BERKELEY LAB
Bringing Science Solutions to the World

BioSciences

PLANT SINGLE-CELL SOLUTIONS FOR ENERGY AND THE ENVIRONMENT

WORKSHOP REPORT

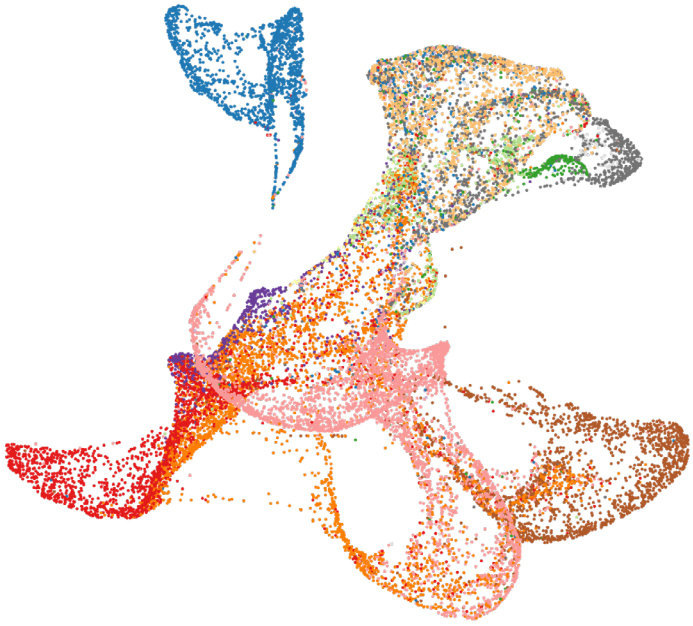
Lawrence Berkeley National Laboratory | January 23, 2020

Organizing Committee: Diane Dickel (chair), Ben Cole, Rex Malmstrom, Jenny Mortimer, Chris Mungall, Ronan O'Malley, Axel Visel

PLANT SINGLE-CELL SOLUTIONS FOR ENERGY AND THE ENVIRONMENT

WORKSHOP REPORT

Lawrence Berkeley National Laboratory | January 23, 2020



Organizing Committee: Diane Dickel, Ben Cole, Rex Malmstrom, Jenny Mortimer, Chris Mungall, Ronan O'Malley, Axel Visel

BERKELEY LAB
Bringing Science Solutions to the World

BioSciences

A U.S. Department of Energy National Laboratory. Managed by the University of California.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Plants are Earth's primary producers of biomass and are increasingly harnessed as sustainable sources of energy and materials. However, to fully realize their potential for bioenergy and biomaterial production, a fundamentally improved understanding of plant responses to different environmental conditions and agricultural practices, as well as a deep functional characterization of both cultivated feedstocks and other ecologically important species is urgently needed. Like all multicellular eukaryotes, plants are composed of multiple specialized cell types that form their tissues and organs, with each cell type having a distinct structure, composition, and function. Illuminating the characteristics of these cell types has the potential to dramatically improve our foundational knowledge of plant biology, along with our ability to engineer plants, fungi, and algae to express desirable traits, optimally grow under specific conditions, or plentifully produce biomaterials. Over the last decade, progress in sequencing and microfluidics technologies, coupled with powerful new computational analysis strategies, has revolutionized the granularity at which multicellular organisms can be studied. In particular, single-cell transcriptomic methods have led to fundamental new insights into animal biology and development, such as the discovery of new cell types and cell type-specific processes in disease. In contrast, the application of single-cell approaches to plants and other organisms important for energy and the environment, including fungi and algae, has been far more limited, largely due to the challenges posed by the presence of polysaccharide walls surrounding the cells of these species.

Lawrence Berkeley National Laboratory hosted a workshop on January 23, 2020 for a diverse group of leaders in functional genomics technologies. In this forum, attendees from academia, the National Laboratories, and local research institutions discussed opportunities afforded by single-cell technologies for energy and environmental science, as well as conceptual and technological grand challenges that must be tackled to apply these powerful approaches to plants, fungi and algae. The workshop included presentations from ten speakers and breakout sessions to explore applications of single-cell technology, including:

1. Powering higher-resolution studies of plant responses to environmental stimuli
2. Improving the functional annotation of genes across tissues and species
3. Optimizing bioproduct and biomaterial production

These discussions highlighted the pressing need to develop better and more comprehensive single-cell technologies, data analysis and visualization tools, and tissue preparation methods for plants, fungi, algae, and bacteria. Attendees expressed overwhelming support for the creation of a centralized, open-access database to house plant single-cell data, analogous to the Human Cell Atlas, and considered how such an effort should balance the need for deep characterization of a few important model species while still capturing the broader diversity in the plant kingdom.

While there was lively discussion and debate about how to maximize the technological and scientific impact of these endeavors, there was overwhelming consensus across all workshop participants that the time is ripe for a bold and concerted effort to harness single-cell methods for the study of environmental and bioenergy-relevant questions. Targeted investments into the development of methods and their application to relevant species, as well as the creation of data resources to support the dissemination of single-cell data across research communities, will rapidly enable groundbreaking insights to propel energy and environmental science forward.

INTRODUCTION

Biomass derived from the growth and harvest of plant feedstocks is a renewable and sustainable resource for the production of energy and materials. The global energy supply increasingly relies on robust and scalable bioenergy resources, which contributes to both energy security and the sustainability of energy production. Likewise, biomaterials derived from plants, algae, and microorganisms are growing in importance for a breadth of applications. Currently available plant feedstocks require substantial amounts of land, water, and mineral resources, and their associated agricultural practices have considerable environmental impacts. To develop a more sustainable bioenergy and biomaterials portfolio for the future, we must significantly advance our understanding of how feedstock crops can be improved to tolerate and thrive in a continuously changing environment.

Critical to this understanding is knowing how the genome of a plant or other environmental organism (e.g., plant-associated bacteria or fungi) contributes to productivity, as this will empower breeding and bioengineering programs to enhance bioenergy and biomaterial production. The genomics era has significantly contributed to this cause by inspiring major investments into exploring the mechanisms underlying complex biological processes, principally through the use of global profiling strategies to measure RNA (transcriptomics), protein (proteomics), or metabolite (metabolomics) levels in plants. For example, recent work has leveraged multiple global profiling tools, such as genome-wide association surveys, transcriptomics, and proteomics to better understand how sorghum, an important bioenergy crop, responds to drought[1,2]. This work uncovered new hypotheses regarding how photosynthesis and the soil environment influence drought tolerance. While these powerful methods have already revealed major insights into the biology of bioenergy feedstock plants, they have been limited to surveys of whole organisms or complex tissues. Plant tissues consist of numerous distinct cell types, each with a specialized function within the context of that tissue or organism. Thus, each cell type will likely exhibit different molecular behaviors in response to an environmental challenge or produce a unique combination of metabolites or other products[3]. However, signals associated with specific cell types are averaged with, and thus diluted by, all of the other cell types present in the sample when profiling whole tissues using conventional bulk methods. Therefore, there is a need to develop molecular profiling methods that can evaluate individual cells or cell types for a more accurate understanding of how plant feedstocks can maintain productivity under environmental stress or to design more rational plant engineering strategies for the sustainable generation of bioproducts. Recently, there has been an explosion in methods that profile global biomolecule expression patterns in individual cells derived from complex tissues, which has revolutionized the way we can study and think about biological organization[4]. A major goal of these single-cell characterization methods is to divide cells from tissues into discrete classes (cell types or states), identify a unique transcriptional profile for each cell type, associate these with specific cell type functions, and define how cell types relate to one another functionally or developmentally (i.e., early, versus late developmental stages). Once the full complement of cell types is well defined, computational methods can be employed to address a wide range of questions, such as: what each cell type produces and how cell types respond to a variety of perturbations (e.g., environmental conditions or genetic mutation). These types of analyses have great potential to yield a more complete understanding of the function of cell populations, their adaptive and plastic properties, and sophisticated molecular toolboxes for biotechnological engineering, such as regulatory sequences that can activate a gene or pathway within a particular cell type after exposure to a specific stimulus. Thus, single-cell characterization technologies comprise a powerful new suite of methods to study biological heterogeneity and promise to deliver a much deeper understanding of how organisms function as a unified collection of cell types.

The most widely used of these new technologies, single-cell RNA sequencing (scRNA-seq), works by using microfluidics and barcoded DNA particles to capture whole transcriptomes of single cells. Cutting-edge scRNA-seq methods can capture expression for tens of thousands of cells in a single experiment[5]. While microfluidics-based methods have the power to profile cell populations en masse, any spatial information (how those cells were organized within the larger tissue) is lost because the tissues must be first dissociated into individual cells for profiling. Newer sequence-based imaging methods (e.g., Slide-seq[6], merFISH[7], FISSEQ[8], Nanostring[9]) hold great promise to impart spatial information to single-cell transcriptomic data. Some of these methods[6,10] work by arraying barcoded particles along a 2-dimensional surface, then exposing this array to a thin tissue section to capture spatially resolved transcriptomes of individual cells, or even subcellular compartments. Others use fluorescently labeled oligonucleotide mixtures that can be manipulated to report the position of hundreds to thousands of transcripts in a single specimen. These methods have been applied to a rapidly growing number of animal tissues, genotypes, and species to build extremely high resolution profiles of gene expression. They have also been used to uncover novel cell types, infer gene regulatory networks, and to understand how developmental processes unfold within highly heterogeneous biological specimens. However, they have not yet been widely applied to plant, fungal, or algal species.

Applications of single-cell methods in mammalian systems: As presented in inspiring talks by Joe Ecker and Yirong Peng, one of the most widely-used applications of scRNA-seq is in profiling cell populations of mammalian neural tissue. In addition to scRNA-seq, single-nucleus methylation profiling (snmC-seq), and single-nucleus open chromatin profiling (snATAC-seq) have been employed to elucidate gene regulatory architecture. As presented by Dr. Ecker, these three technologies are highly complementary, and have been used to identify dozens of distinct brain cell types, in addition to cell type-specific expression for many genes[11–15]. Similar studies that Dr. Peng described have probed single-cell expression patterns in mammalian retinal tissue, and these methods have taken advantage of antibody-based depletion approaches to mitigate population size imbalances between cell types and to enrich for rarer types. This has enabled not only the massive expansion of the number of inferred cell types, but also the characterization of conserved and distinct populations of retinal cells between species (e.g., mouse and chimpanzee). Computational methods for interspecies comparisons developed for this work could be harnessed for comparisons of different plant species either for evolutionary comparisons or to extend functional annotations to less well studied species. In addition to single-cell transcriptomics and epigenomics, single-cell proteomics is also making advances, though this technology is still nascent[16,17].

Recent progress in plant single-cell transcriptomics: While single-cell transcriptomics is now routinely and widely used in animal research programs, it has yet to be firmly rooted in plant or fungal research communities, which limits the ability to leverage this powerful set of tools to address current bioenergy and environmental challenges. Very recently, a number of groups independently addressed this technological gap by performing the first set of scRNA-seq studies on Arabidopsis root cells[18–22]. These studies identified nearly all major expected cell types, and many identified subclasses of cell types that were not previously well defined. Furthermore, these studies were useful in **1)** characterizing complex signaling networks important for root development; **2)** identifying a biphasic switch essential for xylem cell development, as presented by Siobhan Brady[23]; and **3)** detailing developmental progression of the endodermis[18] and hair cells[20] of the Arabidopsis root. As presented by Benjamin Cole, these datasets are being integrated together to form a comprehensive map of plant roots at an unprecedented level of detail (**Figure 1**). Apart from plant roots, there is growing interest in using scRNA-seq technologies to profile the development of other important plant tissues, including leaf, flower, and seed endosperm. Stomata biogenesis and development within the leaf is a particularly intriguing topic, and is the subject of on-going work presented at the workshop by Dominique Bergmann.

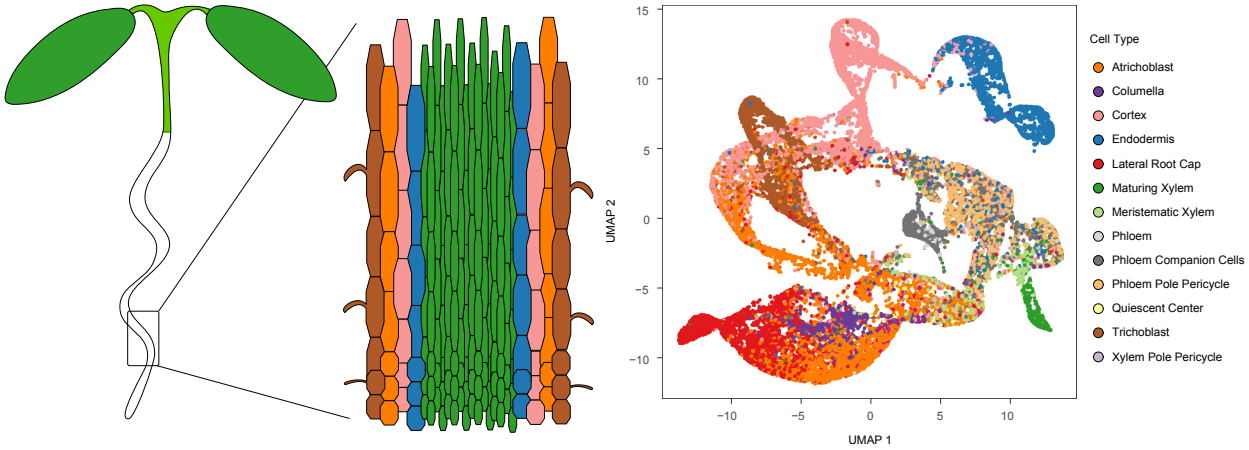


Figure 1: scRNA-seq of Arabidopsis root. Root development has recently been extensively characterized at the single-cell level in a series of scRNA-seq studies of root cell protoplasts. These methods can confidently identify all major cell types within roots, and can begin to shed light on developmental trajectories that underlie root growth. Figure courtesy of Benjamin Cole.

Future goals: Building upon these initial profiling efforts to understand previously uncharacterized cell types, tissues, and species will prove critical in the near future if we want a better understanding of how individual cells behave in stressful environmental conditions, how plants interact with their microbiota (a point raised by John Vogel during his workshop presentation), or how to better engineer plants or fungi for efficient and sustainable bioproduct synthesis, a need described by Blake Simmons. In addition, it is yet unclear how single-cell profiling will be applied to non-model plant or fungal species, but such an expansion would have enormous benefits for biotechnology applications. During his presentation, David Goodstein made the case for single-cell profiling as a way to better annotate gene function in non-model plant species. There is a rising need to move beyond transcriptomics and epigenomics to describe states of individual cells, with a particular emphasis from Seung Yon Rhee on elucidating metabolic pathways. Along these lines, Anup Singh presented nascent, though highly encouraging, single-cell metabolomics, proteomics, and imaging technologies that may be powerful in helping to address this unmet need (**Figure 2**). There has also been recent progress in developing mass spectrometry-based metabolite imaging for spatially profiling metabolite quantities in plants[24]. Further development of this technology will nicely complement advances in other single-cell methods.

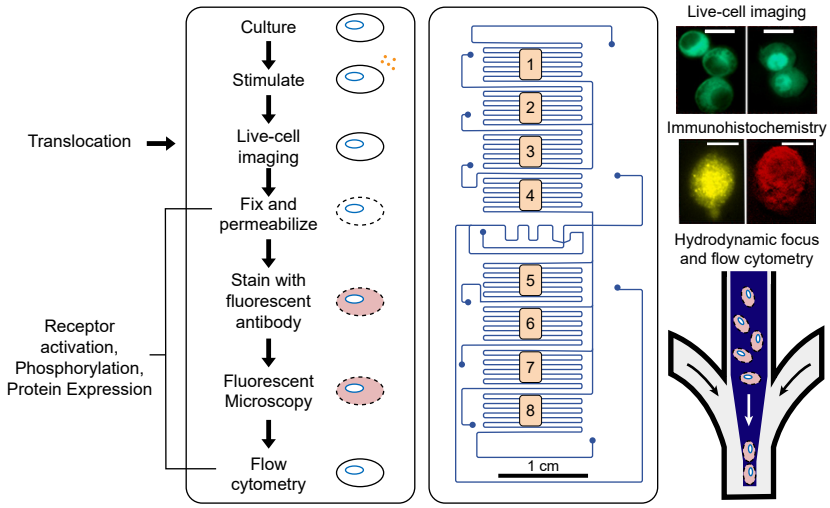


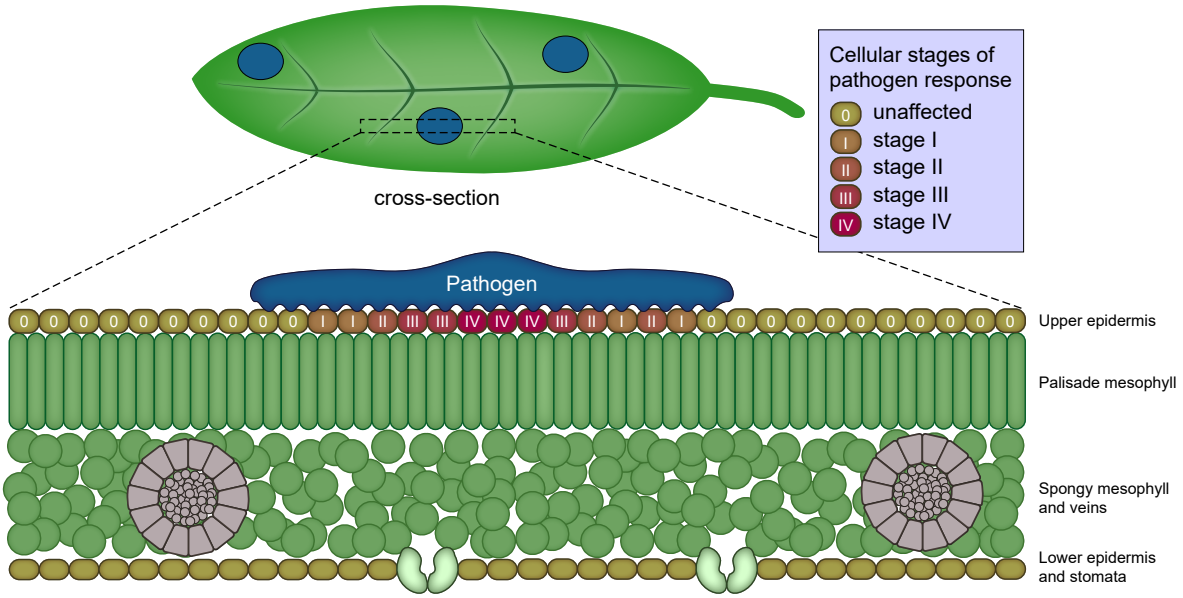
Figure 2. Microfluidic platform for single-cell imaging and protein profiling. Microfluidics devices are capable of integrating multiple profiling methods for single-cell characterization, including live-cell imaging, immunohistochemistry, and more detailed characterization following flow-cytometry. Figure courtesy of Anup Singh.

WORKSHOP FINDINGS

Workshop participants were asked to discuss scientific grand challenges for which single-cell technologies could provide a powerful toolset to further our understanding of fundamental processes relevant to the biology of biomass feedstocks and their interactions with the environment. Specific areas of interest include a detailed understanding of how plants respond to biotic and abiotic environmental factors, opportunities for improved functional annotation of genomes, and applications for the production of bioproducts and biomaterials. In addition, there were substantive discussions about critical technological and analysis methods challenges, the need for a centralized resource to facilitate the sharing of single-cell data, the scope of such a data resource, and potential funding avenues for single cell science in energy and the environment.

Plant Responses to Biotic and Abiotic Interactions

Emerging single-cell technologies are expected to enable impactful discoveries in studies of plant responses to their environment. Examples of interactions with particularly high relevance include pathogenic infections and mutualistic associations with nitrogen-fixing bacteria, as well as abiotic environmental conditions such as drought, heat, or limited nutrient availability. Both pathogenic and commensal microorganisms typically interact with very specific subpopulations of cells in a host plant. For example, arbuscular mycorrhizal fungi specifically target cortical cells of the plant root. Current methods of performing RNA-seq on bulk tissue or cell populations isolated by fluorescence-activated cell sorting (FACS) of reporter-labeled plant lines massively dilute any signal originating from affected cells in the plant (**Figure 3**). Alternatively, construction of INTACT or TRAP lines helps to circumvent this challenge[25] and enables profiling of nuclei and ribosomes, respectively, but requires construction of transgenic plants. Microfluidic-based single-cell RNA-seq, in combination with emerging spatial transcriptomics methods, holds great promise for elucidating cell-specific responses to pathogenic infections or other perturbations. As described below (see **Critical Technological and Analysis Challenges**), the first essential step for tackling this problem is developing sample preparation and fixation methods to obtain accurate and reproducible baseline measurements of plant species of interest. These necessary experiments will help establish the detection power and dynamic ranges for single-cell techniques, which will guide the selection of tractable environmental problems in the near term. Of particular importance for this research area is the development of methods that are capable of capturing RNA molecules from both eukaryotic and prokaryotic organisms in the same experiment, since current methods are limited to eukaryotic cells that have mRNA polyadenylation. While existing methods can, in principle, be used for the interrogation of plant-fungal interactions at single-cell resolution, methods that can be extended to prokaryotes will be essential for fully characterizing plant interactions with multi-species microbial communities. Workshop attendees identified drought as another high-priority focus area in this domain. In addition to the long-term goal of understanding the biological effects of decreased water availability caused by changing environments, single-cell methods could be used in the short term to better understand which experimental systems that are currently used to simulate drought in the lab are the most biologically relevant. With resource investments in developing tissue preparation methods and new technologies, along with the study of targeted scientific questions, single-cell technologies have the potential to revolutionize plant environmental science.



RNA-seq Analysis Resolution	Examples of Methods	Detection Sensitivity for Pathogen-Induced Changes	Biological Insights
whole leaf	bulk tissue RNA-seq	**low** (high dilution of signal)	general pathogen-induced transcriptome changes without cell type information
all upper epidermis	RNA-seq of microdissected epidermis	**moderate** (some dilution of signal)	epidermis-specific pathogen-induced transcriptome changes (diluted)
pathogen-affected upper epidermis cells	RNA-seq of laser-capture-microdissected cells	**high** (relevant cells only)	epidermis-specific pathogen-induced transcriptome changes
stage-specific pathogen-affected cells & many other cell types	single-cell RNA-seq	**very high** (stage-specific)	detailed insight into temporal progression of pathogen-induced transcriptome response

Figure 3: Advantages of using single-cell RNA-seq to study plant-pathogen interactions. Relatively few plant cells interact directly with most pathogens. However, these local interactions often determine disease severity. Thus, understanding gene expression in these few cells could be valuable for enhancing resistance. Unfortunately, bulk tissue RNA-seq greatly dilutes the signal from interacting cells, and signals from genes upregulated throughout the leaf in response to pathogens, like PR genes (not shown), can mask expression changes in the interacting cells. While methods like microdissection can improve the signal-to-noise ratio to a degree, they are labor intensive and not universally applicable to all pathogens, Thus, the increased cellular resolution promised by single-cell RNA-seq could revolutionize our understanding of plant-microbe interactions.

Better Annotation of Plant/Fungal/Algal Gene Function

A second major scientific focus area where single-cell technologies could have substantial impact is in the functional annotation of genes from plants, fungi, and algae. For example, DOE Joint Genome Institute portals (Phytozome, MycoCosm, PhycoCosm) host the sequences of >180 plant genomes (from >100 distinct species), along with >1,600 fungal and >50 algal genomes. Newly sequenced and assembled genomes are run through standardized annotation pipelines, which include using DNA sequence homology to genes in well-studied model species (e.g., Arabidopsis for plants) to infer the function of genes from the newly sequenced species. However, due to the ubiquity of large gene families with similar sequences in plants, identification of exactly homologous gene pairs between species is often challenging. Further complicating this challenge, functional understanding for most genes, even in well-studied species, is lacking. This can be mitigated by the use of gene expression information, in addition to sequence homology. RNA-seq data derived from different bulk plant tissues is already being used to define "expressologs", which are pairs of genes with similar expression profiles across general tissues in the species being compared. With scRNA-seq data, it will be possible to perform such analyses across dozens of cell types, thereby increasing the accuracy of the resulting annotations and inferred gene functions (**Figure 4**). There was general agreement that using this information could result in substantial improvements to plant functional gene annotations, and there was extensive discussion about the power of different study designs. Beyond using quantitative expression information for identification of functional gene homologs between species, nascent technologies for capturing full-length transcripts from single cells (e.g., scIso-seq[26] or Smart-seq3[27]) also have the potential to identify cell type-specific mRNA isoforms, adding another important layer of functional genomic annotation to transcriptome data. Generally, attendees supported the generation of scRNA-seq and/or spatially-resolved transcriptomics data for tissues from a panel of species as a starting point to build analysis tools and assess their utility. The initial panel should include a diverse set of species, including both better studied models (e.g., *Arabidopsis thaliana, Brachypodium distachyon*) and additional species selected for phenotypic or phylogenetic diversity, which would maximize the potential as a general resource. Alternatively, the initial panel might be selected based on more pragmatic criteria, like the availability of tissue preparation methods. Once established, this program could then be scaled to include a much wider diversity of species.

To complement this effort, attendees also supported the development of better cell culture transformation systems for environmental species. Such methods will be essential for performing high-throughput gene functional characterization in environmental species using Perturb-seq[28] screens, or conceptually similar methods such as CROP-seq[29] or CRISP-seq[30]. Perturb-seq, which has been used extensively in cultured human cell lines, combines CRISPR/Cas9-mediated gene knockout with scRNA-seq to elucidate gene regulatory networks, and it could be harnessed to study the importance of different genes and pathways under altered growth conditions, for example in the absence of specific nutrients. These methods could potentially be adapted to plants using a source of relatively homogeneous cells (e.g., leaf mesophyll protoplasts, or protoplasts derived from callus tissue). While these cells may behave differently than they would *in planta,* the high-throughput gene expression manipulation afforded by Perturb-seq and related methods would greatly accelerate gene function prediction and serve as a powerful hypothesis generation tool. Collectively, single-cell technologies performed on a diverse panel of plant species and tissues, along with the application of high-throughput functional screens, could substantially improve our understanding of gene function.

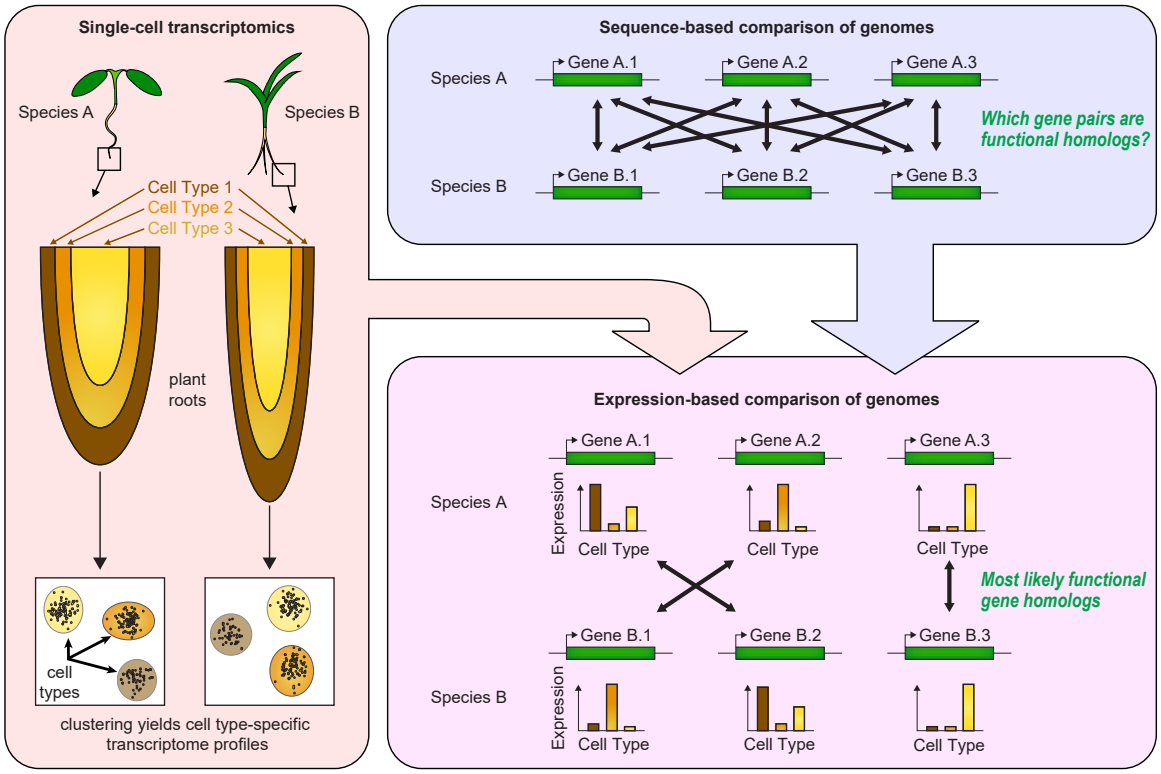


Figure 4: Using single-cell transcriptome data to improve comparative annotation of plant genomes. Expression profiles across multiple cell types derived from single-cell transcriptome data of tissues from different plant species (left), in combination with sequence homology-based comparison of protein sequences (top right), can be used to identify functionally homologous genes across different plant species (bottom right), thereby substantially enhancing the ability to assign functional knowledge from deeply annotated model species correctly to other species that are of interest to bioenergy and biomaterial production.

Improving Production of Bioproducts and Biomaterials

In addition to elucidating a foundational understanding of metabolism in plants and microbes, single-cell data will be important for both discovering natural product pathways and for successfully leveraging genome engineering and synthetic biology methods to produce biomaterials efficiently. Attendees discussed how single-cell techniques could aid in predicting and refactoring biosynthetic pathways, optimizing bioproduction, and generating predictive metabolic models. One important application for single-cell technologies will be in the area of biosynthetic pathway discovery. Some bioproducts produced by plants are synthesized predominantly in one or a few specific cell types (e.g., suberin in root endodermis cells), and biosynthetic pathways are known for only a small subset of plant products. While many types of enzymes can be predicted from genome information based on sequence similarity to related proteins, this information generally is insufficient to understand which genes work as part of a common pathway *in vivo.* For example, sequence similarity often enables robust prediction of enzyme class, such as “hydrolase” or “reductase”, but rarely predicts the substrate(s)[31]. High-throughput single cell metabolomics and proteomics methods would be invaluable for systematically mapping where naturally occurring bioproducts are produced in plant tissues. For those products restricted to specific cellular

populations, cell type-specific expression profiling could be used to narrow down components of a common biochemical pathway by identifying sets of enzymatic genes that are co-expressed in the same cell type (**Figure 5, top**). Additionally, single-cell expression information has the potential to improve bioengineering processes. Specific cell types are likely to provide better host environments than others for bioproduction because of the availability of substrates/cofactors, the absence of inhibitors, or resistance to product toxicity. Single-cell technologies applied to diverse plant tissue types are widely expected to aid identification of cell types that are best for making a product. More importantly, these approaches can also identify promoters or other regulatory elements that can direct expression to those cell types with high specificity, thereby providing crucial building blocks for biosynthetic engineering (**Figure 5, middle**). Finally, single-cell transcriptome profiling can be coupled with single-cell proteomics, antibody labeling or high-throughput microfluidic phenotyping systems (**Figure 2**) using plant protoplasts or unicellular eukaryotes, such as algae. This approach could be used to assess, for example, libraries of cells engineered to overexpress candidate genes/pathways or saturation mutagenesis libraries (**Figure 5, bottom**). By combining single-cell gene expression and phenotyping information, it will be possible to correlate transcript abundance with cellular measurements, enabling a rapid assessment of thousands of genetic manipulations for their phenotypic impact. Example applications for this approach include the search for genes and pathways that increase production of a biomaterial of interest in a given species.

Adapting single-cell technologies (transcriptome, proteome, and metabolome) to fungal and algal species, in addition to plant cell suspension systems, will be particularly important for improving bioproduct and biomaterial production. These methods could additionally provide a foundational understanding of culture population diversity, facilitate pathway optimization through parallelization, elucidate synthesis dynamics, and reveal whether heterogeneous populations are important for synthesis. For instance, synthesis of some bioproducts may require a combination of cell types and a mechanism for transport of metabolites between cell types. The development of methods that allow sampling of multiple different molecule types in parallel (e.g., mRNA AND metabolites) or imaging in combination with molecular profiling, along with sample preparation methods for a diversity of commonly used production species/strains that do not substantially alter cellular phenotype, will be critical for bioproduct synthesis applications.

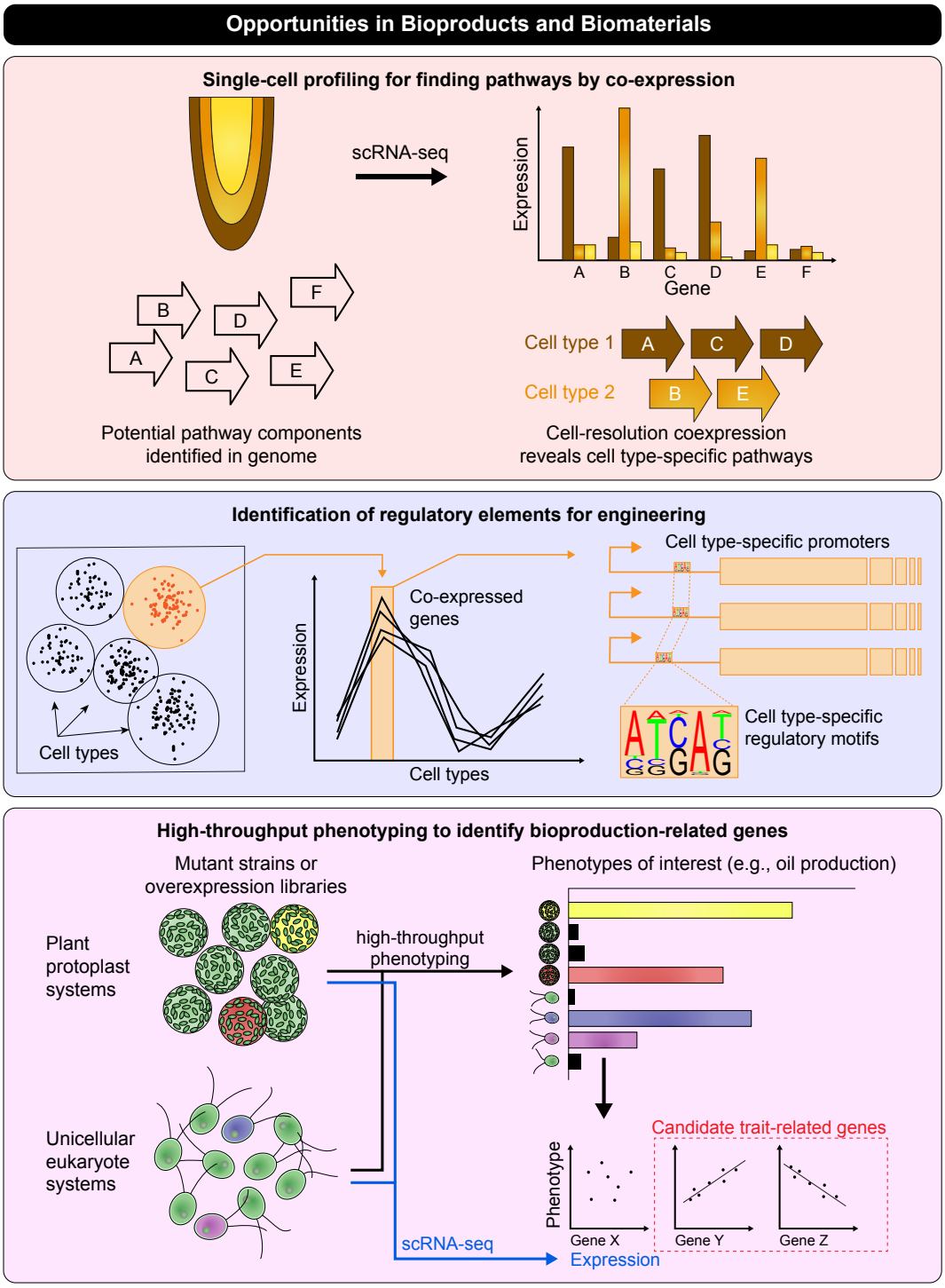


Figure 5: Using single-cell methods in bioproducts and biomaterials applications. Top panel: single-cell resolution data can be used to find genes in biosynthesis pathways by identifying co-expressed genes in individual cells or cell types. **Middle panel:** single-cell expression data can identify cell- and condition-specific building blocks, as genes that co-vary across clusters of cells are likely regulated by common components (e.g., transcription factors). This can be exploited to identify promoters useful for bioengineering applications where production in a specific cell type is desired. **Bottom panel:** improvements to bioproduction targets in plant or algal systems could be achieved through correlating high-throughput phenotyping and single-cell resolution "omics" data. High-throughput analyses of mutant strains or libraries containing engineered biosynthetic clusters could be used to identify or verify which genes and pathways are necessary for the production of specific products and to optimize for higher production yield.

Critical Technological and Analysis Challenges

Tissue preparation: Plant, algal, and fungal species, in contrast to animals, have complex polysaccharide cell walls that must be removed or permeabilized for single-cell characterization. This challenge has substantially hindered the application of these methods to such species. Methods of using enzyme cocktails to remove cell walls (i.e., protoplasting) are available for some species and tissues (e.g., Arabidopsis root), but cell wall composition differs from species to species and even between tissues of the same plant, so these methods are not universally applicable. Additionally, these dissociation/permeabilization methods impart unintended transcriptional or metabolic changes to the cells, and the very large size of some plant cells may inhibit the use of many of the microfluidics-based single-cell technologies commonly applied to mammalian cells. There was widespread agreement that isolation of nuclei, rather than whole cells, and cellular fixation methods (e.g., methanol treatment), likely hold the best promise for future single-cell science in plants and fungi. Several attendees also pointed to the need to revisit and revive historical methods of cell isolation and tissue preparation[32–36]. Overall, developing better tissue, cellular, and nuclear preparation methods for plants, fungi, and algae is an immediate focus area that would broadly enable the application of single-cell methods to environmental and energy science.

New single-cell technologies: There was substantial enthusiasm among attendees for the development of single-cell technologies beyond the commonly used microfluidic scRNA-seq methods. One limitation of the microfluidic methods is that they require tissue dissociation, and spatial information about where a specific cell came from within the tissue is lost. Emerging spatial transcriptomics methods are being widely adopted in the animal biology community, but thus far have been limited to plant shoot tissue[37,38]. Several such methods (e.g., Slide-seq[6]), while not currently single cell-resolution, provide spatially-resolved gene expression for tissue slices. Additionally, methods such as MERFISH[7,39] can reveal gene expression down to specific sub-regions within cells but currently require substantial investments and specialized expertise in microscopy equipment. There was strong interest in the development of high-throughput single-cell transcriptomics methods that could capture information for both eukaryotic and non-eukaryotic organisms at the same time, since current widely used methods are restricted to reading RNA transcripts that have polyadenylation signals, prohibiting their use for profiling bacteria or archaea that are interacting with plants. The study of plant/microbial interaction systems pose the additional challenge of spatial complexity in three dimensions. For instance, bacteria are often non-uniformly distributed across multiple planes when colonizing plants. Recently, methods incorporating polyadenylation enzymes that target mRNA from bacteria have been demonstrated to overcome limitations in prokaryotic transcriptome capture[40,41]. Still other technologies have shown the possibility of describing transcriptomic changes in 3-dimensional space (e.g., FISSEQ[42] and STARmap[43]), while computational methods are being developed to accurately segment 3-dimensional images of plant tissues into their composite cells for detailed analysis[44]. Further application and integration of such methods would substantially benefit the study of plant-microbe interactions in the environment. Currently, many single-cell or spatially resolved transcriptomics methods result in data that is restricted to specific regions of genes (e.g., the 3' end of transcripts for droplet-based scRNA-seq), and there were suggestions that sequencing full-length transcripts would provide useful information about gene isoforms, which may be necessary to address some questions as a complement to the current, higher-throughput methods. Beyond gene expression, there is a strong need for high-throughput single-cell proteomics and metabolomics methods. Such methods are in development but have throughputs that currently lag substantially behind transcriptomics methods or require specialized antibodies, limiting their application to specific panels of proteins[45,46]. Emerging methods such as CITE-seq[47] combine scRNA-seq with antibody labeling to interrogate gene expression and the repertoire of cell surface proteins for individual

cells in the same experiment. In addition to single-cell profiling methods, many meeting attendees pointed to the need for better methods to validate single-cell results, including improved *in situ* hybridization protocols such as single-molecule FISH[48] for plants, as well as faster and more efficient ways to generate reporter lines. Additionally, application of technologies like the 10X-Genomics VISIUM platform could also serve as a powerful validation and discovery tool[38].

Analysis methods: Complementary to novel microfluidics methods and advanced molecular biology reagents and protocols, innovative computational methods employing statistical tools rooted in machine learning have been the third technology pillar that has enabled breakthrough advances in single-cell approaches in recent years. Examples of these approaches include computational strategies to capture "free" information from existing data, including developmental trajectories (pseudotime[49]) and RNA dynamics (RNA velocity[50]). While many of the tools already developed for analysis of single-cell data from mammalian tissues will be applicable to analysis of plant data sets, workshop participants discussed several computational challenges that are unique to plants and would benefit from the development of new tools and databases. For example, as research moves from Arabidopsis roots in the first wave of studies to new species and tissues, how do we know if cell clusters represent true cell types if no high-quality cell type-specific markers are already known? Furthermore, how will typical mapping pipelines for scRNA-seq perform when aligning to transcripts from less well-annotated genomes? To enable robust and valid data analysis, we will need tools for cross-species comparison. In considering the funding landscape for such efforts with a focus on species relevant to bioenergy applications, it was felt that DOE, with its strong history of driving plant and microbial computational tool development, was particularly well positioned to support such efforts.

A Pressing Need for an Open Plant Cell Atlas Data Resource

There was broad consensus among attendees regarding the need for a publicly accessible platform that conforms to the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles (**https://www.go-fair.org/fair-principles/**) for sharing results throughout the plant biology community. Large-scale consortia in the human biology field, such as the Human Cell Atlas and the Encyclopedia of DNA Elements (ENCODE), have made routine the immediate and open sharing of data, often even in advance of publication. In contrast, attendees expressed frustration that research communities that have organized around different plant species often have vastly different data sharing practices, and journal publication requirements alone have proven inadequate to compel consistent data sharing. Seung Yon Rhee, in a talk at the workshop and in a recent perspective[51], outlined a vision for a Plant Cell Atlas that would include not only single-cell transcriptomics data but multi-scale imaging, proteomics, and other data types, as well. A unified portal for single-cell and other data from the plant community would greatly facilitate the widespread movement toward FAIR data principles. There was universal agreement that such a platform, like the Human Cell Atlas, should have international support and accessibility and not be wholly funded by a single country or funding agency. In addition to enabling data sharing, such a resource would have the added benefit of establishing high standards for data quality, enable consistency in data analysis, and provide innovative and powerful data visualization tools. Having a unified platform supporting multiple plant research communities would facilitate solutions to emerging problems, such as how to define analogous cell types between different species. It could also serve as a platform to support the sharing of information beyond results, such as tissue dissociation or preparation protocols.

Deep or Wide?

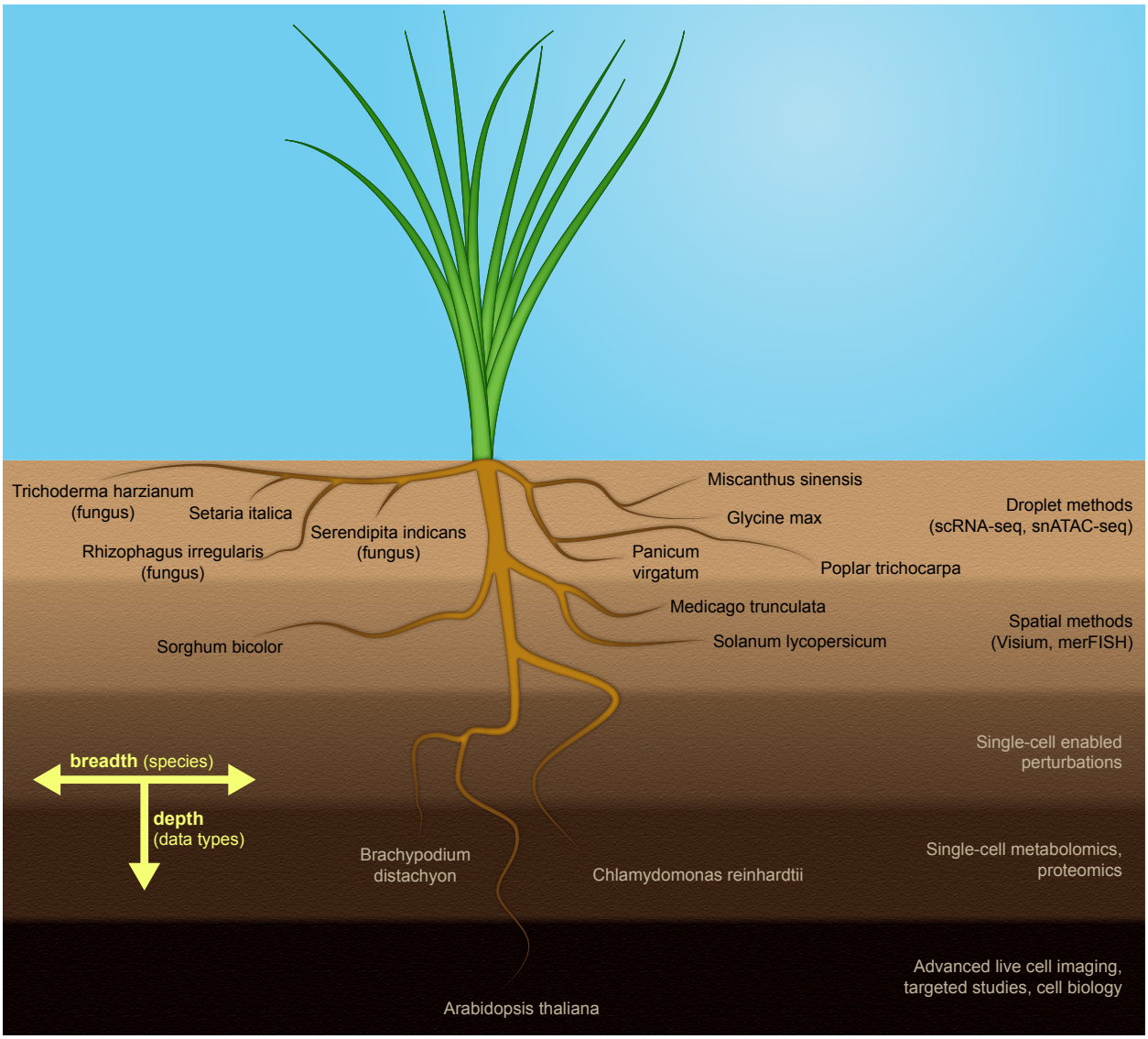


Figure 6: Deep or wide? The schematic root system of a hypothetical plant, covering a large area close to the surface while also penetrating deeper soil layers with some of its roots, provides a visual metaphor for the need to complement "wide and shallow" characterization of many species using a select subset of single-cell assays with "deep and narrow" in-depth studies of select model species using the full arsenal of single-cell methods available.

An overarching question discussed at the workshop is whether the effort to establish a Plant Cell Atlas should focus exclusively on very deep characterization of a single plant species or generation and curation of data from a wide variety of species. Given the existing research and database infrastructure, a single-species effort would almost certainly focus on *Arabidopsis thaliana.* Pragmatically, it makes inherent sense if resources are limited to commit to completing a deep, multimodal characterization of a single species. Having such a dataset would have the best potential for being able to integrate different types of data with machine learning and similar strategies to construct accurate systems-level models of an entire plant. However, this "narrow-and-deep" approach has inherent limitations for understanding aspects of plant biology that Arabidopsis either does not perform or poorly models, such as a C_4-mode of carbon fixation or

many of the general anatomical, physiological, and molecular features that facilitate high-yield biomass production in bioenergy grasses. A "wide-and-shallow" effort to perform a subset of the proposed molecular profiling, such as only single-cell or spatial transcriptomics, on tissues from a larger panel of phylogenetically or phenotypically diverse plant species would be highly complementary to a deep characterization of Arabidopsis (**Figure 6**). This approach would provide critical baseline information about a variety of plant species important for the environment, energy, biosynthesis, and food production. It would establish a centralized open data resource for the research communities working on these species and inform downstream experimental studies, genome annotation, and genetic engineering of these organisms. Ideally, both wide and deep efforts would not be mutually exclusive and would work together to coordinate data production and release through a centralized portal that would broadly serve the plant biology community.

Existing Funding Landscape for Single-Cell Science

Workshop participants discussed and identified a variety of different potential funding sources and other resources to support the development and application of single-cell technologies to energy and environmental research questions. The National Institutes of Health (NIH) and the Chan Zuckerberg Initiative (CZI) have both heavily invested in single-cell technology and analytics development and support, the use of these technologies to further biomedical sciences and for the development of the Human Cell Atlas. As of Spring 2020, NIH has funded >300 different single-cell research projects at a total cost of >$175M. CZI funded the initial pilot projects for the Human Cell Atlas and is leading the development of a data coordination platform and computational tools for this initiative. While unlikely to fund research programs not directly related to biomedicine, NIH and CZI have a substantial interest in generally supporting single-cell technology and computational tool development. Projects on developing single-cell proteomics and metabolomics techniques, along with computational analysis methods, would be of particular interest to NIH and CZI and would be broadly applicable to scientific questions in both biomedicine and bioenergy. The National Science Foundation has to date funded >30 single-cell projects totaling $19.9M. Beyond foundational plant science, NSF has previously supported major plant database initiatives, including The Arabidopsis Information Resource (TAIR). NSF also provided initial workshop funding for the Plant Cell Atlas, signalling a strong interest in this initiative and related science going forward. The USDA has funded a few projects that leverage single-cell technology, though the scope of these projects is limited to animal studies. As a major sponsor of agricultural research, USDA would represent an attractive avenue for supporting plant single-cell research, though likely this would be restricted to USDA mission priority crops.

The Department of Energy, primarily through the Biological and Environmental Research arm of the Office of Science, has made major investments in foundational genomics research on a wide variety of plant, fungal, and algal species relevant to bioenergy and environmental problems. The DOE-funded Joint Genome Institute has begun to actively explore the development and application of single-cell methods for energy and environmental research, with the goal to make such technologies available for JGI users applying them in support of the DOE scientific mission. Historically, the Department of Energy has also supported efforts in applying single-cell genomics in prokaryotic systems. However, workshop participants agreed that these initial investments are unlikely to capture the full potential resulting from applying single-cell methods to plants and fungi for advancing energy and environmental science and highlighted the opportunities that would result from more substantial investments into these technologies for bioenergy and biomaterials applications.

Beyond US governmental agencies, there was strong support for pursuing diversified funding from other governmental agencies, nonprofit organizations, and industry. Organizations with a potential interest in such research included The Bill and Melinda Gates Foundation, Simons Foundation, Genome Canada, the European Research Area Network for Coordinating Action in Plant Sciences (ERA-CAPS), GARNet (UK), Deutsche Forschungsgemeinschaft (DFG, the German Research Foundation), Howard Hughes, Japan Society for the Promotion of Science (JSPS), Israeli Science Foundation, BBSRC (UK), Gordon and Betty Moore Foundation, W.M. Keck Foundation, Sloan Foundation, BASF, Corteva, 10x Genomics, and Illumina.

SUMMARY AND CONCLUSIONS

Over the past several years, an astounding array of new technologies has driven unprecedented advances in the biomedical sciences. New methods that leverage advanced experimental and computational tools provide single-cell resolution transcriptome and epigenome information and are complemented by nascent methods for proteome, metabolome, and spatially-resolved transcriptomics at the single-cell level. Within the past year, we have begun to see some of these same approaches demonstrated in plants, fungi, and algae. While significant technical challenges still need to be overcome before these techniques can be broadly applied to the wide array of species that are of interest to energy and environmental studies, this initial wave of published studies is only a harbinger of the powerful discovery opportunities these methods will enable. Thus, the time is ripe for focused investments into development and adoption of single-cell methods to drive the next wave of biological innovation for energy and environmental science. It was with these new opportunities and challenges in mind that we convened the Plant Single-Cell Solutions for Energy and the Environment Workshop on January 23, 2020, where experts in plant, fungal, and microbial genomics gathered to discuss what these new technologies can bring to energy and environmental science. The workshop identified key challenges and opportunities for single-cell methods in plants, fungi, and algae, along with plant-biome interactions.

Single-cell molecular profiling methods are expected to have the same paradigm-shifting potential for plant and environmental biology as they have already had in the biomedical sciences. As John Vogel (Joint Genome Institute) stated in his presentation, "in plant science, cell type resolution has always been 'the holy grail,'" and single-cell methods are expected to provide a direct window into multiple areas of plant biology:

- **Insight into plant cell type function:** Nearly all biological functions a plant executes *in vivo* occur through the interplay of many different cell types with highly specialized functional profiles. Resolving the molecular blueprint (transcriptome, proteome, metabolome, etc.) of specific cell types across plant tissues will give direct insight into how cells perform their respective specialized roles.
- **Insight into plant development:** Understanding the development of plants is critical for improving traits such as biomass yield. Advanced single-cell transcriptome analysis strategies, such as "pseudotime" and "RNA velocity", enable unprecedented insight into developmental trajectories of cell types during plant development.
- **Understanding plant responses to environmental factors:** Many factors affecting the response of plants to environmental factors, such as pathogens, drought, nutrients, climate, or soil are likely driven by very specific processes taking place only in subsets of their cell types. Single-cell methods will make it possible to deconvolute these responses and assign specific aspects of the organismal response to the cell types they occur in.
- **Functional annotation of plant genes and gene families:** Plants tend to have large gene families and complex, polyploid genomes, which creates major challenges in correctly identifying functional gene homologs across related plant species. Single-cell technologies provide high-resolution gene expression data that can be used to enable correct assignment of functional orthologs across species, making it possible to correctly extrapolate gene function from deeply annotated model species to crops of interest. Methods combining single cell technologies with high-throughput genome engineering could elucidate the function of genes that have not been previously characterized.
- **Identification of targets for bioenergy crop improvement:** As scientists and breeders strive for predictive engineering of plant traits, a detailed understanding of how gene networks are

composed and regulated in response to environmental input at a cell type-level will substantially accelerate progress towards the creation of more productive and sustainable energy crops.

Applying single cell methods to microbial and fungal species, in addition to plants, was seen as another valuable opportunity for discovery:

- By applying these techniques to bacteria, fungi, and plant-microbe interactions, it will be possible to understand in detail how plants and microbes interact in commensal, competitive, and pathogenic relationships.
- Fundamental insights into cell state properties of eukaryotic microbes could be used to improve bioreactor-based production. Single-cell measurements of individuals across a population can capture properties such as life cycle, measure population heterogeneity, distinguish between stochastic and regulated processes, and guide how desired cell states can be selected for through engineering.

Workshop participants acknowledged that applying single-cell genomics to important plant, fungi and microbial systems is likely to face unique challenges, but it was also uniformly agreed that these issues could be addressed through targeted investments in technology development and a data sharing platform. Participants shared their excitement regarding the impact that advances in single-cell technologies will have when widely applied to plants, fungi, and microbes. It was agreed by participants that these types of technologies will be transformative for our understanding of environmental biology and powerful for trait engineering for bioenergy and biomaterials. In conclusion, workshop participants were enthusiastic about the bright future for single-cell molecular technologies and expressed hope that relevant funding agencies will recognize the timely opportunities associated with these emerging methods.

REFERENCES

1. Spindel, J. E. *et al.* Association mapping by aerial drone reveals 213 genetic associations for Sorghum bicolor biomass traits under drought. *BMC Genomics* **19,** 679 (2018).
2. Varoquaux, N. *et al.* Transcriptomic analysis of field-droughted sorghum from seedling to maturity reveals biotic and metabolic responses. *Proc. Natl. Acad. Sci.* **116,** 27124–27132 (2019).
3. Taylor-Teeples, M. *et al.* An *Arabidopsis* gene regulatory network for secondary cell wall synthesis. *Nature* **517,** 571–575 (2015).
4. Han, X. *et al.* Construction of a human cell landscape at single-cell level. *Nature* **581,** 303–309 (2020).
5. Birnbaum, K. D. Power in Numbers: Single-Cell RNA-Seq Strategies to Dissect Complex Tissues. *Annu. Rev. Genet.* **52,** 203–221 (2018).
6. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363,** 1463–1467 (2019).
7. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348,** (2015).
8. Lee, J. H. *et al.* Highly multiplexed subcellular RNA sequencing in situ. *Science* **343,** 1360–1363 (2014).
9. Merritt, C. R. *et al.* Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat. Biotechnol.* **38,** 586–599 (2020).
10. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353,** 78–82 (2016).
11. Raj, B. *et al.* Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36,** 442–450 (2018).
12. Rosenberg, A. B. *et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360,** 176–182 (2018).
13. Kanton, S. *et al.* Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* **574,** 418–422 (2019).
14. Zhong, S. *et al.* Decoding the development of the human hippocampus. *Nature* **577,** 531–536 (2020).
15. Hodge, R. D. *et al.* Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573,** 61–68 (2019).
16. Zhu, Y. *et al.* Single-cell proteomics reveals changes in expression during hair-cell development. *eLife* **8,**.
17. Marx, V. A dream of single-cell proteomics. *Nat. Methods* **16,** 809–812 (2019).
18. Shulse, C. N. *et al.* High-Throughput Single-Cell Transcriptome Profiling of Plant Cell Types. *Cell Rep.* **27,** 2241-2247.e4 (2019).
19. Zhang, T.-Q., Xu, Z.-G., Shang, G.-D. & Wang, J.-W. A Single-Cell RNA Sequencing Profiles the Developmental Landscape of Arabidopsis Root. *Mol. Plant* **12,** 648–660 (2019).
20. Ryu, K. H., Huang, L., Kang, H. M. & Schiefelbein, J. Single-Cell RNA Sequencing Resolves Molecular Relationships Among Individual Plant Cells. *Plant Physiol.* **179,** 1444–1456 (2019).
21. Denyer, T. *et al.* Spatiotemporal Developmental Trajectories in the Arabidopsis Root Revealed Using High-Throughput Single-Cell RNA Sequencing. *Dev. Cell* **48,** 840-852.e5 (2019).

22. Jean-Baptiste, K. *et al.* Developmental and conditional dynamics of gene expression in single root cells of A. thaliana. *bioRxiv* 448514 (2018) doi:10.1101/448514.

23. Turco, G. M. *et al.* Molecular Mechanisms Driving Switch Behavior in Xylem Cell Differentiation. *Cell Rep.* **28,** 342-351.e4 (2019).

24. Kompauer, M., Heiles, S. & Spengler, B. Autofocusing MALDI mass spectrometry imaging of tissue sections and 3D chemical topography of nonflat surfaces. *Nat. Methods* **14,** 1156–1158 (2017).

25. Reynoso, M. A. *et al.* Evolutionary flexibility in flooding response circuitry in angiosperms. *Science* **365,** 1291–1295 (2019).

26. Gupta, I. *et al.* Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat. Biotechnol.* **36,** 1197–1202 (2018).

27. Hagemann-Jensen, M. *et al.* Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat. Biotechnol.* **38,** 708–714 (2020).

28. Dixit, A. *et al.* Perturb-seq: Dissecting molecular circuits with scalable single cell RNA profiling of pooled genetic screens. *Cell* **167,** 1853-1866.e17 (2016).

29. Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14,** 297–301 (2017).

30. Jaitin, D. A. *et al.* Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167,** 1883-1896.e15 (2016).

31. Schnoes, A. M., Brown, S. D., Dodevski, I. & Babbitt, P. C. Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. *PLoS Comput. Biol.* **5,** (2009).

32. Turco, G. M. *et al.* DNA methylation and gene expression regulation associated with vascularization in Sorghum bicolor. *New Phytol.* **214,** 1213–1229 (2017).

33. Rodriguez-Villalon, A. & Brady, S. M. Single cell RNA sequencing and its promise in reconstructing plant vascular cell lineages. *Curr. Opin. Plant Biol.* **48,** 47–56 (2019).

34. Nakazono, M., Qiu, F., Borsuk, L. A. & Schnable, P. S. Laser-Capture Microdissection, a Tool for the Global Analysis of Gene Expression in Specific Plant Cell Types: Identification of Genes Expressed Differentially in Epidermal Cells or Vascular Tissues of Maize. *Plant Cell* **15,** 583–596 (2003).

35. Ohtsu, K. *et al.* Global gene expression analysis of the shoot apical meristem of maize (Zea mays L.). *Plant J.* **52,** 391–404 (2007).

36. Sheen, J. Chapter 21 Methods for Mesophyll and Bundle Sheath Cell Separation. in *Methods in Cell Biology* (eds. Galbraith, D. W., Bohnert, H. J. & Bourque, D. P.) vol. 49 305–314 (Academic Press, 1995).

37. Giacomello, S. & Lundeberg, J. Preparation of plant tissue to enable Spatial Transcriptomics profiling using barcoded microarrays. *Nat. Protoc.* **13,** 2425–2446 (2018).

38. Giacomello, S. *et al.* Spatially resolved transcriptome profiling in model plant species. *Nat. Plants* **3,** 1–11 (2017).

39. Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **116,** 19490–19499 (2019).

40. Blattman, S. B., Jiang, W., Oikonomou, P. & Tavazoie, S. Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. *Nat. Microbiol.* 1–10 (2020) doi:10.1038/s41564-020-0729-6.

41. Kuchina, A. *et al.* Microbial single-cell RNA sequencing by split-pool barcoding. *bioRxiv* (2019) doi:10.1101/869248.

42. Lee, J. H. *et al.* Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc.* **10,** 442–458 (2015).

43. Wang, X. *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361,** (2018).

44. Wolny, A. *et al.* Accurate and Versatile 3D Segmentation of Plant Tissues at Cellular Resolution. *bioRxiv* 2020.01.17.910562 (2020) doi:10.1101/2020.01.17.910562.

45. Slavov, N. Unpicking the proteome in single cells. *Science* **367,** 512–513 (2020).

46. Dou, M. *et al.* High-Throughput Single Cell Proteomics Enabled by Multiplex Isobaric Labeling in a Nanodroplet Sample Preparation Platform. *Anal. Chem.* **91,** 13119–13127 (2019).

47. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14,** 865–868 (2017).

48. Duncan, S., Olsson, T. S. G., Hartley, M., Dean, C. & Rosa, S. A method for detecting single mRNA molecules in Arabidopsis thaliana. *Plant Methods* **12,** 13 (2016).

49. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32,** 381–386 (2014).

50. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560,** 494–498 (2018).

51. Rhee, S. Y., Birnbaum, K. D. & Ehrhardt, D. W. Towards Building a Plant Cell Atlas. *Trends Plant Sci.* **24,** 303–310 (2019).

APPENDIX 1: WORKSHOP AGENDA

January 23, 2020

8:30 – 9:00	Registration
9:00 – 9:15	Welcome and Introductions — **Diane Dickel**
9:15 – 9:45	TALK: Single Cell Epigenomics of the Brain **Joe Ecker,** Salk Institute for Biological Studies
9:45 – 10:15	TALK: Plant Cell Atlas **Seung Yon Rhee,** Carnegie Institution for Science
10:15 – 10:30	**Coffee Break**
10:30 – 11:00	TALK: Seeing with Cell Types in the Retina **Yirong Peng,** UC Los Angeles
11:00 – 12:00	HIGHLIGHT TALKS 1. Integrative Analysis of Plant Single-Cell Transcriptomes **Ben Cole,** LBNL/JGI 2. Single Cell Sequencing and Vascular Development **Siobhan Brady,** UC Davis 3. Microfluidic Platforms for Single Cell Analysis and Manipulation **Anup Singh,** Sandia/JBEI 4. Potential for Single Cell Transcriptomics to Study Plant-Microbe Interactions **John Vogel,** LBNL/JGI 5. Coming to Terms with Potential: scRNA-seq Applied to Environmentally Tuned Development in Plant Leaves **Dominique Bergmann,** Stanford 6. Plant Genome Annotation at the JGI **David Goodstein,** LBNL/JGI 7. Plant and Fungal Single-Cell Analysis at JBEI **Blake Simmons,** LBNL/JBEI
12:00 – 12:15	**Lunch Setup**
12:15 – 13:45	Breakout Sessions Leaders: **Devin Coleman-Derr, Ben Cole, Crysten Blaby-Haas, Jenny Mortimer, Kris Niyogi, Ronan O'Malley**
13:45 – 14:15	**Tour of the New Integrative Genomics Building / Break**
14:15 – 15:00	Breakout Group Summaries — **Devin Coleman-Derr, Crysten Blaby-Haas, Kris Niyogi**
15:00 – 16:00	Discussion and Next Steps — **Axel Visel**
16:00 – 16:15	**Group Photo**
16:15	Close — **Diane Dickel**

APPENDIX 2: WORKSHOP CHARGE QUESTIONS

Topic 1: Plant Responses to Biotic/Abiotic Interactions

Discussion Facilitator: **Devin Coleman-Derr**
Google Doc Facilitator: **Ben Cole**

Grand Challenges

- What are the grand challenges for understanding plant responses to biotic and abiotic stress?
- Do different tissues/cell types have different responses to heat or drought stress?
- What is the diversity of cellular states in plant-colonizing bacterial or fungal populations, and are they different in plant- and non-plant contexts?
- Do plant cell types respond to biotic stress differently?
- Do microbes preferentially affect/infect certain cell types?

Opportunities

- What plant species to focus on for each of the topics above?
- What fungal/bacterial species to focus on to explore pathogenic/commensal/beneficial interactions?
- Are there existing tissue preparation techniques that could be harnessed for emerging spatial transcriptomics methods?
- Possible to conduct experiments in soil or soil-like substrata?
- What technical challenges will be raised by specific species (e.g., bioenergy crops) and tissues?
- How to incorporate single-cell technologies into field or manufactured ecosystem (EcoFab) experiments?

Needed Capabilities/Resources

- How to link specific plant cells and microbes that are directly interacting?
- Strategies to target plant/bacterial interactions that include sampling of bacterial transcriptomes?
- How to balance the need for replicate experiments with the cost of technologies?

Topic 2: Better Annotation of Plant/Fungal/Algal Gene Function

Discussion Facilitator: **Kris Niyogi**
Google Doc Facilitator: **Ronan O'Malley**

Grand Challenges

- Could combining sequence conservation and single-cell expression information improve ortholog assignments and aid in gene functional characterization?

- What species/tissues would be best to start with (Arabidopsis? Brachypodium? other DOE relevant?)?
- Is broad phylogenetic coverage important?

Opportunities

- What are current important technologies (scRNA, scATAC) and what will be needed (spatial transcriptomics, in vivo validation)?
- What is needed to develop tissue dissociation techniques for a wider variety of species and/or will data from nuclei be sufficient?
- Are there plant/fungal/algal systems that are amenable to Perturb-seq or mutant analysis type methods to study gene function?
- Will full length single-cell spliceform information be needed?

Needed Capabilities/Resources

- How to standardize cell type definitions and plant growth stages to allow interspecies comparisons?
- Computational, data, and visualization resources to facilitate interspecies comparisons for outside users
- In vivo validation techniques?

Topic 3: Applications for Bioproducts and Biomaterials

Discussion Facilitator: **Crysten Blaby-Haas**
Google Doc Facilitator: **Jenny Mortimer**

Grand Challenges

- Discovery: Could methods of predicting biosynthesis pathways be aided by cell type specific gene expression to improve prediction in plants and/or other organisms?
- Production: How could plant/algal single-cell production systems benefit from scRNA-seq? (e.g., tobacco cell suspension system)
- Production: How can cell type-specific information aid engineering of plants to increase production of desired bioproducts?
- What other grand challenges in bioproduct production could be addressed by single-cell technologies?

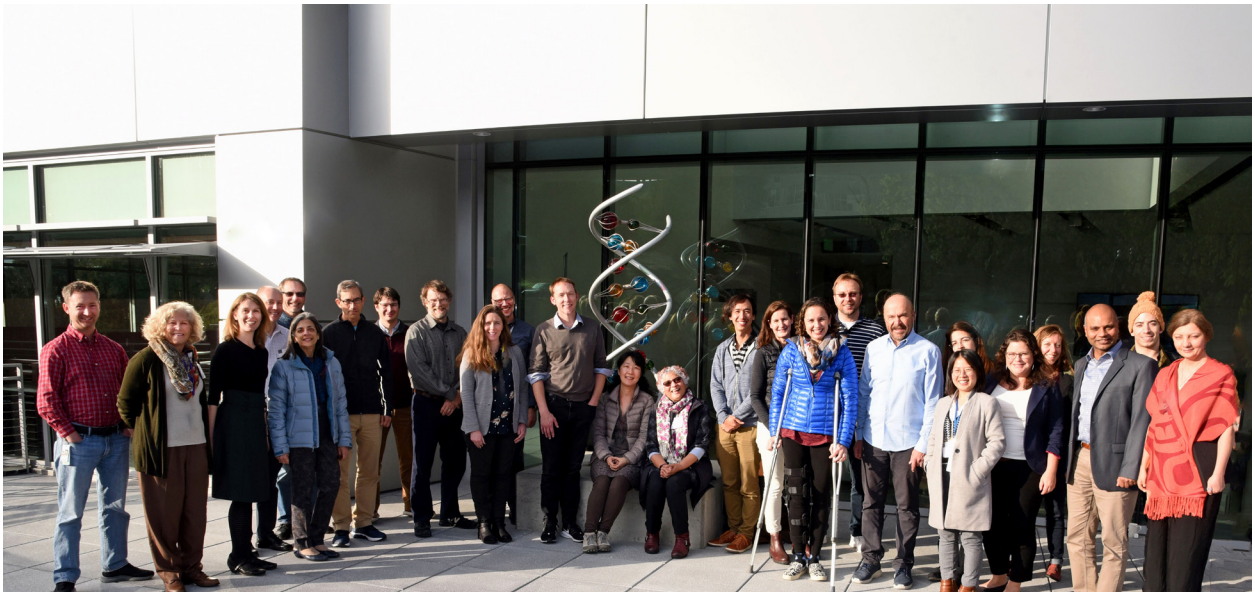
Opportunities

- How can single cell sequencing be combined with other techniques for metabolic modeling such as flux balance analysis, machine learning, etc.?
- What are high priority plant species/tissues for bioproduct prediction? (e.g., trichomes)
- Is there value in using scRNA-seq over existing barcoded/pooled library techniques for optimizing bioproduct synthesis pathways?
- Future partners (e.g., industry? academic?)

Needed Capabilities/Resources

- What optimization is needed for scRNA-seq in non-model species (e.g., cell wall cocktails)? What technical challenges are these species likely to raise?

APPENDIX 3: WORKSHOP PARTICIPANTS



Organizing Committee

Diane Dickel (chair)
Ben Cole
Rex Malmstrom
Jenny Mortimer
Chris Mungall
Ronan O'Malley
Axel Visel

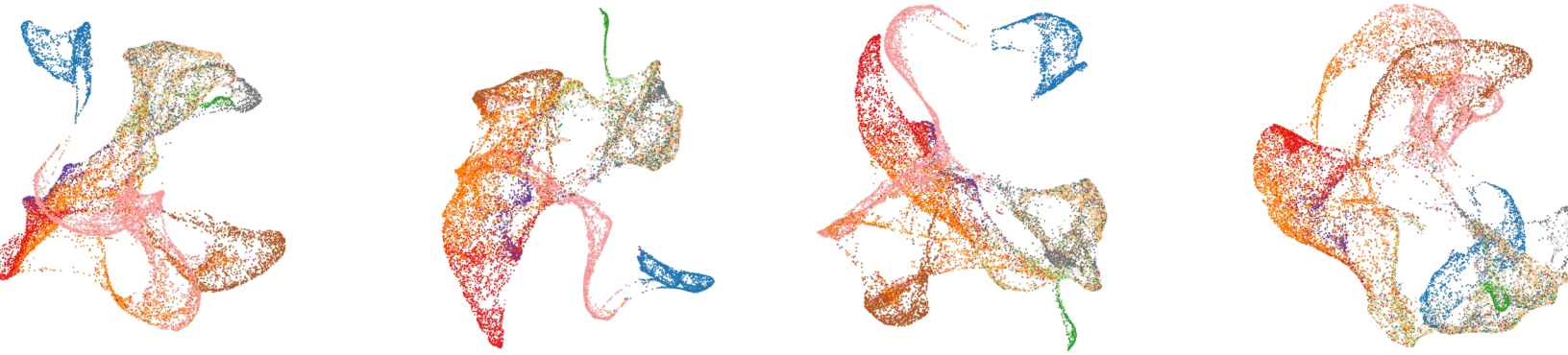
Writing Team

Preparation of this report was led and coordinated by the Writing Team with input from all workshop participants:
Diane Dickel
Ben Cole
Ronan O'Malley
Axel Visel

Editing Support:
Kelly Cobaugh, LBNL
Kianna von Maydell, LBNL

Workshop Participants

Name	Affiliation
Simon Alamos	UC Berkeley
Dominique Bergmann	Stanford University
Crysten Blaby-Haas	Brookhaven National Lab
Ian Blaby	LBNL, Joint Genome Institute
Matt Blow	LBNL, Joint Genome Institute
Kristofer Bouchard	LBNL, Biological Systems and Engineering
Siobhan Brady	UC Davis
Doina Ciobanu	LBNL, Joint Genome Institute
Benjamin Cole	LBNL, Joint Genome Institute
Devin Coleman-Derr	UC Berkeley
Diane Dickel	LBNL, Environmental Genomics and Systems Biology
Joseph Ecker	Salk Institute for Biological Studies
N. Louise Glass	UC Berkeley; LBNL, Environmental Genomics and Systems Biology
David Goodstein	LBNL, Joint Genome Institute
Mona Gouran	UC Davis
Sharon Greenblum	LBNL, Joint Genome Institute
Igor Grigoriev	LBNL, Joint Genome Institute
Samuel Leiboff	UC Berkeley
Rex Malmstrom	LBNL, Joint Genome Institute
Sabeeha Merchant	UC Berkeley
Jenny Mortimer	LBNL, Environmental Genomics and Systems Biology
Nigel Mouncey	LBNL, Joint Genome Institute
Kris Niyogi	UC Berkeley
Trent Northen	LBNL, Environmental Genomics and Systems Biology
Ronan O'Malley	LBNL, Joint Genome Institute
Yirong Peng	UC Los Angeles
Len Pennacchio	LBNL, Joint Genome Institute
Seung Yon Rhee	Carnegie Institution for Science
Jeremy Schmutz	HudsonAlpha Institute for Biotechnology/Joint Genome Institute
Blake Simmons	LBNL, Biological Systems and Engineering
Anup Singh	Sandia National Labs
Neelima Sinha	UC Davis
Axel Visel	LBNL, Joint Genome Institute
John Vogel	LBNL, Joint Genome Institute



BERKELEY LAB
Bringing Science Solutions to the World

BioSciences

A U.S. Department of Energy National Laboratory. Managed by the University of California.

20-EGSB-14021