Spatial and temporal tools for building a human cell atlas

Jonah Cool*, Richard S. Conroyb,c, Sean E. Hanlon*, Shannon K. Hughesd, and Ananda L. Royb

*Chan Zuckerberg Initiative, Redwood City, CA 94063; bOffice of Strategic Coordination, Division of Program Coordination, Planning, and Strategic Initiatives, National Institutes of Health, Bethesda, MD 20852; cCenter for Strategic Scientific Initiatives, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892; dDivision 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 spatio-temporal 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.

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; HIMC, International Human Microbiome Consortium; ImmGen, Immunological Genome Project; IMPC, International Mouse Phenotyping Consortium; LINCS, Library of Integrated Network-Based Cellular Signatures Project; NHP, nonhuman primate; NIH, National Institutes of Health; UV, ultraviolet; ZEBRA, Zebra Finch Expression Brain Atlas.

© 2019 Cool et al. This article is distributed by The American Society for Cell Biology under license from the author(s). Two months after publication it is available to the public under an Attribution-Noncommercial-Share Alike 3.0 Unported Creative Commons License (http://creativecommons.org/licenses/by-nc-sa/3.0). ASCB®, "The American Society for Cell Biology®," and "Molecular Biology of the Cell®" are registered trademarks of The American Society for Cell Biology.
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 (Herzig 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 over-constrained 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.
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 (IHMCI) 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.
