled.

Atlasing Goals

Visualization is acknowledged to be key to single-cell atlas interpretation, yet the evaluation of visualization techniques to display such data remains underexplored. In outlining the guiding principles of the Human Cell Atlas, the consortium stated that "...how best to capture and represent this information requires serious attention."[1] Recently, the field has

highlighted difficulties that have arisen as atlas datasets have become more complex, calling for the development of techniques which, "...ease interpretation and visualization of atlases spanning tissue resolutions and omics layers."[3]

The high-level goals of cell atlasing efforts have been thoroughly outlined (see Table S1 of the Supplementary Material).[1,2,3,16] Multiple perspective papers articulate goals of surveying cell types and their functions across the human body, identifying properties and interactions that are relevant to disease, and linking these findings to genetics. Goals related to medicine are to use single-cell data to stratify patients, to diagnose disease, and to compare drug delivery routes.

Within biology, alternative perspectives have been raised. Regev et al.[1] proposed construction of an atlas with two branches: one focused on cataloging cellular properties and the other focused on spatial organization. Domcke and Shendure[6] called for the field to settle on a primary organizing principle for atlases to better facilitate projection onto an established reference. Adler et al.[17] pointed out several ways of looking at cellular organization with underpinnings in complex systems theory, such as primary-supportive and supplier–consumer. Nevertheless, the outlines of high-level goals serve as primers on the direction of the field and useful reference points during lower level analyses.

Experimental Techniques

Since the invention of the microscope, researchers have been able to observe cells and, by extension, record measurements at single-cell resolution. Such measurements of phenotype reflect the processes that occur at the molecular level within and among cells and their environment. The causal relationship between molecular processes and observed phenotype then explains our interest in molecular biology: this is where interventions (e.g., therapies or perturbations) can be made that affect phenotype. Owing to the complexity of cell biology, the development of such interventions today heavily depends on our ability to make measurements at the molecular scale. As there are several types of such measurements (e.g., for different types of biomolecules), the term 'omics refers to this general class of measurement techniques.

Experimental Modalities

A cell's genome encodes (in DNA) a blueprint for its function. However, this blueprint encodes function much like a circuit or a state machine, whose operation depends on a number of inputs, including surrounding cells and other environmental stimuli, which all then contribute to the observed output (i.e., phenotype). Notably, at any given moment, each cell is actively executing only a subset of its encoded instructions. Thus, knowledge of a cell's *genomic* information alone, while providing a base level of insight into potential programs it could execute, is not a sufficient readout of its state.

There is a continuum of 'omics techniques—hereafter referred to as *experimental modalities*—that target different types of biomolecules at different stages along the path from blueprint to phenotype in cells. Within the single-cell domain, the most widely applied experimental modality to date has been *transcriptomics*. The transcriptome, or the set of RNA transcripts in a cell, sits further along the continuum between DNA and phenotype, as the DNA in the genome encodes genes, which must first be *transcribed* into RNA transcripts. This RNA can then (potentially) be *translated* into proteins. In other words, the transcriptome of a cell is dynamic and provides more information (compared to DNA) about the gene programs that are currently being executed. Single-cell transcriptomics have been widely applied due to several factors, including cost and maturity of statistical methods, but are not necessarily the supreme readout of cell state. Transcript abundances do not have perfect correlation with abundances of corresponding proteins (e.g., due to transcript degradation), a shortcoming which can be addressed through the application of alternative experimental modalities along the continuum.

Before transcription, there are complex regulatory processes that govern whether a given gene will be transcribed in the first place. These regulatory processes are tightly intertwined with the question of cell state and can be measured via *epigenomics* modalities. Downstream of transcription, processes of protein translation, or abundance of the resulting proteins (or posttranslational modifications) can be profiled via *proteomics* techniques. Further along the continuum is *metabolomics*, which measures metabolites, the components (of cellular or environmental origin) of chemical reactions that occur within cells (or upon or outside, extracellularly). Within the field of metabolomics, there are subtypes, such as lipidomics and glycomics, which measure fat and sugar molecules, respectively. These methods provide complementary perspectives into cellular operations and are envisioned to become more broadly applied as they increase in resolution and decrease in cost.

Finally, the -omics suffix is sometimes used to refer to phenotypic-level measurements, for instance, in the

fields of *connectomics* and *pathomics*, which extract morphological measurements at ultrastructural and microscopic scales, respectively. Where alternative modalities such as these produce high-throughput measurements at single-cell resolution, statistical methods can often be reused.

Multiple Experimental Modalities

A logical progression would be to combine multiple of these techniques in the same cell to obtain a richer snapshot. Indeed, researchers have recently developed *multimodal* experimental techniques, typically able to profile two or three modalities in the same cells Vandereyken et al.[18] Currently, not all experimental techniques are compatible with one another (e.g., destructive to the cell). Thus, multimodal techniques are an area of active research. Until the robustness and feasibility of multimodal methods increase, the majority of available single-cell data will remain *unimodal*. As a consequence, computational approaches are being developed to bridge this gap.[19]

Why Single-Cell?

Readouts from 'omics methods have historically not contained a mapping between biomolecules and their cell of origin. Only in recent years has it become possible to preserve cellular identity in 'omics experiments (e.g., via microfluidics). Biologists now refer to the previous techniques as "bulk" (to distinguish from single cell). Bulk is to single cell as a fruit smoothie is to a fruit salad: the former provides a vague sense of its components, but we are unable to distinguish the individual original entities within.[1]

Bulk experiments prevent discernment of whether an observed difference in gene expression (e.g., between two cell groups, say health and disease) is due to a differing cell type composition (i.e., number of cells of each class, which are executing their normal state machine programs) versus due to altered gene expression (i.e., same distribution over cell classes but a change in the state machine programs they are executing).[20] Similarly, a "rare" cell type (which may be relevant to disease) that produces an atypical distribution of transcripts may effectively appear as noise in a bulk experiment. Single-cell methods resolve these discrepancies by mapping observed molecules to their cells of origin. Rood et al.[16] reviewed several examples, including one study that found cell type composition to be predictive of therapeutic response in Crohns disease, while in contrast, another study found coordinated shifts in gene expression across multiple cell types to be predictive of risk for ulcerative colitis.

Spatially Resolved Methods (SR)

It may not be apparent so far that many of the aforementioned 'omics methods do not preserve information about the spatial organization of cells (such experiments require *dissociation* of cells). Continuing the fruit analogy, nonspatial is to SR as a fruit salad is to a fruit tart.[1] In the latter, we observe not only the individual entities, but also how they are organized and their spatial relationships.

SR single-cell measurements can provide insights into signaling among cells and characterize the organization of cells into neighborhoods or functional tissue units (FTUs).[21] The introduction of techniques which preserve spatial context has brought new challenges: the development of spatial *and* multimodal techniques, and considerations of resolution, tissue size, and 3-D profiling.

Temporally Resolved Methods

Cells are dynamic, responding to internal and external stimuli over time. Understanding these dynamics is critical but challenging, as many experimental methods are destructive. One workaround, often used in developmental biology, is to profile cells from separate organisms at various timepoints, providing useful but limited snapshots due to differences between cells and coarse temporal resolution. Several approaches are designed for measuring cells over time: live cell imaging, intravital microscopy, and lineage tracing. On the other hand, algorithmic techniques enable inference of dynamics from standard single-cell 'omics experiments that would otherwise be viewed as snapshots. Pseudotime algorithms predict cell orderings by assuming that similar cells in high-dimensional space are closer developmentally. RNA velocity algorithms model the ratios of different types of RNA transcripts to predict whether a cell is in a steady state versus progressing to a different state. Approaches that enable profiling cells both temporally and spatially could become increasingly important due to efforts to construct virtual cell models.

Atlasing Challenges

Single-cell atlasing efforts present interesting challenges for computer scientists generally—and visualization researchers specifically. First, a diverse group of individuals constitute the *users* who aim to analyze and interpret single-cell atlases, which visualization

collaborators must take into account. These users span from wet lab biologists performing experiments, to clinician–scientists involved in tissue donor recruitment and disease-specific atlas interpretation, to high school students learning about cell biology—none of whom necessarily have programming expertise. On the computationally oriented side, users may span the spectrum of computational biologists, bioinformaticians, software engineers, and visualization experts.

Next, single-cell atlas generation and analysis are data-intensive. Thus, the single-cell field has, from its onset, embraced data science and machine learning techniques.[1] However, the number of cells that can be profiled in a single experiment continues to grow each year, today exceeding one million cells.[22] Imaging-based experiments, in particular those that image in 3-D, now routinely produce terabyte-scale volumes. The number and type of features that can be profiled per cell also continues to increase, resulting in heterogeneous datasets that are not trivial to align with each other.[19] Unlike other data-intensive fields, such as astronomy, data generation efforts for single-cell atlases are largely decentralized, with many labs performing experiments and generating data in the form of collections of files. These files are challenging to merge and catalog not only due to differences in the sets of measured features and presence of experimental variations,[4] but also due to the lack of a global coordinate system for alignment. There is also rich prior knowledge to leverage, including biological literature, ontologies, genome-wide association studies, and findings from model organism studies. While the problem of heterogeneous data within the context of bioinformatics is not new,[23] single-cell atlasing efforts are both accentuating this problem and have resulted in several proposed solutions. Further, a focus on millions of cells as data points highlights longstanding visualization challenges (e.g., overplotting), but also provides opportunities for establishing domain-specific solutions.

SURVEY OF STATIC FIGURES

Single-cell atlas research results have primarily been communicated via publications containing static figures and plots. We conducted a systematic review of static figures that appear in single-cell atlas publications. Such figures represent the current state of visualization in the field, providing insight into the findings that authors consider most interesting or consequential to peers, the journal audience, and the scientific community. Through our contribution of this comprehensive and unbiased analysis, an audience of visualization researchers can be confident in relating current practices in the domain to prior best practices from visualization. We conducted this analysis in pursuit of answers to the following research questions: Which visualization designs are used in single-cell atlas publications and which visualization designs are used with certain data types, experimental techniques, or purposes?

Methodology

We used a keyword search to identify single-cell atlas publications and performed manual qualitative analysis of all subfigures within the main text of the publications. We aimed to capture publications describing recent large-scale single-cell atlas efforts of human cells. We searched the Web of Science (see the Supplementary Material), then filtered the results based on the following eligibility criteria and rationale.

- EC1: Include only published 2020–2024.
- EC2: Include only peer-reviewed.
- EC3: Exclude methods and benchmarking.
- EC4: Exclude reviews and perspectives.
- EC5: Exclude nonhuman studies (e.g., mouse).
- EC6: Exclude nonbiology (e.g., chemistry).
- EC7: Exclude cancer studies.
- EC8: Exclude bulk (nonsingle-cell) studies.
- EC9: Exclude studies of less than 50,000 cells.

Due to the recent nature of the single-cell field, studies published in the last five years (EC1) have generally profiled larger numbers of cells with more advanced experimental techniques. As the term "atlas" to describe single-cell data is not well defined, we set the 50,000 cell threshold (EC9) to ensure that we captured publications which had to grapple with challenges of visualization scalability, such as overplotting. The median number of cells profiled by studies published in 2022 was 50,000; thus, this threshold roughly focuses our survey on the upper half of the study distribution during these years.[22] We aimed to capture papers focusing on a single organism, human (EC5), as we expect that papers which focus on model organisms or cross-species comparison would perform different types of analyses and therefore a different distribution of figure types (e.g., related to evolutionary relationships). Similarly, we expect that studies related to cancer (EC7) will include different analyses and figure types (e.g., related to somatic mutation). While these eligibility criteria limit the scope of our analysis, follow-up surveys could expand these criteria.

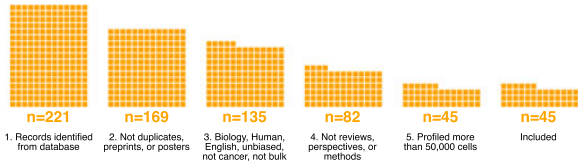

FIGURE 1. Methodology used for literature search, paper identification, and filtering. An initial set of 221 records was filtered down to 45 records matching the inclusion criteria.

We filtered an initial set of 221 matching database records down to 45 publications based on the eligibility criteria (see Figure 1; and Table S2 of the Supplementary Material). We performed qualitative analysis of each subfigure in these publications using an open coding approach. The resulting codes could be grouped into *plot types* (e.g., "pie chart"), *visual properties* (e.g., "highlighting"), *data types* (e.g., "pseudotime"), *experimental techniques and modalities* (e.g., "chromatin accessibility"), and *purpose* (e.g., "comparison"). We define each code in the Supplementary Material.

Results

We applied 85 codes to the 45 papers, across 1846 subfigure quotations, for a total of 6259 code applications. On average, each subfigure was annotated by 3.39 codes (min: 1, max: 11, median 3; see Figure S3 of the Supplementary Material). Each unique code was applied to a median of 23 subfigures and mean of 73.63 (min: 1 and max: 777). The distribution of subfigures per code was right-skewed, with 12 codes being applied to over 100 subfigures each (summing to 4406 code applications) and the remaining 73 codes being applied to fewer than 100 subfigures each (summing to 1853 applications). We provide an interactive web-based survey browser online.[a]

We first examined which codes were most used among all subfigures. The top 12-applied codes were (from most to least): small multiples, custom annotations, scatterplot, DR, nonstandard plot type, spatial, comparison, imaging, heatmap, multiplex, dot plot, and cell type composition (see Figure 2).

We next examined how codes were used across the 45 publications. Limiting analysis to unique papers per code, the distribution was less skewed, with 19 codes each being applied to over half (more than 22)

[a][Online]. Available: https://keller-mark.github.io/sc-star-site/

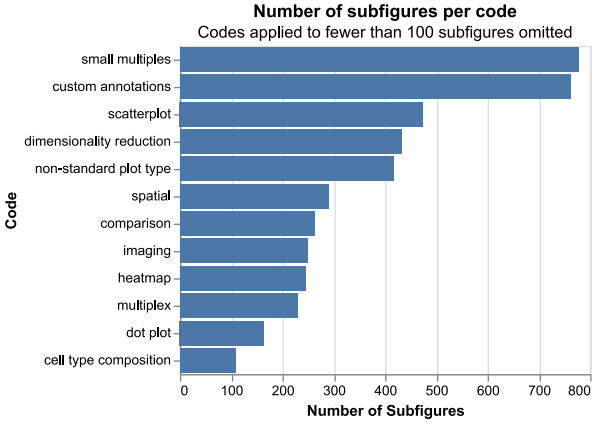

FIGURE 2. Top codes among all subfigures. Data only shown for codes applied to at least 100 subfigures.

of the papers: scatterplot, small multiples, nonstandard plot type, custom annotations, DR, heatmap, cell type composition, comparison, dot plot, spatial, stacked bar plot, imaging, experimental design, multiplex, violin plots, statistical significance, multiscale, bar plot, and gsea (see Figure 3).

Experimental Design Diagrams Were Used in the Majority of Publications

We applied the code "experimental design" to 43 subfigures (2.33% of subfigures) within 31 publications

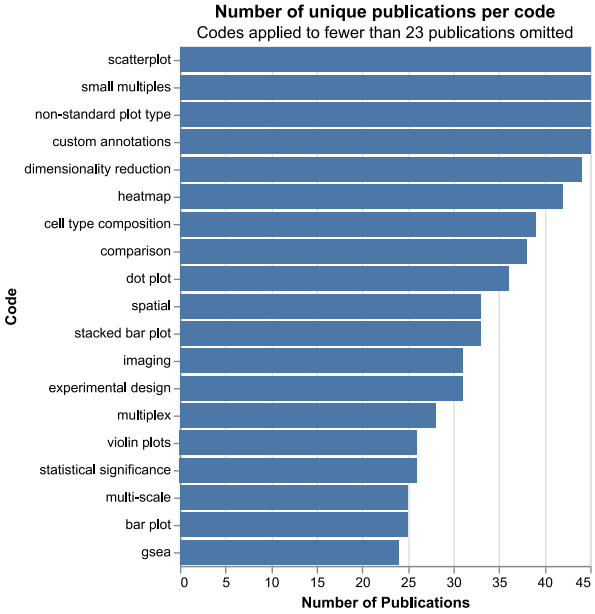

FIGURE 3. Top codes across publications. Data only shown for codes applied in more than half of the 45 publications.

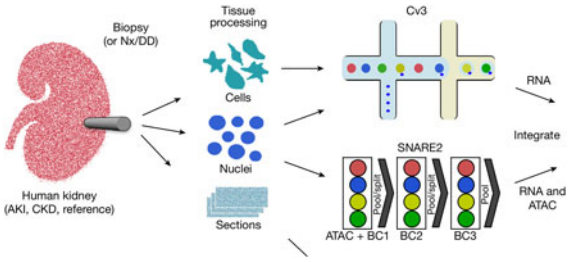

FIGURE 4. Example of code "experimental design" applied to Figure 1(a) from Lake et al.[24] Reprinted (open access).

(68.89% of publications). These diagrams often appeared early within publications to communicate experimental strategies to readers, typically using a combination of flow charts, schematics, and illustrations (see Figure 4). In some cases, diagrams included quantitative information, such as sample counts. For the related code "anatomical diagram," there were 41 subfigures (2.22%) from 22 publications (48.89%). Anatomical diagrams were often used to illustrate the organization and localization of cell types and FTUs within an organ, providing context for references to such entities in the publication.

Four Codes Appeared in all Included Publications

The following codes were applied to at least one subfigure in all 45 publications: small multiples, custom annotations, scatterplot, and nonstandard plot type.

Small Multiples

"Small multiples" was the most applied code, with 777 subfigure applications (42.09% of subfigures). We define small multiples as repeated instances of the same plot type, either within a single subfigure (see Figures 5 and 6) or among subfigure instances in a

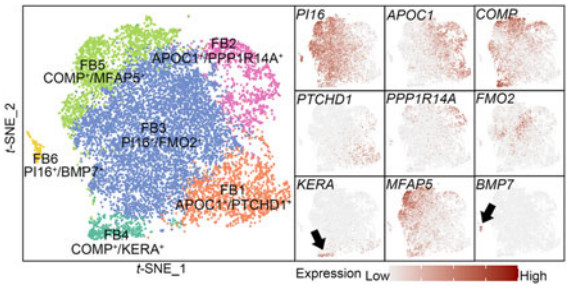

FIGURE 5. Example of code "small multiples" applied to Figure 2(a) from Zhao et al.[25] Reprinted (open access).

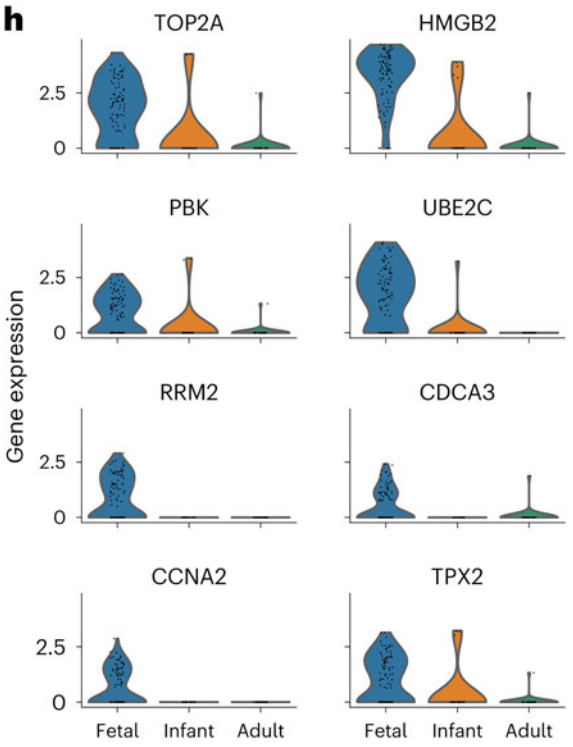

FIGURE 6. Example of code "small multiples" applied to Figure 3h from Chen et al.[26] Reprinted (open access).

parent figure (see Figure 7). The top co-occurring codes were custom annotations (400 co-occurrences; 51.48% of primary occurrences), scatterplot (254; 32.69%), DR (230; 29.60%), spatial (229; 29.47%), and imaging (206; 26.51%). Small multiples were often used to simultaneously visualize the distribution of

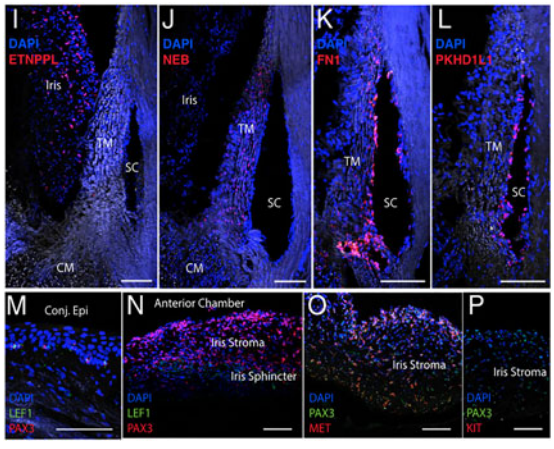

FIGURE 7. Example of "small multiples" applied to Figure (5i)–(5p) from Zyl et al.[26] Reprinted (open access).

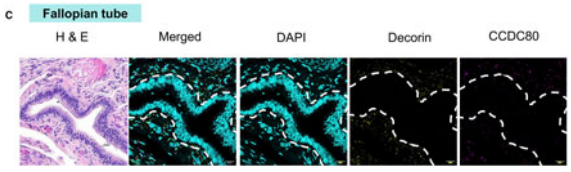

FIGURE 8. Example of code "custom annotations" applied to Figure 3(c) from Lengyel et al.[28] Reprinted (open access). Here, custom annotations take the form of outlined regions.

abundance for a subset of biomarkers across all cells (see Figure 5). Another observed usage was to visualize multiple image channels (corresponding to different biomarkers), enabling reuse of colors and avoiding color blending issues (see Figure 8).

Custom Annotations

The code "custom annotations" was the second most applied code. It was applied to 762 subfigures (41.28%), including at least one subfigure within all 45 analyzed publications. We define "custom annotations" as any nonstandard annotation to a plot. This includes text and region annotations on scatterplots and images, metadata bar plots alongside heatmaps (see Figure 9), and arrows indicating data points of interest. We observe that text annotations on scatterplots to indicate cell type categories are often used in lieu of a legend that would define the color per category (see Figure 10). The top codes that co-occurred with this code were small multiples

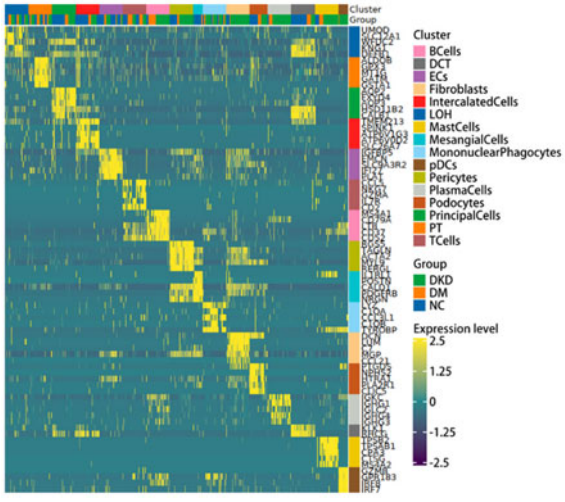

FIGURE 9. Example of code "custom annotations" applied to heatmap Figure 1(d) from Chen et al.[29] Reprinted with permission from John Wiley and Sons.

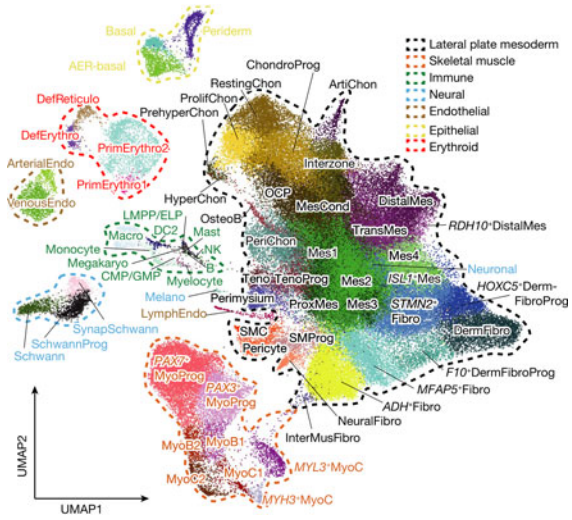

FIGURE 10. Example of code "custom annotations" applied to DR Figure 1(b) from Zhang et al.[30] Reprinted (open access). Here, annotations take the form of text, lines, and outlined regions indicating cell types.

(400; 52.49% of primary occurrences; see Figure 8), scatterplot (307; 40.29%; see Figures 10 and 11), DR (278; 36.48%), spatial (216; 28.35%; see Figure 8), and imaging (194; 25.46%). We hypothesize that custom annotations are widespread in static publication figures in order to reinforce the authors' intended takeaways.

Scatterplot

Scatterplot usage was widespread, with this code applied to 474 subfigures (25.68%). While the majority of scatterplot subfigures displayed DR results, other occurrences included gating plots (23; see Figure 12) and volcano plots (16; see Figure 13), which facilitate identification of specific subpopulations of cells and

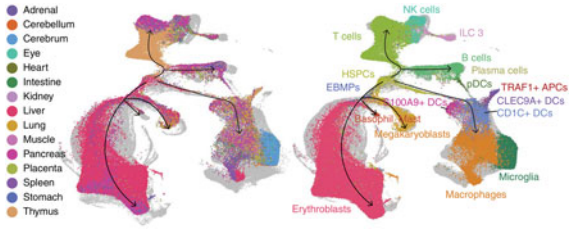

FIGURE 11. Example of code "custom annotations" applied to DR Figure 4(c) from Cao et al.[31] Reprinted with permission from AAAS. Here, custom annotations take the form of text overlays and trajectory arrows.

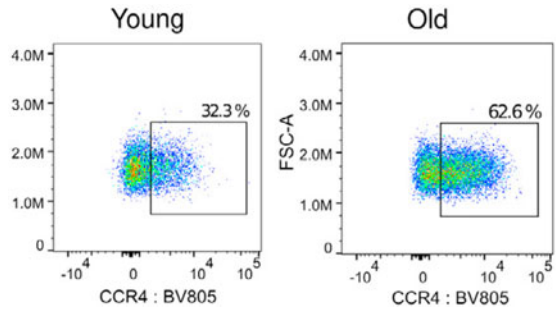

FIGURE 12. Example of "gating scatterplot" applied to Figure 3n from Terekhova et al.[28] Reprinted (open access).

enable identification of differentially expressed genes between groups of cells, respectively.

Nonstandard Plot Type

The code "nonstandard plot type" indicates a plot either not corresponding to a function available in the most widely used single-cell plotting toolkits (ScanPy and Seurat) or having semantics that were atypical among plots with similar designs. In some cases, these correspond to plots designed to communicate a specific study-defined quantity (e.g., a signature of multiple biomarkers, see Figure 14). In other cases, these correspond to designs that did not appear within more than one atlas publication (see Figure 15). This code was applied to 416 subfigures (22.54%). The top ten co-occurring codes were small multiples (127 co-occurrences; 30.53% of primary occurrences), comparison (106; 25.48%), custom annotations (62; 14.90%), bar plot (49; 11.78%), heatmap (40; 9.62%), statistical significance (35; 8.41%), ligand–receptor (34; 8.17%), scatterplot (32; 7.69%), box plot (30; 7.21%), and cell type composition (25; 6.01%). With a larger sample of subfigures, this code could likely be stratified and represents an area for future work.

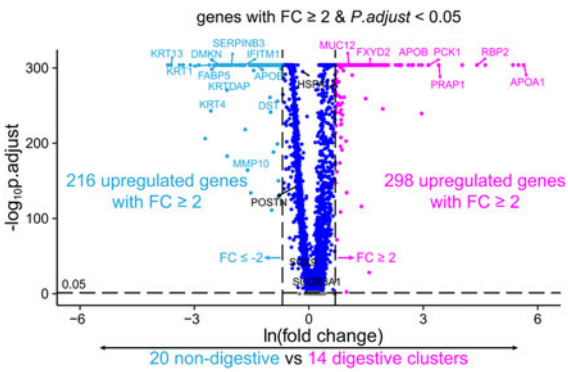

FIGURE 13. Example of code "volcano plot" applied to Figure 5(d) from He et al.[32] Reprinted (open access).

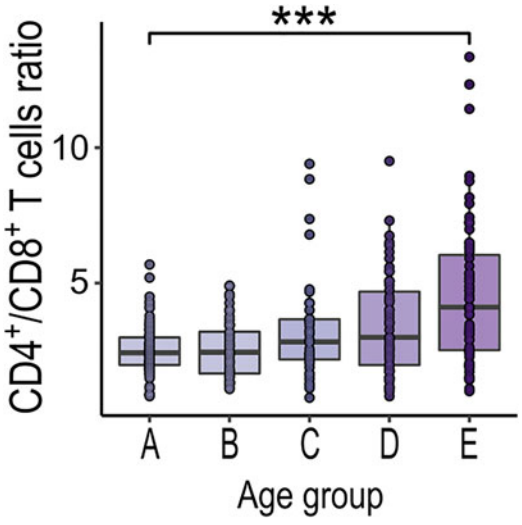

FIGURE 14. Example of "nonstandard plot type" applied to Figure 1(e) from Terekhova et al.[32] Reprinted (open access).

Dimensionality Reduction (DR) Scatterplots Were Used in Almost All Publications

The code "DR" was applied in 44 of the 45 publications to 432 subfigures (23.40%). Almost all DR results were visualized using scatterplots. In the vast majority of the observed DR scatterplots, the points correspond to cells (as opposed to other types of entities).

Points were often colored using a quantitative colormap (e.g., to visualize levels of a biomarker; see Figure 5) or a categorical colormap (e.g., to visualize cell types; see Figure 10). These plots often included text overlays, region outlines, or trajectory arrows.

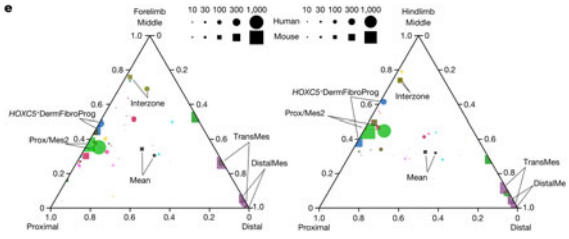

FIGURE 15. Example of "nonstandard plot type" applied to Figure 6(e) from Zhang et al.[30] Reprinted (open access).

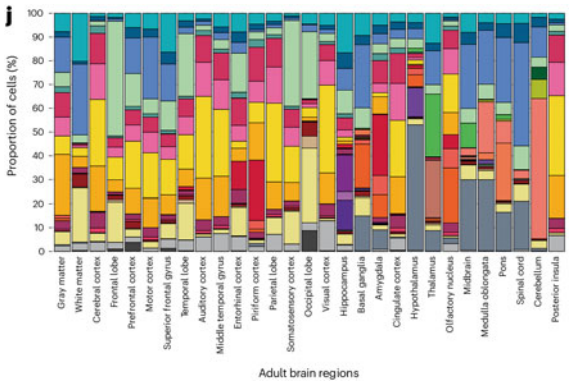

FIGURE 16. Example of code "cell type composition" applied to Figure 1j from Chen et al.[27] Reprinted (open access).

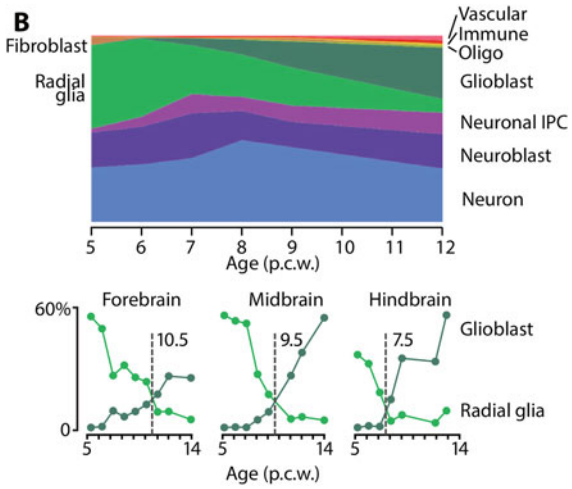

FIGURE 17. Example of code "cell type composition" applied to composite Figure 1(b) containing area and line plots from Braun et al.[34] Reprinted with permission from AAAS.

Observed DR scatterplot visualizations often plotted data subsets, using filtering and highlighting techniques. We also observed 36 instances of sub-clustering (a special case of filtering).

Heatmaps Were Used in Almost All Publications

The code "heatmap" was applied to 244 subfigures (13.22%) in 42 publications. Heatmaps were often used to encode the expression or abundance of a gene or other biomarker across many cells, with metadata bar plot annotations along the axes to indicate categories (see Figure 9). Special cases of heatmaps include "clustermaps" in which a dendrogram is aligned to indicate axis arrangement based on a hierarchical clustering. As clustering provides only a partial leaf ordering, in many cases an optimized leaf ordering procedure is also used. Alternative axis orderings include pseudotime or pseudoaxis. Another observed type of heatmap is the confusion matrix, used to demonstrate model performance (e.g., for cell type prediction; coded as "model performance"). In yet other cases, heatmaps were used to display patterns of cell–cell or ligand–receptor interaction (both coded as "ligand–receptor").

Cell Type Composition was a Common Goal, Often via Stacked Bar Plots

We applied the code "cell type composition" to 108 subfigures (5.85%) from 39 publications. These often (in 76 cases, 70.37%) used a stacked bar plot visualization design, often normalized (see Figure 16), with colors indicating cell types and individual bars stratifying groups of cells (e.g., grouping by tissue region, donor, or condition). We observed a range of cell type granularity, with plots containing from two to tens of cell types. Alternative designs for cell type composition included small multiples of pie charts. In certain cases, conditions were ordered (e.g., in developmental studies) enabling usage of line or area plot designs (see Figure 17).

Comparison was a Common Goal, via a Diversity of Visualization Designs

We applied the code "comparison" to 263 subfigures (14.25%) from 38 publications. We noticed three common types of comparisons: case-versus-control (two conditions), among spatial regions, and among timepoints (e.g., developmental stages). In many cases (144 instances, 54.75%), comparison was visualized using small multiples. In other cases, comparison was visualized using a within-plot condition axis, such as disease status or age (see Figures 7 and 17). We hypothesize that we have underestimated the extent to which comparison was an analysis goal, as our methodology only captures explicitly encoded comparisons.

Dot Plots Were Used in the Majority of Publications

We coded 164 occurrences of "dot plot" (8.88% of subfigures) in 36 publications. Dot plots use the radius and color channels of circular visual marks arranged in a grid to encode two quantitative values along two categorical dimensions (see Figure 18). Typical categorical dimensions are biomarker and cell type. Dot color encodes a measure of biomarker

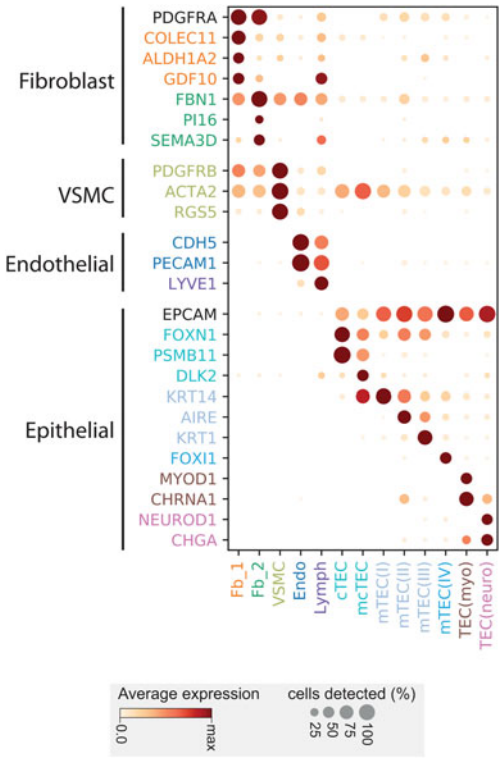

FIGURE 18. Example of code "dot plot" applied to Figure 1(e) from Park et al.[35] Reprinted with permission from AAAS.

abundance among members of the cell type, while radius encodes a proportion of members in which the biomarker level exceeds some threshold. We observed dot plot grids organized according to higher level categories (in line with our observations of heatmap axes).

Other usages of dot plots included those in which the biomarker axis contained pathways (coded as

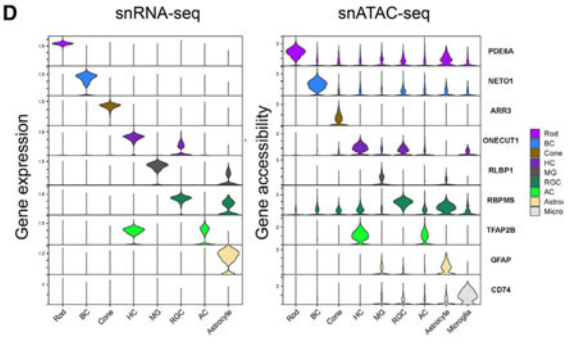

FIGURE 19. Example of code "multiomics" applied to Figure 1 (d) from Liang et al.[36] Reprinted (open access).

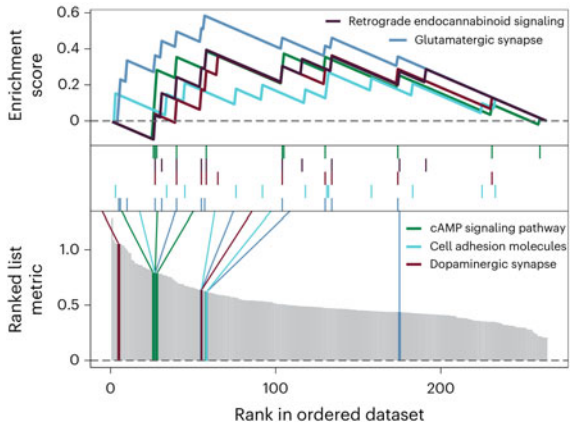

FIGURE 20. Example of code "gsea" applied to Figure 5(f) from Chen et al.[26] Reprinted (open access).

"gsea") and those visualizing ligand–receptor interactions (coded as "ligand–receptor").

Violin Plots Were Often Used to Compare Gene Expression Distributions Among Cell Classes

We applied the code "violin plots" to 75 subfigures (4.06%) from 26 publications. Subcode "violin plot matrix" was applied to 17 subfigures, in which a grid of violin plots share horizontal and vertical axes (similar to the grid of dot plots). Both designs were used to display distributions of biomarker abundance (e.g., gene expression) for multiple cell groups and biomarkers. In some cases, cell groups corresponded to cell types (see Figure 19) yet conditions in others (see Figure 6).

Plots Displaying Gene Set Enrichment Analysis Results Were Used in Approximately Half of the Publications

We applied the code "gsea" to 68 subfigures (3.68%) from 24 publications. This code indicates a plot that encodes the results of a gene set enrichment analysis (GSEA). GSEA identifies biological processes (pathways) associated with specific cell groups or conditions. In some cases, plots encode a full ordered list of genes along with the enrichment score at every index (see Figure 20), yet in other cases a single enrichment score is directly plotted for each pathway (see Figure 21).

Spatial and Multiplexed Imaging Visualizations Were Used in the Majority of Publications

SR visualizations (coded "spatial") appeared in 289 subfigures across 33 publications. This was a broad code, encompassing subcodes including "spatial transcriptomics" and "imaging" (itself including

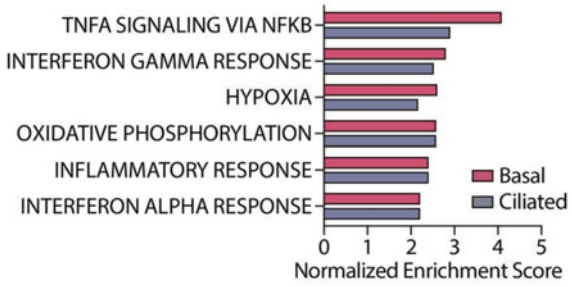

FIGURE 21. Example of code "gsea" applied to Figure 4(e) from Lee et al.[37] Reprinted (open access).

"multiplexed" and "histology" subcodes). Excluding subcodes, the top co-occurring codes with "spatial" were small multiples (229; 79.24%), custom annotations (216; 74.74%), multiscale (87; 30.10%), representative sample (41; 14.19%), and comparison (31; 10.73%). Figure 8 is an example in which many of these codes were applied within a single subfigure: histology, multiplexed, custom annotations, small multiples, multiscale, and representative sample. Common designs for spot-based data included grids of pie charts (see Figure 22) and cell abundance color scales (see Figure 23). We observed direct visualizations of point-based data (see Figure 24) and three instances of subfigures displaying 3-D spatial data.

Patterns Among Multiomics Figures

We applied the code "multiomics" to 31 subfigures (1.68%) from 10 publications. This code indicates a plot which simultaneously displays information from more than one experimental modality, such as both

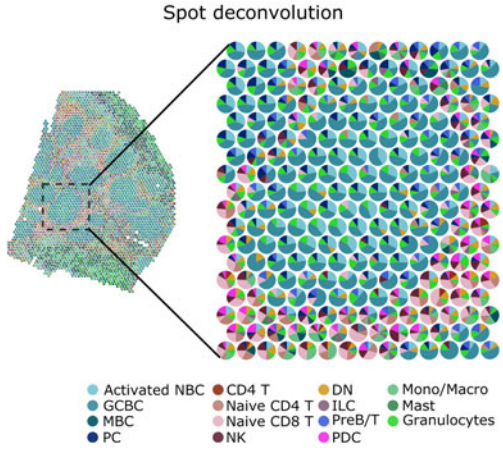

FIGURE 22. Example of "deconvolution" applied to Figure 1g from Massoni-Badosa et al.[38] Reprinted (open access).

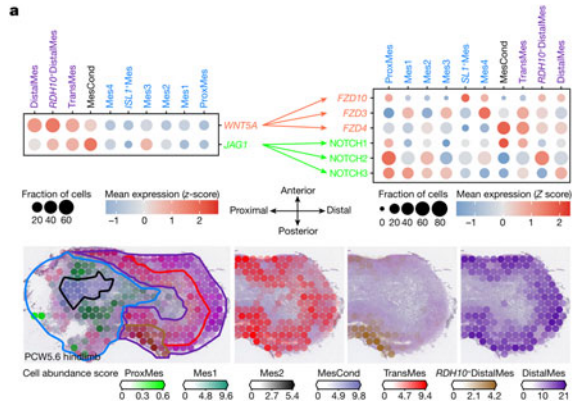

FIGURE 23. Example of code "dot plot" applied to Figure 5(a) from Zhang et al.[29] Reprinted (open access).

transcriptomics (scRNA-seq) and epigenomics (e.g., scATAC-seq). In most cases, different modalities were displayed using small multiples, for example, of box plots (see Figure 25), violin matrix plots (see Figure 19), or scatterplots (see Figure 26). In only a small number of cases were comparisons between modalities directly plotted (see Figure 27).

Patterns Among Cell–Cell and Ligand–Receptor Interaction Figures

Cell–cell interaction refers to communication among cells, enabling coordination of immune responses or other types of signaling. Ligand–receptor interaction is a more specific mechanism where a ligand molecule (e.g., hormone) binds to a receptor protein on a cell's surface (or inside). These interactions can be inferred or predicted from coexpression patterns, ligand–receptor knowledge bases, or proximity.

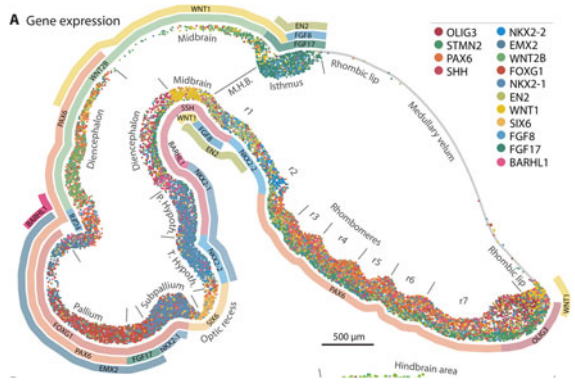

FIGURE 24. Example of "FISH-like" applied to Figure 2(a) from Braun et al.[34] Reprinted with permission from AAAS.

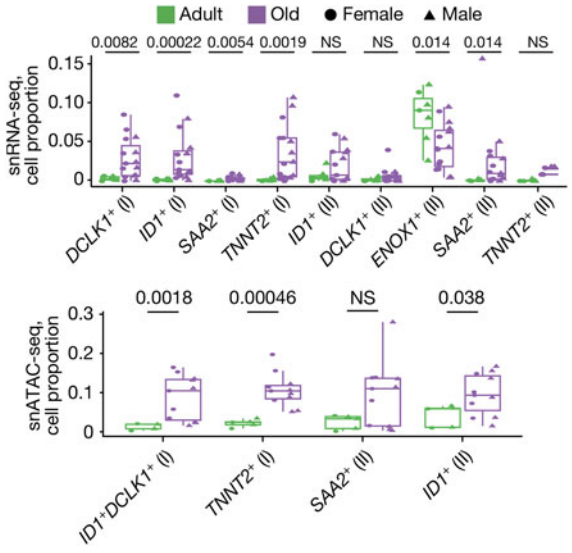

FIGURE 25. Example of code "multiomics" applied to Figure 2(d) from Lai et al.[39] Reprinted (open access).

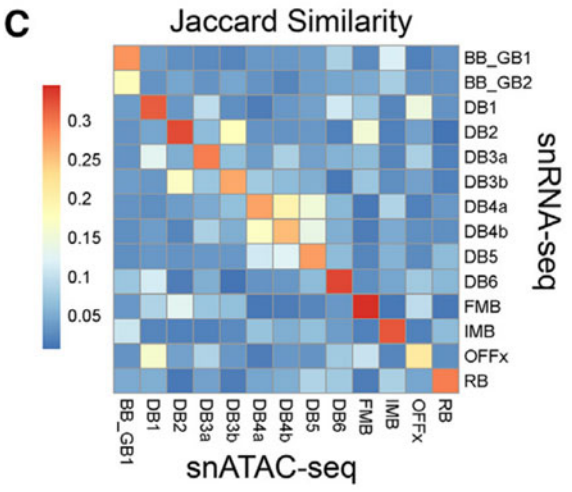

FIGURE 27. Example of code "multiomics" applied to Figure 2(c) from Liang et al.[36] Reprinted (open access).

We used the code "ligand–receptor" to annotate both types of visualizations, applying this code to 73 subfigures (3.95%) from 19 publications. There were a range of visualization designs, including node-link diagrams with circular node arrangements. The subcode "ligand–receptor circle plot" applied to 21 of these from nine publications. In these plots, nodes correspond to cell types and edges indicate interaction strengths between cell type pairs (see Figure 28). Other visualization designs included scatterplots with points corresponding to cell types and axes indicating the strengths of interactions (see Figure 28). Alternatively, visualizations may directly display ligand–receptor interactions for example with dot plots (see Figure 23) or Sankey diagrams.

Patterns Among Pseudotime, Pseudoaxis, and RNA Velocity Figures

Pseudotime inference and RNA velocity algorithms are complementary for studying dynamic cellular processes, and do not require temporally resolved experimental techniques. Many subfigures matched the related codes "pseudotime" (43), "pseudoaxis" (2), and "velocity" (19). Pseudotime trajectories were commonly plotted using arrows layered on top of DR scatterplots (see Figure 11). RNA velocity information was commonly projected into the DR space and plotted using streamlines overlaid on a

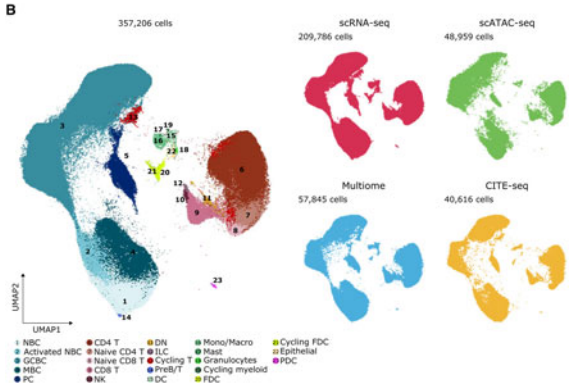

FIGURE 26. Example of code "multiomics" applied to Figure 1(b) from Massoni-Badosa et al.[38] Reprinted (open access).

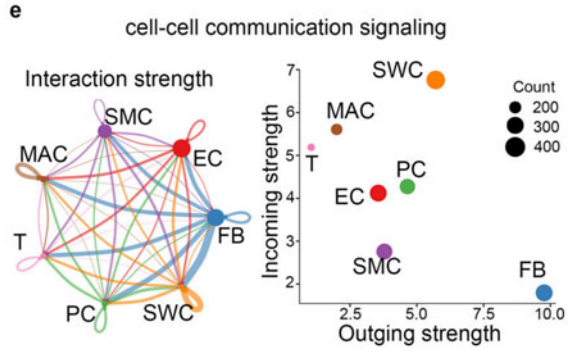

FIGURE 28. Example of code "ligand–receptor" applied to Figure 1(e) from Zhao et al.[25] Reprinted (open access).

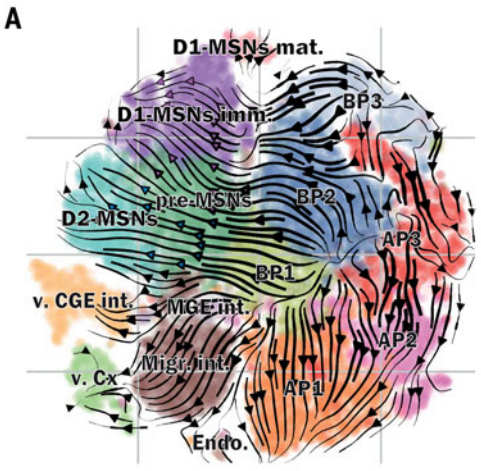

FIGURE 29. Example of code "velocity" applied to Figure 2(a) from Bocchi et al.[40] Reprinted with permission from AAAS.

scatterplot to indicate cell state transitions (see Figure 29). Other pseudotime visualization designs included heatmaps and line charts with pseudotime along the horizontal axis. Similar heatmap designs were used for "pseudoaxis" plots, which aim to order cells along a spatial axis.

DISCUSSION

Our focus on publication figures limits the scope of this survey. We hypothesize that publication figures, especially in the main text, will be biased toward confirmatory (as opposed to exploratory) goals, as they are intended to communicate findings to a journal audience. Included figures will also reflect journal requirements. Future analyses could consider supplementary figures.

Our analysis of figures, while valuable, disregards additional context available. Captions and text references can guide the audience to the analysis goal that motivated the figure's inclusion in the article. Future work should relate subfigures to their corresponding text to investigate whether visualizations align with authors' motivations and conclusions. A related opportunity is to integrate our findings with existing visualization taxonomies.

This survey could benefit from a more semantic approach to paper identification to overcome keyword search limitations. Our eligibility criteria, while justified, likely impacts experimental modalities differently (e.g., more recent techniques, if less scalable, will be underrepresented in included papers). Future surveys should consider that experimental scalability is a moving target and should be considered separately for each modality.

Because many single-cell papers use the same software toolkits for not only data analysis but also visualization, our results are intertwined with the set of plot types they offer. However, the open-source nature of such toolkits might mean that sufficient demand would negate this consideration. We emphasize that visualization contributions should both advance atlasing goals and be made interoperable with established toolkits in the domain.

Lessons for Future Designs

Our results suggest that atlas publications use a relatively small set of visualization designs, with common ways to map from the data abstractions of the domain onto visual properties. Yet the large number of usages of the "nonstandard plot type" code also suggests that publications often aim to communicate findings for which an established chart type does not exist within the domain. In relation to visualization practices for genomics data, this finding is consistent (i.e., commonality of a design, such as genome browser tracks in genomics, yet also usage of many other designs).[13]

Concepts in analysis of single-cell data intersect with those in existing domains studied by visualization researchers, including analysis of electronic health records, genomics, time-series, molecular, and bioimaging data types. However, new paradigms are needed, such as at the single-cell domain's intersections with the fields of genomics and bioimaging, where these fields can rely on coordinate systems (the genome and physical space, respectively) yet consideration of individual cells adds a new (cellular) dimension that lacks an obvious coordinate system.

Multimodal and Spatially Resolved (SR) Designs

While the publications we analyzed included several multimodal and epigenomics studies, our findings lack depth in these areas. Among the figures we surveyed, those displaying multimodal data primarily used small multiples (i.e., one plot instance per modality). Examples of other designs (not captured by our survey methodology) included Figure 2B of Ober-Reynolds et al.[41] (aligned genomic tracks and horizontal violin plots per cell type for chromatin accessibilty and transcriptomics modalities, respectively), Figure 6D of Cusanovich et al.[42] (aligned genomic tracks and differentiation trajectories with highlighted branches for chromatin accessibility data), and Figures 3G and 5F of Zhang et al.[43]

(visualization of chromatin accessibility, gene expression, and splicing events). We reproduce these example figures in Section S8 of the Supplementary Material.

Regarding SR studies, we primarily identified cell–cell interaction and 2-D bioimaging figures, despite the existence of other modalities and enthusiasm around 3-D data. Examples of alternative visual encodings for SR data include line plots for isodepth quantities,[44] dot plots for feature importance quantities,[45] and heatmaps for metagene spatial affinities.[46] As these examples demonstrate, in addition to publications describing datasets, those describing bioinformatic methods can also serve as sources of visualization design inspiration.

Notable Characteristics and Designs

Important characteristics of surveyed figures include filtering and highlighting (e.g., to highlight a particular cell type), small multiples (e.g., to visualize the expression distributions of many genes simultaneously), and custom annotations (e.g., to point out an important image region). Visualization designs that were less prevalent yet we find them to be particularly useful include line plots (e.g., to display the value of a score for different groups of cells, often over developmental time or pseudotime) and non-DR scatterplots (e.g., gating scatterplots, volcano plots, and incoming–outgoing interaction plots). Designs, such as dot plots, volcano plots, ligand–receptor circle plots, deconvolution pie charts, and velocity arrow scatterplots, are notable for their uniquely widespread usage in the single-cell domain.

Interactivity

There are already interactive tools for single-cell data analysis, including several atlas data portals which embed or link to interactive visualizations. While current tools, such as CELLxGENE Discover[47] and Vitessce,[48] support many plot types and interactive features (including dimensionality reduction scatterplots, spatial and imaging views, dot plots, violin plots, and histograms), they currently lack robust support for several of the aforementioned notable characteristics, including small multiples, custom annotations, and non-DR scatterplots. These interactive platforms also overlook surveyed designs, such as anatomical and experimental design diagrams, which can help provide context and data provenance.

A complete survey of interactive tools remains important future work. In the meantime, our findings can inform the development of tools which offer seamless transition between exploration, confirmation, and creation of static figures.

CONCLUSION

This report provides a timely perspective into visualization usage in the single-cell domain. In contrast to surveys that restrict their analysis to visualization venues, we survey the state of visualization in the target domain using scientific publications as a lens. While this means analysis of plots that have not necessarily been published for their visualization merit, this method can teach us about barriers to adoption of visualization best practices or to prompt their revision. By providing biological background and a survey of visualization usage, this report invites visualization researchers to contribute at the intersection of both domains and offers biological researchers a new perspective on visualization.

COMPETING INTERESTS

Nils Gehlenborg is a co-founder and equity owner of Datavisyn.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health under Grant OT2OD026677, Grant OT2OD033758, Grant T15LM007092, Grant R33CA263666, and Grant T32HG002295.

REFERENCES

1. A. Regev et al., "The human cell atlas," *eLife*, vol. 6, Dec. 2017, Art. no. e27041.
2. R. Elmentaite et al., "Single-cell atlases: Shared and tissue-specific cell types across human organs," *Nature Rev. Genet.*, vol. 23, no. 7, pp. 395–410, 2022.
3. K. Hrovatin et al., "Considerations for building and using integrated single-cell atlases," *Nature Methods*, vol. 22, pp. 41–57, Dec. 2024.
4. M. D. Luecken et al., "Benchmarking atlas-level data integration in single-cell genomics," *Nature Methods*, vol. 19, no. 1, pp. 41–50, 2022.
5. J. Lause, P. Berens, and D. Kobak, "The art of seeing the elephant in the room: 2D embeddings of single-cell data do make sense," *PLoS Comput. Biol.*, vol. 20, no. 10, Oct. 2024, Art. no. e1012403.
6. S. Domcke and J. Shendure, "A reference cell tree will serve science better than a reference cell atlas," *Cell*, vol. 186, no. 6, pp. 1103–1114, Mar. 2023.
7. J. E. Rood et al., "Toward a common coordinate framework for the human body," *Cell*, vol. 179, no. 7, pp. 1455–1467, Dec. 2019.

8. B. Cakir et al., "Comparison of visualization tools for single-cell RNAseq data," *NAR Genomic. Bioinf.*, vol. 2, no. 3, 2020, Art. no. lqaa052.
9. L. Zappia and F. J. Theis, "Over 1000 tools reveal trends in the single-cell RNA-seq analysis landscape," *Genome Biol.*, vol. 22, no. 1, 2021, Art. no. 301.
10. A. J. Pretorius, I. A. Khan, and R. J. Errington, "A survey of visualization for live cell imaging," *Comput. Graphics Forum*, vol. 36, no. 1, pp. 46–63, 2017.
11. J. Beyer et al., "A survey of visualization and analysis in high-resolution connectomics," *Comput. Graphics Forum*, vol. 41, no. 3, pp. 573–607, 2022.
12. L. A. Garrison et al., "Trends & opportunities in visualization for physiology: A multiscale overview," *Comput. Graphics Forum*, vol. 41, no. 3, pp. 609–643, 2022.
13. S. Nusrat, T. Harbig, and N. Gehlenborg, "Tasks, techniques, and tools for genomic data visualization," *Comput. Graphics Forum*, vol. 38, no. 3, pp. 781–805, 2019.
14. B. Kozlíková et al., "Visualization of biomolecular structures: State of the Art revisited," *Comput. Graphics Forum*, vol. 36, no. 8, pp. 178–204, 2017.
15. J. E. Rood et al., "The human cell atlas from a cell census to a unified foundation model," *Nature*, vol. 637, no. 8048, pp. 1065–1071, 2025.
16. J. E. Rood et al., "Impact of the human cell atlas on medicine," *Nature Med.*, vol. 28, no. 12, pp. 2486–2496, Dec. 2022.
17. M. Adler, A. R. Chavan, and R. Medzhitov, "Tissue biology: In search of a new paradigm," *Annu. Rev. Cell Devlop. Biol.*, vol. 39, no. 1, pp. 67–89, Oct. 2023.
18. K. Vandereyken et al., "Methods and applications for single-cell and spatial multi-omics," *Nature Rev. Genet.*, vol. 24, no. 8, pp. 494–515, Mar. 2023.
19. R. Argelaguet et al., "Computational principles and challenges in single-cell data integration," *Nature Biotechnol.*, vol. 39, no. 10, pp. 1202–1215, 2021.
20. V. Petukhov et al., "Case-control analysis of single-cell RNA-seq studies," *bioRxiv*, Mar. 2022, doi: 10.1101/2022.03.15.484475.
21. G. Palla, D. S. Fischer, A. Regev, and F. J. Theis, "Spatial components of molecular tissue biology," *Nature Biotechnol.*, vol. 40, pp. 308–318, 2022.
22. V. Svensson, E. da Veiga Beltrame, and L. Pachter, "A curated database reveals trends in single-cell transcriptomics," *Database*, vol. 2020, 2020, Art. no. baaa073.
23. N. F. Samatova et al., "An outlook into ultra-scale visualization of large-scale biological data," in *Proc. 2008 Workshop Ultrascale Visual.*, 2008, pp. 29–39.
24. B. B. Lake et al., "An atlas of healthy and injured cell states and niches in the human kidney," *Nature*, vol. 619, no. 7970, pp. 585–594, Jul. 2023.
25. L. Zhao et al., "Single-cell transcriptome atlas of the human corpus cavernosum," *Nature Commun.*, vol. 13, Jul. 2022, Art. no. 4302.
26. T. van Zyl et al., "Cell atlas of the human ocular anterior segment: Tissue-specific and shared cell types," *PNAS*, vol. 119, Jul. 2022, Art. no. e2200914119.
27. X. Chen et al., "A brain cell atlas integrating single-cell transcriptomes across human brain regions," *Nature Med.*, vol. 30, no. 9, pp. 2679–2691, 2024.
28. M. Terekhova et al., "Single-cell atlas of healthy human blood unveils age-related loss of NKG2C GZMB-CD8 memory T cells and accumulation of type 2 memory T cells," *Immunity*, vol. 56, no. 12, pp. 2836–2854, Dec. 2023.
29. B. Zhang et al., "A human embryonic limb cell atlas resolved in space and time," *Nature*, vol. 635, no. 8039, pp. 668–678, 2024.
30. E. Lengyel et al., "A molecular atlas of the human postmenopausal fallopian tube and ovary from single-cell RNA and ATAC sequencing," *Cell Rep.*, vol. 41, Dec. 2022, Art. no. 111838.
31. J. Cao et al., "A human cell atlas of fetal gene expression," *Science*, vol. 370, no. 6518, Nov. 2020, Art. no. eaba7721.
32. S. He et al., "Single-cell transcriptome profiling of an adult human cell atlas of 15 major organs," *Genome Biol.*, vol. 21, no. 294, Dec. 2020, Art. no. 294.
33. D. Chen et al., "Single-cell RNA-seq with spatial transcriptomics to create an atlas of human diabetic kidney disease," *FASEB J.*, vol. 37, Jun. 2023, Art. no. e22938.
34. E. Braun et al., "Comprehensive cell atlas of the first-trimester developing human brain," *Science*, vol. 382, no. 6667, Oct. 2023, Art. no. eadf1226.
35. J.-E. Park et al., "A cell atlas of human thymic development defines T cell repertoire formation," *Science*, vol. 367, no. 6480, Feb. 2020, Art. no. eaay3224.
36. Q. Liang et al., "A multi-omics atlas of the human retina at single-cell resolution," *Cell Genomic.*, vol. 3, no. 6, Jun. 2023, Art. no. 100298.
37. W. Lee et al., "A single-cell atlas of in vitro multiculture systems uncovers the in vivo lineage trajectory and cell state in the human lung," *Exp. Mol. Med.*, vol. 55, no. 8, pp. 1831–1842, Aug. 2023.
38. R. Massoni-Badosa et al., "An atlas of cells in the human tonsil," *Immunity*, vol. 57, no. 2, pp. 379–399, Feb. 2024.

39. Y. Lai et al., "Multimodal cell atlas of the ageing human skeletal muscle," *Nature*, vol. 629, no. 8010, pp. 154–164, 2024.
40. V. D. Bocchi et al., "The coding and long noncoding single-cell atlas of the developing human fetal striatum," *Science*, vol. 372, no. 6542, May 2021, Art. no. eabf5759.
41. B. Ober-Reynolds et al., "Integrated single-cell chromatin and transcriptomic analyses of human scalp identify gene-regulatory programs and critical cell types for hair and skin diseases," *Nature Genet.*, vol. 55, no. 8, pp. 1288–1300, Jul. 2023.
42. D. A. Cusanovich et al., "A single-cell atlas of in vivo mammalian chromatin accessibility," *Cell*, vol. 174, no. 5, pp. 1309–1324, Aug. 2018.
43. S. Zhang et al., "Simultaneous profiling of RNA isoforms and chromatin accessibility of single cells of human retinal organoids," *Nature Commun.*, vol. 15, no. 1, Sep. 2024, Art. no. 8022.
44. U. Chitra et al., "Mapping the topography of spatial gene expression with interpretable deep learning," *Nature Methods*, vol. 22, pp. 298–309, Jan. 2025.
45. A. B. Esselman et al., "In situ molecular profiles of glomerular cells by integrated imaging mass spectrometry and multiplexed immunofluorescence microscopy," *Kidney Int.*, vol. 107, no. 2, pp. 332–337, Feb. 2025.
46. B. Chidester et al., "SPICEMIX enables integrative single-cell spatial modeling of cell identity," *Nature Genet.*, vol. 55, no. 1, pp. 78–88, Jan. 2023.
47. CZI Cell Science Program et al., "CZ CELLxGENE discover: A single-cell data platform for scalable exploration, analysis and modeling of aggregated data," *Nucleic Acids Re.*, vol. 53, no. D1, pp. D886–D900, Jan. 2025.
48. M. S. Keller et al., "Vitessce: Integrative visualization of multimodal and spatially resolved single-cell data," *Nature Methods*, vol. 22, no. 1, pp. 63–67, 2025.

MARK S. KELLER is currently working toward the Ph.D. degree at Harvard Medical School, Boston, MA, 02115, USA. His research focuses on developing visualization tools for single-cell data. Contact him at mark_keller@hms.harvard.edu.

ERIC MÖRTH is a postdoctoral fellow at Harvard Medical School, Boston, MA, 02115, USA. His research focuses on the visualization of large-scale 3-D tissue images using novel rendering techniques. Contact him at ericmoerth@fas.harvard.edu.

THOMAS C. SMITS is an associate in biomedical informatics at Harvard Medical School, Boston, MA, 02115, USA. His research focuses on performing analyses of data from the Human BioMolecular Atlas Program. Contact him at tsmits@hms.harvard.edu.

SIMON WARCHOL is currently working toward the Ph.D. degree at Harvard University, Allston, MA, 02134, USA. His research interests include visual analytics and visualization for cancer biology applications. Contact him at simonwarchol@g.harvard.edu.

GRACE GUO is a postdoctoral fellow at Harvard University, Allston, MA, 02134, USA. Her research focuses on building human-centered explainability tools for AI, with an emphasis on biomedical domains. Contact her at gguo31@seas.harvard.edu.

QIANWEN WANG is an assistant professor in computer science and engineering at the University of Minnesota, Twin Cities, Minneapolis, MN, 55455, USA. Her research focuses on enhancing interactions between users and AI through visualizations, particularly to address biomedical challenges. Contact her at qianwen@umn.edu.

ROBERT KRUEGER is an assistant professor in computer science and engineering at New York University, Brooklyn, NY, 11201, USA. His research focuses on visual analytics for spatially resolved data with a focus on biomedical applications. Contact him at rk4815@nyu.edu.

HANSPETER PFISTER is An Wang Professor in Computer Science at Harvard University, Allston, MA, 02134, USA. His research interests include scientific visualization, information visualization, and computer graphics. Contact him at pfister@seas.harvard.edu.

NILS GEHLENBORG is an Associate Professor of Biomedical Informatics at Harvard Medical School, Boston, MA, 02115, USA. His research focuses on developing interfaces and algorithms that enable interaction with biomedical data. He is the corresponding author of this article. Contact him at nils@hms.harvard.edu.