Spatial and temporal tools for building a human cell atlas

Jonah Cool[a], Richard S. Conroy[b,*], Sean E. Hanlon[c], Shannon K. Hughes[d], and Ananda L. Roy[b]
[a]Chan Zuckerberg Initiative, Redwood City, CA 94063; [b]Office of Strategic Coordination, Division of Program Coordination, Planning, and Strategic Initiatives, National Institutes of Health, Bethesda, MD 20852; [c]Center for Strategic Scientific Initiatives, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892; [d]Division of Cancer Biology, National Cancer Institute, National Institutes of Health, Rockville, MD 20850

ABSTRACT **Improvements in the sensitivity, content, and throughput of microscopy, in the depth and throughput of single-cell sequencing approaches, and in computational and modeling tools for data integration have created a portfolio of methods for building spatiotemporal cell atlases. Challenges in this fast-moving field include optimizing experimental conditions to allow a holistic view of tissues, extending molecular analysis across multiple timescales, and developing new tools for 1) managing large data sets, 2) extracting patterns and correlation from these data, and 3) integrating and visualizing data and derived results in an informative way. The utility of these tools and atlases for the broader scientific community will be accelerated through a commitment to findable, accessible, interoperable, and reusable data and tool sharing principles that can be facilitated through coordination and collaboration between programs working in this space.**

Monitoring Editor
David G. Drubin
University of California, Berkeley

Received: Apr 16, 2019
Revised: Jul 25, 2019
Accepted: Jul 29, 2019

OPPORTUNITIES

Human development and reproduction create a fascinating biomolecular symphony; a high-fidelity spatiotemporal system capable of going from a single cell to a vast ecosystem of tens of trillions of cells and back through the single-cell bottleneck repeatedly and faithfully. The lifecycle of all multicellular organisms involves dynamic processes that occur across many timescales and spatial contexts, from chromatin reorganization within the nucleus, to the formation of protein–protein interaction networks in the cytoplasm, to intercellular interactions that drive extracellular matrix and tissue remodeling and, finally, to aging of the organism. The suite of modern tools and technologies to study complex spatiotemporal patterns in development and aging is redefining the notion of a cell atlas, and the insights an atlas can provide regarding an organism's ability to maintain homeostasis in the face of diverse perturbations and dysfunctions. Atlas building extends back centuries and historically was rooted in close anatomical observations based on morphology of tissue and localization of microscopic structures. The emergence of quantitative, high-resolution, high-content, high-throughput tools that can be used to observe cells in situ is pushing us toward a deeper understanding of the role of spatiotemporal patterns in tissues and organisms. This is an exciting moment in cellular and molecular biology. Here, we briefly discuss three sets of technologies that are poised to unify diverse atlas efforts and identify some challenges that give rise to the need for coordination and collaboration across the scientific community, including both scientists and funders.

Multiscale microscopy

Organism-wide cell atlas efforts began with the groundbreaking work of Sulston and Horvitz (1977), who methodically tracked the lineage relationship of *Caenorhabditis elegans* hermaphrodites and males to define the lineage and identify their 959 or 1031 somatic cells, respectively. Dramatic performance improvements offered by light-sheet microscopy have taken these observational studies to a new level, tracking the lineage of tens to hundreds of millions of cells over several days in the early development of zebrafish (Keller *et al.*, 2008), *Drosophila* (Tomer *et al.*, 2012), and mammals (Han *et al.*, 2018; McDole *et al.*, 2018). The ability to directly label single cells and track them over time has also significantly improved

DOI:10.1091/mbc.E18-10-0667

*Address correspondence to: Richard Conroy (Richard.Conroy@nih.gov).

Abbreviations used: BICCN, BRAIN Initiative Cell Census Network; BRAIN, Brain Research through Advancing Innovative Neurotechnologies Initiative; CRISPR, clustered regularly interspaced short palindromic repeats; EMAP, e-Mouse Atlas Project; FAIR, findable, accessible, interoperable, and reusable; FISH, fluorescence in situ hybridization; GTEx, Genotype-Tissue Expression Project; HTAN, Human Tumor Atlas Network; HuBMAP, Human BioMolecular Atlas Program; IHEC, International Human Epigenetics Consortium; IHMC, International Human Microbiome Consortium; ImmGen, Immunological Genome Project; IMPC, International Mouse Phenotyping Consortium; LINCS, Library of Integrated Network-Based Cellular Signatures Program; NHP, nonhuman primate; NIH, National Institutes of Health; UV, ultraviolet; ZEBrA, Zebra Finch Expression Brain Atlas.

(Frieda *et al.*, 2017; Takei *et al.*, 2017). In parallel, the toolbox for characterizing clonal and spatiotemporal relationships between cells has grown with genetic markers introduced through viral vectors (Biddy *et al.*, 2018), time-dependent modification of cellular components (Herzog *et al.*, 2017), recombination-activated multicolor fluorescent reporters (Kretzschmar and Watt, 2012), and perturbation techniques such as optogenetics (Johnson *et al.*, 2017). New microscopy methods such as lattice light-sheet imaging (Chen *et al.*, 2014) are not only increasing resolution but also expanding the volume and speed of imaging (Liu *et al.*, 2018), and light-sheet imaging in combination with tissue clearing has opened up imaging to span from the subcellular to the organismal level (Belle *et al.*, 2017). These approaches are giving rise to many opportunities and challenges for linking together ongoing atlas-generating programs that exist at the nuclear (Cusanovich *et al.*, 2018), cellular (Cai *et al.*, 2018), and organismal levels (Regev *et al.*, 2017).

Integrated and multiplexed assays

New technologies are opening additional possibilities for multiplexed measurements within a single sample. Light microscopy is nondestructive and has high spatial and temporal resolution, though it has historically been limited by throughput, lack of quantification, and dependency on the affinity labeling of targets of interest. Recent methods have pushed beyond these limits. UV (Fereidouni *et al.*, 2015) and midinfrared (Yeh *et al.*, 2015) label-free imaging can monitor cellular and subcellular structures. Highly multiplexed methods can measure the transcriptome in tissue sections, not just dissociated cells (Moffitt and Zhuang, 2016; Shah *et al.*, 2018), and dense markers are being applied to the proteome as well (Goltsev *et al.*, 2018; Gut *et al.*, 2018). These advances have resulted in a convergence of imaging and "omics" techniques, where data can be anchored and compared across different assays (Stuart *et al.*, 2019). Complementing these multiplexed approaches, the integration of methods such as tissue clearing (Cai *et al.*, 2019) and expansion microscopy (Gao *et al.*, 2019) into both traditional and highly multiplexed approaches is pushing the current limits on volume, spatial resolution, and molecular depth.

Analytical methods and data integration

Waddington was visionary in presenting spatiotemporal decisions as a manifold that a single cell, or collection of cells, must navigate, conceptualizing development as a quantitative dynamic system (Waddington, 1957). This vision continues to inspire new analytical and computational approaches that are a core component of modern atlas-building efforts. For instance, it is now possible to predict subcellular structures from label-free images (Chen *et al.*, 2018) and the probabilistic fate of the cell can be estimated through analysis of nascent versus mature transcripts (La Manno *et al.*, 2018). The clonality of differentiation can be inferred from genomic scarring (Raj *et al.*, 2018), diverse data types can be normalized and even integrated across conditions, modalities, and species (Butler *et al.*, 2018), and single cells tracked through pseudotemporal molecular analysis (Bendall *et al.*, 2014; Haghverdi *et al.*, 2016; Qiu *et al.*, 2017).

Increasingly, biologists can acquire more high-quality single-cell resolution data than they can analyze. The future holds great promise for sharing and mining rich imaging data filled with features of known, and unknown, significance. This creates a challenge in finding, accessing, interpreting, and extracting knowledge across many data types and experimental protocols. Accordingly, computational biology, bioinformatics, and systems biology sit at the center of most modern atlas efforts, connecting various atlas efforts and enabling diverse communities to utilize the fruits of such efforts. Continued advances in modeling and understanding complex processes such as human development, malignancy, and human homeostasis will help us learn the state space and relationships among them. Additionally, advances in integrative visualization approaches can help make atlases accessible and beneficial to experimental biologists, computational scientists, and clinicians.

CHALLENGES

Each of the promising approaches described above has limits. For multiscale analysis, it remains extremely difficult, and sometimes impossible, to study multiple biomolecules in living human cells with high resolution. For multiplexed assays, there is a need to integrate sparse temporal data collected on live cells and tissue with more detailed molecular snapshots available after fixation. Computational approaches need analytical methods that are scalable, prognostic, and generalizable, and better approaches for establishing ground-truth.

To address these bottlenecks in tracking spatiotemporal dynamics in complex populations of cells, we need better tools. First, experimental conditions, tissue collection and preprocessing times, tissue preservation conditions, composition of matrix and media, and microscope and environmental stability can all significantly impact the quality of data generated (Ferreira *et al.*, 2018). Increasingly sophisticated and automated tissue chip devices that allow for controlled culture of multiple tissues in microscope-friendly multiwell platforms is one promising area that may address this bottleneck (Skardal *et al.*, 2016).

Second, complementary and comprehensive reductionist approaches are needed to link biomolecular spatiotemporal dynamics in individual cells and bridge the spatial scales to cells in a multicellular organism. From rapid cytokine signaling to the essential role of long-lived proteins in cellular structures such as nuclear pores (Toyama *et al.*, 2013), analytical tools and models are nearly always limited to a reductionist approach in space, time, or molecular complexity. Many single-cell analysis techniques do not take the cellular environment and signaling into account, either because they dissociate cells for analysis or because they examine only a few biomarkers in specific cell types. The development and integration of techniques that enable precise fluorescent labeling of individual molecules for readout of gene expression are an exciting advance enabling more comprehensive spatiotemporal analysis, though plenty of challenges and opportunities remain.

Third, the computational field should generate predictive models for complex dynamics and emergent behavior that are not overconstrained by either experiment or theory. This field needs ground-truth data sets for comparing methods, approaches for comparing models such as pseudotemporal analysis with direct experimental results, and more biological input; for example, a deep understanding of receptor–ligand interactions and their dynamics in tissues (Nandagopal *et al.*, 2018).

Finally, like any new science, cell atlas approaches must balance the inclusion of new technologies with a focus on ensuring rigor and reproducibility. Good data stewardship calls for thoughtful curation, annotation, maintenance, and release of data and metadata.

PERSPECTIVE

We, the authors, as part of a larger international community, are working together to synergistically support the development of new tools to systematically build and analyze human cell atlases of normal and diseased tissue. Tables 1 and 2 list some of the technologies and programs currently part of this ecosystem. We believe there exists opportunity for organizations that support similar

Biomolecule	Assay/composition	Spatial resolution	Temporal resolution
DNA	Sequencing	Fluorescence in situ hybridization (FISH), Electron microscopy	CRISPR imaging, Small molecule dyes
RNA	Sequencing	Fluorescent in situ hybridization (FISH), Spatial-encoded sequencing	CRISPR, Tags
Proteins	Mass spectrometry, Immuno-tagged sequencing	Imaging mass spectrometry, Immunofluorescence	Reporters
Small molecules	Mass spectrometry	Imaging mass spectrometry	Indicators
Lipids	Mass spectrometry	Imaging mass spectrometry	Reporters

TABLE 1: General techniques used for identification of specific biomolecules.

Biosample	Development	Normal	Pathology
Cultured cells	Allen Institute for Cell Science	4D Nucleome, LINCS, Human Protein Atlas	Cancer Cell Line Encyclopedia
One organ	LungMAP, NHP Brain Atlas	BRAIN, Kidney Precision Medicine Program, Brain Maps, ZEBrA	Kidney Precision Medicine Program, Gut Cell Atlas
Several organs/ systems	GenitoUrinary Development Molecular Anatomy Project	HuBMAP, ImmGen	Human Tumor Atlas Network, The Cancer Genome Atlas, Clinical Proteomic Tumor Analysis Consortium
Organism-wide	Human Cell Atlas, EMAGE	GTEx, Human Cell Atlas, FANTOM, EMAP	Human Protein Atlas

TABLE 2: Examples of current programs engaged in comprehensive molecular characterization of cells and tissues.

research activities to actively coordinate and collaborate between programs, share techniques, cross-validate results, and develop standards where ones do not exist. For example, the Human Cell Atlas (HCA) initiative, with key support by the Chan Zuckerberg Initiative, is working closely with several National Institutes of Health (NIH)-funded programs, including the BRAIN Initiative Cell Census Network (BICCN), the Human Tumor Atlas Network (HTAN), and the Human Biomolecular Atlas Program (HuBMAP) to develop a common coordinate framework for the human body that will enable integration of data across these programs. Furthermore, the NIH and the HCA will hold a joint meeting in the spring of 2020 to bring many different stakeholder groups together to catalyze discussion among the different communities on reaching a consensus on data formats, metadata standards, and how to realize data and software FAIRness (FAIR: findable, accessible, interoperable, and reusable). In addition, many funders are promoting the use of preprint servers and services for sharing experimental and computational protocols.

This cell atlas ecosystem is also learning from successful international consortia such as the International Human Epigenetics Consortium (IHEC), the International Mouse Phenotyping Consortium (IMPC), and the International Human Microbiome Consortium (IHMC) that coordinate activities in their respective fields. For example, representatives of all funders with a shared interest in building human cell atlases are invited to regular phone calls hosted by the Chan Zuckerberg Initiative and attended by an international mix of private and public funders, including representatives from related NIH programs. Through this forum, we can coordinate funding opportunities and minimize duplication of efforts without sacrificing the autonomy of each funder. We also work toward coordination and collaboration on validating tools, sharing protocols and reagents, developing standards where none currently exist, and making data open and FAIR (Wilkinson *et al.*, 2016). Working together, we believe there are exciting opportunities to support the community as it integrates spatiotemporal data sets of molecular information, cellular states, and tissues in normal and disease contexts.

Generating cohesive multidimensional maps of normal and diseased tissues and providing them in a user-friendly environment for the research and clinical communities will be a key outcome for atlas-generating programs. As these atlas building programs progress, we will reach out to the broader research community to help identify and define use-cases that can drive the generation of approaches for presenting the integrated data sets as unified and interactive atlases. Building the integrated atlases that are accessible, interactive, and include the necessary data will be key for allowing researchers from basic to clinical sciences to ask and answer new questions.

As Dyson (2012) noted, new tools let us discover new "monsters" that we must study with new and more precise tools before we can understand them. Dyson also noted that ideas must go hand-in-hand with tools to drive science forward. For atlas-building programs to reach their potential, they need to inspire the wider scientific research community to generate new concepts and integrated models, which in turn will require new tools to elucidate cellular intricacies.

ACKNOWLEDGMENTS

R.S.C. and A.L.R. acknowledge the support of the Office of Strategic Coordination and the Division of Program Coordination, Planning, and Strategic Initiatives, NIH. We thank Elizabeth Wilder for comments and suggestions but assume sole responsibility for the views expressed herein.

REFERENCES

Belle M, Godefroy D, Couly G, Malone SA, Collier F, Giacobini P, Chedotal A (2017). Tridimensional visualization and analysis of early human development. Cell 169, 161–173. e112.

Bendall SC, Davis KL, Amir el-AD, Tadmor MD, Simonds EF, Chen TJ, Shenfeld DK, Nolan GP, Pe'er D (2014). Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. Cell 157, 714–725.

Biddy BA, Kong W, Kamimoto K, Guo C, Waye SE, Sun T, Morris SA (2018). Single-cell mapping of lineage and identity in direct reprogramming. Nature 564, 219–224.

Butler A, Hoffman P, Smibert P, Papalexi E, Satija R (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 36, 411–420.

Cai Y, Hossain MJ, Hériché JK, Politi AZ, Walther N, Koch B, Wachsmuth M, Nijmeijer B, Kueblbeck M, Martinic-Kavur M, *et al.* (2018). Experimental and computational framework for a dynamic protein atlas of human cell division. Nature 561, 411–415.

Cai R, Pan C, Ghasemigharagoz A, Todorov MI, Förstera B, Zhao S, Bhatia HS, Parra-Damas A, Mrowka L, Theodorou D, *et al.* (2019). Panoptic imaging of transparent mice reveals whole-body neuronal projections and skull-meninges connections. Nat Neurosci 22, 317–327.

Chen J, Ding L, Viana MP, Hendershott MC, Yang R, Mueller IA, Rafelski SM (2018). The Allen Cell Structure Segmenter: a new open source toolkit for segmenting 3D intracellular structures in fluorescence microscopy images. bioRxiv, 491035.

Chen BC, Legant WR, Wang K, Shao L, Milkie DE, Davidson MW, Janetopoulos C, Wu XS, Hammer JA, Liu Z, *et al.* (2014). Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution. Science 346, 1257998.

Cusanovich DA, Hill AJ, Aghamirzaie D, Daza RM, Pliner HA, Berletch JB, Filippova GN, Huang X, Christiansen L, DeWitt WS, *et al.* (2018). A single-cell atlas of in vivo mammalian chromatin accessibility. Cell 174, 1309–1324. e1318.

Dyson FJ (2012). History of science. Is science mostly driven by ideas or by tools? Science 338, 1426–1427.

Fereidouni F, Mitra AD, Demos S, Levenson R (2015). Microscopy with UV Surface Excitation (MUSE) for slide-free histology and pathology imaging. In Optical Biopsy XIII: Toward Real-Time Spectroscopic Imaging and Diagnosis, International Society for Optics and Photonics, Vol. 9318, p. 93180F.

Ferreira PG, Muñoz-Aguirre M, Reverter F, Sá Godinho CP, Sousa A, Amadoz A, Sodaei R, Hidalgo MR, Pervouchine D, Carbonell-Caballero J, *et al.* (2018). The effects of death and post-mortem cold ischemia on human tissue transcriptomes. Nat Commun 9, 490.

Frieda KL, Linton JM, Hormoz S, Choi J, Chow KH, Singer ZS, Budde MW, Elowitz MB, Cai L (2017). Synthetic recording and in situ readout of lineage information in single cells. Nature 541, 107–111.

Gao R, Asano SM, Upadhyayula S, Pisarev I, Milkie DE, Liu TL, Singh V, Graves A, Huynh GH, Zhao Y, *et al.* (2019). Cortical column and whole-brain imaging with molecular contrast and nanoscale resolution. Science 363, eaau8302.

Goltsev Y, Samusik N, Kennedy-Darling J, Bhate S, Hale M, Vazquez G, Black S, Nolan GP (2018). Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. Cell 174, 968–981. e915.

Gut G, Herrmann MD, Pelkmans L (2018). Multiplexed protein maps link subcellular organization to cellular states. Science 361, eaar7042.

Haghverdi L, Buttner M, Wolf FA, Buettner F, Theis FJ (2016). Diffusion pseudotime robustly reconstructs lineage branching. Nat Methods 13, 845–848.

Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, Saadatpour A, Zhou Z, Chen H, Ye F, *et al.* (2018). Mapping the mouse cell atlas by Microwell-seq. Cell 172, 1091–1107. e1017.

Herzog VA, Reichholf B, Neumann T, Rescheneder P, Bhat P, Burkard TR, Wlotzka W, von Haeseler A, Zuber J, Ameres SL (2017). Thiol-linked alkylation of RNA to assess expression dynamics. Nat Methods 14, 1198–1204.

Johnson HE, Goyal Y, Pannucci NL, Schupbach T, Shvartsman SY, Toettcher JE (2017). The spatiotemporal limits of developmental Erk signaling. Dev Cell 40, 185–192.

Keller PJ, Schmidt AD, Wittbrodt J, Stelzer EH (2008). Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. Science 322, 1065–1069.

Kretzschmar K, Watt FM (2012). Lineage tracing. Cell 148, 33–45.

La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lönnerberg P, Furlan A, *et al.* (2018). RNA velocity of single cells. Nature 560, 494–498.

Liu TL, Upadhyayula S, Milkie DE, Singh V, Wang K, Swinburne IA, Mosaliganti KR, Collins ZM, Hiscock TW, Shea J, *et al.* (2018). Observing the cell in its native state: imaging subcellular dynamics in multicellular organisms. Science 360, eaaq1392.

McDole K, Guignard L, Amat F, Berger A, Malandain G, Royer LA, Turaga SC, Branson K, Keller PJ (2018). *In toto* imaging and reconstruction of post-implantation mouse development at the single-cell level. Cell 175, 859–876. e833.

Moffitt JR, Zhuang X (2016). RNA imaging with multiplexed error-robust fluorescence *in situ* hybridization (MERFISH). Methods Enzymol 572, 1–49.

Nandagopal N, Santat LA, LeBon L, Sprinzak D, Bronner ME, Elowitz MB (2018). Dynamic ligand discrimination in the Notch signaling pathway. Cell 172, 869–880. e819.

Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat Methods 14, 979–982.

Raj B, Wagner DE, McKenna A, Pandey S, Klein AM, Shendure J, Gagnon JA, Schier AF (2018). Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. Nat Biotechnol 36, 442–450.

Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, Bodenmiller B, Campbell P, Carninci P, Clatworthy M, *et al.* (2017). Science forum: the human cell atlas. Elife 6, e27041.

Shah S, Takei Y, Zhou W, Lubeck E, Yun J, Eng CH, Koulena N, Cronin C, Karp C, Liaw EJ, *et al.* (2018). Dynamics and spatial genomics of the nascent transcriptome by Intron seqFISH. Cell 174, 363–376. e316.

Skardal A, Shupe T, Atala A (2016). Organoid-on-a-chip and body-on-a-chip systems for drug screening and disease modeling. Drug Discov Today 21, 1399–1411.

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM III, Hao Y, Stoeckius M, Smibert P, Satija R (2019). Comprehensive integration of single cell data. Cell 177, 1888–1902.e21.

Sulston JE, Horvitz HR (1977). Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans.* Dev Biol 56, 110–156.

Takei Y, Shah S, Harvey S, Qi LS, Cai L (2017). Multiplexed dynamic imaging of genomic loci by combined CRISPR imaging and DNA sequential FISH. Biophys J 112, 1773–1776.

Tomer R, Khairy K, Amat F, Keller PJ (2012). Quantitative high-speed imaging of entire developing embryos with simultaneous multiview light-sheet microscopy. Nat Methods 9, 755–763.

Toyama BH, Savas JN, Park SK, Harris MS, Ingolia NT, Yates JR 3rd, Hetzer MW (2013). Identification of long-lived proteins reveals exceptional stability of essential cellular structures. Cell 154, 971–982.

Waddington CH (1957). The Strategy of Genes, London: George Allen & Unwin, Ltd.

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018.

Yeh K, Kenkel S, Liu JN, Bhargava R (2015). Fast infrared chemical imaging with a quantum cascade laser. Anal Chem 87, 485–493.