

Opinion

Towards Building a Plant Cell Atlas

Seung Y. Rhee,^{1,*} Kenneth D. Birnbaum,² and David W. Ehrhardt^{1,*}

Enormous societal challenges, such as feeding and providing energy for a growing population in a dramatically changing climate, necessitate technological advances in plant science. Plant cells are fundamental organizational units that mediate the production, transport, and storage of our primary food sources, and they sequester a significant proportion of the world's carbon. New technologies allow comprehensive descriptions of cells that could accelerate research across fields of plant science. Complementary to the efforts towards understanding the cellular diversity in human brain and immune systems, a Plant Cell Atlas (PCA) that maps molecular machineries to cellular and subcellular domains, follows their dynamic movements, and describes their interactions would accelerate discovery in plant science and help to solve imminent societal problems.

Plant Science Is Needed to Address Global Challenges

Climate change is the single biggest threat to global human health, and is expected to cause about a quarter of a million deaths globally from malnutrition, infectious diseases, and heat stress (www.who.int/news-room/fact-sheets/detail/climate-change-and-health). We need innovations in science and technology to tackle global sustainability, food security, and human health. Plants not only form the foundation of terrestrial ecosystems and the human food supply but also provide renewable energy and essential medicines. Plant research is therefore fundamental to meeting these challenges [1]. Recent innovations in systems biology, sensors and biosensors, data science, artificial intelligence, gene editing, precision breeding, and the microbiome position us to make unprecedented discoveries in plant science to enable new solutions.

Information Gaps in Plant Cell Science and the Need for a Plant Cell Atlas

Cells were first described in slices of cork by Robert Hooke in 1665 [2]. Nearly 200 years later, Schleiden's and Schwann's investigations of plant and animal microanatomy [3,4] led to the theory that these cells were in fact the fundamental organizational units of life [5]. A comprehensive understanding of plant cell structure and function at a molecular level is essential to uncover the mechanisms that plants use to produce the services we depend upon. Advances in genetics, molecular biology, biochemistry, and microscopy have produced detailed pictures of important cellular components, pathways, and mechanisms in higher plants, but many structural and functional features of plant cells remain to be understood at a molecular level, and it is highly likely that important features and compartments remain to be discovered and elucidated. For example, plants synthesize a variety of natural products that are consumed for sustenance and medicinal purposes, but the cellular organization of many pathways is not understood and we do not know whether enzymes in a metabolic pathway are generally physically linked to provide a metabolic tunnel for efficient channeling of metabolites. Proteins have been observed to cluster into speckles in the nucleus and in microdomains in the cell membrane, but we do not know the composition, diversity, or indeed function of most of these subcellular structures. Membrane trafficking pathways in plant cells are highly elaborated, but the diversity of trafficking compartments and their functional organization are poorly understood. Plant cells are

Highlights

Plant cells are the fundamental organizational unit that mediate the production, transport, and storage of our primary food sources and sequester a significant proportion of the world's carbon.

A comprehensive and quantitative understanding of how plant cells are organized and function is crucial in advancing plant science.

Recent advances in microscopy, mass spectrometry, artificial intelligence, sequencing, and nanotechnology make this the right time for the creation of a PCA initiative.

¹Carnegie Institution for Science, Department of Plant Biology, Stanford, CA 94305, USA

²New York University, Department of Biology, New York, NY 10003, USA

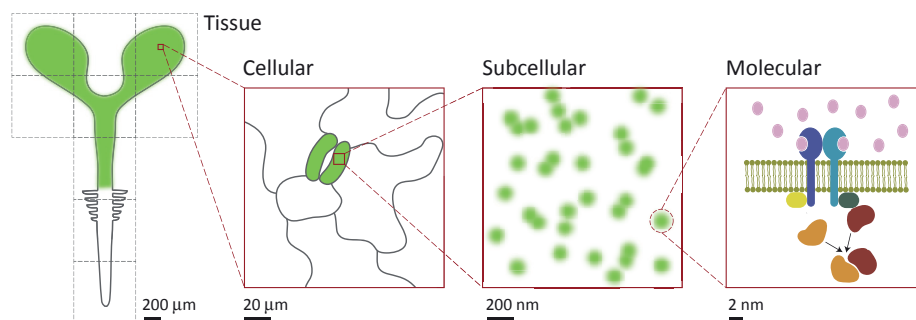
*Correspondence: srhee@carnegiescience.edu (S.Y. Rhee) and dehrhardt@carnegiescience.edu (D.W. Ehrhardt).

connected by cell–cell channels (plasmodesmata), but little is understood of their molecular structure, biogenesis, and regulation. These are only a few examples of broad unanswered questions that could impact many fields of basic and applied plant research. Finally, the functions of substantial proportions of genomes are unknown by either experimentation or prediction based on sequence similarity [6]. Knowledge about where these proteins reside in the cell, the other proteins they interact with, and how they behave as the environment changes will yield important clues about their functions and the mechanisms and pathways they are part of, including mechanisms and pathways not previously discovered.

Plant Cell Atlas – Goals and Potential Impact

We propose here the building of a Plant Cell Atlas (PCA), with the goal of creating a community resource that comprehensively describes the state of the various cell types found in plants and that incorporates information on nucleic acids, proteins, and metabolites at increasingly higher resolutions (Figure 1). At its core, a PCA will map cellular and subcellular protein localization patterns, track the dynamics and various interactions between proteins, identify the molecular components of different cellular substructures, discern complete states and transitions of specialized cell types, and integrate these different types of data to generate testable models of cellular function [7]. What the integrated data should look like is a big open question (see Outstanding Questions). The dynamic, spatiotemporal information contained in the Human Mitotic Cell Atlas [8] is one potential model. A related possibility is a virtual plant where users can zoom in and out from an organ down to the micro- and nanoscales where different objects are visible at each zoom level, as in Google Maps (Figure 1). Features that appear at any given level could be linked to genome databases where detailed, quantitative information can be obtained. Another possibility is to develop functional models of plant cells that enable simulations to be made at the level of signaling, growth, differentiation, and metabolism. The PCA initiative will need to leverage emerging fields, such as data science, proteomics, single-cell profiling, imaging, and **nanotechnology** (see Glossary), together with data visualization innovations, to create a high-resolution, molecular, temporal, and spatial map of the plant cell.

A PCA will facilitate the elucidation of basic, longstanding questions in biology (see Outstanding Questions) that relate directly to grand challenges in food security and climate change. For example, how do we get the maximal output of specialized plant cells that produce useful products (e.g., nutrients, oils, natural products) while reducing environmental impact? In addition, a Cell Atlas can provide insights into how viral or bacterial pathogens change cell properties, addressing crucial issues such as how to minimize herbivory and pathogens in stressful environments? Efforts to



Trends in Plant Science

Figure 1. Various Scales of the Plant Cell Components That Need to Be Mapped and Integrated in the Plant Cell Atlas.

Glossary

Active learning: an ML framework where the learner algorithm prioritizes data that will be most beneficial for training, typically based on the degree of uncertainty, and that iteratively improves itself from experimental or manual annotation of the unlabeled data.

Deep learning: a type of ML approach which uses neural networks and relies on large amounts of data to predict patterns and classify items.

Machine learning (ML): a type of artificial intelligence approach in computer science that uses statistics and information theory to predict entities and relationships from examples.

Nanotechnology: science, engineering, and technology conducted at scales of less than 100 nm.

Semantic zooming: visualizing qualitative changes of information content as a function of scale.

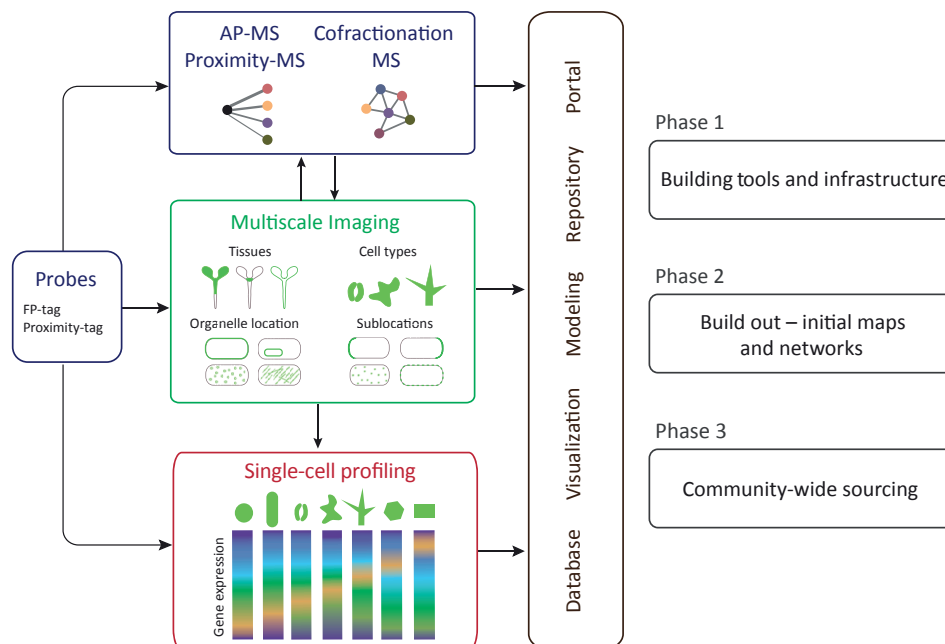
understand cellular specialization in C4 plants can, for example, address issues of crop efficiency and help to feed a growing population while reducing the carbon footprint of agriculture. Moreover, basic questions about the totipotency of plant cells relate directly to long-term growth in plants and the ability to take advantage of gene-editing techniques by efficient regeneration of a variety of crops.

Technological Components of the PCA

To build a PCA, several methodologies will need to be employed and integrated. They include large-scale gene tagging and transformation to reveal the cellular and subcellular localization of proteins, imaging technologies to understand the dynamic behaviors of those proteins, large-scale proteomics to understand the interactions among proteins and the cellular machines they comprise, single-cell profiling to assess the variety and range of cell types, cell states, and their transitions, and finally analytic and visualization tools to integrate and model these data to generate testable models of cellular function (Figure 2). We envisage the creation of the PCA as proceeding in stages, starting with an initial tool-building and proof-of-concept stage, a build-out phase where these tools are deployed to build up initial maps and networks of information, and then a phase where multiple projects across the community contribute to and build the Atlas, using infrastructure created and expanded in phases one and two. In the sections below we highlight some recent advances as well as limitations in key technologies necessary to establish PCA as a community.

Gene Tagging, Transformation, and Expression

Visualizing the *in vivo* localization patterns of proteins requires protein tagging with markers, most commonly fluorescent tags [9]. For a PCA project, tags need to be sufficiently bright to allow detection of small protein complexes, monomerized to avoid artifacts by self-association, and not easily attenuated by protonation of the fluorophore. The latter is important because



Trends in Plant Science

Figure 2. Core Activities, Relationships, and Phases for Building a Plant Cell Atlas (PCA) Community. The diagram focuses on activities related to proteins, but the broader PCA initiative will include other molecules including RNAs,

many plant organelles are acidic. Several fluorescent proteins have been engineered to meet these criteria, including mCitrine [10]. The choice of where the fusion is made and the nature of the peptide link between the tag and target can also be crucial. A reasonable proposal [11] is to employ a flexible peptide linker for all fusions and to use sequence analysis to assess the likelihood N- or C-terminal targeting to direct the choice of where to tag. Overexpression can mask many localization patterns and produce significant artifacts if the machineries that process and localize proteins become saturated. Therefore, stable transformations using native expression sequences are greatly preferred. To preserve native expression, *in situ* tagging would be ideal, such as has been accomplished using CRISPR in animal cells [12], but efficient *in situ* tagging remains to be demonstrated in higher plants. Although transient expression can be compromised by overexpression, it can also allow much greater throughput. A controllable promoter to titrate expression, such as a steroid-inducible promoter, could mitigate the overexpression problem. Exciting new methods have been developed using nanoparticles to aid cellular transformation, including electrospray [13], magnetofection [14], and nanostraw-mediated electroporation [15]. If adapted to plant tissue, these techniques may allow large-scale transient transformation of probes into differentiated plant tissue, and even extend studies to a wide range of species with specialized cell types.

Imaging

In vivo and *in planta* observation of fluorescently tagged proteins requires sensitive detection, rapid acquisition, and optical sectioning. Instruments such as spinning disk confocal microscopes that use highly efficient detectors [e.g. electron-multiplying charge-coupled device (EMCCD) and scientific complementary metal-oxide-semiconductor (sCMOS)] and multipoint scanning currently meet these needs. The ability to mitigate the light-scattering effects of highly refractile cells walls, especially for analysis of cells below the epidermis, is also important for imaging plant tissues. Using water, glycerol, and the recently developed silicon oil objectives can improve results when imaging cells through thick plant tissue. Many new and advanced imaging modalities, including super-resolution microscopies [9] to resolve optical probes below the diffraction limit, will enable the visualization of novel complexes and interactions. Although super-resolution microscopies reveal new biological detail using light imaging, a revolution in cryo-electron microscopy (CryoEM) techniques, including correlated fluorescence-CryoEM, add molecular information to 3D EM imaging, bringing visualization of molecular organization in cells to the nanoscale [9,16]. Imaging can also explore protein-protein interactions and dynamic protein relationships in living cells using advanced techniques, including cross-correlation spectroscopy (CCS) [17], scanning fluorescence correlation spectroscopy (FCS) [18], Forster resonance energy transfer (FRET) [19], bioluminescence energy transfer (BRET) [20], and bimolecular fluorescence complementation (BiFC) [21]. All these data will be valuable in helping to identify potential protein partners, complexes, and networks to build a comprehensive PCA.

Proteomics

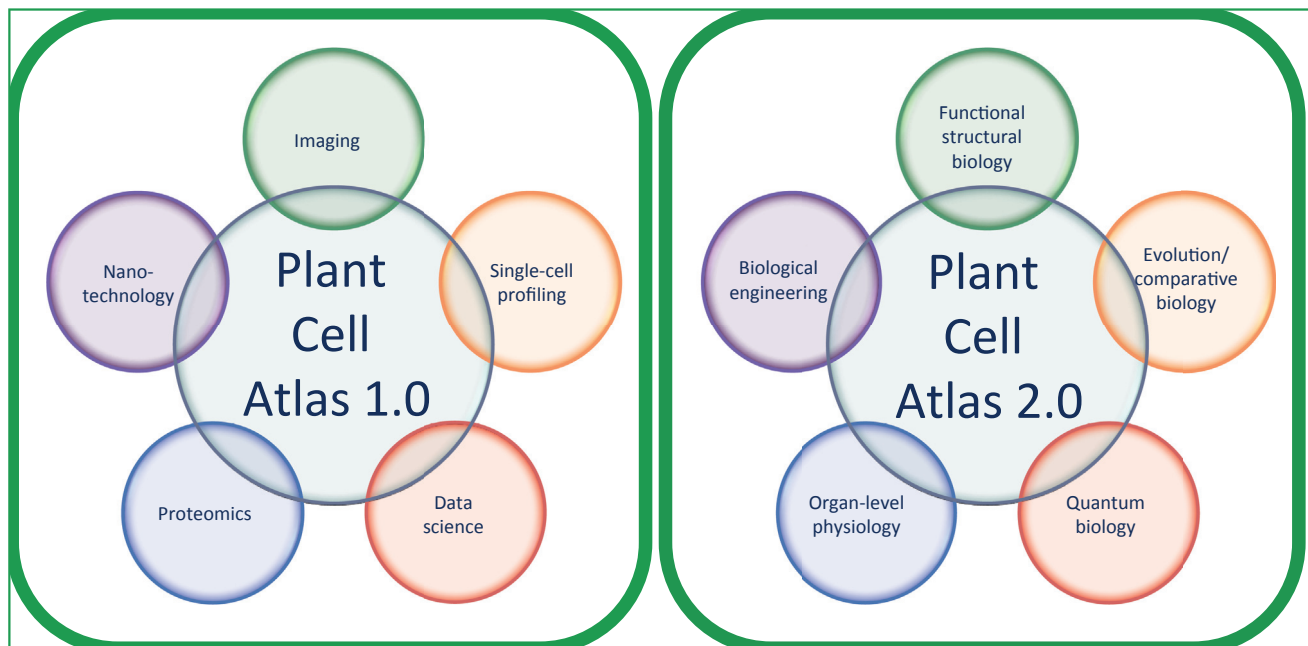
Technical innovations in proteomics, especially large-scale methods for affinity purification MS (AP-MS) [22,23], cofractional MS (CF-MS) [24,25], and two-hybrid screening [26,27] have allowed the collection of protein-protein data on an unprecedented scale. These studies have revealed new components of known protein complexes and, excitingly, entirely new complexes whose functions remain to be determined [28]. Traditionally, large-scale protein interaction assays required high protein-protein affinity to identify interaction partners. However, many functionally important protein interactions are not of high affinity, such as kinase-substrate interactions. A family of methods, known as proximity labeling, utilize an enzyme to label nearby proteins with a high-affinity molecular tag. This technique enables detection of weak protein interactions and even positional proximity, allowing the possibility to explore spatial and

functional relationships among proteins that were inaccessible by previous methods [29,30]. New possibilities for determining organellar proteomes have opened up by targeting the labeling enzyme itself to the organelle [31]. If combined with cell type-specific expression, this technique could also enable determination of organelle proteomes by cell type or even cell state. Building maps of protein localization together with interaction data will be strongly synergistic and mutually reinforcing.

RNA Sequencing

Single-cell RNA-seq profiling and related analyses that interrogate chromatin, DNA, and even protein levels present new opportunities for defining complex cellular states. These techniques have evolved rapidly from sequencing hundreds of cells [32–34] to combinatorial barcoding techniques that permit processing tens of thousands of cells in one experiment [35,36]. The dramatic increase in sampling scale can be used to overcome increased noise [37], enhancing the ability to fine-map cellular states and define the transcriptional differences among cellular subtypes. In addition, the ability to conduct sequence-enabled single-cell molecular analyses provides new opportunities to address plant-specific questions, such as what cell-specific regulatory responses mediate the ability of the plant cell to alter their state during development and in response to external cues [38,39]. One key advantage of single-cell analyses at the genome scale is the ability to discover previously undescribed heterogeneity in cellular states that allow a description of the early regulatory responses in coping with stress or developmental transitions.

The indeterminate growth of plant meristems means that cells at all developmental stages are present in a single organ, providing a progression of cellular maturation in a single set of samples. This property is ideal for single-cell sequencing approaches that reconstruct



Trends in Plant Science

Figure 3. Fields Needed to Establish the Plant Cell Atlas Community.

developmental trajectories from single-cell assays. Models that take advantage of this fine-scale time series could be used to identify key steps in meristem organization that, for example, control the specification of cells that load nutrients or mediate growth to optimize yield. In crop physiology, the function of differentiated cells is often a crucial question. Along those lines, newer combinatorial barcoding techniques have demonstrated the feasibility of labeling mRNAs in single nuclei [36]. The latter technique offers some promise to assay cells that have been recalcitrant to cell-wall digestion, such as mature plant cells or those with specialized secondary cell walls.

Data Science

Advances in computing and communications technologies have opened an era of big data and artificial intelligence which is transforming all sectors of our societies and cultures [40,41]. The PCA should leverage and extend big data technologies such as parallel computing, data management infrastructure, data analytics, and **machine learning** (ML). The PCA initiative has various applications for ML, including detecting and delineating compartments, tracking movements, and extracting morphological characteristics. The rapid increase in throughput, scale, and dimensionality of image capture will enable us to chart out the plant cell at unprecedented detail but will require image-processing tools that are easy to use, adaptable, and able to take in multiple dimensions and parameters. ML is ideally suited for this type of task [42]. From the PCA initiative we will likely see many new cellular compartments and structures, which may not easily be defined by a few parameters. In addition, we may identify various states or novel subdomains of known compartments. Because there would be no prior knowledge for these novel structures and compartments, unsupervised ML methods need to be applied. These include simple clustering to categorize data, **active learning** [42], and neural network-based **deep learning** [40] to identify new structures and compartments [43].

Innovation in data visualization will be crucial in every phase of the project such as data acquisition, quantification, pattern recognition, and interpretation. Much progress has been made towards visualizing high-dimensional data in the past decade [44]. An innovative way of visualizing complex data at multiple scales is **semantic zooming**. Unlike geometric zooming that increases objects in size, semantic zooming changes the content of the object at different scales. Semantic zooming has been successfully implemented by Google Maps and, although some efforts have been made in biology [7,45,46], the full potential of this tool has yet to be applied. The PCA project would be an ideal vehicle for semantic zooming (Figure 1). To semantically connect objects at different scales, the entities need to be organized in a hierarchy. The cellular component domain of Gene Ontology [47] and Cell Ontology [48] would serve as an excellent starting point for mapping these entities.

Open-source initiatives such as ImageJ and FIJI [49] have greatly enabled useful sharing of image-analysis tools and have significantly increased access to image analysis. However, these tools for image processing and quantification are still largely manually driven using macros and plug-ins that are customized for particular datasets or experimental setups. Customization and automation can be performed by modifying the open-source code and integrating these tools with environments such as Matlab [50]. However, it is often difficult to find the right software or analysis functions: even if they exist, they can still be challenging to adapt and implement in other environments. Further, these solutions are typically not optimized and are inefficient, often limiting their application from large-scale datasets and computationally intense analyses. Therefore, there are big opportunities for software engineering to play a pivotal role in the PCA initiative.

Concluding Remarks and Future Perspectives

To successfully implement the PCA initiative, we need to build a community that brings together scientists from fields such as imaging, nanotechnology, single-cell profiling, data science, and proteomics (Figure 3). The community needs a forum, perhaps via a website and social media links, to connect various projects whose datasets could contribute to the PCA network. The community also needs to tackle topics such as data sharing and release policy, public and internal data tracking, and data visualization tools. Connecting with projects such as the Human Cell Atlas [51] and OpenWorm (www.openworm.org) to learn from their experiences would be important. An obvious first step is to convene a workshop to bring in key stakeholders and scientists to discuss the vision, approaches, and bottlenecks. How to engage the public in this initiative should also be an important part of the discussion.

Going beyond the initial phase, we envisage the next-generation PCA to encompass emerging technologies (Figure 3). For example, functional structural biology such as X-ray free electron lasers [52], large-volume imaging [53], and dynamic transmission EM [54] have the potential to transform the way we visualize the molecular events inside cells. In addition, the diversity of cellular structures and functions across the spectrum of species would require evolutionary and comparative biology bringing this field to the micro- and nano-scale. Exciting developments in biological engineering could make possible the engineering of artificial organelles or the use of biological machineries and structures as an inspiration for manufacturing synthetic systems [55]. Going beyond genes and proteins, metabolites should be incorporated into the PCA. Finally, the next-generation PCA would enable the probing of organ- and organism-level physiology at molecular resolution, enabled by modeling relationships and functions between cells, which would open doors for modeling plant behavior under various environmental scenarios (see Outstanding Questions).

Acknowledgments

We thank Heather Meyer, Hannah Vahldick, Benjamin Jin, and Suryatapa Jha for their comments on the manuscript. Work in the laboratory of S.Y.R. is supported by a Carnegie Institution for Science endowment and by grants from the National Science Foundation (IOS-1546838, IOS-1026003), the Department of Energy (DE-SC0008769, DE-SC0018277), and the National Institutes of Health (1U01GM110699-01A1). The laboratory of D.W.E. is supported by the Carnegie Institution for Science endowment and grants from the National Institutes of Health (RO1-GM123259-01), the National Science Foundation (NSF MCB-1413254), and the Department of Energy (DE-AR0000825).

References

- Jones, A.M. *et al.* (2008) The impact of *Arabidopsis* on human health: diversifying our portfolio. *Cell* 133, 939–943
- Hooke, R. (1665) *Micrographia, or, Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses: With Observations and Inquiries Thereupon*, Royal Society of London
- Schleiden, M.J. (1838) Contributions to our knowledge of phyto-genesis. *Arch. Anat. Physiol. Wiss. Med.* 13, 137–176
- Schwann, T. (1838) Ueber die analogie in der structur und dem wachstum der thiere und pflanzen. *Neue Not Geb Nat Heil Jan*, 33–36; Feb, 25–29; Apr, 21–23
- Schwann, T. (1839) *Mikroskopische Untersuchungen über die Uebereinstimmung in der Struktur und dem Wachstum der Thiere und Pflanzen*, Verlag der Sander'schen Buchhandlung
- Rhee, S.Y. and Mutwil, M. (2014) Towards revealing the functions of all genes in plants. *Trends Plant Sci.* 19, 212–221
- Waese, J. *et al.* (2017) ePlant: visualizing and exploring multiple levels of data for hypothesis generation in plant biology. *Plant Cell* 29, 1806–1821
- Cai, Y. *et al.* (2018) Experimental and computational framework for a dynamic protein atlas of human cell division. *Nature* 561, 411–415
- Shaw, S.L. and Ehrhardt, D.W. (2013) Smaller, faster, brighter: advances in optical imaging of living plant cells. *Annu. Rev. Plant Biol.* 64, 351–375
- Shaner, N.C. *et al.* (2005) A guide to choosing fluorescent proteins. *Nat. Methods* 2, 905–909
- Tian, G.W. *et al.* (2004) High-throughput fluorescent tagging of full-length *Arabidopsis* gene products in planta. *Plant Physiol.* 135, 25–38
- Roberts, B. *et al.* (2017) Systematic gene tagging using CRISPR/Cas9 in human stem cells to illuminate cell organization. *Mol. Biol. Cell* 28, 2854–2874
- Abayadeh, M. *et al.* (2017) Electrospayed chitosan nanoparticles: facile and efficient approach for bacterial transformation. *Int. Nano Lett.* 7, 291–295
- Zhao, X. *et al.* (2017) Pollen magnetofection for genetic modification with magnetic nanoparticles as gene carriers. *Nat. Plants* 3, 956–964
- Cao, Y. *et al.* (2018) Universal intracellular biomolecule delivery with precise dosage control. *Sci. Adv.* 4, eaat8131
- Zanetti, G. *et al.* (2013) The structure of the COPII transport-vesicle coat assembled on membranes. *eLife* 2, e00951

Outstanding Questions

How does a cell know when to start and stop dividing, growing, or differentiating? How does a cell sense the change in its physical and chemical environment? How does a new cell type arise over evolutionary scales?

How is long-term growth maintained in stem cells and the meristem?

What is the diversity of plant cell responses to pathogens and other stresses, and how does the diversity of response relate to the resistance of an individual cell?

What property enables specific plant cells to embark on regeneration?

What is the diversity and function of intracellular trafficking compartments?

How do plant cells coordinate their developmental maturation in the meristem?

How is the plant plasma membrane organized, both structurally and dynamically, at the nano- to the microscale?

How is the interface between the plasma membrane and the cell wall organized, and what functions does this interface mediate for development and defense?

How is biosynthesis and function of the cell wall organized and regulated?

How does cellular scale organization and function arise from nanoscale organization and function?

How are cell-cell channels regulated and functionally diversified?

How do plant cells communicate with each other over short and long distances?

How do plant cells establish and position functional domains at their periphery to regulate cytoskeletal organization, morphogenesis, polarized intercellular communication, and developmental patterning?

17. Schwille, P. *et al.* (1997) Dual-color fluorescence cross-correlation spectroscopy for multicomponent diffusional analysis in solution. *Biophys. J.* 72, 1878–1886
18. Petrasek, Z. *et al.* (2010) Scanning FCS for the characterization of protein dynamics in live cells. *Methods Enzymol.* 472, 317–343
19. Förster, T. (1948) Zwischenmolekulare energiewanderung und fluoreszenz. *Annalen der Physik* 437, 55–75
20. Xie, Q. *et al.* (2011) Bioluminescence resonance energy transfer (BRET) imaging in plant seedlings and mammalian cells. *Methods Mol. Biol.* 680, 3–28
21. Kerppola, T.K. (2008) Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. *Annu. Rev. Biophys.* 37, 465–487
22. Hein, M.Y. *et al.* (2015) A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* 163, 712–723
23. Huttlin, E.L. *et al.* (2015) The BioPlex network: a systematic exploration of the human interactome. *Cell* 162, 425–440
24. Kirkwood, K.J. *et al.* (2013) Characterization of native protein complexes and protein isoform variation using size-fractionation-based quantitative proteomics. *Mol. Cell. Proteomics* 12, 3851–3873
25. Havugimana, P.C. *et al.* (2012) A census of human soluble protein complexes. *Cell* 150, 1068–1081
26. Rolland, T. *et al.* (2014) A proteome-scale map of the human interactome network. *Cell* 159, 1212–1226
27. Rual, J.F. *et al.* (2005) Towards a proteome-scale map of the human protein–protein interaction network. *Nature* 437, 1173–1178
28. Drew, K. *et al.* (2017) Integration of over 9,000 mass spectrometry experiments builds a global map of human protein complexes. *Mol. Syst. Biol.* 13, 932
29. Han, S. *et al.* (2018) Proximity labeling: spatially resolved proteomic mapping for neurobiology. *Curr. Opin. Neurobiol.* 50, 17–23
30. Chen, C.L. and Perrimon, N. (2017) Proximity-dependent labeling methods for proteomic profiling in living cells. *Wiley Interdiscip. Rev. Dev. Biol.* 6, e272
31. Rees, J.S. *et al.* (2015) Protein neighbors and proximity proteomics. *Mol. Cell. Proteomics* 14, 2848–2856
32. Picelli, S. *et al.* (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* 10, 1096–1098
33. Hashimshony, T. *et al.* (2012) CEL-Seq: single-cell RNA-seq by multiplexed linear amplification. *Cell Rep.* 2, 666–673
34. Jaitin, D.A. *et al.* (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343, 776–779
35. Rosenberg, A.B. *et al.* (2018) Scaling single cell transcriptomics through split pool barcoding. *Science* 360, 176–182 <http://science.sciencemag.org/content/360/6385/176>
36. Cao, J. *et al.* (2017) Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* 357, 661–667
37. Ziegenhain, C. *et al.* (2017) Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* 65, 631–643 e634
38. Gifford, M.L. *et al.* (2008) Cell-specific nitrogen responses mediate developmental plasticity. *Proc. Natl. Acad. Sci. U. S. A.* 105, 803–808
39. Dinney, J.R. *et al.* (2008) Cell identity mediates the response of *Arabidopsis* roots to abiotic stress. *Science* 320, 942–945
40. Ching, T. *et al.* (2018) Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface* 20170387
41. O'Driscoll, A. *et al.* (2013) 'Big data', Hadoop and cloud computing in genomics. *J. Biomed. Inform.* 46, 774–781
42. Sommer, C. and Gerlich, D.W. (2013) Machine learning in cell biology – teaching computers to recognize phenotypes. *J. Cell Sci.* 126, 5529–5539
43. Christiansen, E.M. *et al.* (2018) *In silico* labeling: predicting fluorescent labels in unlabeled images. *Cell* 173, 792–803
44. Liu, S. *et al.* (2017) Visualizing high-dimensional data: advances in the past decade. *IEEE Trans. Vis. Comput. Graph.* 23, 1249–1268
45. Zhukova, A. and Sherman, D.J. (2015) Mimoza: web-based semantic zooming and navigation in metabolic networks. *BMC Syst. Biol.* 9, 10
46. Hu, Z. *et al.* (2007) Towards zoomable multidimensional maps of the cell. *Nat. Biotechnol.* 25, 547–554
47. The Gene Ontology, C (2017) Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* 45, D331–D338
48. Diehl, A.D. *et al.* (2016) The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. *J. Biomed. Semantics* 7, 44
49. Schindelin, J. *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682
50. Hiner, M.C. *et al.* (2017) ImageJ-MATLAB: a bidirectional framework for scientific image analysis interoperability. *Bioinformatics* 33, 629–630
51. Regev, A. *et al.* (2017) The Human Cell Atlas. *eLife* 6, e27041
52. Johansson, L.C. *et al.* (2017) A bright future for serial femtosecond crystallography with XFELs. *Trends Biochem. Sci.* 42, 749–762
53. Ji, N. *et al.* (2016) Technologies for imaging neural activity in large volumes. *Nat. Neurosci.* 19, 1154–1164
54. Campbell, G.H. *et al.* (2014) Time resolved electron microscopy for in situ experiments. *Appl. Phys. Rev.* 1, 041101
55. Wintle, B.C. *et al.* (2017) A transatlantic perspective on 20 emerging issues in biological engineering. *eLife* 6, e30247

What are the best mechanisms for funding the PCA initiative?

What are the best ways to engage the general public with the PCA initiative?

Which organism(s) and cell type(s) are most appropriate to start the PCA initiative with?

How can position information be mapped to cell-type information from single-cell profiling analyses?

What is the best way to visualize and access the integrated data?